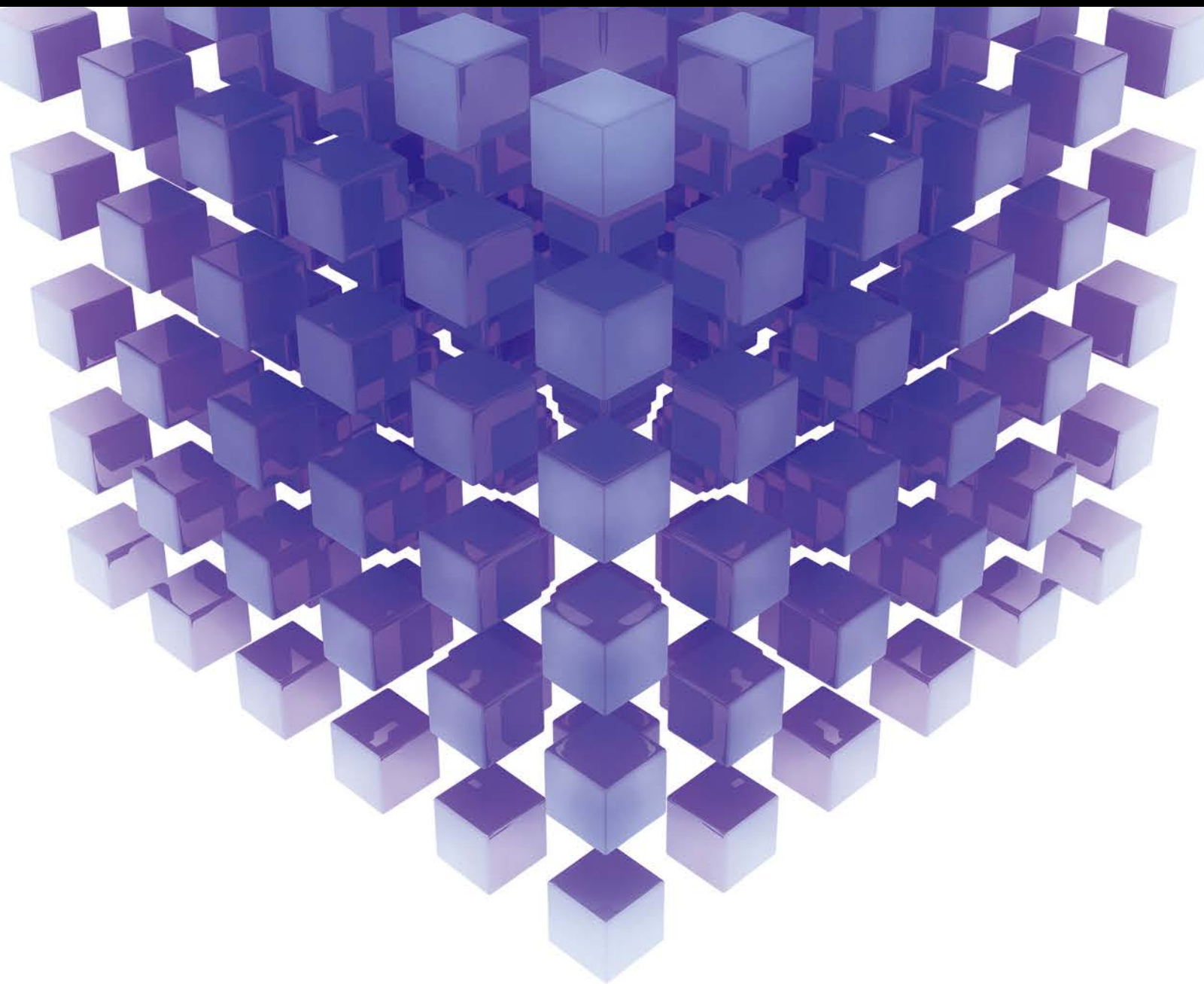# Modeling and Control Problems in Sustainable Transportation and Power Systems

Guest Editors: Xiaosong Hu, Shengbo Eben Li, and Dongsuk Kum

Modeling and Control Problems in Sustainable Transportation and Power Systems

Modeling and Control Problems in Sustainable Transportation and Power Systems

Guest Editors: Xiaosong Hu, Shengbo Eben Li, and Dongsuk Kum

Editorial Board

Contents

Editorial

Modeling and Control Problems in Sustainable Transportation and Power Systems

Xiaosong Hu,[1,2,3] Shengbo Eben Li,[4] and Dongsuk Kum[5]

[1]*The State Key Laboratory of Mechanical Transmissions, Chongqing University, Chongqing 400044, China*
[2]*College of Automotive Engineering, Chongqing University, Chongqing 400044, China*
[3]*Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA*
[4]*The State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China*
[5]*The Cho Chun Shik Graduate School for Green Transportation, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Republic of Korea*

Correspondence should be addressed to Xiaosong Hu; xiaosonghu@ieee.org

Sustainable transportation and power systems have the great potential to significantly reduce hydrocarbon consumptions, pollutant emissions, and carbon footprint [1]. The two key energy sectors are intimately coupled. For instance, large-scale deployment of electrified vehicles is considerably beneficial to renewable ways of power generation by including wind and solar energy sources. Such good synergies between the transportation and power sectors expedite a revolutionary transition towards a clean, high-efficiency, and affordable energy future. The inevitably increasing system complexity and the desire to make the most of sustainable energy systems constitute a major incentive to leverage advanced approaches of system modeling, simulation, control, optimization, diagnosis, and prognostics. In system-level analysis, there are various modeling/control challenges surrounding transportation electrification (e.g., alternative-energy powertrains), intelligent transportation system (e.g., autonomous driving and connected vehicles), smart grid (e.g., microgrids and renewables integrations), and vehicle-traffic-grid-building interactions. A wealth of optimization methods has been proposed to devise sophisticated energy management strategies for tackling such challenges, including dynamic programming (DP) [2], equivalent consumption minimization strategy (ECMS) [3], Pontryagin's minimum principle (PMP) [4], convex programming [5], and model predictive control [6]. In component-level analysis, there are also plentiful modeling/control challenges surrounding energy storage systems (e.g., batteries, ultracapacitors, fuel cells, flywheels, and hybrid storage schemes), electric machine/motor drive systems, V2I/V2V communication, and power electronic circuits, and so forth. For example, diverse approaches have been utilized to establish high-fidelity battery/ultracapacitor models, including offline particle swarm optimization (PSO) algorithm, genetic algorithm, fractional-order calculus, and online extended Kalman filtering. A plethora of battery/ultracapacitor state estimation and charging control techniques has also been reported in the recent literature [7–9].

The main goal of this special issue is to provide a professional platform sharing timely, advanced solutions to modeling, simulation, control, optimization, and fault diagnosis/prognosis of sustainable transportation and power systems. Particular emphasis is placed on system-level/component-level modeling, optimization, control, and fault diagnosis/prognosis.

For this special issue, 84 submissions have been totally received. After rigorous review processes, 37 manuscripts have been ultimately accepted, which cover a broad range of key modeling/control problems in sustainable transportation and power systems. A brief snapshot of them is given below.

W. Song et al. proposed an intention-aware autonomous driving decision making algorithm for an uncontrolled

intersection scenario. Y. Ji et al. developed a feedback gating control policy for mitigating network congestion by adjusting signal timings of gating intersections, based on a macroscopic fundamental diagram. J. Heng et al. devised a hybrid power load forecasting method, which comprises empirical mode decomposition, Cuckoo search algorithm, and wavelet neural network. P. D. U. Coronado and H. Ahuett-Garza synthesized a rule-based control strategy for a dual motor propulsion system with a differential transmission. K. Hwang et al. reported an autonomous coil alignment system using fuzzy steering control for electric vehicles with dynamic wireless charging. Z. Chen et al. discussed an online energy management strategy of plug-in hybrid electric vehicles for prolongation of all electric range based on dynamic programming. H. Li et al. proposed an optimization scheduling model for wind power and thermal power with energy storage system considering carbon emission trading. S. Zhu et al. introduced a genetic algorithm-based micro-scale vehicle emission model. X. Gong et al. presented a bus travel time deviation analysis using automatic vehicle location data and structural equation modeling. J. Zhao et al. delivered a dynamic model of Kaplan turbine regulating system which is suitable for power system assessment. X. Wu et al. proposed a multistage CC-CV charging protocol for Li-ion batteries. C. Peng et al. presented a wind power and load forecasting method based on frequency control approach for a wind-diesel-battery hybrid power system. S. Salas-Duarte et al. evaluated a trapezoidal predictive controller for a four-wire active power filter for utility equipment of metro railway power-land substations. L. Pan and C. Zhang designed a high power density integrated charger for electric vehicles with active ripple compensation. C. T. Calafate et al. handled traffic flow classification issue in traffic management as a service. D. Wang et al. used polarity comparison principle to implement ultra-high-speed travelling wave protection of transmission line, based on empirical mode decomposition. C. Lu et al. investigated intelligent ramp control for incident response by means of Dyna-Q architecture. Y.-N. Zhao et al. proposed an entropy-cost function evaluation method for unmanned ground vehicles. H. Wang et al. presented dynamic modeling and control strategy optimization of a hybrid electric tracked vehicle. T. Zhao et al. devised a robust online SOC estimator for a lithium-ion battery pack based on error sensitivity analysis. J. Liu et al. built a modified quasi-steady state model of DC system in transient stability simulation under asymmetric faults. K. Yu et al. assessed model predictive energy management for connected hybrid electric vehicles. T. Chen et al. conducted a loss prediction and thermal analysis of surface-mounted brushless AC PM machines for electric vehicle application considering driving duty cycle. Y. Li et al. presented a geometric-process-based battery management optimization policy for an electric bus. H. Tao et al. constructed a small-signal model of marine electromagnetic detection transmitter controlled-source circuit. J. Wang et al. studied the effect of the integrated service mode and travel time uncertainty on taxis network equilibrium. J. Gao and H. He compared several nonlinear filtering methods for SOC estimation of a $Li_4Ti_5O_{12}$ Li-ion battery. P. Jiang and Q. Dong proposed a new hybrid model for wind speed prediction using an intelligent optimization algorithm and a data denoising method. F. Chang et al. elucidated fault characteristics and control strategies of multiterminal high voltage direct current transmission based on modular multilevel converter. Z. Shi et al. presented an improved macro model of traffic flow with the consideration of ramps and numerical tests. S. Toosi et al. studied operational modes and control strategies of a multidirectional MC for battery-based systems. X. Meng et al. presented a hybrid MMC topology with DC fault ride-through capability for MTDC transmission system. W. Yang et al. proposed a multiperiod vehicle lease planning policy for urban freight consolidation network. X. Li et al. evaluated robustness of some SOC estimation algorithms for EV Li-ion batteries against modeling errors and measurement noise. C. Piao et al. proposed a Li-ion battery cell-balancing algorithm for battery management system based on real-time outlier detection. C. Chu et al. presented a temporal-spatial analysis of traffic congestion based on modified CTM. C. Zhang et al. examined an improved adaptive control of static synchronous compensator in power systems.

Acknowledgments

References

[1] X. Hu, C. M. Martinez, and Y. Yang, "Charging, power management, and battery degradation mitigation in plug-in hybrid electric vehicles: a unified cost-optimal approach," *Mechanical Systems and Signal Processing*, 2016.

[2] L. Li, C. Yang, Y. Zhang, L. Zhang, and J. Song, "Correctional DP-based energy management strategy of plug-in hybrid electric bus for city-bus route," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 7, pp. 2792–2803, 2015.

[3] C. Musardo, G. Rizzoni, Y. Guezennec, and B. Staccia, "A-ECMS: an adaptive algorithm for hybrid electric vehicle energy management," *European Journal of Control*, vol. 11, no. 4-5, pp. 509–524, 2005.

[4] N. Kim, S. Cha, and H. Peng, "Optimal control of hybrid electric vehicles based on Pontryagin's minimum principle," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1279–1287, 2011.

[5] X. Hu, J. Jiang, B. Egardt, and D. Cao, "Advanced power-source integration in hybrid electric vehicles: multicriteria optimization approach," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7847–7858, 2015.

[6] H. Borhan, A. Vahidi, A. M. Phillips, M. L. Kuang, I. V. Kolmanovsky, and S. Di Cairano, "MPC-based energy management of a power-split hybrid electric vehicle," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 3, pp. 593–603, 2012.

[7] X. Hu, R. Xiong, and B. Egardt, "Model-based dynamic power assessment of lithium-ion batteries considering different operating conditions," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 3, pp. 1948–1959, 2014.

[8] W. Waag, C. Fleischer, and D. U. Sauer, "Critical review of the methods for monitoring of lithium-ion batteries in electric and hybrid vehicles," *Journal of Power Sources*, vol. 258, pp. 321–339, 2014.

[9] P. Keil and A. Jossen, "Charging protocols for lithium-ion batteries and their impact on cycle life—an experimental study with different 18650 high-power cells," *Journal of Energy Storage*, vol. 6, pp. 125–141, 2016.

Research Article

A Hybrid Forecasting Model Based on Empirical Mode Decomposition and the Cuckoo Search Algorithm: A Case Study for Power Load

Jiani Heng,[1,2] Chen Wang,[2] Xuejing Zhao,[2] and Jianzhou Wang[1]

[1]*School of Statistics, Dongbei University of Finance and Economics, Dalian, Liaoning 116025, China*
[2]*School of Mathematics and Statistics, Lanzhou University, Lanzhou, Gansu 730000, China*

Correspondence should be addressed to Chen Wang; chenwang15@lzu.edu.cn

Power load forecasting always plays a considerable role in the management of a power system, as accurate forecasting provides a guarantee for the daily operation of the power grid. It has been widely demonstrated in forecasting that hybrid forecasts can improve forecast performance compared with individual forecasts. In this paper, a hybrid forecasting approach, comprising Empirical Mode Decomposition, CSA (Cuckoo Search Algorithm), and WNN (Wavelet Neural Network), is proposed. This approach constructs a more valid forecasting structure and more stable results than traditional ANN (Artificial Neural Network) models such as BPNN (Back Propagation Neural Network), GABPNN (Back Propagation Neural Network Optimized by Genetic Algorithm), and WNN. To evaluate the forecasting performance of the proposed model, a half-hourly power load in New South Wales of Australia is used as a case study in this paper. The experimental results demonstrate that the proposed hybrid model is not only simple but also able to satisfactorily approximate the actual power load and can be an effective tool in planning and dispatch for smart grids.

1. Introduction

In a power system, the short-term power load forecasting is very important for the stable operation of the system. Accurate forecasting is a guarantee in the development of preventive maintenance plans, which include generator safeguards, power system reliability estimation, and scheduling dispatch [1, 2]. High-accuracy power load forecasts improve the economic and social benefits of power grid management, which reduce generation costs, improve the security of power systems, and help administrators develop optimal plans. Moreover, accurate load forecasting is crucial in forecasts of the power price in power markets [3]. Therefore, developing power load forecasting techniques to achieve accurate, simple, and fast load forecasts is necessary. Thus far, many short-term power load forecasting methods have been proposed, and these methods can be mainly divided into three categories: conventional methods, modern forecasting methods, and hybrid forecasting methods. Conventional methods include multiple linear regression analysis [4, 5], time series

[6, 7], state space models [8], general exponential smoothing [9], and knowledge-based methods. However, these methods cannot provide appropriate nonlinear mathematical relationships to express actual power loads. The primary modern forecasting methods are intelligent evolutionary algorithms [10, 11], expert systems [12, 13], neural networks [14–17], and fuzzy inference [18]. Intelligent algorithms and neural networks obtain good performance because of their clear patterns, easy implementation, and strong ability to address the problem. Hybrid forecasting methods, proposed to avoid the shortcomings that exist in individual forecasting methods, have become increasingly prevalent [19, 20]. A detailed introduction of the three categories is given below.

The deduction processes of traditional forecasting methods are rigorous, and most of them are based on traditional mathematics theories such as statistics, calculus, and modeling by subjective data analysis [21]. The main idea of trend extrapolation technology is to look for the trend of data changes, according to the trend equation, to forecast future data. The method is simple, and, especially for smooth power

load changes, it can achieve a good prediction effect. Its deficiency is that its precision is greatly influenced by the random load component [22]. The regression analysis method is often applied to short-term load forecasting [23]. This method has many advantages such as a simple principle and better quality of data which leads to better precision; however, the selection of the main factors affecting power load in the model is difficult as many factors that affect the forecasting accuracy are hard to quantify. This model is lacking any self-study capability, and the input variable and output variable cannot be revised automatically [24]. With years of development, the time series forecasting method has become a mature theory method and has been applied to power load forecasting [25]. The basic time series prediction models mainly include AR, MA, and ARMA [26]. Although the time series forecasting method has advantages such as only requiring a small volume of historical data and a small amount of calculation and the fast speed of its calculation, this method has certain limitations such as its inability to reflect the influence of meteorological factors and how its forecasting accuracy will decrease with the increase of the prediction step [27]. ANN [17] is a type of nonlinear simulation of the human brain information processing system with an intelligent processing process; for an inaccurate variation trend, this method also has a good ability to adapt, is able to grasp information and keep on learning, and has good knowledge reasoning and self-optimization [28]. An expert system is a computer system based on the knowledge of the programming approach, mainly a software system, and the main components of an expert system include the inference engine of the system, the expert knowledge base, the explain interface, and the knowledge acquisition module. An expert system is a program that has decision-making capabilities based on reasoned knowledge; however, this method is limited by whether the expert knowledge is complete [29]. The grey forecasting method is an important technique in grey theory, and it uses approximate differential equations to describe future tendencies for a time series [30]. The limitation of this method is that the greater the dispersion degree of data, the worse the forecasting accuracy. Although traditional forecasting methods and forecasting methods based on intelligent computing have their respective applications, it is difficult to achieve better results when using one of them by itself [31]. In the literature related to forecasting [31–34], the forecasting results are not quite as good with any single forecasting model. The primary reason is that single forecasting models cannot extract the complicated factors encountered in reality.

Due to the limitations of the forecasting capacity of a single model, it cannot always be optimal in all cases. In this paper, a novel hybrid model was developed with the hope of obtaining more accurate power load forecasting results. The proposed hybrid wind speed forecasting model can be grouped into four steps. Firstly, we used the empirical model decomposition technique, which represents a nonstationary data analysis technique, to reconstruct the original wind speed series. Secondly, a WNN model was employed to create the power load forecasting, and the parameters in the WNN

model were tuned by the CSA. The simulation results illustrate that the hybrid model is an effective method in power load forecasting. The main contributions of this paper are summarized as follows:

(1) The CSA algorithm is applied to choose the optimal initial weight in the WNN model, which always leads to unstable forecasting error.

(2) In the field of power load forecasting, the proposed hybrid model is manifested as a valid method with efficient computation and satisfactory forecasting accuracy.

(3) Considering the skewness and kurtosis of the forecasting accuracy distribution, the developed forecasting availability is proposed as an effective evaluation criterion for model selection in the power load forecasting field.

This paper is organized as follows. First, we outline the concept of models used in this paper, including empirical model decomposition, WNN, CSA, BPNN, GABPNN, and EMD-CSAWNN. Second, the modeling processes of the methods mentioned above are introduced. Simulation results are presented and analyzed. Finally, the overall conclusion is included.

2. Methodology

In this section, the required individual tools will be presented concisely, including the empirical model decomposition technique, BPNN, the WNN model, and the CSA and GA algorithms. Moreover, the proposed hybrid approach will be described in detail. In addition, the structure of the feedforward neural network will be confirmed.

2.1. Empirical Mode Decomposition. Empirical model decomposition was proposed by Karthikeyan and Kumar as an adaptive method for nonstationary time series analysis, and it is now widely used. It can be applied to any type of signal decomposition [35]. Thus, it has obvious advantages in processing nonstationary and nonlinear series. The foundation of this technique is to decompose a time series into a finite set of several IMFs and a residue [36].

Definition 1. All IMFs are defined to satisfy the following conditions: (1) the number of local extreme points and the number of zero crossings must be equal or at least differ by only one; (2) the mean value of the upper envelope and lower envelope is zero.

Definition 2. The stoppage criterion determined is defined as

$$\mathrm{SD}_k = \frac{\sum_{t=0}^{T} \left| h_{1k-1}(t) - h_{1k}(t) \right|^2}{\sum_{t=0}^{T} h_{1k-1}^2(t)}. \quad (1)$$

The sifting process stops when SD_k is smaller than a pregiven value. The process of decomposition is over when the value of SD_k is between 0.2 and 0.3. Additional details of the
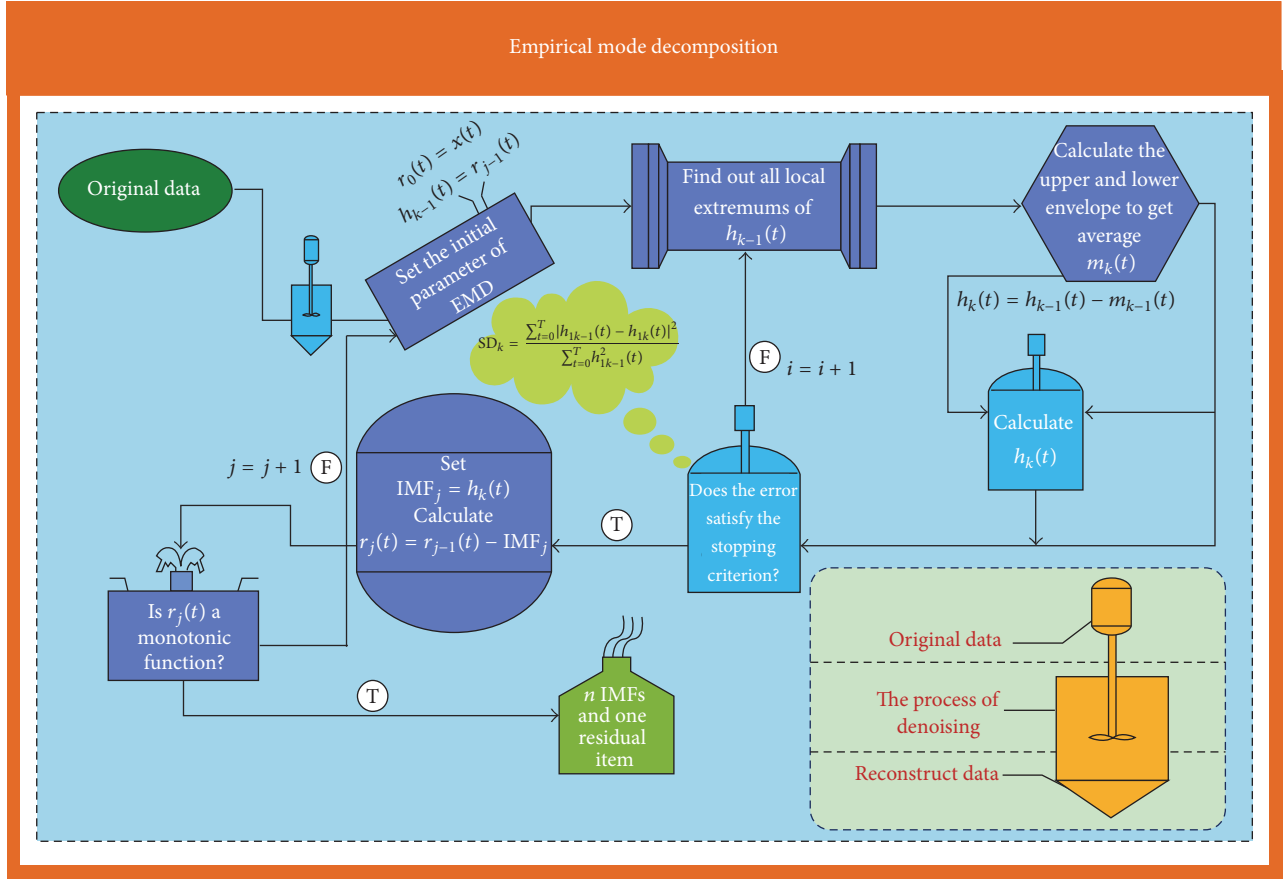
FIGURE 1: The construction of empirical model decomposition.

empirical model decomposition technique are illustrated in Figure 1.

2.2. Artificial Neural Network (ANN)

2.2.1. Confirmation of the Structure of the Network.
The ANN has received considerable attention as a powerful computational tool for forecasting in many fields since 1980. ANN models always outperform statistical models because of their ability to map the inputs onto outputs via simple computation [37]. We discuss the feed-forward neural network in this paper because of its strong learning ability and simple structure. The determination of the network structure is as follows [38].

Definition 3. Given an arbitrary continuous function $f : [0, 1]^n \rightarrow R^m$, f can be accurately approximated using a three-layer forward neural network realization. The first layer of the network is the input layer, containing n neurons. The middle layer is the hidden layer, containing $2n + 1$ neurons. The third layer is the output layer, which has m neurons.

Definition 4. Let ϕ be bounded continuous monotone in function, let K be the bounded compact subsets, and let $K \subset R^n$, $f(X) = f(x_1, x_2, \ldots, x_n)$ be a real continuous function in $f(X) \in K$. $\forall \varepsilon > 0$, $\exists$ integer N, real constant C_i, θ_i ($i = 1, 2, \ldots, N$) and W_{ij} ($i = 1, 2, \ldots, N$), such that $\widehat{f}_{ij}(x_1, x_2, \ldots x_n) = \sum_{i=1}^{N} C_i \phi(\sum_{j=1}^{n} W_{ij} x_j - \theta_i)$ satisfying $\max |\widehat{f}(x_1, \ldots, x_n) - f(x_1, \ldots, x_n)| < \varepsilon$. Note that, $\forall \varepsilon > 0$, $\exists$ a three-layer network structure, and the output function of the hidden layer is $\phi(x)$, the output function of the input and output layer is linear, and the total relationship of I/0 is $f(x_1, \ldots, x_n)$, such that $\max |\widehat{f}(x_1, \ldots, x_n) - f(x_1, \ldots, x_n)| < \varepsilon$.

Proof. Because $f(X)$ ($X = (x_1, \ldots, x_n)$) is a continuous function, $f(X) \in K$, $K \subset R^n$, assume that $f(X)$ is a bounded C^∞ function. Based on the Paley-Wiener Theorem [39], the Fourier transform $F(W)$ ($W = (W_1, \ldots, W_n)$) is a real analytic function of $f(X)$, $\forall$ a constant C_N, such that $|F(W)| \leq C_N(1 + |W|)^{-N}$. We define such a function:

$$I_A(x_1, \ldots, x_n) = \int_{-A}^{A} \cdots \int_{-A}^{A} \phi \left(\sum_{i=1}^{n} x_i W_i - W_0 \right) \frac{1}{(2\pi)^n \phi(1)} F(W_1, \ldots, W_n) \exp(iW_0) dW_0 \cdots dW_n,$$

$$I_{\infty,A}(x_1,\ldots,x_n) = \int_{-A}^{A} \cdots \int_{-A}^{A} \left[\int_{-\infty}^{\infty} \phi\left(\sum_{i=1}^{n} x_i W_i - W_0\right) \frac{1}{(2\pi)^n \phi(1)} F(W_1,\ldots,W_n) \exp(iW_0) \right] dW_0 \cdots dW_n,$$

$$J_A(x_1,\ldots,x_n) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(W_1,\ldots,W_n) \exp\left(i\sum_{i=1}^{n} x_i W_i\right) dW_1 \cdots dW_n.$$

$$(2)$$

Using an estimation of $F(\cdot)$, $\lim_{A\to\infty} J_A(x_1,\ldots,x_n) = f(x_1,\ldots,x_n)$ can be proved: that is, J_A is uniform convergence. Thus, $\lim_{A\to\infty} I_{\infty,A}(x_1,\ldots,x_n) = f(x_1,\ldots,x_n)$. That is, $\forall \varepsilon > 0, \exists A > 0$, such that $\max_{x\in R^n} |I_{\infty,A}(x_1,\ldots,x_n) - f(x_1,\ldots,x_n)| < E/2$. Moreover, for $A' > 0$, let $I_{A',A}(x_1,\ldots,x_n) = \int_{-A}^{A} \cdots \int_{-A}^{A} [\int_{-A'}^{A} \phi(\sum_{i=1}^{n} x_i W_i - W_0) \cdot (1/(2\pi)^n \phi(1)) F(W_1,\ldots,W_n) \cdot \exp(iW_0)] dW_0 \cdots dW_n$. For $\varepsilon > 0$, $A' > 0$ can be found such that $\max_{x\in K} |I_{A,A}(x_1,\ldots,x_n) - I_{\infty,A}(x_1,\ldots,x_n)| < \varepsilon/2$. That is to say, $f(X)$ is a uniform approximation by the integral $I_{A',A}(X)$, $I_{A',A}(X) \in K$. If $A_i > 0$ $(i = 1, 2, \ldots, m)$, K is a bounded subset, $K \subset R^n$, and $h_i(x_1,\ldots,x_m; t_1,\ldots,t_m)$ is a uniformly continuous function of $[-A_1, A_1] \times \cdots \times [-A_m, A_m] \times K$. There are functions:

$$H(t) = \int_{-A_1}^{A_1} \cdots \int_{-A_m}^{A_m} h(x_1,\ldots,x_m; t_1,\ldots,t_m) \, dx_1$$

$$\cdots dx_m,$$

$$H_N(t) = \frac{2A_1 \cdots 2A_m}{N^m} \times \sum_{k_1\cdots k_m=0}^{N-1} h\left(-A_1 \right.$$

$$\left. + \frac{k_1 \cdot 2A_1}{N} \cdots A_m + \frac{k_m \cdot 2A_m}{N} t_1 \cdots t_m \right).$$

$$(3)$$

$\forall \varepsilon > 0, \exists N_0, N_0$ is a natural number, when $N \geq N_0$, such that

$$\max_{l\in K} |H(t) - H_N(t)| < \varepsilon. \qquad (4)$$

Thus, $I_{A',A}(X)$ is a uniform approximate on K,

$$\phi\left(\sum_{i=0}^{n} x_i W_i - W_0\right) = \phi\left(\sum_{i=1}^{n} \frac{W_i x_i}{\delta} - W_0 + \alpha\right)$$

$$- \phi\left(\sum_{i=1}^{n} \frac{W_i x_i}{\delta} - W_0 - \alpha\right).$$

$$(5)$$

The formula above can be achieved by a three-layer neural network. Thus, $f(X)$ can be approximated by that neural network. $\qquad\square$

Definition 5. $\forall \varepsilon > 0$ and $f : [0,1]^n \to R^m$, there exists a three-layer structure that can approximate f in any square error precision of ε.

The definition above proves that, $\forall f : [0,1]^n \to R^m$, we can use a feed-forward neural network with a three-layer structure $n \times (2n + 1) \times m$ to approximate it accurately.

Thus, this part not only proves the existence of the mapping network but also demonstrates the network structure of the mapping. In summary, this paper adopts the three-layer neural network as the basis neural network.

2.2.2. BPNN. BPNN is a type of multilayer feed-forward neural network with an error back propagation learning process. The structure of BPNN is illustrated in Figure 2. Details of BPNN are introduced in [40].

2.2.3. WNN. WNN, a feed-forward network, is generally multilayer [41]. It is widely applied in signal processing because of its advantages of the localization property and generalization ability [42]. The structure of WNN is shown in Figure 2.

2.3. Neural Network Optimized by an Intelligence Algorithm. The intelligent optimization algorithm provides an efficient and powerful mathematical tool for optimizing the initial weights and thresholds of the ANN [43].

2.3.1. CSA. Cuckoo search is a heuristic swarm intelligence algorithm inspired by the behavior of the obligating brood parasitism of cuckoo species [44]. CSA is utilized in this paper for its stronger capability of global optimization [43].

Definition 6. To simulate the mode of cuckoo breeding, three idealized assumptions are presented, as follows: (1) each cuckoo selects nest randomly and dumps only one egg at a time, (2) the eggs with high quality will be carried over to the next generation, and (3) the available nest number n is fixed, and the probability of the host bird discovering the exotic egg is p_α.

Definition 7. The Lévy flight model simulates the process of the nest-seeking characteristic of cuckoo, and the update formula of the path and location is as follows:

$$X_i^{\text{iter}+1} = X_i^{\text{iter}} + \alpha \oplus L(\lambda). \qquad (6)$$

X_i^{iter} represents the location of nest i at generation iter, α is the step length-controlled factor, $\oplus$ is the point-to-point multiplication, and Levy(λ) obeys the Lévy distribution with a random search path of parameter λ:

$$L(\lambda) = 0.01 \frac{u}{|v|} \left(X_i^{\text{iter}} - X_*^{\text{iter}}\right). \qquad (7)$$
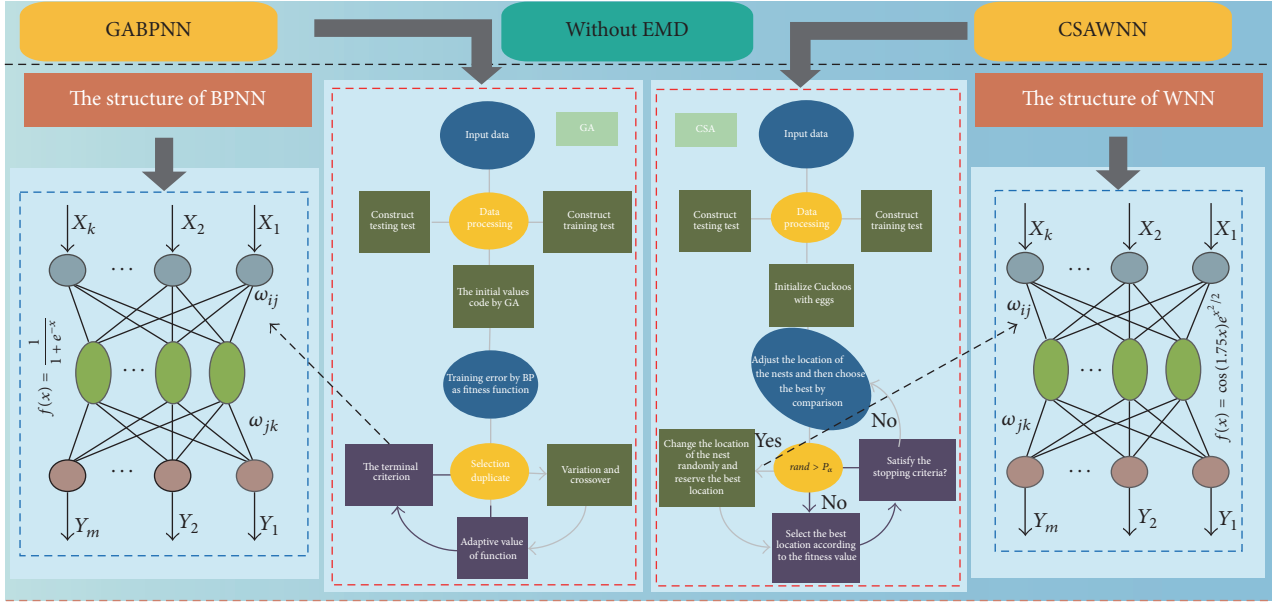
FIGURE 2: The construction of GABPNN and CSAWNN.

Here $u = t^{-\lambda}$, $1 < \lambda < 3$, u and v obey the normal distribution, $u \sim N(0, \sigma_u^2)$, $v \sim N(0, 1)$, $\sigma_u = [\Gamma(\lambda)\sin(0.5\pi(\lambda - 1))/2^{(\lambda-2)/2}\Gamma(0.5\lambda)(\lambda - 1)]^{1/(\lambda-1)}$, X_*^{iter} denotes the location of the best nest at generation iter, and Γ is the standard gamma function with unbounded variance and mean of the probability distribution.

2.3.2. GA. GA is a population-based optimization algorithm that simulates natural genetic mechanisms and biological evolutionism. It possesses a capacity for powerful global optimization [45]. The principle of GA relies on a random process, which is constituted by the processes of selection, crossover, and mutation [46]. The implementation process is shown in Figure 2.

2.4. GABPNN. The primary mechanism of GABPNN is composed of three parts: GA optimization, determination of the BPNN structure, and forecasting covered by BPNN [47]. The pseudo-code for GABPNN is as shown in Algorithm 1.

2.5. EMD-CSAWNN. In this paper, the proposed model, which incorporates the empirical model decomposition technique into the WNN model based on CSA, is adopted for short-term power load forecasts. Empirical model decomposition represents a self-adaptive decomposition technique to decompose short-term power load series into several IMFs and one residual item. WNN is adopted as a forecasting engine in the proposed approach because of its powerful approximation and high computation speed. Additionally, to avoid the deficiencies of WNN such as its unstable structure, CSA is used to initialize and determine the weights and thresholds of WNN, thereby imparting an outreach capacity to WNN. Figure 3 illustrates the general structure of the hybrid power load forecasting method. The pseudo-code for

the algorithm of the EMD-CSAWNN model is as shown in Algorithm 2.

3. Experiments and Evaluations

Applications of the proposed hybrid approach and five comparison models are shown in this subsection. All algorithms are operated on the given platform: 3.20 GHz CPU, 8.00 GB RAM, Windows 7, and MATLAB R2012a. Meanwhile, taking into account the randomness factors and to make sure the final results are reliable and independent from the initial weights, we carry out each ANN experiment 50 times and then take the average value.

3.1. Region Description and Data Collection. Australia has plentiful coal and wind resources across its coastline and land. A power load data set from NSW, which is the state with the largest population and highest levels of industrialization and urbanization in Australia, is employed to validate each model. In this paper, power load data are collected randomly from NSW from January 12, 2009, to March 8, 2009, which includes eight weeks of data. Among them, the data from January 12, 2009, to February 1, 2009, is used as a training set to obtain the appropriate model, and data from February 2, 2009, to March 1, 2009, is used as the testing set. The power load data from January 12, 2009, to February 1, 2009, together with their statistical measures, that is, minimum, maximum, mean, and standard deviations, are shown in Figure 4(c); the standard deviations are all above 1800, which implies that the power load series fluctuates significantly with the minimum/maximum of week one, week two, and week three, which are 6049.94/13518.06 (MWh), 6280.27/13326.47, and 6375.84/13096.37, respectively. This can be intuitively observed from the amplitude and frequency of the series fluctuation, which can change from very high to low values and vice versa.

```
(1)  GENERATE iter = 0  /* Calculate the fitness of each individual in the population. */
(2)  DO WHILE iter ≤ iter_max
(3)      iter = iter + 1  /* Record the best fitness values and the average fitness values. */
(4)      FOR EACH 1 ≤ i ≤ P DO  /* The process of the select operation. */
(5)          p_i^iter = f_i^iter(·)/ Σ_{i=1}^P f_i^iter(·)
(6)          FOR EACH 1 ≤ j ≤ s DO  /* The process of the crossover and mutation operations. */
(7)              w_kj^iter = w_kj^iter(1 − b) + w_lj^iter b  w_lj^iter = w_lj^iter(1 − b) + w_kj^iter b
```

$$
(8) \quad w_{ij}^{\text{iter}} = \begin{cases} w_{ij}^{\text{iter}} + \left(w_{ij}^{\text{iter}} - w_{\max}^{\text{iter}}\right) r_2 \left(1 - \text{iter}/G_{\max}\right) & r \geq 0.5 \\ w_{ij}^{\text{iter}} + \left(w_{\min}^{\text{iter}} - w_{ij}^{\text{iter}}\right) r_2 \left(1 - \text{iter}/G_{\max}\right) & r < 0.5 \end{cases}
$$

```
(9)  END FOR; END WHILE
(10) W_best = [ ω_jk  ω_ij  a_j  b_k ]  /* Initial weights and thresholds of BPNN by the obtained best individual. */
(11) x_k = (x_k − x_min)/(x_max − x_min), k ∈ [1, q]  /* The process of training data normalization. */
(12) iter = 0 /* Adjust the weights and threshold of BPNN according to the forecast error. */
(13) DO WHILE iter ≤ iter_max
(14)     iter = iter + 1
(15)     FOR EACH 1 ≤ i ≤ n DO
(16)         FOR EACH 1 ≤ j ≤ n_Hidden DO
(17)             H_j = f(Σ_{i=1}^n ω_ij x_i − a)  /* Calculate the outputs of the hidden layer. */
(18)             FOR EACH 1 ≤ k ≤ m DO
(19)                 O_k = Σ_{j=1}^l H_j ω_jk − b_k  /* Calculate the outputs of the output layer. */
(20)                 e_k = yn(k) − O_k  /* Calculate the forecast error in the output layer. */
(21)                 ω_ij = ω_ij + ηH_j(1 − H_j)x(i) Σ_{k=1}^m ω_jk e_k  /* Update the connection weights */
(22)                 ω_jk = ω_jk + ηH_j e_k
(23)                 a_j = a_j + ηH_j(1 − H_j) Σ_{k=1}^m ω_jk e_k  b_k = b_k + e_k  /* Update the threshold. */
(24) END FOR; END WHILE
(25) RETURN x̂_f = (x̂(q + 1), x̂(q + 2),…, x̂(q + p))
```

ALGORITHM 1: GABPNN.

Figure 4(b) exhibits 1008 power load data points from January 12, 2009, to February 1, 2009, divided into three groups, with 336 data points in every group. Because the power load data of NSW are collected once every half hour, each day includes 48 data points. On different days of the week, daily life and human economic production usually have different behaviors; thus, the characteristics of the load are different on different days. To minimize forecasting error as much as possible, we forecast the load of different days in the week separately. In this paper, the cycles of data division are seven days; the first three weeks of Monday data, January 12, January 19, and January 26, are employed to forecast the next Monday load on February 2, 2009. Accordingly, the data on January 13, January 20, and January 27 are employed to forecast the power load on February 3, 2009. The rest can be conducted in the same manner. The structures of the training and testing sets are illustrated in Figure 3(b).

3.2. Evaluation Metrics. Forecasting accuracy is an important criterion for evaluating a forecasting model. In this paper, the basic error calculation method is as follows:

$$
\text{MAE} = \frac{1}{T}\sum_{t=1}^T \left|\text{AE}_t\right| = \frac{1}{T}\sum_{t=1}^T |x_t - \hat{x}_t|. \tag{8}
$$

Here, T is the number of data points; the formula AE represents the absolute error between the observed value and the

forecasting value at time t, x is the observed value, and $\hat{x}_t$ is the forecasting value at time t. To avoid a positive or negative offset in forecasting error, solve the problem for which positive and negative forecasting error cannot be added, adding to the absolute value of the error, and take the average in the end. This error belongs to the comprehensive index in error analysis:

$$
\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^T \text{AE}_t^2} = \sqrt{\frac{1}{T}\sum_{t=1}^T (x_t - \hat{x}_t)^2}. \tag{9}
$$

RMSE is the square root of the mean square error. It also belongs to the comprehensive index in error analysis. Take the square of the absolute error AE; thus, the role of great values in the error will be strengthened, improving the sensitivity of the indicators, which is the prevailing reason in the error analysis:

$$
\text{MAPE} = \frac{1}{T}\sum_{t=1}^T \left|\frac{x_t - \hat{x}_t}{x_t}\right| \times 100\%, \tag{10}
$$

where the symbolic meaning is as above. The indicator is the average of the absolute error. The index is one of the comprehensive indexes in error analysis that usually occupies a very important position in the analysis and forecasting performance of the model.

(1) /* Initialize $SD_0 = 1$, $r_0(t) = x(t)$, $h_{k-1}(t) = r_{j-1}(t)$, $k = 1$, $j = 1$.*/
(2) **WHILE** $r_j(t)$ is not a monotonic function **DO**
(3)   **WHILE** $SD_k > 0.3$ **DO** /* Find all local maxima and minima of $h_{k-1}(t)$ by cubic spline.*/
(4)   /* Produce the upper and lower envelopes expressed as $e_{max}(t)$ and $e_{min}(t)$.*/
(5) $m_k(t) = [e_{max}(t) + e_{min}(t)]/2$, $h_k(t) = h_{k-1}(t) - m_{k-1}(t)$, $k = k + 1$
(6)   **END WHILE**
(7) $IMF_j = h_k(t)$, $r_j(t) = r_{j-1}(t) - IMF_j$, $j = j + 1$
(8) **END WHILE**
(9) $x_k = \sum_{i=2}^{n} IMF_i + r_n$, $k \in [1, q]$
(10) **GENERATE** iter = 0
(11) **DO WHILE** iter $\leq$ iter$_{max}$
(12)   iter = iter + 1   /* Generate a cuckoo egg X_i by taking a Lévy flight from a random nest.*/
(13)   **FOR EACH** $1 \leq i \leq n$ **DO**/*Calculate the fitness value $f_i(\cdot)$ and select the candidate. */
(14) $X_i^{iter+1} = X_i^{iter} + \alpha \oplus L(\lambda)$
(15)   **IF** $f^{iter}(\cdot) \leq f^{iter+1}(\cdot)$ **DO** /* Replace the worse location of the nest by the better one.*/
(16) $X^{iter+1} = X^{iter}$
(17)   **ELSE** $X^{iter+1} = X^{iter+1}$
(18)   rand = randn()
(19)   **IF** rand $> P_\alpha$ **DO** /* Discard the worst solution by P_α and produce a new solution.*/
(20) $X_i^{iter+1} = X_i^{iter} + r(\chi_j^{iter} - \chi_k^{iter})$
(21)   **ELSE** $X^{iter+1} = X^{iter+1}$
(22) **END IF; END FOR; END WHILE**
(23) $X_{best} = [\omega_{jk} \quad \omega_{ij} \quad a_j \quad b_j]$ /*Initial weights and thresholds of WNN by obtaining the best individual. */
(24) $x_k = (x_k - x_{min})/(x_{max} - x_{min})$, $k \in [1, q]$ /* The process of training data normalization. */
(25) iter = 0 /* Adjust the weights and threshold of WNN according to the forecast error. */
(26) **DO WHILE** iter $\leq$ iter$_{max}$
(27)   iter = iter + 1
(28)   **FOR EACH** $1 \leq i \leq n$ **DO**
(29)   **FOR EACH** $1 \leq j \leq n_{Hidden}$ **DO**
(30) $h(j) = h_j((\sum_{i=1}^{k} \omega_{ij} x_i - b_j)/a_j)$ /* Calculate the outputs of the hidden layer. */
(31)   **FOR EACH** $1 \leq k \leq m$ **DO**
(32) $y(k) = \sum_{i=1}^{n_H} \omega_{ij} h(i)$ /* Calculate the outputs of the output layer. */
(33) $e_k = yn(k) - y(k)$ /* Calculate the forecast error in the output layer. */
(34) $\omega_{i+1,k} = \omega_{i,k} + \Delta\omega_{i+1,k}$ /* Update the connection weights */
(35) $a_k = a_k + \Delta a_k$ $b_k = b_k + \Delta b_k$ /* Update the threshold.*/
(36) $\Delta\omega_{i+1,k} = -\eta(\partial e_k/\partial\omega_{i,k})$ $\Delta a_k = -\eta(\partial e_k/\partial a_k)$ $\Delta b_k = -\eta(\partial e_k/\partial b_k)$
(37) **END FOR; END WHILE**
(38) **RETURN** $\hat{x}_f = (\hat{x}(q + 1), \hat{x}(q + 2), \ldots, \hat{x}(q + p))$

ALGORITHM 2: EMD-CSAWNN.

To determine the degree of correlation between different forecasting model results with observed values, GRA [48] is employed in this paper.

Definition 8 (determining the grey relational coefficient). $\{x_0(i)\}_{i=1}^n$ is the reference sequence, $\{x_j(i)\}_{i=1}^n$ is the comparative sequence, $j = 1, 2, \ldots, m$, and the relational coefficient of x_0 and x_j in point k is represented as

$$\xi_j(k) = \frac{\min_j \min_k |x_0(k) - x_j(k)| + \rho \max_j \max_k |x_0(k) - x_j(k)|}{|x_0(k) - x_j(k)| + \rho \max_j \max_k |x_0(k) + x_j(k)|}. \quad (11)$$

Here, ρ is the distinguishing coefficient, and $\rho \in [0, 1]$, usually $\rho = 0.5$.

Definition 9 (grey relational degree). By focusing the degree of $\xi_j(k)$ at utter points, the algorithm on the grey relational degree is

$$r_i = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k). \quad (12)$$

Considering the generation capacity of the proposed hybrid model, four statistical indices are employed as evaluation metrics to measure the forecasting accuracy, MAE, RMSE, MAPE, and GRA. MAPE, MAE, and RMSE measure the mean performance, and GRA illustrates how well the forecasted data points fit the trend.

4. Results and Analysis

The experiments were divided into three parts, Experiment 1, Experiment 2, Experiment 3, Experiment 4, and
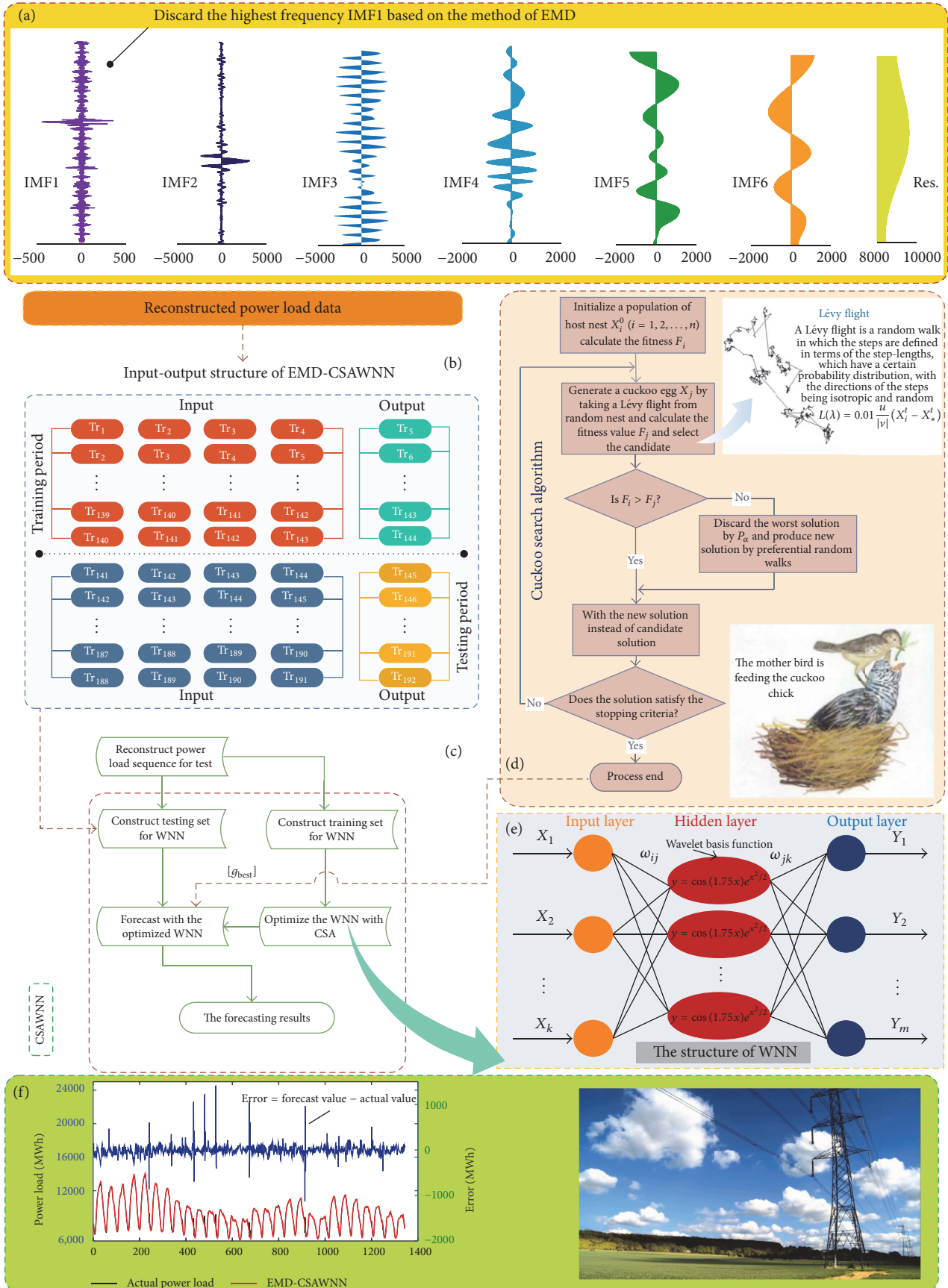
FIGURE 3: The construction of EMD-CSAWNN. (a) IMFs and the residual decomposed by empirical model decomposition. (b) The structure of the training and testing sets of the neural network. (c) The procedure of WNN combined with CSA. (d) The procedure of CSA. (e) The structure of WNN. (f) The forecasting sequence and error by using the EMD-CSAWNN model.

(a)

(b)

| Time | Power load of New South Wales from Jan. 12 to Feb. 1, 2009 | | | |
	Mean (MWh)	Standard deviation	Min (MWh)	Max (MWh)
Week one	9115.66	1895.18	6049.94	13518.06
Week two	9753.53	2017.32	6280.27	13326.47
Week three	9753.53	1854.42	6375.84	13096.37

(c)
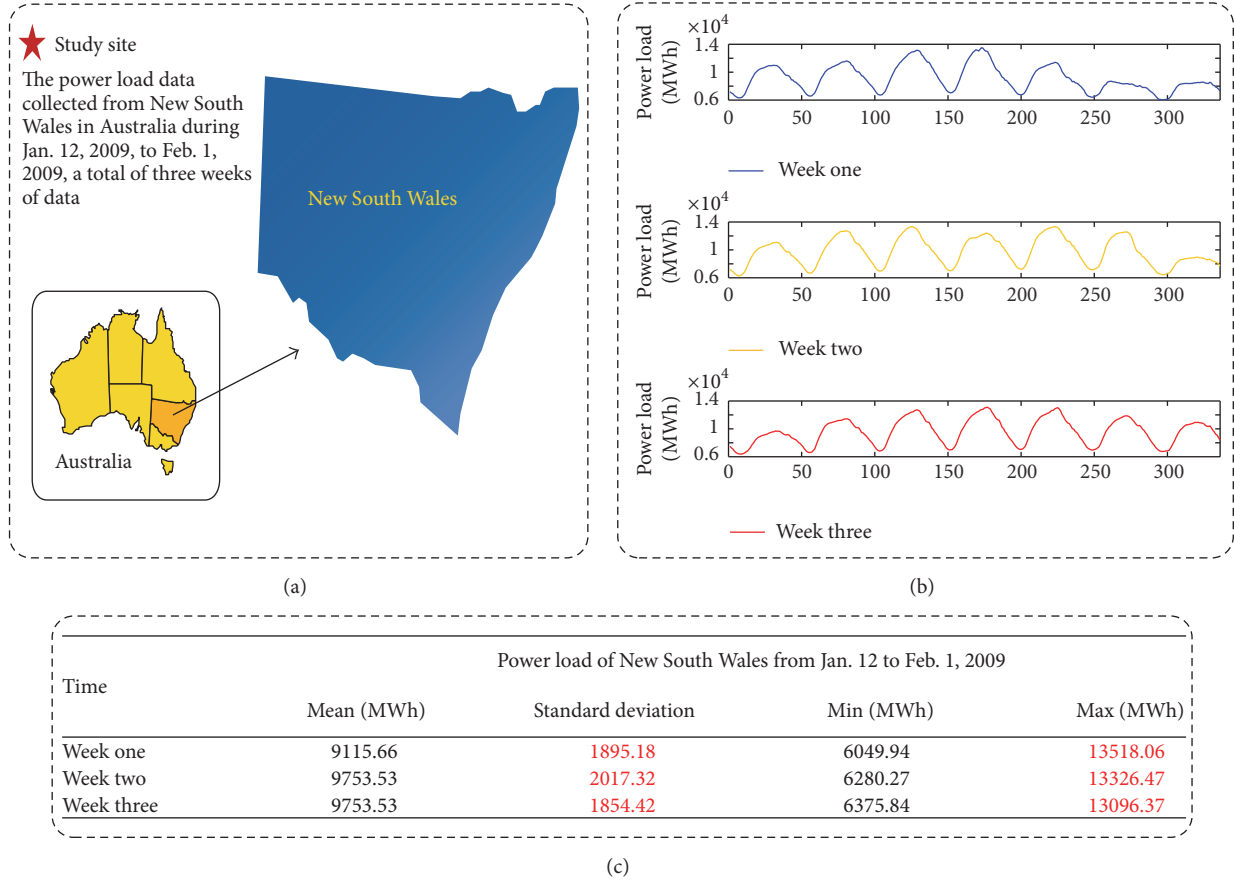
FIGURE 4: Description of observations in New South Wales of China. (a) Location of the study site. (b) The original power load series from January 12 to February 1, 2009. (c) The statistical measures for the power load.

Experiment 5. Experiment 1, the primary experiment, aims to compare the performance of different models in one day (February 2, 2009) and, meanwhile, to compare the ARIMA and ANN models. Experiment 2 aims to compare the performance of different models between weekdays and weekends. Experiment 3 aims to compare the performance of different models in 28 days (from February 2, 2009, to March 1, 2009). In order to further validate the hybrid model we proposed Experiments 4 and 5 which are studied in this paper. In addition, ENN (Elman neural network) and RBFNN (radial basis function neural network) are utilized in further comparison. Experiment 4 aims at globally testing the proposed hybrid model by using real power load data on Thursday, April 24, 2008, and Tuesday, April 29, 2008, which is selected from New South Wales of Australia randomly. Experiment 5 aims to prove the general applicability of hybrid model; thus, the power load data on Saturday, June 28, 2008, and Monday, June 30, 2008, from Victoria of Australia are chosen for forecasting.

Experiment 1. According to determination of the network structure, $n \times (2n + 1) \times m$ as the structure of the three-layer neural network will achieve better forecasting results. The network structure of the ANN model is 4-9-1 because of its four nodes in the input layer and one node in the output layer.

The iteration time is set to 100, the learning rate is set to 0.01, and the training requirement accuracy is set to 0.0004. To forecast the power load on Monday, February 2, the historical values from the Mondays of the first three weeks, January 12, January 19, and January 26, are chosen. Test results (MAE, RMSE, MAPE, and GRA) are presented in Tables 1 and 2, Figures 5, 6, and 7. Each individual model exhibits its best performance at a special time. For example, Figure 6 shows that BPNN provides the lowest MAPE value at 3:00 among all of the individual models, and CSAWNN yields the highest accuracy forecasting value from 1:00 to 3:00 among all of the individual models, whereas the maximum error is with the BPNN forecasting model on February 2, with a MAPE value of 10.12%. This result is due to the unstable initialization of the ANN. The result of the original data of empirical model decomposition is shown in Figure 3(a). The noise in the data is eliminated by using the empirical model decomposition technique; in this paper, the IMF1 is a high-frequency sequence with small values, which can be regarded as interference factors. As a result, the rest of the IMFs and the remainder term can be constructed as the training input of the CSAWNN model. Table 1 shows the experimental results for Monday of six types of forecasting models. The average values of MAPE for six models on February 2, 1.6% are 1.54%, 1.97%, 1.37%, 1.02%, and 1.94%, respectively; as shown in Table 2,

TABLE 1: Comparison of power load forecasting result by using different methods in Feb. 2.

Time	Actual value (MWh)	BPNN Forecasting Value	MAPE (%)	GABPNN Forecasting Value	MAPE (%)	WNN Forecasting Value	MAPE (%)	CSAWNN Forecasting Value	MAPE (%)	EMD-CSAWNN Forecasting Value	MAPE (%)	ARIMA Forecasting Value	MAPE (%)
0:00	8086.20	7267.78	10.12	7471.59	7.60	7828.93	3.18	7708.89	4.67	8033.14	0.66	8146.33	0.74
2:00	7132.07	7381.04	3.49	7193.94	0.87	7303.83	2.41	7150.74	0.26	7013.90	1.66	7298.17	2.33
4:00	7038.50	7023.50	0.21	7010.09	0.40	6470.25	8.07	6938.94	1.41	6938.97	1.41	6899.89	1.97
6:00	8803.29	8561.04	2.75	8611.39	2.18	8159.83	7.31	8612.24	2.17	8668.60	1.53	8619.45	2.09
8:00	10646.56	10778.68	1.24	10416.06	2.16	10890.92	2.30	10474.09	1.62	10707.41	0.57	10692.11	0.43
10:00	11822.68	11877.37	0.46	11811.56	0.09	11975.11	1.29	11880.45	0.49	11787.15	0.30	12002.49	1.52
12:00	12397.22	12329.59	0.55	12453.22	0.45	12328.77	0.55	12432.34	0.28	12326.00	0.57	12597.51	1.62
14:00	12824.82	12638.87	1.45	12918.86	0.73	12708.23	0.91	12929.78	0.82	12599.61	1.76	13091.61	2.08
16:00	13088.14	12810.33	2.12	13065.94	0.17	12903.89	1.41	13171.34	0.64	12830.13	1.97	12949.91	1.06
18:00	12001.48	11857.34	1.20	12090.61	0.74	12148.48	1.22	11500.36	4.18	11987.11	0.12	11776.37	1.88
20:00	10898.97	10827.46	0.66	10896.10	0.03	10967.20	0.63	10679.01	2.02	10778.06	1.11	10705.75	1.77
22:00	9398.37	9516.59	1.26	9665.37	2.84	9415.90	0.19	9706.03	3.27	9451.51	0.57	9446.46	0.51
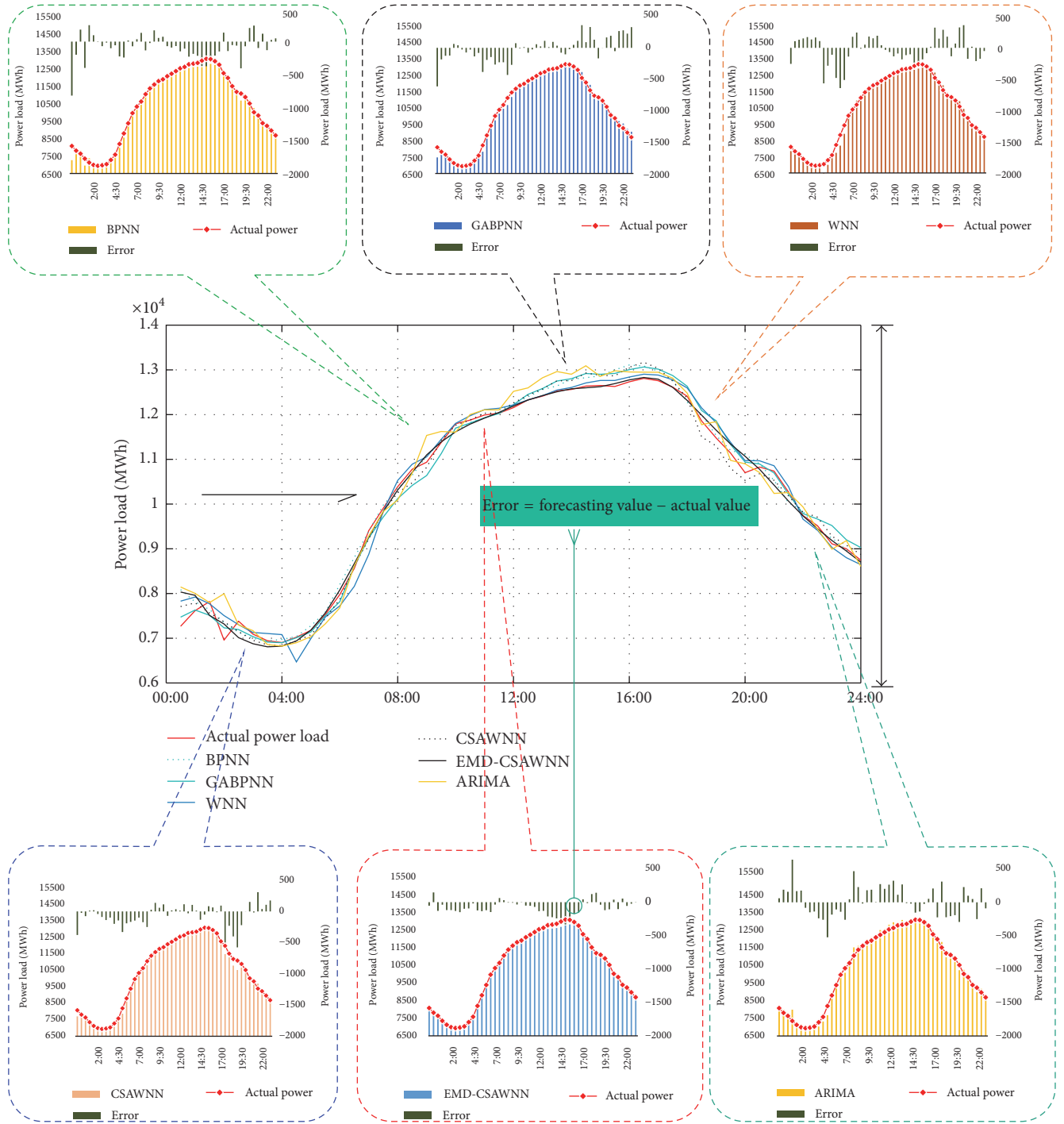
FIGURE 5: The forecasting results and actual values for February 2.

the MAPE afforded by the hybrid model decreased by 48.22% compared with the maximum average value of MAPE. In addition, it is shown that the value of MAPE offered by the hybrid model is more stable than that of the other proposed models, and the maximum value of MAPE is 2.54%. By comparing the hybrid model with the other models, it is shown that the hybrid model can provide high and stable forecasting accuracy.

Experiment 2. Figure 8 and Table 3 describe the comparison of six models with values on weekdays and weekends on different evaluation metrics. On weekdays, the best performance model is the hybrid model and the value of MAPE is 0.82%; on the contrary, the worst is the ARIMA model, whose value of MAPE is 1.48%. The MAPE offered by the hybrid model is 44.59% lower than that offered by the ARIMA model. On weekends, the value of MAPE offered by the hybrid model

Table 2: Comparison of power load forecasting result by using different methods in Feb. 2.

Model	MAPE (%)		
	Maximum	Minimum	Average
BPNN	10.12	0.05	1.6
GABPNN	7.6	0.03	1.54
WNN	8.07	0.12	1.97
CSAWNN	5.19	0.01	1.37
EMD-CSAWNN	2.54	0.03	1.02
ARIMA	8.67	0.03	1.94
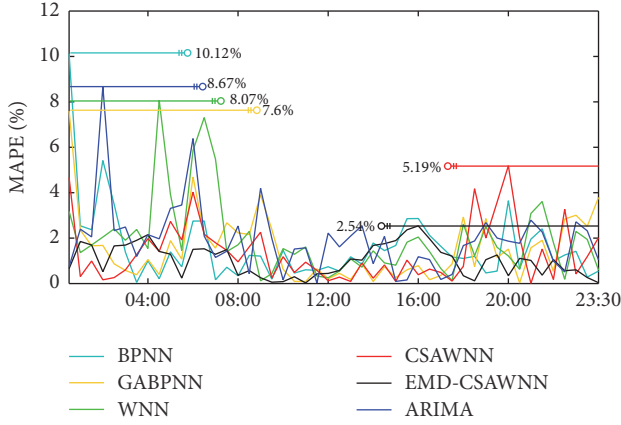
Figure 6: The BPNN, GABPNN, WNN, CSAWNN, EMD-CSAWNN, and ARIMA forecasting MAPE (%) for February 2.

is 0.68%, which outperforms all other models and is 49.25% lower than that of the worst model, ARIMA. Figure 8 indicates that, except for the criterion RMSE, the weekdays outperform the weekends when utilizing WNN and the criterion GRA, the weekdays outperform the weekends when utilizing the BPNN, GABPNN, WNN, and CSAWNN models, and the other performances on weekends are all better than those on weekdays when utilizing the six models. The values of MAPE offered by the six models on weekends are 1.23%, 1.05%, 1.27%, 1.13%, 0.68%, and 1.34%, which are 5.38%, 10.25%, 0.78%, 8.13%, 17.07%, and 9.46% lower than the corresponding weekday values, respectively. This illustrates that the forecasting results on weekends outperform those of weekdays.

Experiment 3. Figures 9 and 10 illustrate the forecasting results of Experiment 3. In comparison with the BPNN, WNN, GABPNN, CSWNN, and ARIMA models, the forecasting results offered by the hybrid model are more accurate. The forecasting results are shown in Figure 10, and detailed prediction results are shown in Table 4. The maximum values of MAPE offered by the six models are 1.28%, 1.14%, 1.27%, 1.2%, 0.78%, and 1.44%, respectively. The maximum value of MAPE offered by the ARIMA model is 4.4% over 28 days. Meanwhile, on all days of the test, the average values of MAE, MAPE, and RMSE offered by the hybrid model are all smaller than those of the other models. The average value of MAPE offered by the hybrid model over four weeks is 0.78%,

and the highest decrease is 39.06% compared with the other ANN models. This indicates that the hybrid model is an effective power load forecasting approach. The GRA result is shown in Table 5. In addition, on March 1, the GRA of GABPNN is higher than that of the hybrid model; on the remaining 27 days, the values of GRA offered by the hybrid model are higher than those of the other five models. According to the average value of GRA over the 28 days, the forecasting effects of all six forecasting models are increasing in the following order: ARIMA, BPNN, which is the WNN, CSAWNN, GABPNN, and EMD-CSAWNN, which concludes that the effect of the hybrid forecasting model is the best model among the six forecasting models.

The higher the power load forecasting accuracy, the lower the economic cost, which has actual economic significance [49]. As is illustrated in this case, the ANN optimized by the intelligence algorithm after denoising provides a better power load forecasting effect.

Experiment 4. The power load on Thursday, April 24, 2008, and Tuesday, April 29, 2008, from New South Wales of Australia is used to globally testing the proposed hybrid model. The results are shown as in Tables 6 and 7. To forecast the power load on Thursday, April 24, 2008, the historical values from the Thursdays of the first three weeks, April 3, April 10, and April 17, are chosen, respectively. To forecast the power load on Tuesday, April 29, 2008, the historical values from the Thursdays of the first three weeks, April 8, April 15, and April 22, are chosen, respectively. Test results (AE, MAE, RMSE, and MAPE) are presented in Tables 6 and 7 and part (a) of Figure 11. Table 6 indicates that EMD-CSAWNN has the highest accuracy forecasting results on Thursday, April 24, 2008; the maximum, minimum, and average MAPE values are 1.2355% at 2:00, 1.1836 at 14:00, and 1.2025%, respectively. The second-highest to sixth-highest accurate models are GABPNN, BPNN, RBFNN, CSAWNN, WNN, ENN, and ARIMA with average MAPE values of 1.6436%, 1.7795%, 1.8314%, 2.3373%, 2.4141%, 3.1381%, and 5.5872%, respectively. Table 7 indicates that EMD-CSAWNN still yields the highest accuracy forecasting value from among all of other models mentioned in this paper when forecasting power load on Tuesday, April 29, 2008; the maximum, minimum, and average MAPE values are 1.9844% at 14:00, 1.6033% at 12:00, and 1.7811%, respectively. According to the average MAPE value, CSANN is the second most accurate model, GABPNN is the third most accurate model, RBFNN is the fourth most accurate model, WNN is the fifth most accurate model, BPNN is the sixth most accurate model, ENN is the fifth most accurate model, and ARIMA is the sixth most accurate model with average MAPE values of 2.5827%, 2.8472%, 2.8473%, 3.1717%, 3.2725%, 3.5151%, and 3.7549%, respectively. As shown in Table 6, the average MAPE afforded by the hybrid model decreased by 78.48% compared with the maximum average MAPE value. In Table 7, the average MAPE afforded by the hybrid model decreased by 52.57% compared with the maximum average MAPE value. In addition, it is shown that the value of MAPE offered by the hybrid model is more stable than that of the other proposed models. By comparing
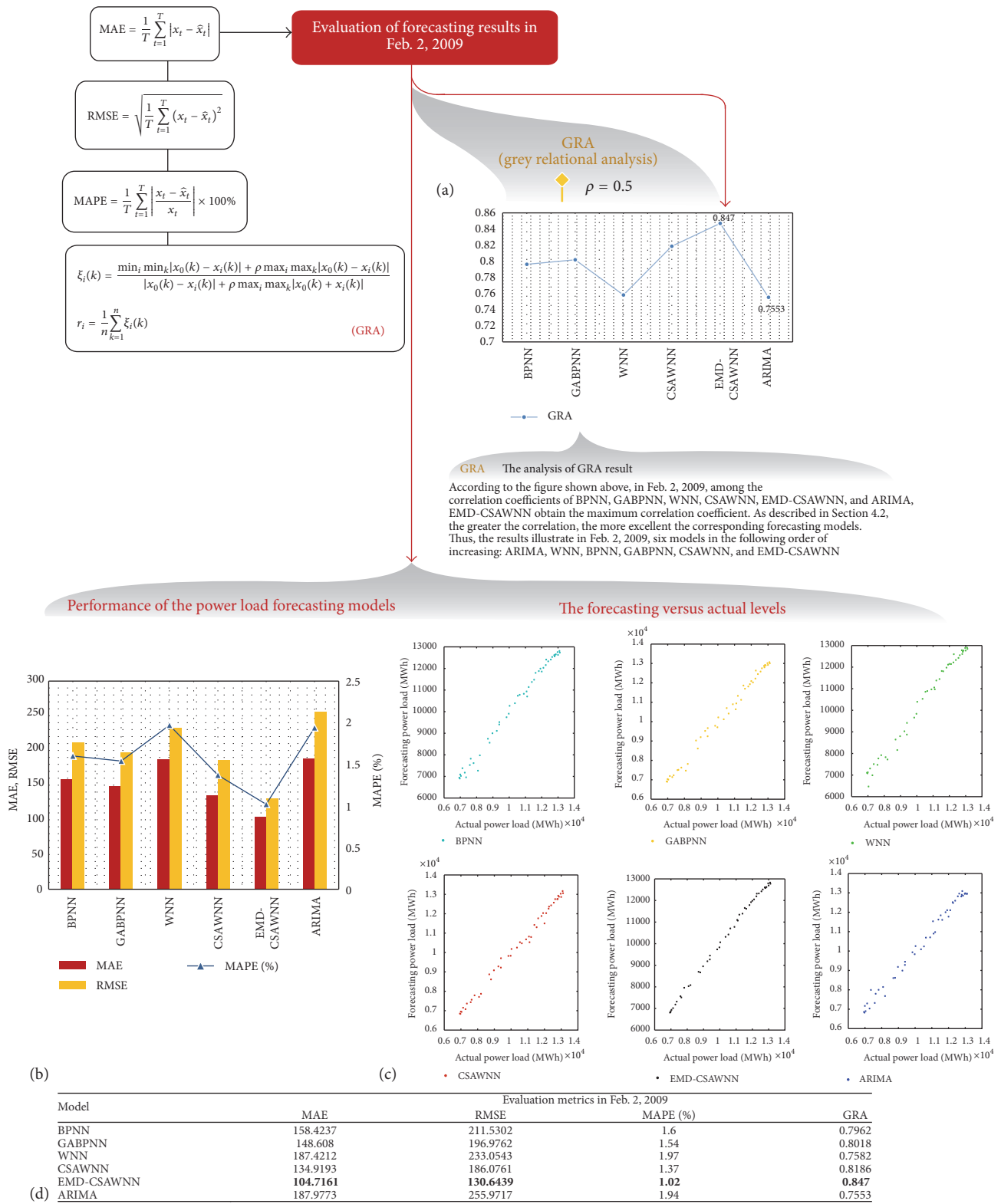
FIGURE 7: The comprehensive evaluation of forecasting models for February 2. (a) The comparison of GRA by using six different models. (b) The comparison of MAE, RMSE, and MAPE by using six different models. (c) The scatterplot of forecasting versus actual levels by using six different models. (d) The evaluation results of six different models. The value in bold is the best value of each evaluation index.

TABLE 3: Comparison of power load forecasting evaluation by using different methods between weekdays and weekends.

Model	Date	Evaluation criterion			
		MAPE (%)	RMSE	MAE	GRA
BPNN	Weekdays	1.30	164.97	118.03	0.8654
	Weekends	1.23	148.35	101.40	0.8600
GABPNN	Weekdays	1.17	154.94	106.05	0.8781
	Weekends	1.05	137.21	88.33	0.8779
WNN	Weekdays	1.28	162.09	114.33	0.8688
	Weekends	1.27	162.83	108.65	0.8581
CSAWNN	Weekdays	1.23	157.05	109.05	0.8737
	Weekends	1.13	138.86	94.96	0.8689
EMD-CSAWNN	Weekdays	0.82	101.56	72.77	0.9087
	Weekends	0.68	87.00	55.19	0.9148
ARIMA	Weekdays	1.48	184.23	128.98	0.8539
	Weekends	1.34	169.06	109.09	0.8548

GABPNN, BPNN, CSAWNN, WNN, ARIMA, ENN, and RBFNN, it is shown that the hybrid EMD-CSAWNN model is better than that of single models.

Experiment 5. The power load on Saturday, June 28, 2008, and Monday, June 30, 2008, from Victoria of Australia is used to further prove that the proposed hybrid model can improve the performance of power load forecasting in different cases. The historical values from the Saturday of the first three weeks, June 7, June 14, and June 21, are chosen in order to forecast the power load on Saturday, June 28, 2008. Moreover, the historical values from the Monday of the first three weeks, June 9, June 16, and June 23, are chosen in order to forecast the power load on Monday, June 30, 2008. The experimental results are presented in Tables 8 and 9 and part (b) of Figure 11. For Saturday (June 28, 2008) data, the average MAPE values for the BPNN, GABPNN, WNN, CSAWNN, ARIMA, ENN, RBFNN, and EMD-CSAWNN models are 3.3224%, 2.8656%, 3.2301%, 2.5248%, 3.7716%, 3.3890%, 3.1754%, and 1.7699%, respectively. The maximum MAPE values for the six models are 3.5467% at 16:00, 3.3061% at 18:00, 3.5751% at 4:00, 2.9527% at 16:00, 4.2174% at 0:00, 3.7591% at 20:00, 3.3889% at 10:00, and 2.0503% at 14:00, respectively, and the minimum MAPE values are 2.9644% at 20:00, 2.2710% at 12:00, 2.7591% at 20:00, 2.1818% at 20:00, 3.4718% at 4:00, 3.1379% at 18:00, 2.8405% at 22:00, and 1.5798% at 10:00, respectively. The differences between the maximum and minimum MAPE values for the models are 0.5823%, 1.0351%, 0.816%, 0.1936%, 0.7456%, 0.6212%, 0.5484%, and 0.4705%. For Monday (June 30, 2008) data, the average MAPE values for the BPNN, GABPNN, WNN, CSAWNN, ARIMA, ENN, RBFNN, and EMD-CSAWNN model are 3.2747%, 2.8415%, 3.0806%, 2.5422%, 3.7656%, 3.4113%, 3.1982%, and 1.7312%, respectively. The maximum MAPE values for the six models are 3.9915% at 22:00, 3.4117% at 2:00, 3.6561% at 6:00, 2.8144% at 0:00, 4.1783% at 22:00, 3.6738% at 8:00, 3.5262% at 0:00, and 2.0361% at 12:00, respectively, and the minimum MAPE values are 3.0113% at 20:00, 2.3376% at 22:00, 2.4491% at 16:00, 2.1398% at 20:00, 3.2795% at 0:00, 3.2235% at 4:00, 2.9466% at 22:00, and 1.4521% at 10:00, respectively. The differences

between the maximum and minimum MAPE values for the models are 0.9802%, 1.0741%, 1.207%, 0.6746%, 0.8988%, 0.4503%, 0.5796%, and 0.584%. Therefore, the proposed hybrid EMD-CSAWNN model is not only the most accurate but also the most stable of the investigated forecasting models.

5. Discussion

In this section, we discuss two important evaluation metrics, convergence speed and degree of certainty [50], offered by the GABPNN and CSAWNN models to determine a more practical forecasting model by considering reality factors such as forecasting stability and calculation time. The results illustrate that the CSAWNN model is more practical than the GABPNN model in forecasting power load. In addition, we propose forecasting availability to analyze and evaluate the quality of power load forecasting.

5.1. Convergence Speed. The computational complexity of evolutionary algorithms and swarm intelligence still remains a challenging issue; here, we use convergence speed as one of the evaluation metrics to examine the forecasting performance of GABPNN and CSAWNN. We obtain the computation time of the best fitness by analyzing the convergence speed of GA and CSA for use in comparative evaluation of optimization algorithm performance. However, the exploration and development are always two competing goals, and the conflict would exist between the convergence speed and forecasting accuracy. We define performance less than 10^{-5} as the convergence criteria.

We take the data from January 12, January 19, and January 26 as an example to illustrate the convergence speed of GA and CSA; Figure 12 shows the results of the comparison of evolutions among GA and CSA with different population sizes. We observed that the fitness values monotonically decrease as the iterations increase. In addition, when the iterations are less than 100, the larger the population size, the faster the convergence speed. We also observed that CSA has better convergence speed than GA. At iteration 20, the

TABLE 4: Comparison of power load forecasting evaluation by using different methods from Feb. 2 to Mar. 1.

Time	BPNN			GABPNN			WNN			CSAWNN			EMD-CSAWNN			ARIMA		
	MAE	MAPEE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)	RMSE
Feb. 2	334.27	4.22	211.53	304.59	3.85	196.98	278.96	3.55	233.05	280.07	3.55	186.08	190.32	2.41	130.64	344.95	4.40	225.97
Feb. 3	126.81	1.70	126.63	96.89	1.29	125.59	114.76	1.55	116.64	92.33	1.25	127.49	100.06	1.35	78.43	184.62	2.48	128.64
Feb. 4	115.99	1.67	117.88	106.40	1.53	121.43	121.62	1.75	135.78	89.46	1.30	116.67	100.59	1.46	85.54	97.21	1.40	134.84
Feb. 5	62.90	0.93	209.87	55.75	0.82	177.65	73.07	1.09	141.15	83.43	1.26	142.02	63.85	0.96	115.75	88.74	1.32	169.61
Feb. 6	71.29	1.07	205.84	57.27	0.85	165.60	70.96	1.06	139.43	81.86	1.24	185.20	57.25	0.85	98.40	88.15	1.32	184.44
Feb. 7	84.41	1.21	226.10	60.33	0.86	198.48	86.40	1.24	285.05	82.14	1.19	203.72	75.10	1.07	175.72	121.60	1.75	211.26
Feb. 8	130.54	1.71	123.10	101.31	1.33	145.93	142.19	1.85	123.60	128.98	1.70	120.03	65.08	0.87	67.65	262.37	3.47	124.33
Feb. 9	118.07	1.43	150.69	108.27	1.31	146.03	115.14	1.38	169.23	104.84	1.28	134.63	70.86	0.84	97.30	107.14	1.31	180.04
Feb. 10	107.55	1.21	161.91	96.10	1.06	139.39	128.24	1.48	156.85	123.78	1.39	142.14	61.45	0.69	96.35	145.55	1.70	182.63
Feb. 11	93.31	1.01	192.90	111.84	1.17	177.37	95.98	1.03	172.67	118.18	1.29	219.21	53.49	0.58	83.04	118.46	1.30	221.91
Feb. 12	109.95	1.14	226.38	104.93	1.08	194.03	100.88	1.04	209.32	130.04	1.34	221.86	37.33	0.39	98.76	89.20	0.91	236.75
Feb. 13	81.85	0.82	253.47	58.74	0.59	231.68	92.82	0.94	249.81	79.82	0.79	243.81	43.70	0.44	127.79	61.35	0.62	310.47
Feb. 14	64.30	0.62	141.32	52.72	0.51	124.09	73.70	0.71	152.33	69.55	0.66	124.22	35.49	0.35	66.87	58.42	0.56	176.15
Feb. 15	55.99	0.56	99.81	39.82	0.40	93.03	50.17	0.49	102.20	55.17	0.55	120.02	32.92	0.33	57.69	47.36	0.46	118.80
Feb. 16	79.41	0.77	157.56	66.18	0.63	151.21	71.83	0.66	174.28	61.84	0.60	147.61	40.62	0.39	151.27	63.28	0.60	183.95
Feb. 17	82.06	0.77	116.51	59.34	0.55	147.71	82.63	0.73	128.10	56.92	0.54	131.90	46.33	0.41	97.02	54.36	0.50	142.02
Feb. 18	77.09	0.71	134.92	60.71	0.57	123.87	72.46	0.68	123.32	46.97	0.45	129.04	57.85	0.53	91.79	66.87	0.62	148.78
Feb. 19	96.09	0.89	132.50	87.60	0.82	129.83	99.07	0.92	132.46	83.92	0.79	132.63	58.39	0.52	83.16	71.12	0.68	171.76
Feb. 20	96.07	0.88	146.61	90.49	0.84	140.51	93.47	0.87	132.77	68.44	0.66	104.60	87.30	0.77	93.94	52.59	0.49	148.44
Feb. 21	97.27	0.88	186.13	82.81	0.74	185.30	107.84	0.98	192.47	76.45	0.69	179.31	77.07	0.70	48.00	95.04	0.87	263.72
Feb. 22	131.13	1.28	111.39	109.41	1.11	105.24	128.56	1.26	117.27	93.69	0.95	93.12	61.20	0.61	53.84	138.42	1.38	105.10
Feb. 23	103.86	0.98	132.92	74.92	0.72	150.01	93.02	0.92	168.00	89.48	0.86	173.36	62.09	0.62	93.66	113.45	1.17	163.20
Feb. 24	156.24	1.59	122.25	146.41	1.49	148.16	132.83	1.34	139.12	146.53	1.45	117.79	67.67	0.70	94.62	135.60	1.37	150.05
Feb. 25	152.98	1.59	125.02	146.55	1.51	144.41	162.06	1.67	120.09	146.17	1.50	126.54	99.86	1.05	87.54	174.77	1.82	152.51
Feb. 26	177.81	1.94	128.21	153.29	1.66	118.80	183.35	2.00	157.75	188.12	2.06	141.24	69.22	0.76	82.76	188.15	2.05	192.72
Feb. 27	162.42	1.84	145.73	184.00	2.08	110.68	191.75	2.19	162.54	173.50	1.95	122.48	67.46	0.77	110.58	177.14	2.01	159.88
Feb. 28	100.06	1.18	147.69	120.10	1.41	122.11	106.96	1.25	150.10	96.13	1.12	132.03	55.77	0.64	90.29	121.94	1.43	174.16
Mar. 1	102.17	1.27	104.78	90.90	1.13	74.49	85.07	1.06	92.83	92.96	1.16	101.33	58.62	0.73	62.75	184.40	2.27	111.80
Maximum value	334.27	4.22	253.47	304.59	3.85	231.68	278.96	3.55	285.05	280.07	3.55	243.81	190.32	2.41	175.72	344.95	4.40	310.47
Minimum value	55.99	0.56	99.81	39.82	0.40	74.49	50.17	0.49	92.83	46.97	0.45	93.12	32.92	0.33	48.00	47.36	0.46	105.10
Average value	113.28	1.28	160.39	100.99	1.14	150.09	112.71	1.27	162.30	105.03	1.20	152.07	67.75	0.78	97.62	123.29	1.44	180.03

TABLE 5: Comparison of power load forecasting GRA by using different methods from Feb. 2 to Mar. 1.

Time	BPNN GRA	GABPNN GRA	WNN GRA	CSAWN GRA	EMD-CSAWNN GRA	ARIMA GRA
Feb. 2	0.7962	0.8018	0.7582	0.8186	0.8470	0.7553
Feb. 3	0.7587	0.7532	0.7756	0.7626	0.8121	0.7468
Feb. 4	0.7132	0.7077	0.7216	0.7429	0.7611	0.7076
Feb. 5	0.7225	0.7381	0.7672	0.7717	0.8032	0.7305
Feb. 6	0.7660	0.7991	0.8194	0.7800	0.8598	0.8027
Feb. 7	0.8555	0.8680	0.7952	0.8431	0.8839	0.8381
Feb. 8	0.7507	0.7184	0.7562	0.7504	0.8191	0.7541
Feb. 9	0.7622	0.7682	0.7495	0.7840	0.8352	0.7554
Feb. 10	0.7644	0.8003	0.7750	0.7993	0.8376	0.7573
Feb. 11	0.8297	0.8608	0.8479	0.8035	0.9109	0.8032
Feb. 12	0.7949	0.8509	0.8439	0.8074	0.9018	0.8169
Feb. 13	0.8262	0.8606	0.8396	0.8493	0.9050	0.8156
Feb. 14	0.7628	0.8064	0.7800	0.8067	0.8856	0.8058
Feb. 15	0.7392	0.7565	0.7716	0.7674	0.8262	0.7317
Feb. 16	0.8147	0.8298	0.8093	0.8291	0.8520	0.7859
Feb. 17	0.7927	0.7751	0.7656	0.7755	0.7964	0.7577
Feb. 18	0.7908	0.8114	0.7938	0.8043	0.8400	0.7791
Feb. 19	0.8254	0.8199	0.8227	0.8234	0.8756	0.7884
Feb. 20	0.7249	0.7719	0.7641	0.7901	0.7907	0.7340
Feb. 21	0.8555	0.8756	0.8473	0.8676	0.9334	0.8174
Feb. 22	0.7267	0.7600	0.7412	0.7624	0.8356	0.7547
Feb. 23	0.8258	0.8255	0.7912	0.7951	0.8670	0.7872
Feb. 24	0.7511	0.7630	0.7404	0.7642	0.8316	0.7363
Feb. 25	0.7932	0.7916	0.8259	0.8200	0.8331	0.7758
Feb. 26	0.7628	0.7983	0.7316	0.7402	0.8058	0.7063
Feb. 27	0.7440	0.7870	0.7203	0.7615	0.8257	0.7306
Feb. 28	0.8012	0.8331	0.7970	0.8011	0.8666	0.7607
Mar. 1	0.7681	0.8433	0.7974	0.7667	0.8214	0.7775
Maximum value	0.8555	0.8756	0.8479	0.8676	0.9334	0.8381
Minimum value	0.7132	0.7077	0.7203	0.7402	0.7611	0.7063
Average value	0.8673	0.8811	0.8692	0.8756	0.9127	0.8576

convergence of CSAWNN obtained the best speed in a population of 50, and the convergence of GABPNN at the iteration between 60 and 80, at the iteration 60, obtained the best speed.

5.2. Degree of Certainty. The forecasting results of optimization algorithm-NN are also usually different for each experiment because of the probability mechanism of the optimization algorithm. However, in the actual forecasting field, the future values are not known; thus, we cannot obtain which experiment will obtain the best result. Hence, we use these evaluation metrics to determine the randomness.

We defined the degree of certainty as

$$\mathrm{DC}\,(\sigma) = \frac{\sqrt{\sum_{i=1}^{m}\left(\sigma_i - \overline{\sigma}\right)^2}}{m},\qquad(13)$$

where m is the number of experiments, σ_i is the value of ith forecasting experiment on σ, and $\overline{\sigma}$ is the average value of all

n experiments. It is clear that a smaller DC can bring a higher degree of certainty.

The scatterplot in Figure 13 indicates the MAPE and GRA distributions of different results of 100 experiments for February 2 by, respectively, using GABPNN and CSAWNN. Although the minimum value of MAPE of GABPNN is smaller than that of CSAWNN and DC(σ) of CSAWNN, both are smaller than those of GABPNN. DC (MAPE) and DC (GRA) of CSAWNN are 0.0014 and 0.00016, and the DC (MAPE) and DC (GRA) of GABPNN are 0.0016 and 0.00019. Thus, CSAWNN is a better forecasting method than GABPNN in the actual forecasting field.

5.3. Forecasting Availability. Forecasting availability can be measured not only by the square sum of forecasting error but also by the mean and mean squared deviation of the forecasting accuracy. In certain practical circumstances, the skewness and kurtosis of the distribution of forecasting accuracy need

Table 6: Comparison of power load forecasting result by using different methods on Thursday, Apr. 24, 2008, from New South Wales.

	Time	0:00	2:00	4:00	6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	Average
BPNN	AE	-83.6894	-49.7095	42.2092	70.1483	-1.1418	81.1176	71.0641	-6.3979	-83.6375	70.2015	-74.2667	68.5518	-10.0773
	MAE	139.0489	107.2347	115.6315	153.4157	181.8889	201.3444	189.9489	186.2829	187.2521	163.6832	160.6981	127.2364	161.7255
	MSE	20126.81	11597.8	13753.41	23870.75	33776.23	41545.23	37131.99	34961.75	35301.52	27696.51	25909.11	16227.94	27452.27
	MAPE	1.7927	1.5953	1.6655	1.7118	1.8221	1.9574	1.8205	1.7860	1.8178	1.7268	1.7671	1.5004	1.7795
GABPNN	AE	-70.5662	-61.4181	14.0903	-5.1011	70.2694	-93.49	62.9971	-5.0262	-175.466	9.4920	87.4908	14.0313	3.7326
	MAE	133.1993	118.7066	120.9694	156.3609	168.3905	175.0163	159.27	175.0882	175.4656	142.1779	169.1021	135.8504	149.0696
	MSE	17861.21	14198.39	14907.39	24782.75	28896.06	30696.38	25765.8	30889.71	31099.72	20391.04	28608.41	18670.9	22988.39
	MAPE	1.7178	1.7651	1.7397	1.7443	1.687962	1.7025	1.5259	1.6792	1.6996	1.4971	1.8594	1.6012	1.6436
WNN	AE	-94.9552	11.0611	86.5661	109.4117	129.0671	-120.256	-133.083	-6.5916	-114.111	-121.37	114.9819	-97.3915	1.46853
	MAE	189.6161	163.6107	182.5346	226.7625	249.978	261.0484	274.6263	247.5517	241.1515	223.7553	230.5013	215.3979	219.2263
	MSE	36049.23	26940.16	33456.37	51459.93	62579.29	68421.2	75544.98	61478.9	58389.42	50583.32	53412.08	46583	49476.49
	MAPE	2.4456	2.4343	2.6249	2.5296	2.5013	2.5395	2.6342	2.3707	2.3368	2.3564	2.5336	2.5407	2.4141
CSAWNN	AE	-21.5793	74.9257	-92.6562	-121.433	-242.688	-4.1947	112.5843	-99.8578	108.9015	-117.485	9.2803	18.8441	0.6051
	MAE	187.2812	162.3261	159.9964	230.9195	242.688	234.3061	254.4321	244.3948	207.1397	244.4175	200.3291	213.4906	212.3467
	MSE	35548.74	26522.19	25927.18	53405.7	60342	55413.93	65290.47	60673.03	43149.33	60030.08	40539.89	45950.18	46910.12
	MAPE	2.4171	2.4145	2.3015	2.5755	2.4308	2.2807	2.4378	2.3433	2.0082	2.5742	2.2039	2.5175	2.3373
ARIMA	AE	-215.208	165.0423	7.8473	8.6388	-561.165	316.2328	-273.573	24.7401	313.4157	212.8736	210.3343	-0.2804	11.1686
	MAE	482.1711	333.4499	400.0869	495.5127	561.1652	628.2508	580.8741	607.6131	624.4892	463.577	453.6873	432.3235	507.6683
	MSE	234075.7	112032	1620067.4	257238.9	322221.9	396897.3	341886.4	369861.6	394541.8	216189.5	206839.9	190582.5	268746.7
	MAPE	6.2181	4.9614	5.7564	5.5391	5.6139	6.1119	5.5704	5.8302	6.0641	4.8836	4.9889	5.0989	5.5872
ENN	AE	-134.3503	-116.9724	26.1762	-9.5096	134.8336	-177.9787	121.4220	-9.2200	-334.7105	17.6777	166.4174	26.0920	7.1333
	MAE	254.0135	226.2179	230.6157	298.0677	321.2936	333.8439	304.7279	334.1142	334.7105	272.1785	321.8414	259.5614	284.6262
	MSE	64935.4893	51544.7643	54128.8887	89996.8881	105084.4741	116677.8252	94240.0696	112436.8781	131110.3356	74693.6074	103626.7463	68177.7863	83742.8823
	MAPE	3.2760	3.3639	3.3167	3.3253	3.2205	3.2476	2.9197	3.2042	3.2422	2.8661	3.5391	3.0595	3.1381
RBFNN	AE	-73.9591	-57.3006	1.4359	-115.7463	-11.4324	-103.9494	-69.3536	80.0282	145.0666	-90.0793	-80.4687	-4.8787	-31.7198
	MAE	122.8559	107.0627	125.1804	199.8651	161.9459	204.8784	190.0733	161.0021	145.0666	178.5049	152.1063	180.7679	160.7758
	MSE	15669.7874	11588.6074	15777.1354	40494.5143	26595.7994	43189.1618	37331.3576	26588.7852	21332.3809	31894.7300	23497.2824	33025.1939	27248.7280
	MAPE	1.7706	1.5996	1.4716	2.0627	1.5745	1.9708	1.8099	1.6528	1.5475	2.0890	2.0250	2.4029	1.8314
EMD-CSAWNN	AE	-5.8199	-40.1789	3.2302	-53.9029	0.3157	56.8411	-1.4835	-2.3791	-58.6926	-55.7161	-59.25	55.1392	-9.7616
	MAE	92.57402	83.04634	83.72504	106.387	120.3944	124.3859	128.0547	123.5589	124.0705	116.6259	111.4998	102.3642	109.1557
	MSE	8608.774	6911.941	7025.682	11341.53	14539.97	15513.36	16423.82	15278.06	15453.49	13635.16	12451.61	10503	12183.36
	MAPE	1.1938	1.2355	1.2045	1.1866	1.2037	1.2096	1.2272	1.1836	1.2016	1.2293	1.2265	1.2071	1.2025

Table 7: Comparison of power load forecasting result by using different methods on Tuesday, Apr. 29, 2008, from New South Wales.

	Time	0:00	2:00	4:00	6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	Average
BPNN	AE	4.3235	104.427	1.9326	−169.582	7.9721	−163.878	−189.155	−9.9777	−133.868	−23.2861	22.6951	−8.1229	−9.6227
	MAE	217.8733	215.1561	232.6085	296.2496	297.4614	329.2282	349.862	321.6289	310.8799	328.0564	307.5204	253.7871	293.1005
	MSE	47552.31	46537.35	54319.34	88796.61	89366.89	109897.5	122842.7	104957.9	98292.89	108873.5	96080.46	64784.3	88590.6
	MAPE	2.8131	3.2020	3.3602	3.3071	3.0174	3.2299	3.4431	3.2158	3.1271	3.4705	3.4071	3.0137	3.2725
GABPNN	AE	−128.819	116.37	−93.8387	5.9302	−25.0426	−157.356	2.0908	145.7456	−282.984	11.7947	129.7324	−11.5832	−16.1558
	MAE	260.8535	181.5997	191.8484	262.899	271.7663	306.0667	293.922	285.5415	282.9845	279.7753	244.2308	206.0163	254.9086
	MSE	68557.96	35322.65	39010.9	71358.5	76860.8	96621.05	91438.31	82785.86	8193717	80203.93	6019701	43169.73	68551.07
	MAPE	3.3665	2.7102	2.7624	2.9323	2.7553	3.0142	2.8718	2.8589	2.8581	2.9660	2.7122	2.4488	2.8472
WNN	AE	−102.792	−89.2578	95.7712	−95.0562	15.7301	−8.0755	−132.3	110.6793	127.7917	−302.974	170.0601	33.8715	23.1926
	MAE	237.6748	198.5173	194.0934	276.8971	362.0218	334.4987	301.0269	303.3078	290.4403	302.9738	315.3741	285.342	284.0601
	MSE	57399.3	40265.36	39263.54	79362.39	131879.1	114268.9	91690.29	94464.47	85736.09	9566719	101536.4	83146.33	85230.43
	MAPE	3.0684	2.9532	2.8117	3.0867	3.6747	3.2822	2.9631	3.0161	2.9222	3.2051	3.4881	3.3871	3.1717
CSAWNN	AE	2.7198	2.9868	−88.3877	122.0205	−143.045	−12.6708	−13.6472	4.8180	110.5905	−114.276	1.9674	−127.205	0.9651
	MAE	205.4517	170.8708	173.308	216.7274	244.0602	287.2786	252.2365	266.4932	268.5673	259.4214	244.6435	229.1077	231.6409
	MSE	42279.49	29713.52	30554.35	47370.17	60689.6	82774.86	63979.43	72549.97	72983.45	67791.63	602676	52777.79	55535.78
	MAPE	2.6528	2.5465	2.4986	2.4191	2.4748	2.8256	2.4824	2.6584	2.6966	2.7463	2.7141	2.7266	2.5827
ARIMA	AE	22.0044	278.4617	−1.9584	−222.701	−204.82	−201.995	−411.791	−53.4355	193.9193	−232.404	−15.6347	3.8278	−48.5181
	MAE	276.2362	278.4617	255.0419	411.837	374.2884	384.3065	411.7913	355.8697	329.2014	404.3898	374.5698	299.1434	336.7285
	MSE	78244.45	78470.78	65671.56	170074.2	144274.5	151190.6	173159.5	133331.2	112870.9	165704.4	141249.1	90175.37	118788.7
	MAPE	3.5681	4.1448	3.6876	4.5914	3.8021	3.7762	4.0691	3.5361	3.2986	4.2747	4.1591	3.5501	3.7549
ENN	AE	−122.3804	−106.1050	112.1774	−124.5511	11.2302	−5.7063	−154.8530	138.7430	150.4820	−334.3572	13.6915	25.3581	23.7370
	MAE	266.7030	226.0328	225.6456	309.9669	380.8511	365.7942	342.4249	342.0588	332.0910	334.3572	337.4444	308.6425	314.8147
	MSE	71574.4601	51519.8987	51615.5316	97406.4037	145424.6816	135137.5738	117847.8733	118529.0844	111042.3847	113745.3418	115039.8651	96156.4770	102413.9259
	MAPE	3.4432	3.3631	3.2647	3.4560	3.8651	3.5920	3.3700	3.4068	3.3416	3.5383	3.7352	3.6650	3.5151
RBFNN	AE	−128.8194	116.3700	−93.8387	5.9302	−25.0426	−157.3556	2.0909	145.7456	−282.9845	11.7948	129.7324	−11.5832	−16.1558
	MAE	260.8535	181.5997	191.8484	262.8990	271.7663	306.0667	293.9220	285.5415	282.9845	279.7753	244.2308	206.0163	254.9086
	MSE	68557.9649	35322.6478	39010.8992	71358.4965	76860.7961	96621.0544	91438.3143	82785.8632	81937.1664	80203.9338	60197.0099	43169.7279	68551.0659
	MAPE	3.3665	2.7103	2.7625	2.9324	2.7553	3.0142	2.8718	2.8589	2.8582	2.9660	2.7123	2.4489	2.8473
EMD-CSAWNN	AE	71.8585	19.0263	119.4944	−0.2347	−85.8361	21.7152	162.9778	198.7374	12.5729	0.7221	7.0831	11.6529	13.8229
	MAE	135.0333	124.2513	119.4944	159.9286	166.286	188.3798	162.9778	198.7374	191.4925	164.075	155.0245	145.8072	159.2127
	MSE	18544.32	15861.56	14935.85	25684.91	28077.06	3638753	26760.25	39816.23	38064.54	28373.25	24954.2	21683.46	26609.38
	MAPE	1.7439	1.8477	1.7241	1.7831	1.6889	1.8577	1.6033	1.9844	1.9379	1.7341	1.7161	1.7319	1.7811

Table 8: Comparison of power load forecasting result by using different methods on Saturday, Jun. 28, 2008, from Victoria.

	Time	0:00	2:00	4:00	6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	Average
BPNN	AE	−74.2069	88.5026	164.3428	−94.2384	−108.445	99.3887	−131.822	−97.1293	132.6263	−9.9181	−199.556	95.9906	−25.4772
	MAE	161.8614	176.3894	164.3428	175.3241	219.8255	225.0625	233.3815	229.9635	234.7046	233.2327	199.5555	182.4864	206.2306
	MSE	26380.04	31503.46	27247.69	31227.59	48517.87	51693.03	54905.29	54093.61	55631.95	54814.4	39922.9	33457.23	43708.18
	MAPE	2.9997	3.3342	3.4049	3.1834	3.3696	3.3340	3.5174	3.4406	3.5467	3.1652	2.9644	3.0498	3.3224
GABPNN	AE	95.3661	152.9973	70.1585	95.8643	21.4256	−10.3011	65.8767	6.8887	−16.0091	−14.2224	97.4645	19.2111	3.9264
	MAE	163.6398	152.9973	142.3091	180.6853	192.6446	181.9683	151.09	198.873	168.6475	242.7444	202.9146	183.0895	178.1422
	MSE	27233.94	23592.27	20427.45	33187.61	37879.35	34166.35	23242.55	40431.09	28919.45	60044.79	41793.54	34130.68	33502.97
	MAPE	3.0324	2.8936	2.9494	3.2736	2.9262	2.7029	2.2710	3.0080	2.5461	3.3061	3.0207	3.0576	2.8656
WNN	AE	175.8896	27.8588	−7.9285	86.3077	139.9775	201.231	90.7231	39.9104	18.8402	−108.905	−114.363	193.1549	4.4674
	MAE	175.8896	156.0152	172.6755	164.069	207.6373	201.231	193.4483	205.2855	183.8711	259.0394	186.1753	193.1549	200.6168
	MSE	31845.68	25360.07	30586.03	27917.8	45875.7	42312.96	38650.67	43776.7	34700.63	69347.15	35949.54	38952.51	42677.08
	MAPE	3.2604	2.9507	3.5751	2.9766	3.1492	2.9988	2.9427	3.1425	2.7626	3.5188	2.7591	3.2385	3.2301
CSAWNN	AE	−44.652	−131.152	−63.0165	11.4112	−9.5339	−0.7664	171.5844	193.7814	7.9337	106.7022	3.6645	−69.7819	12.6984
	MAE	124.0692	131.1523	114.0833	146.3165	167.5085	165.3564	171.5844	193.7814	195.2323	191.7276	146.8918	141.5059	156.8164
	MSE	15804.77	17589.3	13292.63	21642.06	29014.79	27346.81	30261.31	37723.03	38246.53	36979.81	21623.75	20248.95	25425.16
	MAPE	2.2995	2.4781	2.3604	2.6579	2.5319	2.4803	2.5786	2.9203	2.9527	2.6113	2.1818	2.3631	2.5248
ARIMA	AE	18.5613	124.0827	−14.7831	111.8472	−20.6124	−19.7243	120.3048	26.1011	−146.313	−172.198	23.3302	145.1471	18.9775
	MAE	227.3289	204.722	167.4229	212.6136	230.4365	247.6076	230.858	253.9037	267.5722	296.6621	255.3748	241.6215	233.6259
	MSE	52394.21	43176.64	28355.68	45620.29	53698.66	62570.82	54191.74	65290.6	72666.23	90860.47	67238.24	59345.33	56855.39
	MAPE	4.2174	3.8741	3.4718	3.8504	3.5598	3.7094	3.4799	3.8433	4.0557	4.0217	3.7627	4.0433	3.7716
ENN	AE	−98.995	−95.8081	17.1172	6.0116	90.0902	−141.101	100.5493	4.7625	−229.624	18.1665	126.2839	11.7423	6.5262
	MAE	189.6312	190.0733	171.6852	196.7854	231.2569	235.0732	210.1064	229.5182	229.6243	230.8097	254.2102	198.2688	210.274
	MSE	36109.8	36258.98	29924.18	39256.18	56148.35	56021.74	44222.47	53380.62	52793.31	53676.67	64753.52	39571.84	45444.62
	MAPE	3.5163	3.5974	3.5538	3.5618	3.4650	3.4901	3.1873	3.4500	3.4851	3.1379	3.7591	3.3164	3.3890
RBFNN	AE	−98.2027	−73.2415	57.3594	77.1316	9.4619	96.7952	101.7499	−14.9842	−97.2306	97.9996	−110.216	86.7962	−11.4032
	MAE	172.1148	156.3662	146.8716	171.0673	214.0821	228.8974	212.5963	212.2372	213.3772	228.0278	213.4688	169.8724	196.9359
	MSE	30159.71	24592.22	21954	29579.24	47567.12	53850.57	45459.14	45699.94	45988.64	52566.64	45607.02	28877.88	39869.92
	MAPE	3.1912	2.9543	3.0386	3.0942	3.2266	3.3889	3.2246	3.1832	3.2214	3.1122	3.1605	2.8405	3.1754
EMD-CSAWNN	AE	35.0621	51.5938	41.9313	−56.3883	−10.8655	51.1997	−42.3959	84.1310	70.1839	82.8233	46.4874	−55.1894	6.6501
	MAE	91.3676	100.3773	81.2863	92.1573	108.8744	104.7747	119.9606	137.012	110.6269	141.76	109.7464	115.5855	109.712
	MSE	8629.51	10214.74	6858.997	8885.95	12543.16	11063.31	14958.06	19196.88	12798.02	20368.17	12301.46	13873.76	12612.17
	MAPE	1.6951	1.8968	1.6898	1.6576	1.6298	1.5798	1.8118	2.0503	1.6639	1.9325	1.6329	1.9192	1.7699

Table 9: Comparison of power load forecasting result by using different methods on Monday, Jun. 30, 2008, from Victoria.

	Time	0:00	2:00	4:00	6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00	22:00	Average
BPNN	AE	−97.1358	−5.79918	−77.632	−69.5865	−5.12235	181.7938	−108.862	172.1967	−11.7716	−112.91	−1.05112	−5.29961	7.270652
	MAE	180.8076	174.7874	160.6262	161.5628	166.0493	181.7938	187.9272	172.1967	176.0733	215.9977	187.979	197.6079	181.9883
	MSE	32946.72	30612.98	26035.71	26290.49	27633.14	33390.49	35789.18	29752.65	31224.95	46849.69	35447.84	39567.68	33681.6
	MAPE	3.2431	3.2773	3.3784	3.3739	3.1979	3.2166	3.3433	3.0728	3.0901	3.2467	3.0113	3.9915	3.2747
GABPNN	AE	87.0447	182.023	62.5079	−12.5589	−92.6525	−84.4608	−25.4557	19.48	−22.7953	76.9831	78.4895	−6.6962	4.1017
	MAE	148.3886	182.023	116.5217	112.0699	155.8314	181.3435	153.1462	162.2208	182.3206	182.7003	175.2353	132.1649	158.2286
	MSE	23338.89	33529.04	14089.3	12900.14	24927.85	33303.2	24526.58	26867.37	34101.29	35242.19	31203.34	17935.79	26253.68
	MAPE	2.6633	3.4117	2.4520	2.3381	3.005	3.2113	2.7189	2.8955	3.1883	2.7492	2.8096	2.3376	2.8415
WNN	AE	−3.4921	−74.9511	−134.269	−80.2326	−142.363	64.7736	−111.042	−89.0593	139.566	14.1565	−82.5293	66.4498	−4.1007
	MAE	153.3954	185.6073	134.2689	174.9947	142.3635	174.6803	186.5582	193.4015	139.566	212.2396	156.4929	167.1148	171.6629
	MSE	23561.14	35229.8	18364.93	31218.27	21459.47	33650.21	37245.82	38184.73	19503.57	47310.2	24810.93	28778.48	31298.91
	MAPE	2.7532	3.4749	2.8191	3.6561	2.7458	3.0828	3.3176	3.4628	2.4491	3.1823	2.5057	2.9566	3.0806
CSAWNN	AE	−5.0573	−3.8213	1.3263	41.3898	54.9301	−66.0719	14.6333	−3.6448	149.7019	80.7116	3.7387	−73.6169	8.0382
	MAE	156.8892	136.6022	125.5879	113.7244	116.9543	143.8786	140.1588	139.9815	149.7019	159.8907	133.4934	137.4735	141.3812
	MSE	24786.49	18999.83	15803.74	13261.64	13725.91	20966.75	20021.99	19790.28	22684.1	25822.94	17849.43	18934.4	20481.5
	MAPE	2.8144	2.5621	2.6401	2.3753	2.2519	2.5478	2.5001	2.5018	2.6261	2.4020	2.1398	2.4302	2.5422
ARIMA	AE	14.7023	24.8103	26.0758	96.5552	−103.598	67.3243	1.5328	−12.8999	229.6626	254.5464	−133.048	125.5915	−0.7485
	MAE	183.057	217.205	158.1086	187.75	186.6591	198.3604	205.3784	213.5527	229.6626	254.5464	240.3672	236.3162	209.7854
	MSE	34965.65	47841.28	26107.02	35347.8	35329.16	41301.58	42893.69	46719.97	54654.13	65818.94	58199.06	55984.16	45718.98
	MAPE	3.2795	4.0777	3.3209	3.9201	3.5982	3.5138	3.6743	3.8162	4.0242	3.8204	3.8526	4.1783	3.7656
ENN	AE	−88.6146	−85.0312	75.6176	−69.7223	4.5320	−0.8129	−87.546	82.3113	88.4143	−227.985	3.5052	12.0972	8.3585
	MAE	186.9891	175.8838	153.515	161.1809	190.7898	196.0601	185.4341	186.5996	187.2131	227.9848	223.304	199.3213	189.5443
	MSE	35060.02	31090.34	23853.32	26154.44	36507.3	38584.26	34554.06	35020.38	35263.23	52336.47	50161.92	39941.99	36505.16
	MAPE	3.3574	3.2973	3.2235	3.3669	3.6738	3.4690	3.3024	3.3301	3.2811	3.4287	3.5763	3.5237	3.4113
RBFNN	AE	−98.586	97.2085	−72.7937	0.0161	−10.8099	−94.714	−1.0184	92.2613	−182.302	5.0740	100.423	−5.2195	−12.4652
	MAE	196.4346	166.1881	149.5233	155.7541	163.1654	186.8753	180.256	179.795	182.3024	217.8852	194.234	166.6086	177.8178
	MSE	38661.36	28328.97	22697.24	24518.37	27001.22	35355.86	32968.16	32671.75	33413.99	47850.86	37890.5	27884.44	32243.32
	MAPE	3.5262	3.1117	3.1447	3.2520	3.1402	3.3037	3.2137	3.2056	3.2051	3.2732	3.1130	2.9466	3.1982
EMD-CSAWNN	AE	−4.1652	−40.9803	−54.073	−28.5539	−49.4364	−46.9202	−1.3708	−51.2203	−61.0243	58.5634	44.2871	55.8282	6.1509
	MAE	99.0617	81.542	83.7323	70.3801	80.9749	82.0208	114.0354	111.8617	115.3834	129.6408	110.32	105.1976	96.6039
	MSE	10306.84	6696.25	7299.67	5046.59	6865.14	6815.57	13126.18	12711.91	13468.79	17077.42	12871.21	11278.86	9834.31
	MAPE	1.7771	1.5295	1.7617	1.4701	1.5594	1.4521	2.0361	2.0032	2.0227	1.9479	1.7703	1.8607	1.7312

↑ Best performance
↓ Worst performance
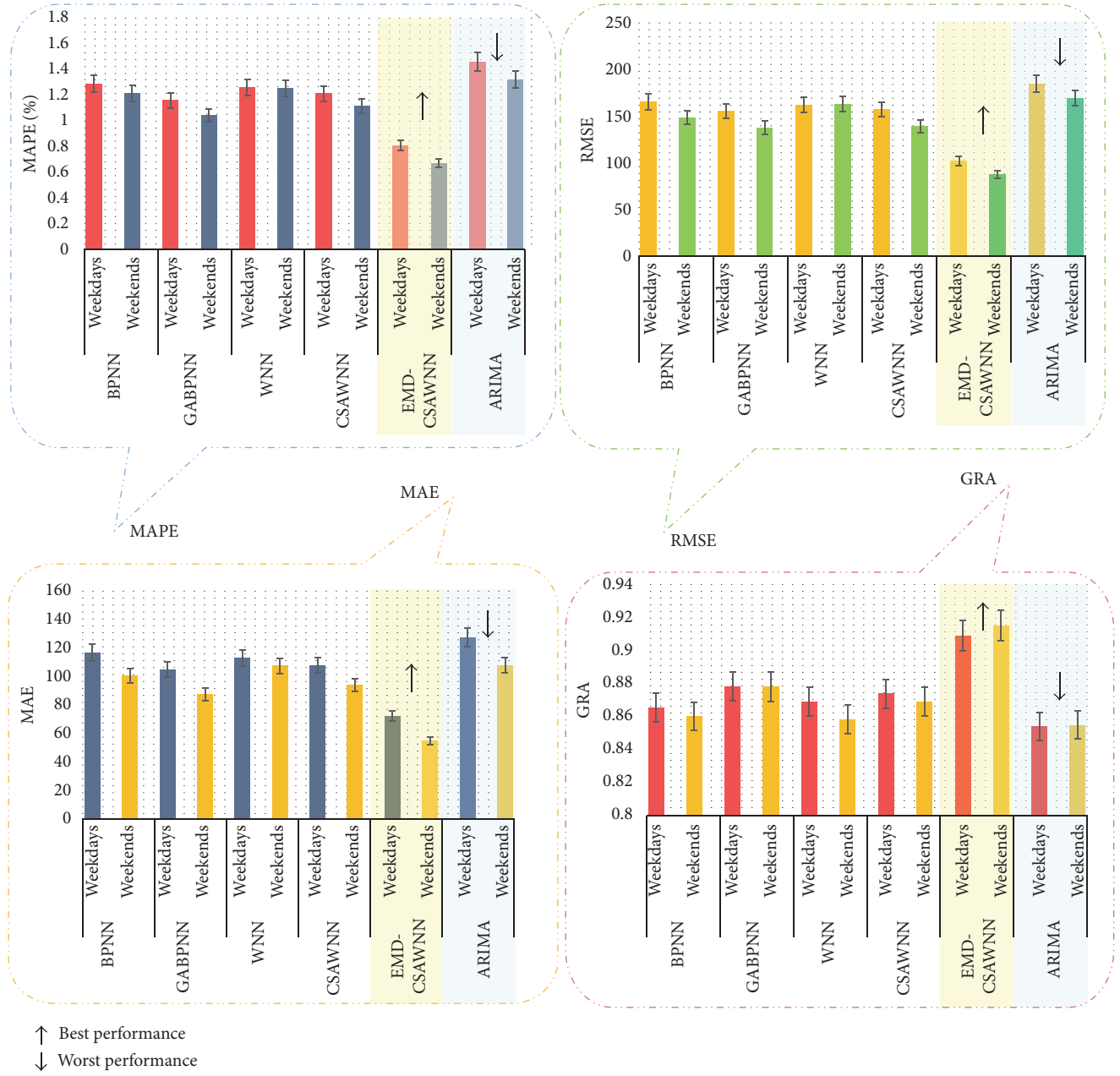
FIGURE 8: The BPNN, GABPNN, WNN, CSAWNN, EMD-CSAWNN, and ARIMA forecasting results between weekdays and weekends.

further consideration; on that basis, this section will give a general discrete form of forecasting availability [51].

Definition 10. Let

$$
e_{it} = \begin{cases} -1, & \dfrac{(x_t - x_{it})}{x_t} > 1 \\ \dfrac{(x_t - x_{it})}{x_t}, & -1 \le \dfrac{(x_t - x_{it})}{x_t} > 1 \\ 1, & \dfrac{(x_t - x_{it})}{x_t} > 1. \end{cases} \tag{14}
$$

e_{it} denotes the relative forecasting error of ith forecasting method at time t, $i = 1, 2, \ldots, m$, $t = 1, 2, \ldots, N$. The matrix $E = (e_{it})_{m \times N}$ is called the relative error matrix of the forecasting model.

Definition 11. $A_{it} = 1 - |e_{it}|$ is called the forecasting accuracy of ith forecasting method at time t, $i = 1, 2, \ldots, m$, and $t = 1, 2, \ldots, N$.

Definition 12. $m_i^k = \sum_{t=1}^{N} Q_t A_{it}^k$ is called kth-order forecasting availability unit of ith forecasting method, k is a positive integer, $i = 1, 2, \ldots, m$, $\{Q_t, t = 1, 2, \ldots, N\}$ is the discrete
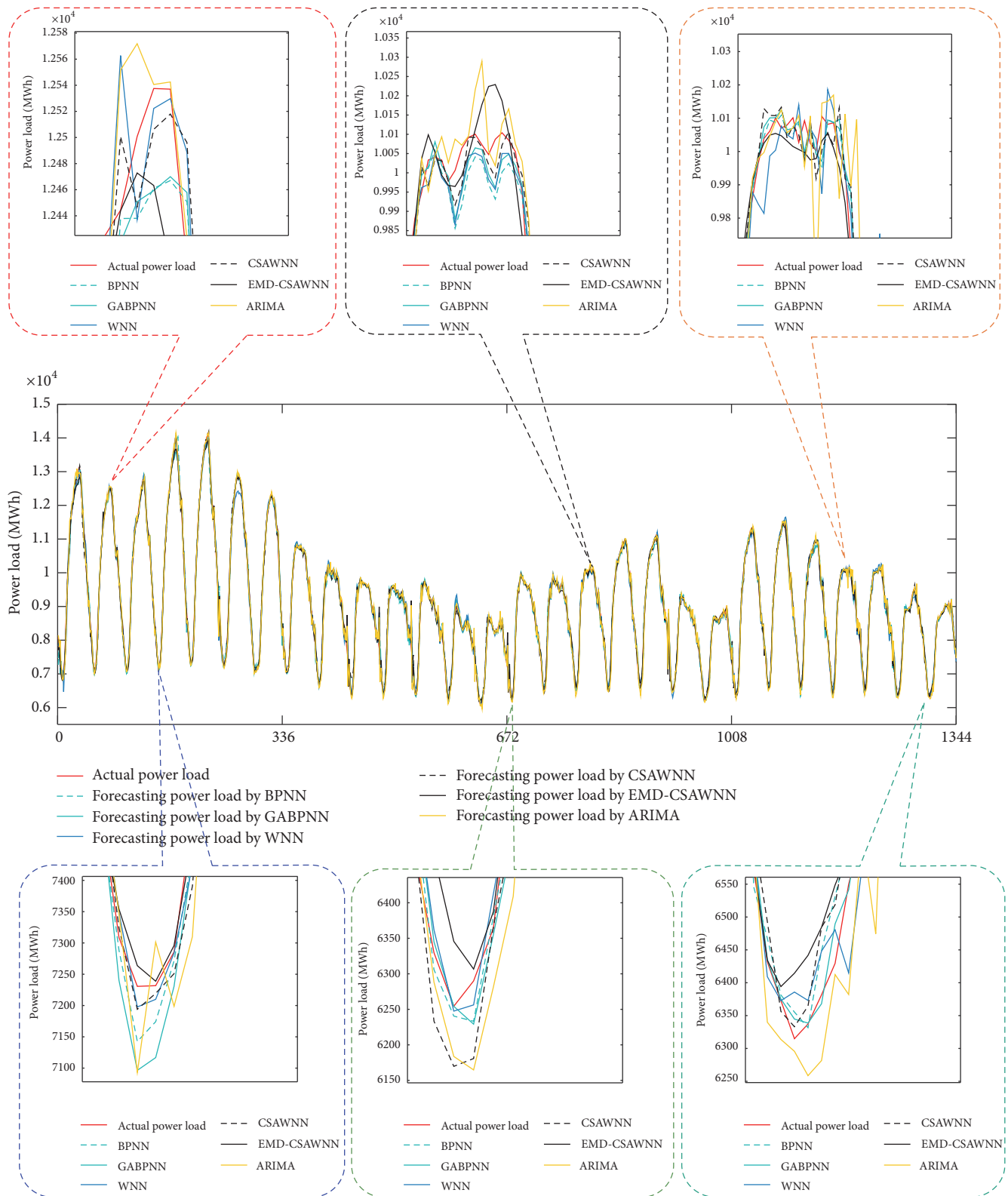
FIGURE 9: The forecasting results and actual values from February 2 to March 1.
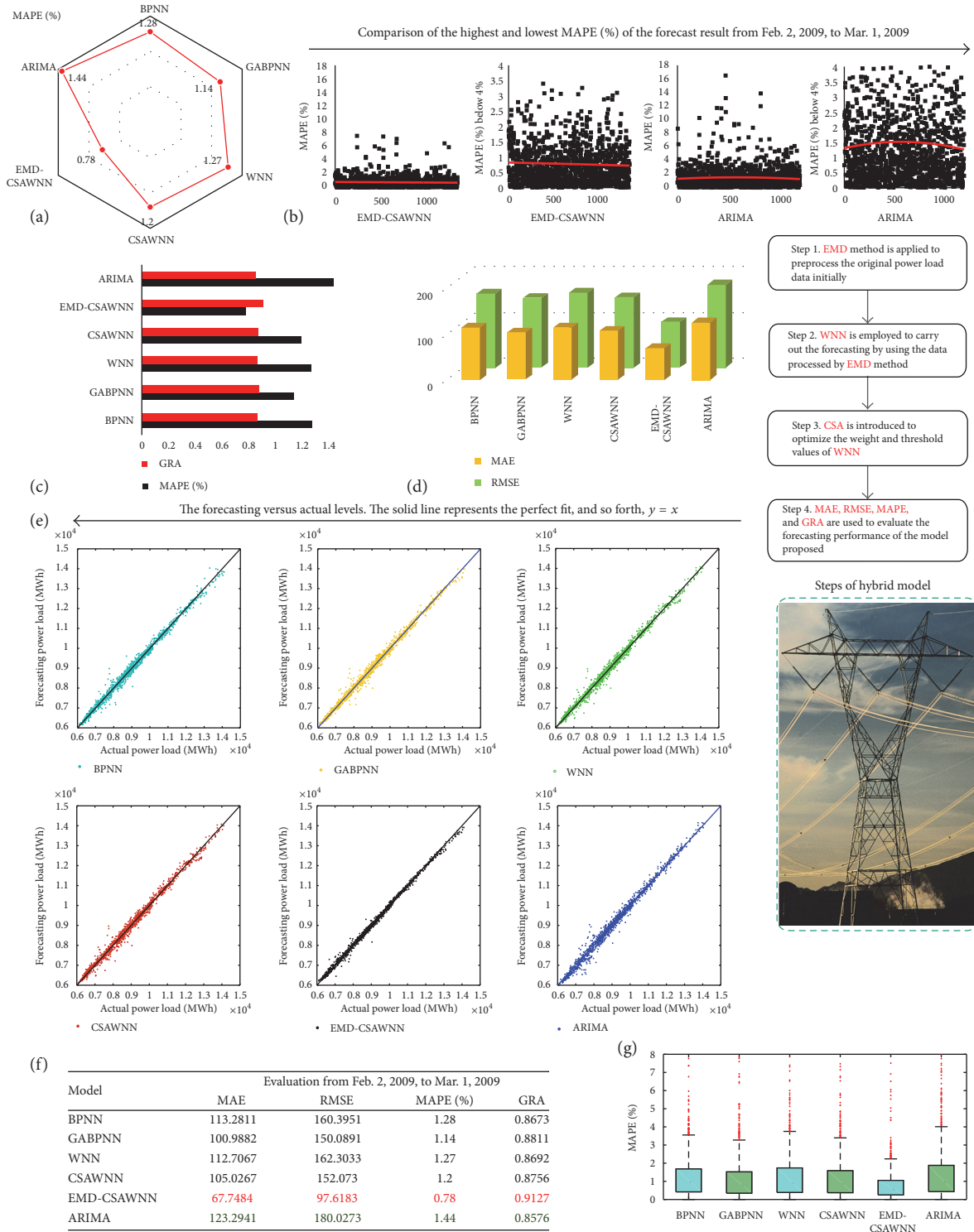
FIGURE 10: The comprehensive evaluation of forecasting models from February 2 to March 1. (a) The radar diagram of MAPE by using six different models. (b) Comparison of forecasting MAPE by using the EMD-CSAWNN and ARIMA models. The red line is the polynomial regression line. (c) The comparison of GRA and MAPE by using six different models. (d) The comparison of MAE and RMSE by using six different models. (e) The scatterplot of forecasting versus actual levels by using six different models. The solid line represents the perfect fit: that is, $y = x$. (f) Evaluation results of six different models. The red font is the best value of every evaluation index; the green font is the worst. (g) The comparison of the box plot by using six different models. The whiskers in the box plot indicate the primary range for the data, in which the lowest data are 1.5 times the interquartile range of the lower quartile and the highest data are 1.5 times the interquartile range of the upper quartile. The outliers, which are not included between the whiskers, are represented by the red crosshair.

(a)

(b)
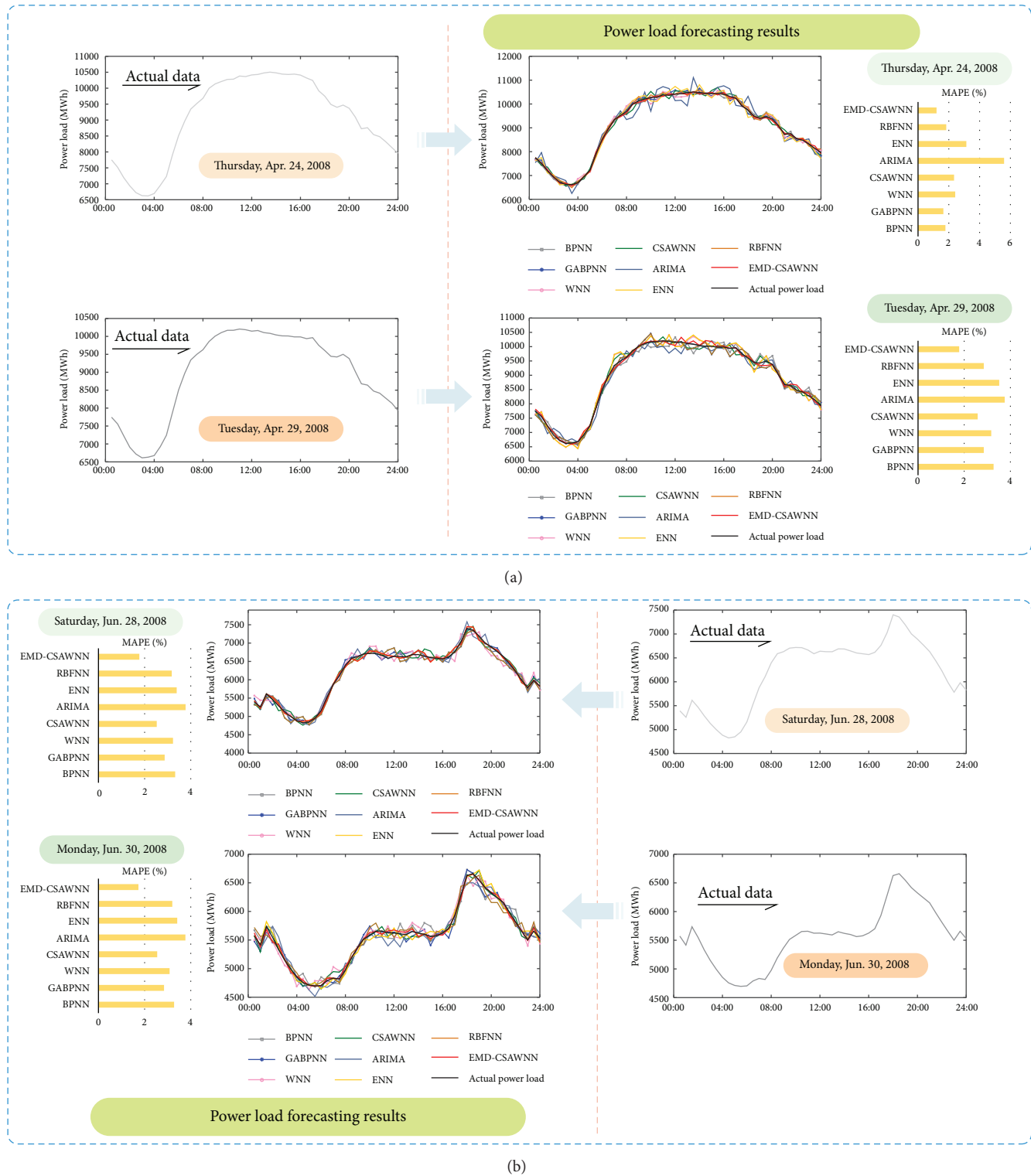
FIGURE 11: The comprehensive evaluation of forecasting models in Experiments 4 and 5. (a) Comparison of forecasting results by eight models on Thursday, April 24, 2008, and Tuesday, April 29, 2008, from New South Wales of Australia. (b) Comparison of forecasting results by eight models on Saturday, June 28, 2008, and Monday, June 30, 2008, from Victoria of Australia.
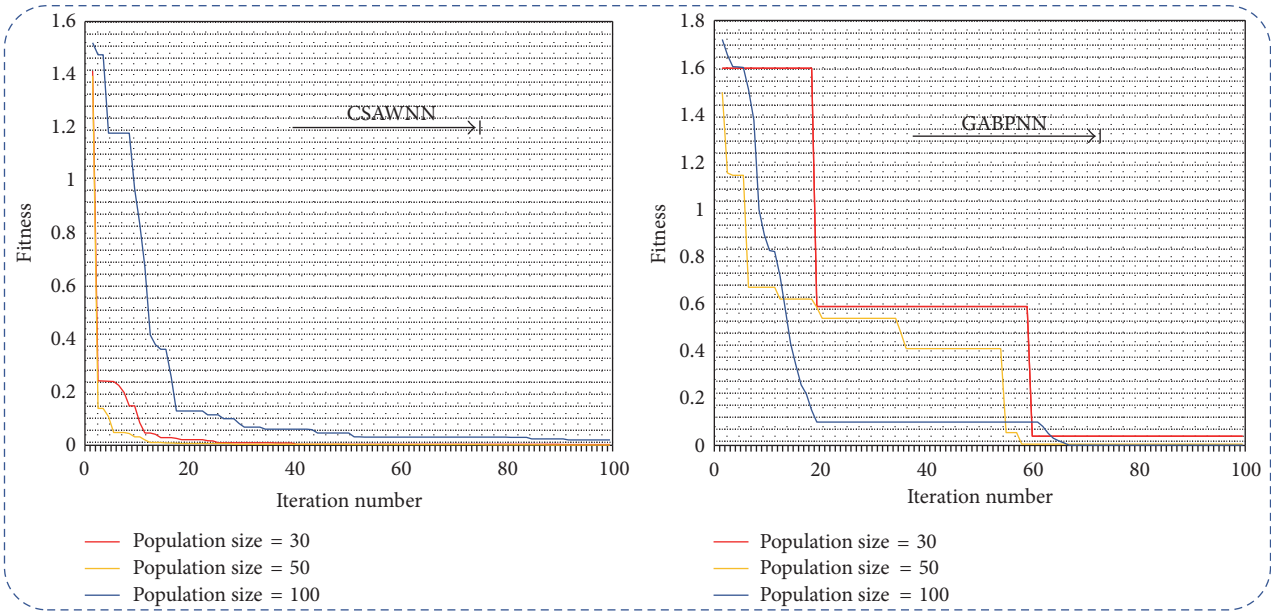
FIGURE 12: A convergence speed comparison of the fitness values between GA and CSA with different population sizes.
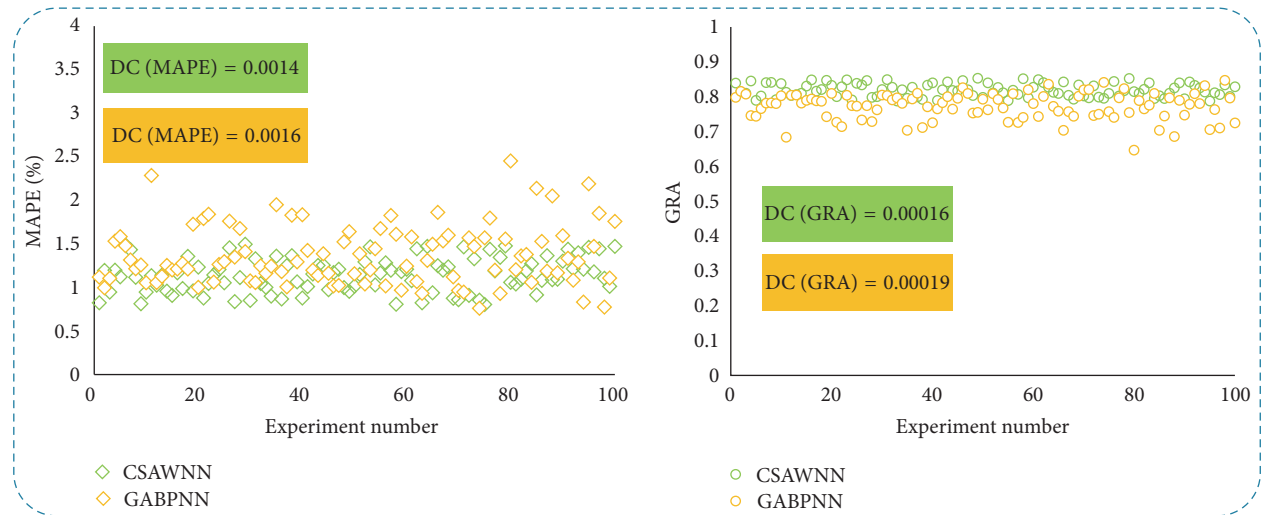
FIGURE 13: The DC (MAPE) and DC (GRA) of 100 experiments by comparison of GABPNN and CSAWNN.

probability distribution of m types of methods at time t, and $\sum_{t=1}^{N} Q_t = 1, \ Q_t > 0$.

Especially if the priori information of the discrete probability distribution of m types of methods is unknown, we define $Q_t = 1/N, \ t = 1, 2, \ldots, N$.

Definition 13. m_i^k is called the k-order forecasting availability unit of ith forecasting method, and H is a continuous function of a certain k unit. $H(m_i^1, m_i^2, \ldots, m_i^k)$ is called the k-order forecasting availability of ith forecasting method.

Definition 14. When $H(x) = x$ is a continuous function of one variable, $H(m_i^1) = m_i^1$ is the 1-order forecasting availability of ith forecasting method. When $H(x, y) = x(1 - \sqrt{y - x^2})$

is a continuous function of two variables, $H(m_i^1, m_i^2) = m_i^1(1 - \sqrt{m_i^2 - (m_i^1)^2})$ is the 2-order forecasting availability of ith forecasting method.

Especially if the first decimal of $H(m_i^1, m_i^2, \ldots, m_i^k)$ is the same, we define $H(m_i^1, m_i^2, \ldots, m_i^k) = \{10H(m_i^1, m_i^2, \ldots, m_i^k)\}$, where $\{\cdot\}$ denotes the fractional part.

Definition 14 illustrates that the 1st-order forecasting availability is the expectation forecasting accuracy sequence. The 2nd-order forecasting availability is the difference between the expectation and standard deviation of the forecasting accuracy sequence. We use the forecasting availability to evaluate the power load forecasting results in this paper. Through Figure 14, we obtain that the 1st-order and 2nd-order
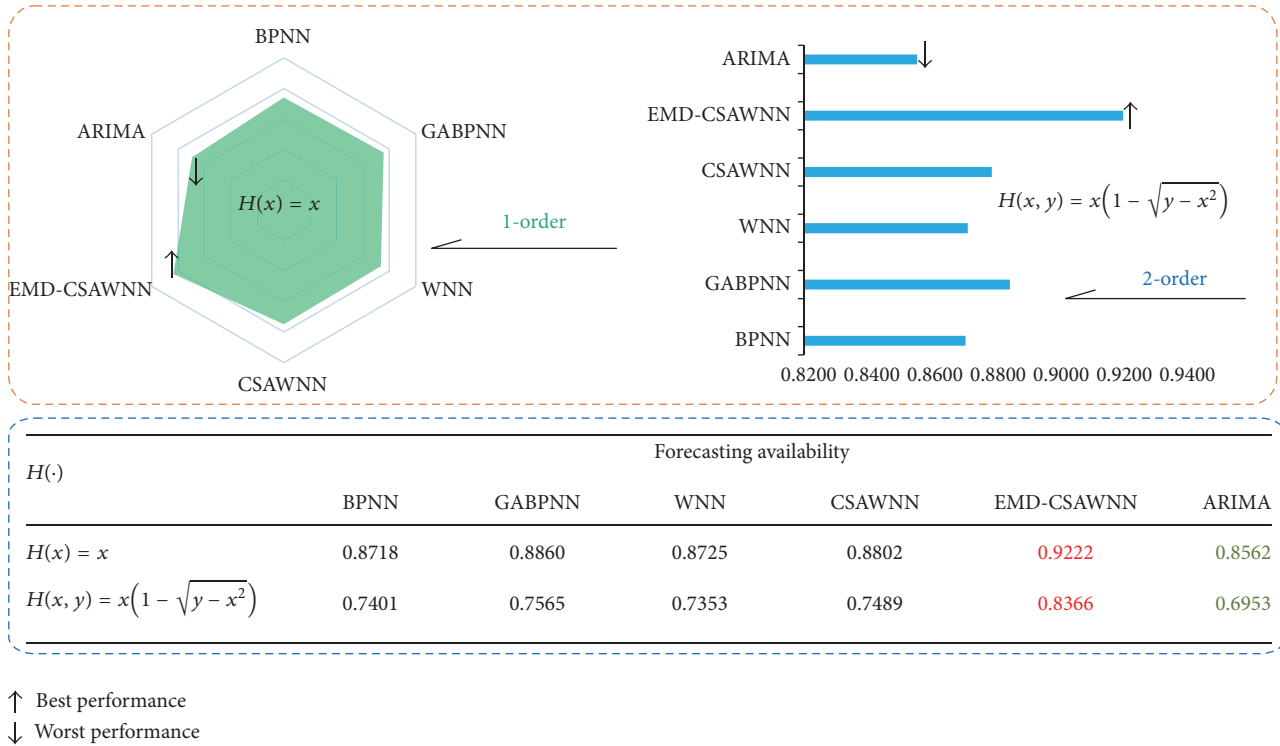
FIGURE 14: The forecasting availability of BPNN, GABPNN, WNN, CSAWNN, EMD-CSAWNN, and ARIMA.

forecasting availability offered by the hybrid model are 0.9222 and 0.8366, respectively, which outperform those of the others; this evaluation result corresponds to the previous evaluation criterion. Thus, the hybrid model is a more valid model than the others.

6. Conclusion

The one-day-ahead power load forecasting is an extremely important problem in power load planning, secure operation, and energy expenditure economy. Assessment of the power load as accurately and quickly as possible is the primary objective in power load forecasting. However, power load is affected by various uncertain factors such as climate change and the social environment, which may lead to difficulty in obtaining accurate power load forecasts. The accuracy of traditional individual forecasting methods, which lack denoising, is not satisfactory for power load forecasting. Herein, a hybrid EMD-CSAWNN model for short-term power load forecasting is developed. The empirical model decomposition technique is applied to reduce the high-frequency items. On the basis that WNN can handle the data with nonlinear features, the ensemble forecasting method is adopted to overcome the uncertainty of the outcomes that can be attributed to the randomness of the initialization of the single WNN. Moreover, we use the CSA to optimize the parameter in the ensemble forecasting model. Experimental studies of power load forecasting in NSW demonstrate that the hybrid model has higher precision than conventional forecasting models. The proposed EMD-CSAWNN model can provide efficient

computation and satisfactory forecasting accuracy for this type of data. Therefore, the developed hybrid approach is suggested for broad application in power load forecasts or even other fields such as wind speed and traffic flow forecasts.

Abbreviations

BPNN: Back Propagation Neural Network
WNN: Wavelet Neural Network
GA: Genetic Algorithm
CSA: Cuckoo Search Algorithm Back Propagation Neural
GABPNN: Network Optimized by Genetic Algorithm
CSAWNN: Wavelet Neural Network Optimized by Cuckoo Search Algorithm
AR: Autoregressive model
MA: Moving average model
ARIMA: Autoregressive integrated moving average model
IMF: Intrinsic mode function
ANN: Artificial Neural Network
PSO: Particle swarm optimization algorithm
ACO: Ant colony optimization algorithm
SAO: Simulated annealing optimization algorithm
MAE: Mean absolute error
RMSE: Root mean square error
MAPE: Mean absolute percentage error
GRA: Grey relational analysis
q: The number of sample data points used to build the NN model

p:	The number of data points to be forecasted in the NN model
ω:	The connection weights between the neurons of the input layer, hidden layer, and output layer, with values belonging to $[-1, 1]$
$f(\cdot)$:	The excitation function of the hidden layer
η:	The learning rate of the NN, which is used to adjust the weights and thresholds of the NN
iter_{max}:	The maximum number of iterations
P:	The population size of the initial population space
b:	A random number belonging to $[0, 1]$
r:	A random number belonging to $[0, 1]$
$w_{max}^{iter}, w_{min}^{iter}$:	The higher and lower bounds of the value of the gene w_{ij}^{iter}
G_{max}:	The maximum number of generations
P_α:	The possibility of finding an exotic egg by the nest master
X_*^{iter}:	The location of the optimum nest in generation iter
$\chi_j^{iter}, \chi_k^{iter}$:	Two random numbers in generation iter.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] S. Kouhi, F. Keynia, and S. Najafi Ravadanegh, "A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection," *International Journal of Electrical Power and Energy Systems*, vol. 62, pp. 862–867, 2014.

[2] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 498–505, 2001.

[3] Y. Zhang and G. Luo, "Short term power load prediction with knowledge transfer," *Information Systems*, vol. 53, pp. 161–169, 2015.

[4] N. An, W. Zhao, J. Wang, D. Shang, and E. Zhao, "Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting," *Energy*, vol. 49, no. 1, pp. 279–288, 2013.

[5] I. Moghram and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1484–1491, 1989.

[6] M. T. Hagan and S. M. Behr, "The time series approach to short term load forecasting," *IEEE Transactions on Power Systems*, vol. 2, no. 3, pp. 785–791, 1987.

[7] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[8] V. Dordonnat, S. J. Koopman, M. Ooms, A. Dessertaine, and J. Collet, "An hourly periodic state space model for modelling French national electricity load," *International Journal of Forecasting*, vol. 24, no. 4, pp. 566–587, 2008.

[9] W. R. Christiaanse, "Short-term load forecasting using general exponential smoothing," *IEEE Transactions on Power Apparatus and Systems*, vol. 90, no. 2, pp. 900–911, 1971.

[10] W.-C. Hong, "Hybrid evolutionary algorithms in a SVR-based electric load forecasting model," *International Journal of Electrical Power & Energy Systems*, vol. 31, no. 7-8, pp. 409–417, 2009.

[11] V. H. Hinojosa and A. Hoese, "Short-term load forecasting using fuzzy inductive reasoning and evolutionary algorithms," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 565–574, 2010.

[12] P. K. Dash, A. C. Liew, S. Rahman, and G. Ramakrishna, "Building a fuzzy expert system for electric load forecasting using a hybrid neural network," *Expert Systems with Applications*, vol. 9, no. 3, pp. 407–421, 1995.

[13] S. Rahman and O. Hazim, "Load forecasting for multiple sites: development of an expert system-based technique," *Electric Power Systems Research*, vol. 39, no. 3, pp. 161–169, 1996.

[14] D. K. Chaturvedi, A. P. Sinha, and O. P. Malik, "Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network," *International Journal of Electrical Power and Energy Systems*, vol. 67, pp. 230–237, 2015.

[15] S. Kouhi and F. Keynia, "A new cascade NN based method to short-term load forecast in deregulated electricity market," *Energy Conversion and Management*, vol. 71, pp. 76–83, 2013.

[16] P. Li, Y. Li, Q. Xiong, Y. Chai, and Y. Zhang, "Application of a hybrid quantized Elman neural network in short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 749–759, 2014.

[17] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001.

[18] R. Mamlook, O. Badran, and E. Abdulhadi, "A fuzzy inference model for short-term load forecasting," *Energy Policy*, vol. 37, no. 4, pp. 1239–1248, 2009.

[19] M. Hanmandlu and B. K. Chauhan, "Load forecasting using hybrid models," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 20–29, 2011.

[20] M. Q. Raza and Z. Baharudin, "A review on short term load forecasting using hybrid neural network techniques," in *Proceedings of the International Conference on Power and Energy (PECon '12)*, pp. 846–851, IEEE, Kota Kinabalu, Malaysia, December 2012.

[21] M. El-Telbany and F. El-Karmi, "Short-term forecasting of Jordanian electricity demand using particle swarm optimization," *Electric Power Systems Research*, vol. 78, no. 3, pp. 425–433, 2008.

[22] A. B. Nutt, R. C. Lenz Jr., H. W. Lanford, and M. J. Cleary, "Data sources for trend extrapolation in technological forecasting," *Long Range Planning*, vol. 9, no. 1, pp. 72–76, 1976.

[23] M. M. Tripathi, K. G. Upadhyay, and S. N. Singh, "Short-term load forecasting using generalized regression and probabilistic neural networks in the electricity market," *The Electricity Journal*, vol. 21, no. 9, pp. 24–34, 2008.

[24] A. D. Papalexopoulos and T. C. Hesterberg, "A regression-based approach to short-term system load forecasting," *IEEE Transactions on Power Systems*, vol. 5, no. 4, pp. 1535–1547, 1990.

[25] N. Ding, Y. Bésanger, and F. Wurtz, "Next-day MV/LV substation load forecaster using time series method," *Electric Power Systems Research*, vol. 119, pp. 345–354, 2015.

[26] O. Valenzuela, I. Rojas, F. Rojas et al., "Hybridization of intelligent techniques and ARIMA models for time series prediction," *Fuzzy Sets and Systems*, vol. 159, no. 7, pp. 821–845, 2008.

[27] H. Nie, G. Liu, X. Liu, and Y. Wang, "Hybrid of ARIMA and SVMs for short-term load forecasting," *Energy Procedia*, vol. 16, pp. 1455–1460, 2012.

[28] T. Yalcinoz and U. Eminoglu, "Short term and medium term power distribution load forecasting by neural networks," *Energy Conversion and Management*, vol. 46, no. 9-10, pp. 1393–1405, 2005.

[29] C. J. Bennett, R. A. Stewart, and J. W. Lu, "Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system," *Energy*, vol. 67, pp. 200–212, 2014.

[30] H. S. Chen and W. C. Chang, "A study of optimal grey model GM (1, 1)," *Journal of the Chinese Grey System Association*, vol. 1, no. 2, pp. 141–145, 1998.

[31] L. Xiao, J. Wang, X. Yang, and L. Xiao, "A hybrid model based on data preprocessing for electrical power forecasting," *International Journal of Electrical Power and Energy Systems*, vol. 64, pp. 311–327, 2015.

[32] J. X. Che and J. Z. Wang, "Short-term load forecasting using a kernel-based support vector regression combination model," *Applied Energy*, vol. 132, pp. 602–609, 2014.

[33] N. Liu, Q. Tang, J. Zhang, W. Fan, and J. Liu, "A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids," *Applied Energy*, vol. 129, pp. 336–345, 2014.

[34] S. Bahrami, R.-A. Hooshmand, and M. Parastegari, "Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm," *Energy*, vol. 72, pp. 434–442, 2014.

[35] L. Karthikeyan and D. N. Kumar, "Predictability of nonstationary time series using wavelet and EMD based ARMA models," *Journal of Hydrology*, vol. 502, pp. 103–119, 2013.

[36] G. K. Sharma, A. Kumar, T. Jayakumar, B. P. Rao, and N. Mariyappa, "Ensemble Empirical Mode Decomposition based methodology for ultrasonic testing of coarse grain austenitic stainless steels," *Ultrasonics*, vol. 57, pp. 167–178, 2015.

[37] A. Baliyan, K. Gaurav, and S. K. Mishra, "A review of short term load forecasting using artificial neural network models," *Procedia Computer Science*, vol. 48, pp. 121–125, 2015.

[38] R. Chao, *[Ph.D. thesis]*, Lanzhou University, 2013.

[39] N. B. Andersen, "Real Paley-Wiener theorems and Roe's theorem associated with the Opdam-Cherednik transform," *Journal of Mathematical Analysis and Applications*, vol. 427, no. 1, pp. 47–59, 2015.

[40] K. Liu, W. Y. Guo, X. L. Shen, and Z. F. Tan, "Research on the forecast model of electricity power industry loan based on GA-BP neural network," *Energy Procedia*, vol. 4, pp. 1918–1924, 2012.

[41] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Transactions on Neural Networks*, vol. 3, no. 6, pp. 889–898, 1992.

[42] Y. Lu, N. Zeng, Y. Liu, and N. Zhang, "A hybrid wavelet neural network and switching particle swarm optimization algorithm for face direction recognition," *Neurocomputing*, vol. 155, pp. 219–224, 2015.

[43] N. M. Nawi, A. Khan, and M. Z. Rehman, "A new back-propagation neural network optimized with cuckoo search algorithm," in *Computational Science and Its Applications—ICCSA 2013: 13th International Conference, Ho Chi Minh City, Vietnam, June 24–27, 2013, Proceedings, Part I*, vol. 7971 of *Lecture Notes in Computer Science*, pp. 413–426, Springer, Berlin, Germany, 2013.

[44] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems," *Engineering with Computers*, vol. 29, no. 1, pp. 17–35, 2013.

[45] F. Yu and X. Xu, "A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network," *Applied Energy*, vol. 134, pp. 102–113, 2014.

[46] K. Liu, W. Guo, X. Shen, and Z. Tan, "Research on the gforecast model of electricity power industry loan based on GA-BP neural network," *Energy Procedia*, vol. 14, pp. 1918–1924, 2012.

[47] M. Smith, "Modeling and short-term forecasting of new south wales electricity system load," *Journal of Business and Economic Statistics*, vol. 18, no. 4, pp. 465–478, 2000.

[48] S. Qin, F. Liu, J. Wang, and B. Sun, "Analysis and forecasting of the particulate matter (PM) concentration levels over four major cities of China using hybrid models," *Atmospheric Environment*, vol. 98, pp. 665–675, 2014.

[49] D. Bunn and E. D. Farmer, *Comparative Models for Electrical Load Forecasting*, John Wiley & Sons, Chichester, UK, 1985.

[50] L. Liu, Q. Wang, J. Wang, and M. Liu, "A rolling grey model optimized by particle swarm optimization in economic prediction," *Computational Intelligence*, 2014.

[51] H. Y. Chen and D. P. Hou, "Research on superior combination forecasting models based on forecasting effective measure," *Journal of University of Science and Technology of China*, vol. 32, no. 2, pp. 172–180, 2002.

Research Article

Feedback Gating Control for Network Based on Macroscopic Fundamental Diagram

YangBeibei Ji,[1] **Chao Mo,**[1] **Wanjing Ma,**[2] **and Dabin Liao**[3]

[1]*School of Management, Shanghai University, 99 Shangda Road, Baoshan, Shanghai 200444, China*
[2]*The Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China*
[3]*Wuhan Planning and Design Institute, Wuhan 430010, China*

Correspondence should be addressed to Wanjing Ma; mawanjing@tongji.edu.cn

Empirical data from Yokohama, Japan, showed that a macroscopic fundamental diagram (MFD) of urban traffic provides for different network regions a unimodal low-scatter relationship between network vehicle density and network space-mean flow. This provides new tools for network congestion control. Based on MFD, this paper proposed a feedback gating control policy which can be used to mitigate network congestion by adjusting signal timings of gating intersections. The objective of the feedback gating control model is to maximize the outflow and distribute the allowed inflows properly according to external demand and capacity of each gating intersection. An example network is used to test the performance of proposed feedback gating control model. Two types of background signalization types for the intersections within the test network, fixed-time and actuated control, are considered. The results of extensive simulation validate that the proposed feedback gating control model can get a Pareto improvement since the performance of both gating intersections and the whole network can be improved significantly especially under heavy demand situations. The inflows and outflows can be improved to a higher level, and the delay and queue length at all gating intersections are decreased dramatically.

1. Introduction

Traffic congestion in urban road networks is still a problem of modern society although it has been studied in a variety of ways during the past decades. Due to the high unpredictability of choices of travelers, realistic modeling and efficient control of urban road networks remain a big challenge. By increasing road capacity or reducing traffic demand, we can change the situation of the congested networks, while the provision of new infrastructure is usually not a feasible solution, and we should make full use of the existing infrastructure. There is a vast literature of congestion dynamics, control, and spreading in urban road networks traffic systems with different traffic modes and the gating control is an import practical tool to realize the network control strategy.

Gating control is frequently employed against oversaturation of significant or sensitive links, arterials, or urban network parts [1]. The idea is to protect links from oversaturation, whereby the level or duration of gating may depend on real-time measurements from the protected links. Gal-Tzur proposed a strategy which employs the concept of metering for small congested networks with one critical intersection [2]. However, gating is usually employed in an ad hoc way (based on engineering judgment) regarding the specific gating policy and quantitative details, which may lead to insufficient or unnecessarily strong gating actions. Moreover, the optimal condition of network is very important for gating control, and how to define the optimal condition remains the biggest obstacle to making use of gating control in reality.

Recently, Daganzo proposed a new concept of macroscopic relationship between average network flow and density, which is called the Macroscopic Fundamental Diagram (MFD) [3]. MFDs describe the relationship between inflows, outflows, and number of vehicles in the network in a very

clear and simple way, which provide an efficient tool for control and mitigate network congestion based on detected data. This characterization of the traffic state makes it possible to diagnose the emergence of congestion and choose measures to mitigate traffic problems by redirecting flows to areas with spare capacity [4].

Since then, the basic property and influence factors of MFD have been done. Leclercq et al. [5] showed that the MFDs are an envelope of possible traffic states. Christine and Ladier [6] use all the data available for a medium-size French city to explore the impact of heterogeneity on the existence of a MFD. They studied the impact of differences between the surface and highway network, distance between loop detectors and traffic signals within the surface network, and differences between penetrating and ring roads within the highway network [6]. Geroliminis and Sun show that different freeway subnetworks do not have a well-defined MFD with low scatter because the aggregated patterns do not just exhibit some high degree of random scatter [7]. Gayah and Daganzo studied the effect of turning volume on MFD with the help of simulation [8]. Ji and Geroliminis also explore the spatial and temporal characteristic of congestion in urban networks [9]. Zhang et al. explore MFD of a network under three different adaptive traffic control modes (SCATS-L, SCATS-F, and SOTL) [10]. Mühlich et al. [11] found from simulation that when traffic density becomes higher, gridlock may occur: queues remain on the intersections after the end of a green phase, blocking the traffic flow for a following signal phase. If this happens, the average flow diminishes with growing traffic densities. Haddad and Geroliminis analyzed the stability of traffic control for two-region urban cities [12] and designed a robust perimeter controller for an urban region [13]. Gayah et al. [14] examined the impact of locally adaptive traffic signals on network stability and the MFD by a family of signal control strategies.

The application of MFD to control a network is also a very important branch. Geroliminis and Daganzo proposed a control rule based on the MFD concept that maximizes the network outflow; however, the proposed method cannot be directly employed for practical use in urban networks [15]. Li et al. introduced a fixed-time signal timing perimeter control by exploiting the MFD, albeit without adaptation to the prevailing real-time traffic conditions [16]. Aboudolas and Geroliminis developed a PI regulator for multiple regions. However, MPC calls for sufficiently accurate model and external disturbance predictions, which may be a serious impediment for practicable control [17]. Keyvan-Ekbatani et al. proposed a generic real-time feedback-based gating concept, which exploits the urban MFD for smooth and efficient traffic control operations, with an application to the network of Chania and Greece [18], which is different from Aboudolas and Geroliminis [17] that researched the gating control for only one urban network, while Aboudolas and Geroliminis [17] separated the whole network, and investigating the boundary control occurs at the intertransfers between neighborhood reservoirs.

Keyvan-Ekbatani et al. demonstrated that efficient feedback-based gating is actually possible with much less real-time measurements, that is, at lower implementation
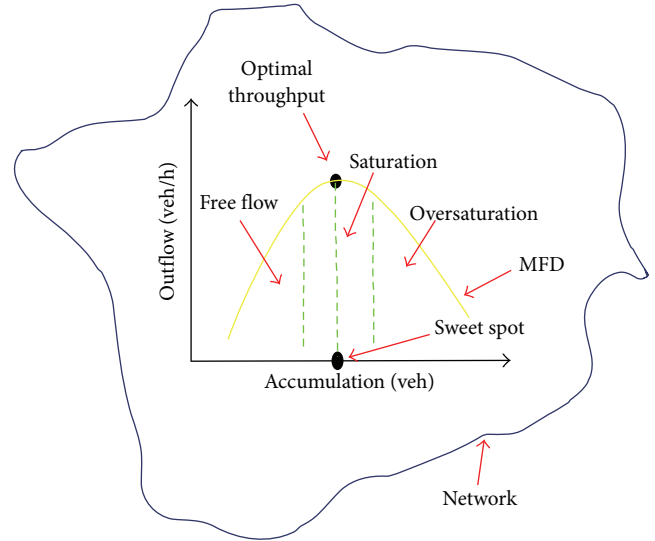
Figure 1: MFD of network.

cost [19]. Keyvan-Ekbatani et al. further proposed a multiple concentric-boundary gating strategy, which implements the aforementioned feedback-based gating strategy, along with considering the heterogeneity of a large-scale urban network like San Francisco, USA [20]. Aboudolas and Geroliminis extended the basic feedback approach for application to multiple subnetworks with separate individual MFD in a heterogeneous urban network [17]. Yildirimoglu et al. [21] explored the effect of route choice behavior on MFD modeling in case of heterogeneous urban networks by extending two MFD-based traffic models with different granularity of vehicle accumulation state and route choice behavior aggregation.

Despite the informative results offered by previous studies, many issues on network gating control have not been sufficiently addressed. The network related data such as dynamic origin-destination (OD) needed by classical gating control model, especially, is not easy to be obtained [3]. The optimization of signal timings has not been explicitly addressed in very limited research on MFD-based network control. How to balance between the performance of gating intersections and target network is also a remaining problem. In response to the above needs, this paper proposed a new feedback gating control strategy with a model to distribute allowed inflows among gating intersections based on MFD.

The paper is organized as follows. The gating control model based on MFD is developed in Section 2. Section 3 presents evaluation and analysis results of the proposed gating control model based on simulation. Conclusions and future research needs are summarized in the last section.

2. Gating Control Model

2.1. The Control Strategy. MFD reflects the relation between the vehicle accumulation in network and outflow of network. It approximates a parabola going downward. Figure 1 show the MFD of a network, the horizontal axis is the total vehicle
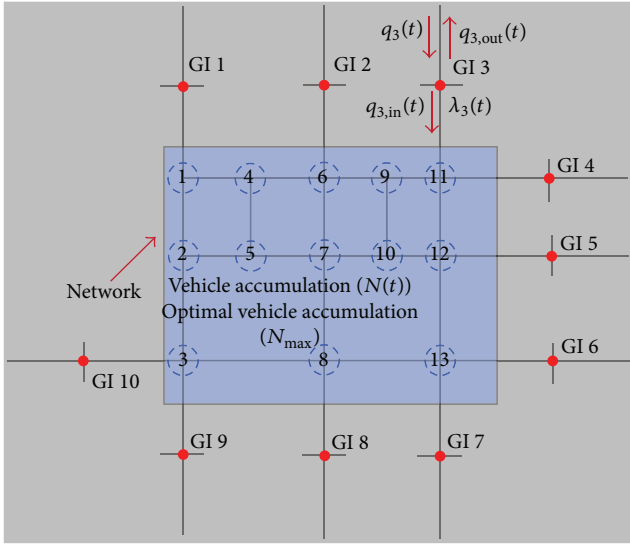
Figure 2: Control parameter of model.

Table 1: Time parameters for 60-second cycle for 13 intersections.

Intersections	1	2	3	4	5	6	7	8	9	10	11	12	13
West-east	36	38	25	18	13	30	33	25	19	19	42	32	32
North-south	20	18	30	37	42	25	22	31	37	37	13	23	23

Table 2: Time parameters for 80-second cycle for 13 intersections.

Intersections	1	2	3	4	5	6	7	8	9	10	11	12	13
West-east	48	51	33	24	17	40	44	33	25	25	56	43	43
North-south	27	24	40	49	56	33	29	41	49	49	17	31	31

Table 3: Time parameters for 100-second cycle for 13 intersections.

Intersections	1	2	3	4	5	6	7	8	9	10	11	12	13
West-east	60	63	42	30	22	50	55	42	32	32	70	53	53
North-south	33	30	50	62	70	42	37	52	62	62	22	38	38

Table 4: Time parameters for 120-second cycle for 13 intersections.

Intersections	1	2	3	4	5	6	7	8	9	10	11	12	13
West-east	72	76	50	36	26	60	66	50	38	38	84	64	64
North-south	40	36	60	74	84	50	44	62	74	74	26	46	46

Table 5: Time parameters for 140-second cycle for 13 intersections.

Intersections	1	2	3	4	5	6	7	8	9	10	11	12	13
West-east	84	89	58	42	30	70	77	58	44	44	98	75	75
North-south	47	42	70	86	98	58	51	72	86	86	30	54	54

in the network and the vertical axis is the total vehicles that leave the network or arrive at the destination in the network. When the vehicle accumulation in network is less than the sweet spot, the outflow of network increases with the increase of the accumulation and then arrives at the optimal throughput. If the vehicle accumulation in network is more than the sweet spot the outflow of network decreases with the increase of the accumulation. Hence, we should take some actions to control the number of vehicles entering the network to prevent network congestion (e.g., maintain the accumulation around the sweet spot to maximize the outflow).

MFD of network can be divided into three parts: the free flow condition, the saturation condition, and the oversaturation condition. During the free flow condition, more vehicles can get in network, while in the saturation condition we should control the number of vehicles getting in network to keep the accumulation around the sweet spot and then the outflow around the maximum throughput; if the condition of network is in the oversaturation condition we must control the vehicle strictly to reduce the accumulation to prevent network from getting blocked.

In order to control input vehicle, a feedback gating control model is proposed. The blue circle in Figure 2 represents the boundary of the network and the red dots represent the gating intersections (GI); all vehicles get in/out of network through gating intersections and intersections on the boundary (i.e., signalized intersection on the boundary of the network); $q_{i,\text{out}}(t)$ represent the outflow of signalized intersection i at time t represented by the red solid line in Figure 2; $q_i(t)$ represents the demand of signalized intersection i at time t, $C_i(t)$ represents the total capacity of all lane groups heading to the network at intersection i, and $q_{i,\text{in}}(t)$ represents the permitted number of vehicles that gets in network through signalized intersection i at time t just like the blue dash line; $\lambda_i(t)$ represent the controls parameter of signalized intersection i at time t just like the yellow solid line; $N(t)$

represent the accumulation of vehicles in network at time t and N_{max} represent the sweet spot of network.

2.2. The Control Model

2.2.1. The Control Objective.
In order to prevent network from getting blocked and serve more vehicles, a suitable control objective is to maximize the total vehicles that leave the network (we do not consider the vehicle arriving at the destination in the network). With this objective, the optimal policy is to allow as many vehicles to enter the network as possible without allowing the accumulation to reach oversaturation condition. Hence if we keep the accumulation around sweet spot, then the outflow is around the optimal throughput, and more vehicles will leave the network which means that more vehicles can get in the network. The control objective as follows:

$$\max \quad Z = \sum_i \sum_t q_{i,\text{out}}(t) = \int_{t_0}^{t} G(N(t))\, dt, \quad (1)$$

where $G(x)$ is the expression of MFD.

The signal control time table is calculated using Synchro and the signal time parameters are shown in Tables 1–5.

2.2.2. The Feedback Gating Control Model.
Equation (1) is not easy to be implemented in practice. Hence, we transfer this problem to a feedback control scheme. When we get the MFD

of a network we know the sweet spot of the network; namely, we know the optimal number of vehicles in the network. The number of vehicles entering, leaving, and remaining in the network timely can be calculated based on detectors located on the boundary intersections. Then we can calculate the number of vehicles which can be allowed to get in the network.

Note that $q_{oi}(t)$ includes both internal (off-street parking for taxis and pockets for private vehicles) and external noncontrolled inflows. The conservation equation for the network is

$$\frac{d(N(t))}{d(t)} = Q(t) + q_{oi}(t) - q_{\text{out}}(t), \tag{2}$$

$$q_{\text{out}}(t) = G(N(t)). \tag{3}$$

Combine (2) and (3); we get

$$\frac{d(N(t))}{d(t)} = Q(t) + q_{oi}(t) - G(N(t)). \tag{4}$$

The linearization yields may be linearized around an optimal steady state by Taylor expansion. Combining with the research target, the best condition of network is the optimal steady state. When the network is in the best condition, the number of vehicles in the network will reach the best condition, the number of vehicles outflowing the network will be the maximum, the number of vehicles outflowing the network from signal intersections will be the maximum, and the number of vehicles outflowing the network from uncontrolled intersections will be the maximum and satisfy the following conditions:

$$\overline{Q} + \overline{q}_{oi} = \overline{q}_{\text{out}},$$

$$\Delta Q(t) = Q(t) - \overline{Q},$$

$$\Delta q_{oi}(t) = q_{oi}(t) - \overline{q}_{oi}, \tag{5}$$

$$\Delta N(t) = N(t) - \overline{N},$$

$$\Delta q_{\text{out}}(t) = q_{\text{out}}(t) - \overline{q}_{\text{out}} = G(N(t)) - G(\overline{N}).$$

The results of linearization are

$$\frac{d(\Delta N(t))}{d(t)} = A\Delta Q(t) + B\Delta q_{oi}(t) - C\Delta N(t),$$

$$A = \left.\frac{\partial F}{\partial Q}\right|_{Q(t)=\overline{Q}, q_{oi}(t)=\overline{q}_{oi}, N(t)=\overline{N}},$$

$$B = \left.\frac{\partial F}{\partial q}\right|_{Q(t)=\overline{Q}, q_{oi}(t)=\overline{q}_{oi}, N(t)=\overline{N}}, \tag{6}$$

$$C = \left.\frac{\partial F}{\partial N}\right|_{Q(t)=\overline{Q}, q_{oi}(t)=\overline{q}_{oi}, N(t)=\overline{N}}.$$

From (2) to (6), we get

$$\frac{d(\Delta N(t))}{d(t)} = \Delta Q(t) + \Delta q_{oi}(t) - \Delta N(t) G'(\overline{N}). \tag{7}$$

Formula (7) can be denoted by

$$\frac{d(\Delta N(t))}{d(t)} = D(\Delta Q(t) + \Delta q_{oi}(t)) + E\Delta N(t). \tag{8}$$

Formula (8) is temporal continuity, which can be discretized by Euler formula:

$$\Delta N(t+1) = Y\Delta N(t) + Z(\Delta Q(t) + \Delta q_{oi}(t)),$$

$$Y = e^{ET} = e^{-G(\overline{N})T},$$

$$Z = \frac{Y-1}{D}E = (Y-1)E \tag{9}$$

$$= \left(1 - e^{-G(\overline{N})T}\right)G(\overline{N}).$$

Based on feedback control logic, the optimal inflow of network at t is given by

$$Q(t) = Q(t-1) - K_P[N(t) - N(t-1)] + K_I(\overline{N} - N(t)). \tag{10}$$

Meanwhile, the external demand of each boundary intersection should also be taken into account. The total actual inflow at time t can be calculated by the following equation:

$$q_{i,\text{in}}(t) = \min \begin{cases} Q(t) \times \alpha_i \\ q_i(t). \end{cases} \tag{11}$$

2.2.3. The Inflow Distribution. In order to determine the inflows at each boundary intersection evenly and minimize the impacts of gating control on the performance of gating intersections, both the demand and capacity of each boundary intersection are taken into account. A parameter α_i is used to represent the ratio of inflows assigned to the intersection i; it can be calculated by the following equation:

$$\alpha_i = \frac{q_i(t)^2[S_i - S_i(t)]}{\sum_i q_i(t)^2[S_i - S_i(t)]}. \tag{12}$$

When we get the $q_{i,\text{in}}(t)$ of each intersection, we can adjust the green time of each intersection according to the data of detector.

Then the green time of the lane group heading to network at boundary intersection i can be calculated as

$$t_i(t) = q_{i,\text{in}}(t) \times \frac{T}{S_i}. \tag{13}$$

We also take the constraint condition of maximum and minimum green time into consideration, as follows:

$$t_{g,\min} \le t_i(t) \le t_{g,\max}. \tag{14}$$

3. Simulation and Analysis

3.1. Network Description and Simulation Setup. In order to test the feedback gating control model, we build a microsimulation network by VISSIM. The network includes 13 intersections and 8 main roads; traffic signals are all multiphase fixed-time operating on a common cycle length of 120 s, shown in Figure 2 by the blue dotted line circle. The red point represents the gating intersections at the border of the protected network, which could be the intersections upstream of the controlled network.

We consider vehicles entering the network if they pass the gating intersections from outside and getting out of the network if they pass the gating intersections to the destination. The loop detectors have been installed at gating intersections to collect the number of vehicles getting in/out of the network.

Nine-hour simulation with time-dependent demand is carried out to test the performance of the proposed feedback control logic. The traffic demand on each entry lane of the boundary intersections is evenly varied from 0 to 900 vehicles in one cycle of 30 minutes. The traffic demand on each internal entrance is evenly varied from 0 to 450 vehicles in one cycle of 30 minutes. The inflow, outflow, and number of vehicles in the network are collected every 120 s.

3.2. Macroscopic Fundamental Diagrams. Through the timely data from detectors, we can get the MFD of the test area which reflects the relation of the outflow and vehicle accumulation in the network. The MFD's y-axis represents the number of vehicles leaving the test area per control cycle (outflow), while the x-axis represents the number of vehicles in the test area (accumulation). The outflow and the vehicle accumulation in the network are obtained from the (emulated) loop measurements via the following equations:

$$\text{Outflow}\,(t) = \sum_i q_{i,\text{out}}\,(t),$$
$$\text{accumulation} = \sum_i \sum_t \left(q_{i,\text{in}}\,(t) - q_{i,\text{out}}\,(t) \right), \tag{15}$$

where $q_{i,\text{out}}(t)$ is the number of vehicles leaving the network from intersection i at time t; $q_{i,\text{in}}(t)$ is the number of vehicles getting in the network from intersection i at time t. Then we get the MFD of the test network as shown in Figure 3.

In Figure 3, at the beginning of simulation, the outflow per control cycle increases with the increase of the number of vehicles in the test area (accumulation); when the accumulation is around 800 (veh) and outflow around 225 (veh/cycle), the outflow does not increase with the increase of the accumulation; it is in a relatively steady state, while the outflow will decrease with the increase of the accumulation steadily until the outflow is 0 (veh/cycle) and the accumulation is 2500 (veh). Although there is a fluctuation when the accumulation is around 1200 (veh), it is not unstable. Hence the test area will operate at optimal status when the accumulation is around 800 (veh) and the outflow around 225 (veh/cycle), and this is our optimal state
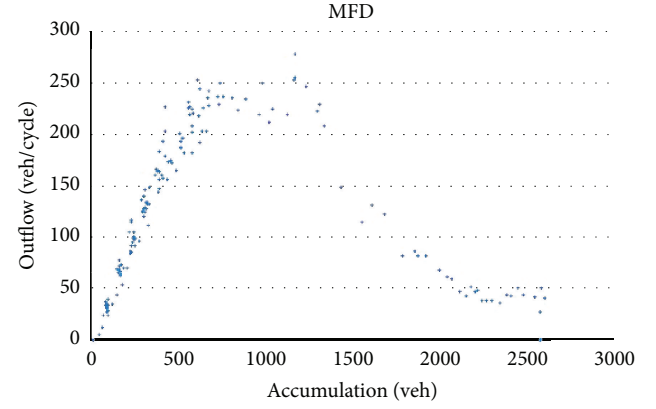
FIGURE 3: The MFD of the test area.

- ◆ Fixed-time
- ▲ Classical actuated control
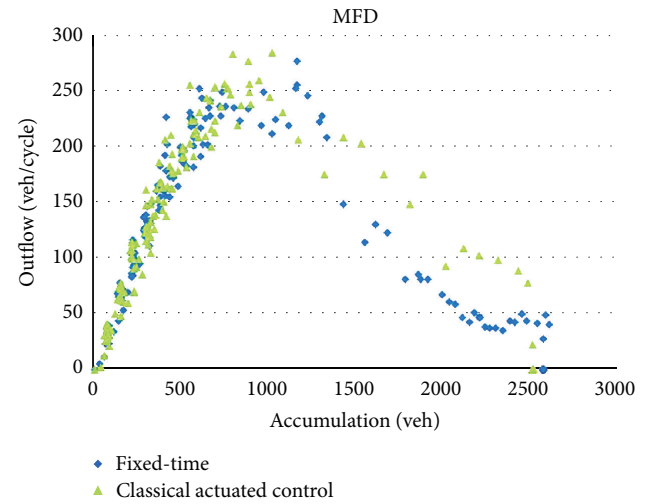
FIGURE 4: Influence factors of MFD.

which ought to be reached by the proposed feedback control method.

3.3. Experiment Design. In order to test the feedback gating control and explore the influence of different control strategy of intersection in the test area, the following two cases are considered.

Case 1 (fixed-time control). The control plans are optimized by Synchro.

Case 2 (classical actuated control). The classical control plans illustrated in text book [22] are used.

As shown in Figure 4, different control strategies have different influence on MFD. Generally speaking, comparing with the MFD of fixed-time, actuated control improves the maximum outflow slightly. The maximum outflow changes from 250 (veh/cycle) to 280 (veh/cycle) while the accumulation in network remains the same.

(a) Occupation of uncontrolled situation at the beginning

(b) Occupation of controlled situation at the beginning

(c) Occupation of uncontrolled situation in the middle

(d) Occupation of controlled situation in the middle

(e) Occupation of uncontrolled situation at the end

(f) Occupation of controlled situation at the end
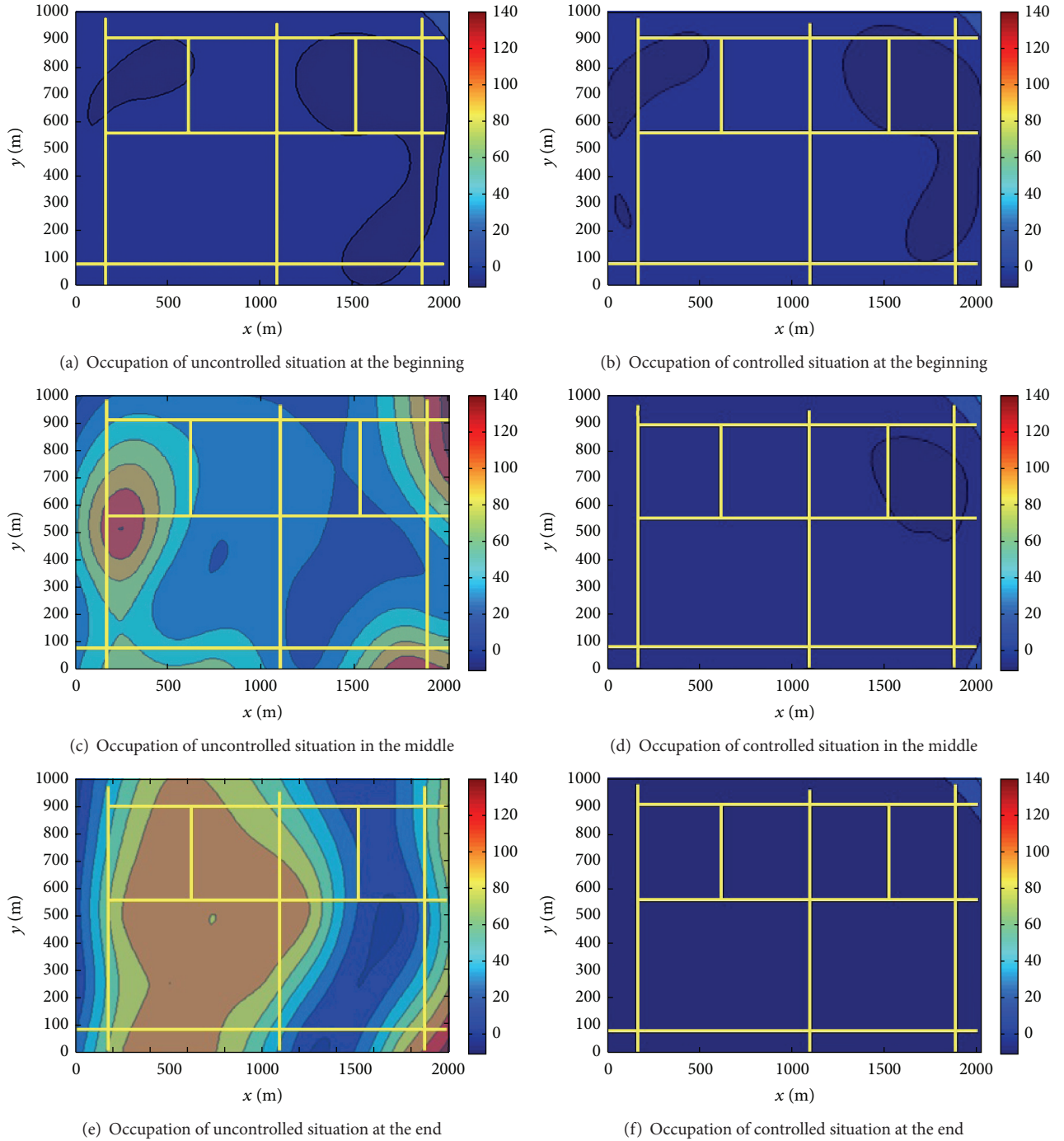
FIGURE 5: The occupation of uncontrolled/controlled during the simulation.

4. Simulation Results

4.1. Network and Gating Intersection Performance Analysis. In order to reveal the evolution process of the testing network without and with feedback gating control under fixed-time and classical actuated control cases, the occupation of the network is extracted as shown in Figure 5, and the total number of inflows, the queue length, and delay at gating intersection are shown in Figures 6(a)–6(f).

In Figure 5, we choose three periods of simulation to analyze the occupancy of network: the beginning of simulation, the middle of simulation, and the end of simulation time. The yellow line represents the roads in testing network.

As shown in Figures 5(a), 5(c), 5(e), 5(b), 5(d), and 5(f), the occupation of the network without control increases as time goes by while occupation of the network under control does not change significantly. Comparing Figures

(a) Inflow under fixed-time

(b) Inflow under actuated control

(c) Queue length under fixed-time

(d) Queue length under actuated control

(e) Delay under fixed-time

(f) Delay under actuated control
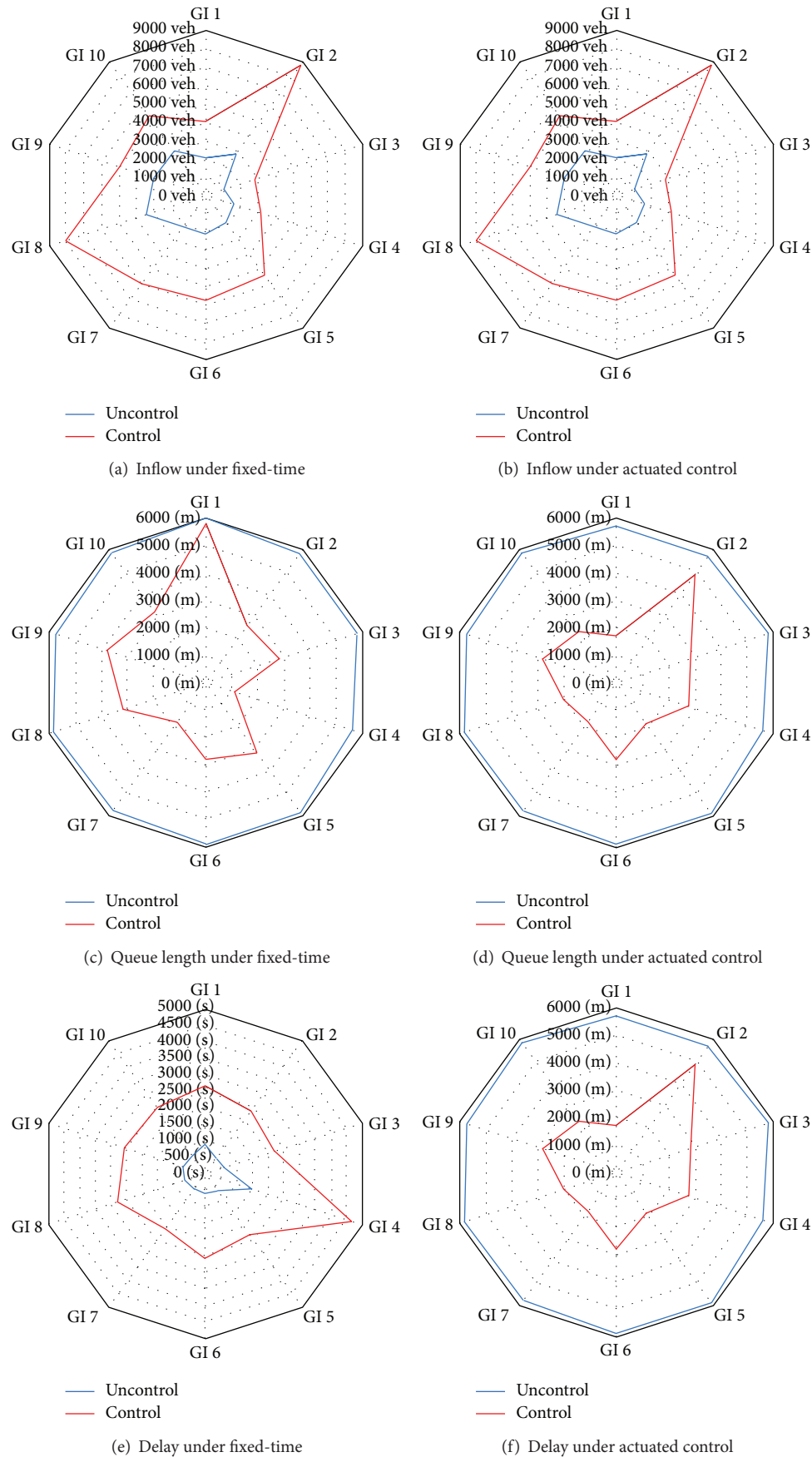
Figure 6: The total number of inflows, queue length, and delay of each approach.

(a) Fixed-time intersection in test area

(b) Actuated control intersection in test area
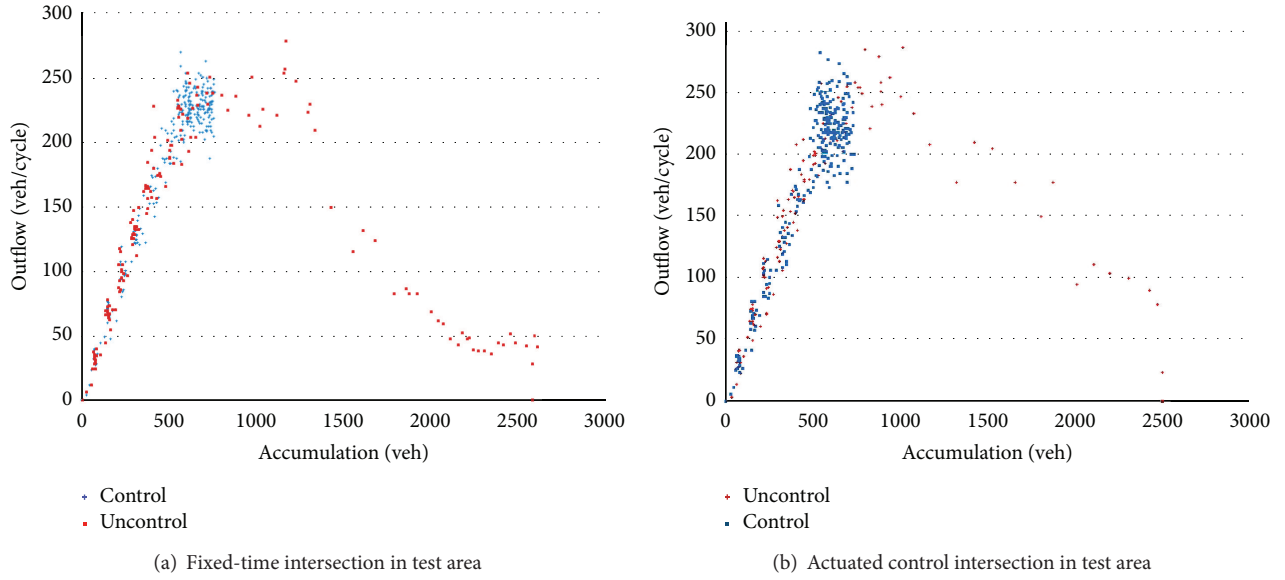
Figure 7: The MFD of two situations.

Table 6: The index of the test area under two situations.

Index	Fixed-time intersection in test area			Actuated control intersection in test area		
	Control	Without Control	Optimized proportion	Control	Without control	Optimized proportion
Average delay time per vehicle (s)	923	3704	75%	928	3767	75%
Total delay time (h)	13756	26966	49%	13698	26881	49%
Number of vehicles that have left the network	50211	20318	147%	49711	19874	150%
Total travel time (h)	18985	29167	35%	18873	29038	35%

5(a) and 5(b), the occupation is the same for both situations which means that at the beginning of simulation the network does not have difference under controlled and uncontrolled condition. The obvious differences shown in Figures 5(c) and 5(d) reflect that the test network without control becomes congested in some parts while this phenomenon does not appear in controlled condition since its occupation is still low. As time goes by, congestion in the test network without control keeps spreading and almost covers most of the network. However, congestion does not appear in the test network under control condition from the very beginning to the end.

Figures 6(a) and 6(b) show that the total amount of inflows from each gating intersection under the feedback control is far more than that under the situation of without control. Figures 6(c), 6(e), 6(d), and 6(f) reflect the same tendency of queue length and delay. Hence, the results of Figure 6 validate that the inflows are properly allocated to all gating intersections, and the proposed gating control policy reaches a Pareto improvement since the performance of all gating intersections is improved in terms of inflow rate, delay, and queue length.

In order to reveal the average improvements reached by the proposed feedback gating control policy, four performance measures including average delay, total delay, outflow, and total travel time are used to further evaluate the impacts of the proposed feedback control method. As shown in Table 6, all the performance indexes of network with control are fully superior to the situation without control. Moreover, there is no significant difference between fixed-time control and the actuated control, which demonstrates that the influence of the control policy implemented in the intersections within the network on MFD is negligible.

4.2. The Comparing of MFD. The MFDs of the test network with and without feedback control are presented in Figure 7. It can be seen that before accumulation arrives at the sweep spot, the two MFDs under the conditions of control and uncontrol are the same. With the increasing of accumulation, the network under the feedback gating control only displays the left half of the total MFD and reaches a stabile state around the optimal state. In contrast, the network without gating control has a complete MFD and the outflow approaches 0 as the accumulation exceeds the sweet point of the network.

4.3. The Comparing of Inflow/Outflow/Accumulation. The impacts of feedback control on inflow rate, outflow rate, and accumulation of vehicles in the network are presented in Figure 8. It can be seen from Figures 8(a) and 8(b) that, at the beginning of simulation, the inflow, outflow, and accumulation increase as the increase of demand outside of gating intersection, and there is no difference between feedback control and no control situations. However, before the network reaches sweet point, the state of network with and without feedback control exhibits different features. The maximum inflow rate under feedback control is around 250 (veh/cycle) while that under no control condition is a little bit higher, around 300 (veh/cycle). After 13000 s, the inflow under control is remaining at 200–250 (veh/cycle) while that under no control situation gradually decreases to 0 (veh/cycle), and the network finally gets blocked. The outflow of the network shows similar features as shown in Figures 8(c) and 8(d).

As shown in Figure 8(c), the vehicles accumulation rate of the network is the same for both control and uncontrol situations. After around 15000 s, the differences of vehicle accumulation rate under two situations becomes significant. The vehicle accumulation rate gradually turns to be stabilized at around 800 (veh). In contrast, the vehicle accumulation rate under no control situation increased rapidly and finally stabilized at around 2500 (veh), and the network gets blocked at that time. Moreover, both actuated control and fixed-time control exhibit similar features. The analysis results validate that the proposed feedback control performs well in terms of increasing inflow and outflow of networks and preventing network congestion.

5. Conclusion

Based on the recently proposed concept of an operational urban MFD, this paper proposed a feedback gating control policy which can be used to mitigate network congestion by adjusting signal timings of gating intersections. The objective of the feedback gating control model is to maximize the outflow and distribute the allowed inflows according to external demand and capacity of each gating intersection. An example network is used to test the performance of proposed feedback gating control model. Two types of background signalization types for the intersections within the test network, fixed-time and actuated control, are considered. Through extensive simulation based analysis, this study has reached the following tentative conclusions:

(1) The proposed feedback gating control model can reach a Pareto improvement since the performance of both gating intersections and the whole network are improved.

(2) Under the feedback gating control, the inflows and outflows can be stabilized at a very high level instead of decreasing all the way to 0 as the increase of external demand.

(3) Compared with no control case, especially under heavy demand situations, the delay and queue length at all gating intersections are decreased dramatically. Moreover, the total inflows are distributed among all gating intersections properly. In this sense, the proposed approach reaches another Pareto improvement in terms of balancing performance of gating intersections.

Note that this paper has presented preliminary evaluation results for the proposed model and only one small network is used for testing. More extensive numerical experiments or simulation tests will be conducted to assess the effectiveness of the proposed model under various traffic demand patterns and network geometry configurations. Another possible extension to this study is to study a more comprehensive control policy for a network with multiple subnetworks.

The Variables and Their Definitions

$N(t)$:   The number of vehicles in the network at time t (veh)

$Q(t)$:   The number of vehicles inflowing the network from controlled intersection at time t (veh)

$q_{oi}(t)$:   The number of vehicles inflowing the network from uncontrolled intersection at time t (veh)

$q_{out}(t)$:   The number of vehicles outflowing the network at time t (veh)

$G(N(t))$:   The calculating formulae of optimum curve regression model of MFD

$\overline{N}$:   The optimal number of vehicles in the network (veh)

$\overline{q}_{out}$:   The maximum number of vehicles outflowing the network (veh)

$\overline{Q}$:   The maximum number of vehicles inflowing the network from controlled intersection (veh)

$\overline{q}_{oi}$:   The maximum number of vehicles outflowing the network from controlled intersection (veh)

$\Delta Q(t)$:   The increment number of vehicles inflowing the network from controlled intersection at time t (veh)

$\Delta q_{oi}(t)$:   The increment number of vehicles inflowing the network from uncontrolled intersection at time t (veh)

$\Delta q_{out}(t)$:   The increment number of vehicles outflowing the network at time t (veh)

$\Delta N(t)$:   The increment number of vehicles in the network at time t (veh)

T: The cycle of controlled intersection (s)

K_I: Integral gains of the feedback control

α_i: The distribution ratio of the number of vehicles inflowing the network from controlled intersection at time t

$q_i(t)$:   The demand of controlled intersection i at time t (veh)

(a) Inflow rate in fixed-time

(b) Inflow rate in actuated control

(c) Outflow rate in fixed-time

(d) Outflow rate in actuated control

(e) Accumulation in fixed-time

(f) Accumulation in actuated control
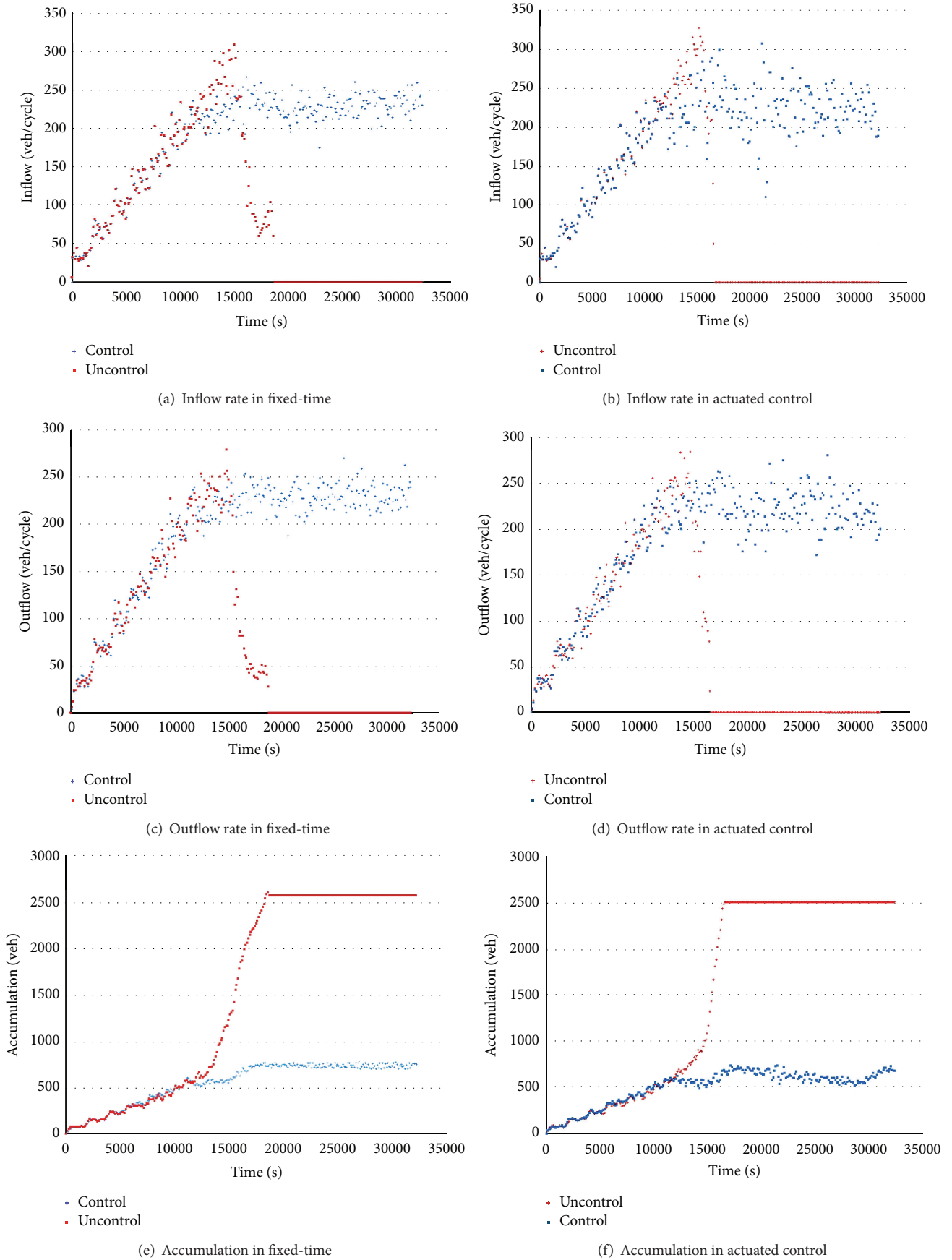
FIGURE 8: The comparing of under control and without control situations.

$q_{i,\text{in}}(t)$: The number of vehicles allowed to inflow to the network from controlled intersection i at time t (veh)

$q_{i,\text{out}}(t)$: The number of vehicles allowed to outflow to the network from intersection i at time t (veh)

S_i: The capacity of intersection i under the maximum green time (veh/cycle)

$S_i(t)$: The capacity of intersection i at time t (veh/cycle)

$t_i(t)$: The green time of intersection i at time t (s)

$t_{g,\text{max}}$: The maximum green time (s)

$t_{g,\text{min}}$: The minimum green time (s)

K_P: Proportional gains of the feedback control.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] D. Bretherton, G. Bowen, and K. Wood, "Effective urban traffic management and control: recent developments in SCOOT," in *Proceedings of the 92th Annual Meeting of Transportation Research Board*, Washington, DC, USA, 2003.

[2] A. Gal-Tzur, D. Mahalel, and J. N. Prashker, "Signal design for congested networks based on metering," *Transportation Research Record*, no. 1398, pp. 111–118, 1993.

[3] C. F. Daganzo, "Urban gridlock: macroscopic modeling and mitigation approaches," *Transportation Research Part B: Methodological*, vol. 41, no. 1, pp. 49–62, 2007.

[4] J. Ortigosa, M. Menendez, and H. Tapia, "Study on the number and location of measurement points for an MFD perimeter control scheme: a case study of Zurich," *EURO Journal on Transportation and Logistics*, vol. 3, no. 3-4, pp. 245–266, 2014.

[5] L. Leclercq, N. Chiabaut, and B. Trinquier, "Macroscopic fundamental diagrams: a cross-comparison of estimation methods," *Transportation Research Part B: Methodological*, vol. 62, pp. 1–12, 2014.

[6] B. Christine and C. Ladier, "Exploring the impact of the homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams," Transportation Research Record 2124, Transportation Research Board of the National Academies, Washington, DC, USA, 2009.

[7] N. Geroliminis and J. Sun, "Hysteresis phenomena of a Macroscopic Fundamental Diagram in freeway networks," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 9, pp. 966–979, 2011.

[8] V. V. Gayah and C. F. Daganzo, *Exploring the Effect of Turning Maneuvers and Route Choice on a Simple Network*, UC Berkeley Center for Future Urban Transport, A Volvo Center of Excellence, 2010.

[9] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1639–1656, 2012.

[10] L. Zhang, T. M. Garoni, and J. de Gier, "A comparative study of Macroscopic Fundamental Diagrams of arterial road networks governed by adaptive traffic signal systems," *Transportation Research Part B: Methodological*, vol. 49, pp. 1–23, 2013.

[11] N. Mühlich, V. V. Gayah, and M. Menendez, "Use of microsimulation for examination of macroscopic fundamental diagram hysteresis patterns for hierarchical urban street networks," *Transportation Research Record*, vol. 2491, pp. 117–126, 2015.

[12] J. Haddad and N. Geroliminis, "On the stability of traffic perimeter control in two-region urban cities," *Transportation Research Part B: Methodological*, vol. 46, no. 9, pp. 1159–1176, 2012.

[13] J. Haddad and A. Shraiber, "Robust perimeter control design for an urban region," *Transportation Research Part B: Methodological*, vol. 68, pp. 315–332, 2014.

[14] V. V. Gayah, X. Gao, and A. S. Nagle, "On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram," *Transportation Research Part B: Methodological*, vol. 70, pp. 255–268, 2014.

[15] N. Geroliminis and C. F. Daganzo, "Existence of urban-scale macroscopic fundamental diagrams: some experimental findings," *Transportation Research Part B: Methodological*, vol. 42, no. 9, pp. 759–770, 2008.

[16] Y. Li, J. Xu, and L. Shen, "A perimeter control strategy for oversaturated network preventing queue spillback," *Procedia—Social and Behavioral Sciences*, vol. 43, pp. 418–427, 2012.

[17] K. Aboudolas and N. Geroliminis, "Perimeter and boundary flow control in multi-reservoir heterogeneous networks," *Transportation Research Part B: Methodological*, vol. 55, pp. 265–281, 2013.

[18] M. Keyvan-Ekbatani, A. Kouvelas, I. Papamichail, and M. Papageorgiou, "Exploiting the fundamental diagram of urban networks for feedback-based gating," *Transportation Research Part B: Methodological*, vol. 46, no. 10, pp. 1393–1403, 2012.

[19] M. Keyvan-Ekbatani, M. Papageorgiou, and I. Papamichail, "Urban congestion gating control based on reduced operational network fundamental diagrams," *Transportation Research Part C: Emerging Technologies*, vol. 33, pp. 74–87, 2013.

[20] M. Keyvan-Ekbatani, M. Yildirimoglu, N. Geroliminis, and M. Papageorgiou, "Multiple concentric gating traffic control in large-scale urban networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2141–2154, 2015.

[21] M. Yildirimoglu, M. Ramezani, and N. Geroliminis, "Equilibrium analysis and route guidance in large-scale networks with MFD dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 59, pp. 404–420, 2015.

[22] W. Bin and L. Ye, *Traffic Management and Control*, China Communications Press, Beijing, China, 2009.

Research Article

Intention-Aware Autonomous Driving Decision-Making in an Uncontrolled Intersection

Weilong Song, Guangming Xiong, and Huiyan Chen

School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Guangming Xiong; xiongguangming@bit.edu.cn

Autonomous vehicles need to perform social accepted behaviors in complex urban scenarios including human-driven vehicles with uncertain intentions. This leads to many difficult decision-making problems, such as deciding a lane change maneuver and generating policies to pass through intersections. In this paper, we propose an intention-aware decision-making algorithm to solve this challenging problem in an uncontrolled intersection scenario. In order to consider uncertain intentions, we first develop a continuous hidden Markov model to predict both the high-level motion intention (e.g., turn right, turn left, and go straight) and the low level interaction intentions (e.g., yield status for related vehicles). Then a partially observable Markov decision process (POMDP) is built to model the general decision-making framework. Due to the difficulty in solving POMDP, we use proper assumptions and approximations to simplify this problem. A human-like policy generation mechanism is used to generate the possible candidates. Human-driven vehicles' future motion model is proposed to be applied in state transition process and the intention is updated during each prediction time step. The reward function, which considers the driving safety, traffic laws, time efficiency, and so forth, is designed to calculate the optimal policy. Finally, our method is evaluated in simulation with PreScan software and a driving simulator. The experiments show that our method could lead autonomous vehicle to pass through uncontrolled intersections safely and efficiently.

1. Introduction

Autonomous driving technology has developed rapidly in the last decade. In DARPA Urban Challenge [1], autonomous vehicles showed their abilities for interacting in some typical scenarios such as Tee intersections and lane driving. In 2011, Google released its autonomous driving platforms. Over 10,000 miles of autonomous driving for each vehicle was completed under various traffic conditions [2]. Besides, many big automobile companies also plan to launch their autonomous driving product in the next several years. With these significant progresses, autonomous vehicles have shown their potential to reduce the number of traffic accidents and solve the problem of traffic congestions.

One key challenge for autonomous vehicles driven in the real world is how to deal with the uncertainties, such as inaccuracy perception and unclear motion intentions. With the development of intelligent transportation system (ITS), the

perception uncertainty could be solved through the vehicle2X technology and the interactions between autonomous vehicles can be solved by centralized or decentralized cooperative control algorithms. However, human-driven vehicles will still be predominance in a short time and the uncertainties of their driving intentions will still be retained due to the lack of "intention sensor." Human drivers anticipate potential conflicts, continuously make decisions, and adjust their driving behaviors which are often not rational. Therefore, autonomous vehicles need to understand human drivers' driving intentions and choose proper actions to behave cooperatively.

In this paper, we focus on solving this problem in an uncontrolled intersection scenario. The uncontrolled intersection is a complex scenario with high accident rate. In US, stop signs can be used to normalize the vehicles' passing sequence. However, this kind of signs is rarely used in China and the right first traffic laws are often broken by some
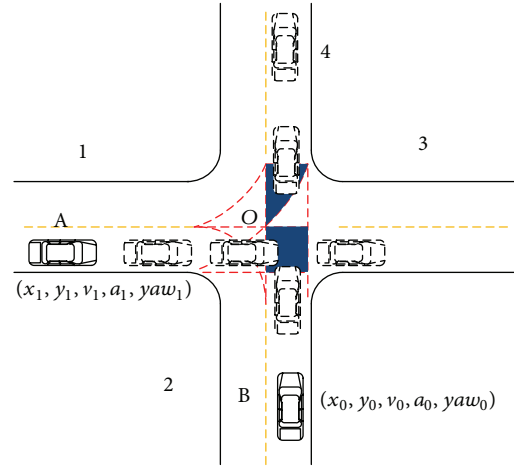
Figure 1: A motivation example. Autonomous vehicle B is going straight, while human-driven vehicle A has three potential driving directions: going straight, turning right, or turning left. If vehicle A turns right, it will not affect the normal driving of autonomous vehicle B. But the other maneuvers including turning left and going straight will lead to a passing sequence problem. Besides, if they have potential conflict, autonomous vehicle B will simulate the trajectories of vehicle A in a prediction horizon and gives the best actions in the current scenario. The vehicles drawn by dash lines are the future prediction positions. The red dash lines are the virtual lane assumption used in this paper, which means that the vehicles are considered to be driven inside the lane. The dark blue area is the potential collision region for these two cars.

aggressive drivers. Perception failures, misunderstandings, and wrong decisions are likely to be performed by human drivers. In such cases, even with stop signs, the "first come, first served" rule is likely to be broken. Besides, human driving behaviors are likely to change as time goes on. With these uncertain situations, specific layout, and the traffic rules, when autonomous vehicles approach an intersection, they should have potential ability to recognize the behavior of other vehicles and give a suitable corresponding behavior considering future evolution of the traffic scenario (see Figure 1).

With these requirements, we propose an intention-aware decision-making algorithm for autonomous driving in an uncontrolled intersection in this paper. Specifically, we first use easily observed features (e.g., velocity and position) and continuous hidden Markov model (HMM) [3] to build the intention prediction model, which outputs the lateral intentions (e.g., turn right, turn left, and go straight) for human-driven vehicles and longitudinal behavior (e.g., the yielding status) for related vehicles. Then, a generative partially observable Markov decision process (POMDP) framework is built to model the autonomous driving decision-making process. This framework is able to deal with the uncertainties in the environment, including human-driven vehicles' driving intentions. However, it is intractable to compute the optimal policy for general POMDP due to its complexity. We make reasonable approximations and assumptions to solve this problem in a low computational way. A human-like policy generation mechanism is used to compute the potential policy set. A scenario prediction mechanism is used to simulate the future actions of human-driven vehicles based on their lateral and longitudinal intentions and the proper reward functions are designed to evaluate each strategy. Traffic time, safety, and laws are all considered to get the final reward equations. The proposed method has been well

evaluated during simulation. The main contributions of this paper are as follows:

(i) Modeling a generative autonomous driving decision-making framework considering uncertainties (e.g., human driver's intention) in the environment.

(ii) Building intention prediction model using easily observed parameters (e.g., velocity and position) for recognizing the realistic lateral and longitudinal behaviors of human-driven vehicles.

(iii) Using reasonable approximations and assumption to build an efficient solver based on the specific layout in an uncontrolled intersection area.

The structure of this paper is as follows. Section 2 reviews the related work and two-layer HMM-based intention prediction algorithm is discussed in Section 3. Section 4 models general autonomous driving decision-making process in a POMDP, while the approximations and the simplified solver are described in Section 5. In Section 6, we evaluate our algorithm in a simulated uncontrolled intersection scenario with PreScan software and a driver simulator. Finally, the conclusion and future work are discussed in Section 7.

2. Related Work

The decision-making module is one of the most important components of autonomous vehicles, connecting environment perception and vehicle control. Thus, numerous research works are performed to handle autonomous driving decision-making problem in the last decade. The most common method is to manually define specific driving rules corresponding to situations. Both finite state machines (FSMs) and hierarchical state machines (HSMs) are used to evaluate situations and decide in their framework [4–6]. In DARPA Urban Challenge (DUC), the winner Boss

used a rule-based behavior generation mechanism to obey the predefined driving rules based on the obstacle vehicles' metrics [1, 6]. Boss was able to check vehicle's acceleration abilities and the spaces to decide whether merging into a new lane or passing intersections is safe. Similarly, the decision-making system of "Junior" [7], ranking second in DUC, was based on a HSM with manually defined 13 states. Due to the advantages including implementing simply and traceability, this framework is widely used in many autonomous driving platforms. However, these approaches always use constant velocity assumptions and lack considering surrounding vehicles future reactions to host vehicle's actions. Without this ability, the driving decisions could have potential risks [8].

In order to consider the evolution of future scenario, the planning and utility-based approaches have been proposed for decision-making. Bahram et al. proposed a prediction based reactive strategy to generate autonomous driving strategies [9]. A Bayesian classifier is used to predict the future motion of obstacle vehicles and a tree-based searching mechanism is designed to find the optimal driving strategy using multilevel cost functions. However, the surrounding vehicles' reactions to autonomous vehicles' actions are not considered in their framework. Wei et al. proposed a comprehensive approach for autonomous driver model by emulating human driving behavior [10]. The human-driven vehicles are assumed to follow a proper social behavior model and the best velocity profiles are generated in autonomous freeway driving applications. Nonetheless, their method does not consider the motion intention of human-driven vehicles and only targets in-lane driving. In their subsequent work, Wei et al. modeled traffic interactions and realized autonomous vehicle social behavior in highway entrance ramp [11]. The human-driven vehicles' motion intentions are modeled by a Bayesian model and the human-driven vehicles' future reactions are introduced, which is based on the yielding/not-yielding intentions at the first prediction step. Autonomous vehicles could perform social cooperative behavior using their framework. However, they do not consider the intention uncertainty over prediction time step.

POMDPs provide a mathematical framework for solving the decision-making problem with uncertainties. Bai et al. proposed an intention-aware approach for autonomous driving in scenarios with many pedestrians (e.g., in campus) [12]. In their framework, the hybrid A^* algorithm is used to generate global path, while a POMDP planner is used to control the velocity of the autonomous vehicle solving by an online POMDP solver DESPOT [13]. Brechtel et al. presented a probabilistic decision-making algorithm using continuous POMDP [14]. They focus on dealing with the uncertainties of incomplete and inaccurate perception in the intersection area, while our goal is to deal with the uncertain intentions of human-driven vehicles. However, the online POMDP solver always needs large computation resource and consumes much time [15, 16], which limits its use in real world autonomous driving platform. Ulbrich and Maurer designed a two-step decision-making algorithm to reduce the complexity of the POMDP in lane change scenario [17]. Eight POMDP states are manually defined to simplify the problem in their framework. Cunningham et al. proposed a

multipolicy decision-making method in lane changing and merging scenarios [18]. POMDPs are used to model the decision-making problem in their paper, while multivehicle simulation mechanism is used to generate the optimal high-level policy for autonomous vehicle to execute. However, the motion intentions are not considered.

Overall, the autonomous driving decision-making problem with uncertain driving intention is still a challenging problem. It is necessary to build an effective behavior prediction model for human-driven vehicles. Besides, it is essential to incorporate human-driven vehicles' intentions and behaviors into autonomous vehicle decision-making system and generate suitable actions to ensure autonomous vehicles drive safely and efficiently. This work addresses this problem by first building a HMM-based intention prediction model, then modeling human-driven vehicle's intentions in a POMDP framework, and finally solving it in an approximate method.

3. HMM-Based Intention Prediction

In order to pass through an uncontrolled intersection, autonomous vehicles should have the ability to predict the driving intentions of human-driven vehicles. Estimating driver's behavior is very difficult, because the state of a vehicle driver is in some high-dimensional feature space. Instead of using driver related features (e.g., gas pedal, brake pedal, and drivers' vision), easily observed parameters are used to build the intention prediction model in this paper.

The vehicle motion intention I considered in this paper is divided into two aspects, lateral intention $I_{lat} \in \{I_{TR}, I_{TL}, I_{GS}, I_S\}$ (i.e., turn right, turn left, go straight, and stop) and longitudinal intention $I_{lon} \in \{I_{Yield}, I_{NYield}\}$. The lateral intention is a high-level driving maneuver, which is determined by human drivers' long term decision-making process. This intention is not always changed in the driving process and determines the future trajectory of human-driven vehicles. In particular, the intention of stop is treated as a lateral intention in our model because it can be predicted only using data from human-driven vehicle itself. However, the longitudinal intention is a cooperative behavior only occurring when it interacts with other vehicles. We will first describe the HMM and then formulize our intention prediction model in this section.

3.1. HMM. A HMM consists of a set of N finite "hidden" states and a set of M observable symbols per state. The state transition probabilities are defined as $\mathbf{A} = \{a_{ij}\}$, where

$$a_{ij} = P[q_{t+1} = j \mid q_t = i], \quad 1 \le i, j \le N. \tag{1}$$

The initial state distribution is denoted as $\pi = \{\pi_i\}$, where

$$\pi_i = P[q_1 = i], \quad 1 \le i \le N. \tag{2}$$

Because the observation symbols are continuous parameters, we use Gaussian Mixture Model (GMM) [19] to represent their probability distribution functions (pdf):

$$b_j(\mathbf{o}) = \sum_{k=1}^{M} c_{jk} \mathcal{N}(\mathbf{o} \mid \mu_{jk}, \Sigma_{jk}), \qu 1 \le j \le N, \tag{3}$$

where c_{jk} represents the mixture coefficient in the jth state for the kth mixture. $\mathcal{N}$ is the pdf of a Gaussian distribution with mean μ and covariance Σ measured from observation $\mathbf{o}$. Mixture coefficient c satisfies the following constraints:

$$\sum_{k=1}^{M} c_{jk} = 1, \tag{4}$$

where $c_{jk} > 0$, $1 \le j \le N$, $1 \le k \le M$.

And

$$\int_{-\infty}^{+\infty} b_j (\mathbf{o}) \, d\mathbf{o} = 1, \quad 1 \le j \le N. \tag{5}$$

Then a HMM could be completely defined by hidden states N and the probability tuples $\boldsymbol{\lambda} = (\pi, A, C, \mu, \Sigma)$.

In the training process, we use the Baum-Welch method [20] to estimate model parameters for different driver intention I. Once the model parameters corresponding to different driver intention have been trained, we can perform the driver's intention estimation in the recognition process. The prediction process for lateral intentions can be seen in Figure 2.

3.2. HMM-Based Intention Prediction Process. Given a continuous HMM, the intention prediction process is divided into two steps. The first step focused on the lateral intention. The training inputs of each vehicle's lateral intention model in time t are defined as $B_{\text{lateral}} = \{L, v, a, \text{yaw}\}$, where L is the distance to the intersection, v is the longitudinal velocity, a is the longitudinal acceleration, and yaw is the yaw rate, while the output of this model is the motion intentions $I_{\text{lat}} \in \{I_{\text{TR}}, I_{\text{TL}}, I_{\text{GS}}, I_{\text{S}}\}$. The corresponding HMMs can be trained, including $\boldsymbol{\lambda}_{\text{TR}}$, $\boldsymbol{\lambda}_{\text{TL}}$, $\boldsymbol{\lambda}_{\text{GS}}$, and $\boldsymbol{\lambda}_{\text{S}}$.

The next step is about longitudinal intention. This probability could be decomposed based on the total probability formula:

$$\begin{aligned}
P\left(I_{\text{Yield}} \mid B\right) &= \sum P\left(I_{\text{Yield}} \mid I_{\text{lat}}, B\right) P\left(I_{\text{lat}} \mid B\right) \\
&= P\left(I_{\text{Yield}} \mid I_{\text{TR}}, B\right) P\left(I_{\text{TR}} \mid B\right) \\
&\quad + P\left(I_{\text{Yield}} \mid I_{\text{TL}}, B\right) P\left(I_{\text{TL}}\right) \\
&\quad + P\left(I_{\text{Yield}} \mid I_{\text{GS}}, B\right) P\left(I_{\text{GS}} \mid B\right) \\
&\quad + P\left(I_{\text{Yield}} \mid I_{\text{S}}, B\right) P\left(I_{\text{S}} \mid B\right),
\end{aligned} \tag{6}$$

where B is the behavior data including B_{lateral} and B_{lon}.

In this process, we assume that the lateral behavior I_{lat} is predicted correctly by a deterministic HMM in the first step, and therefore I_{lat} is determined by the lateral prediction result $I_{\text{latPredict}}$, where $P(I_{\text{lat}} \mid B, I_{\text{lat}} = I_{\text{latPredict}}) = 1$ and $P(I_{\text{lat}} \mid B, I_{\text{lat}}! = I_{\text{latPredict}}) = 0$. And (6) is reformulated by

$$\begin{aligned}
P\left(I_{\text{Yield}} \mid B\right) &= P\left(I_{\text{Yield}} \mid B, I_{\text{latPredict}}\right) P\left(I_{\text{latPredict}} \mid B\right) \\
&= P\left(I_{\text{Yield}} \mid B, I_{\text{latPredict}}\right).
\end{aligned} \tag{7}$$

The problem is changed to model $P(I_{\text{Yield}} \mid B, I_{\text{latPredict}})$. The features used in longitudinal intention prediction are

$B_{\text{lon}} = \{\Delta v, \Delta a, \Delta \text{DTC}\}$, where $\Delta v = v_{\text{social}} - v_{\text{host}}$, $\Delta a = a_{\text{social}} - a_{\text{host}}$, and $\Delta \text{DTC} = \text{DTC}_{\text{social}} - \text{DTC}_{\text{host}}$. DTC means the distance to the potential collision area. The output of the longitudinal intention prediction model is longitudinal motion intention $I_{\text{lon}} \in \{I_{\text{Yield}}, I_{\text{NYield}}\}$.

Instead of building a generative model, we use a deterministic approach to restrict $P(I_{\text{Yield}} \mid B, I_{\text{latPredict}})$ as 0 or 1. Thus, two types of HMMs named $\boldsymbol{\lambda}_{Y,I_{\text{lat}}}$, $\boldsymbol{\lambda}_{N,I_{\text{lat}}}$ are trained where $I_{\text{lat}} \in \{I_{\text{TR}}, I_{\text{TL}}, I_{\text{GS}}, I_{\text{S}}\}$. Two test examples for lateral and longitudinal intention prediction are shown in Figures 3 and 4. Through these two figures, we can find that our approach can recognize human-driven vehicle's lateral and longitudinal intention successfully.

4. Modeling Autonomous Driving Decision-Making in a POMDP Framework

For the decision-making process, the key problem is how to design a policy to perform the optimal actions with uncertainties. This needs to not only obtain traffic laws but also consider the driving uncertainties of human-driven vehicles. Facing potential conflicts, human-driven vehicles have uncertain probabilities to yield autonomous vehicles and some aggressive drivers may violate the traffic laws. Such elements should be implemented into a powerful decision-making framework. As a result, we model autonomous driving decision-making problem in a general POMDP framework in this section.

4.1. POMDP Preliminaries. A POMDP model can be formulized as a tuple $\{\mathcal{S}, \mathcal{A}, T, \mathcal{Z}, O, R, \gamma\}$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is the action space, and $\mathcal{Z}$ denotes a set of observations. The conditional function $T(s', a, s) = \Pr(s' \mid s, a)$ models transition probabilities to state $s' \in \mathcal{S}$, when the system takes an action $a \in \mathcal{A}$ in the state $s \in \mathcal{S}$. The observation function $O(z, s', a) = \Pr(z \mid s', a)$ models the probability of observing $z \in \mathcal{Z}$, when an action $a \in \mathcal{A}$ is taken and the end state is $s' \in \mathcal{S}$. The reward function $R(s, a)$ calculates an immediate reward when taking an action a in state s. $\gamma \in [0, 1]$ is the discount factor in order to balance the immediate and the future rewards.

Because the system contains partially observed state such as intentions, a belief $b \in \mathcal{B}$ is maintained. A belief update function τ is defined as $b' = \tau(b, a, z)$. If the agent takes action a and gets observation z, the new belief b' is obtained through the Bayes' rule:

$$b'\left(s'\right) = \eta O\left(s', a, z\right) \sum_{s \in S} T\left(s, a, s'\right) b\left(s\right), \tag{8}$$

where $\eta = 1 / \sum_{s' \in S} O(s', a, z) \sum_{s \in S} T(s, a, s') b(s)$ is a normalizing constant.

A key concept in POMDP planning is a policy, a mapping π that specifies the action $a = \pi(b)$ at belief b. To solve the POMDP, an optimal policy π^* should be designed to maximize the total reward:

$$\pi^* = \arg\max_{\pi} \left(E\left(\sum_{t=0}^{\infty} \gamma^t R\left(s_t, \pi\left(b_t\right)\right) \mid b_0, \pi \right) \right), \tag{9}$$

where b_0 is marked as the initial belief.
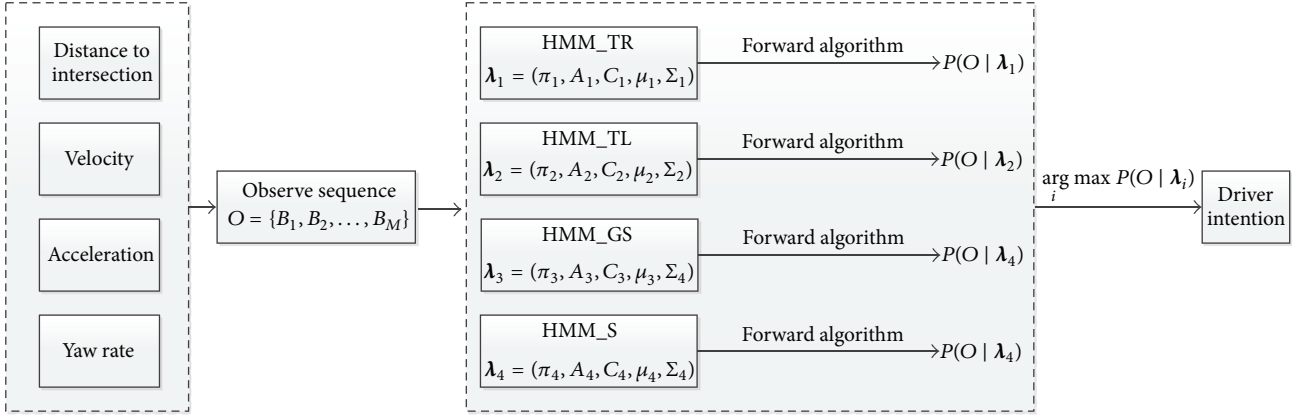
FIGURE 2: Prediction process for HMM. The observed sequence will be evaluated by four HMMs. Forward algorithm is used to calculate the conditional probabilities and the intention corresponding to the largest value will be considered as the vehicle's intention.
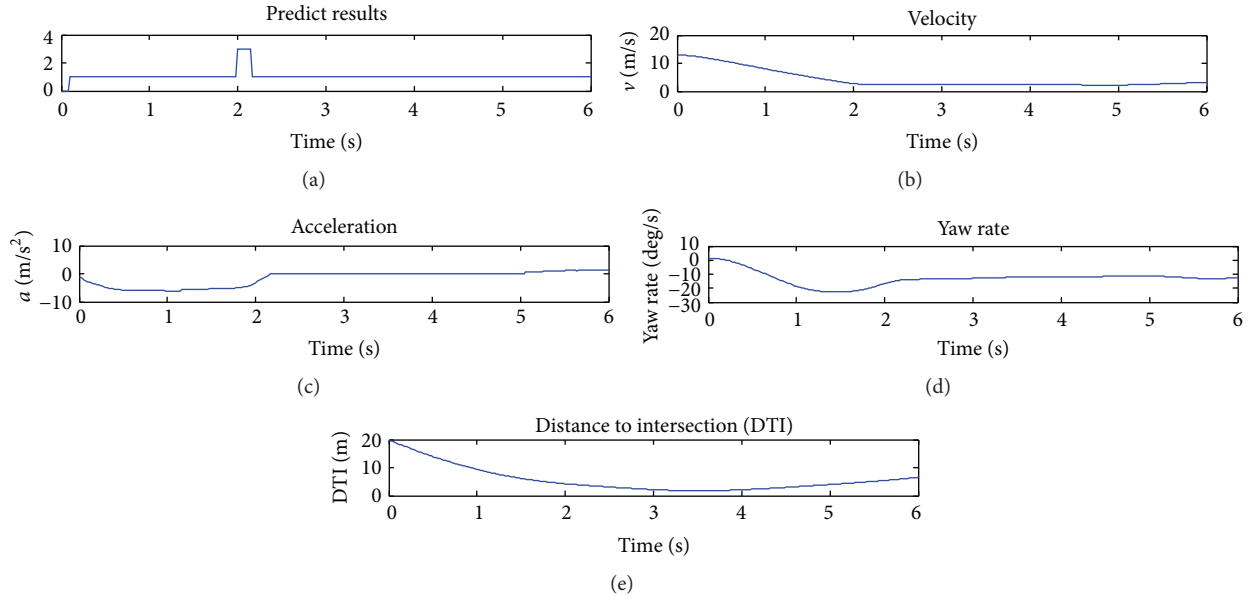
FIGURE 3: Lateral intention prediction example. The true intention of human-driven vehicle is to turn left in this scenario. In the first figure, the value 1 of the y label means turn left, 2 means turn right, 3 represents go straight, and 4 corresponds to stop.

4.2. State Space. Because of the Markov property, sufficient information should be contained in the state space $\mathcal{S}$ for decision-making process [14]. The state space includes the vehicle pose $[x, y, \theta]$, velocity v, the average yaw rate $\mathrm{yaw}_{\mathrm{ave}}$, and acceleration a_{ave} in the last planning period for all the vehicles. For the human-driven vehicles, the lateral and longitudinal intentions $[I_{\mathrm{lat}}, I_{\mathrm{lon}}]$ also need to be contained for state transition modeling. However, the road context knowledge is static reference information so that it will be not added to the state space.

The joint state $s \in \mathcal{S}$ could be denoted as $s = [s_{\mathrm{host}}, s_1, s_2, \ldots, s_N]^T$, where s_{host} is the state of host vehicle (autonomous vehicle), s_i, $i \in \{1, 2, 3, \ldots, N\}$, is the state of human-driven vehicles, and N is the number of human-driven vehicles involved. Let us define metric state $x =$ $[x, y, \theta, v, a_{\mathrm{ave}}, \mathrm{yaw}_{\mathrm{ave}}]^T$, including the vehicle position, heading, velocity, acceleration, and yaw rate. Thus, the state of host vehicle can be defined as $s_{\mathrm{host}} = x_{\mathrm{host}}$, while the human-driven vehicle state s_i is $s_i = [x_i, I_{\mathrm{lat},i}, I_{\mathrm{lon},i}]^T$. With the advanced perception system and V2V communication technology, we assume that the metric state x could be observed. Because the sensor noise is small and hardly affects decision-making process, we do not model observation noise for the metric state. However, the intention state cannot be directly observed, so it is the partially observable variables in our paper. The intention state should be inferred from observation data and predictive model over time.

4.3. Action Space. In our autonomous vehicle system, the decision-making system is used to select the suitable tactical
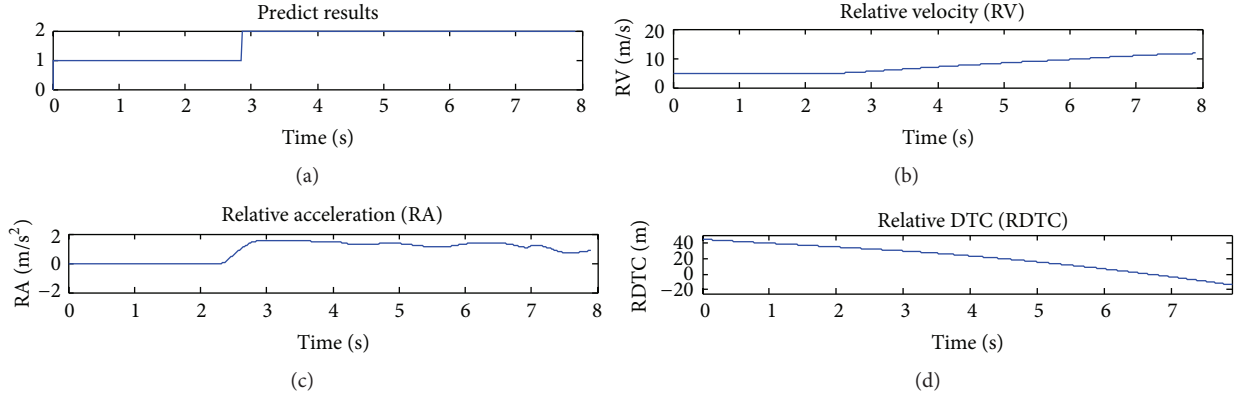
FIGURE 4: One example of predicting longitudinal intentions. This example is based on the scenario of Figure 1 and two vehicles both go straight. The value 1 of y-axis in the first figure denotes the intention of yielding, while 2 represents not yielding. In the first 2.8 s, the intention is yielding. After that, due to the acceleration action and less relative DTC, autonomous vehicle could understand human-driven vehicle's not-yielding intention.

maneuvers. Specifically, in the intersection area autonomous vehicles should follow a global reference path generated by path planning module. The decision-making module only needs to generate acceleration/deceleration commands to the control layer. As the reference path may not be straight, the steering control module can adjust the front wheel angle to follow the reference path. Therefore, the action space $\mathscr{A}$ could be defined as a discrete set $\mathscr{A} = [\text{acc}, \text{dec}, \text{con}]$, which contains commands including acceleration, deceleration, and maintaining current velocity.

4.4. Observation Space. Similar to the joint state space, the observation z is denoted as $z = [z_{\text{host}}, z_1, z_2, \ldots, z_N]^T$, where z_{host} and z_i are the host vehicle and human-driven vehicle's observations, respectively. The acceleration and yaw rate can be approximately calculated by speed and heading in the consecutive states.

4.5. State Transition Model. In state transition process, we need to model transition probability $\Pr(s' \mid s, a)$. This probability is determined by each targeted element in the scenario. So the transition model can be calculated by the following probabilistic equation:

$$\Pr\left(s' \mid s, a\right) = \Pr\left(s'_{\text{host}} \mid s_{\text{host}}, a_{\text{host}}\right) \prod_{i=1}^{N} \Pr\left(s'_i \mid s_i\right). \quad (10)$$

In the decision-making layer, we do not need to consider complex vehicle dynamic model. Thus, the host vehicle's motion $\Pr(s'_{\text{host}} \mid s_{\text{host}}, a_{\text{host}})$ can be simply represented by the following equations given action a:

$$x' = x + \left(v + \frac{a\Delta t}{2}\right) \Delta t \cos\left(\theta + \Delta\theta\right),$$

$$y' = y + \left(v + \frac{a\Delta t}{2}\right) \Delta t \sin\left(\theta + \Delta\theta\right),$$

$$\theta' = \theta + \Delta\theta,$$

$$v' = v + a\Delta t,$$

$$\text{yaw}'_{\text{ave}} = \frac{\Delta\theta}{\Delta t},$$

$$a'_{\text{ave}} = a.$$

$$(11)$$

Thus, the key problem is converted to compute $\Pr(s'_i \mid s_i)$, the state transition probability of human-driven vehicles. Based on the total probability formula, this probability can be factorized as a sum in whole action space:

$$\Pr\left(s'_i \mid s_i\right) = \sum_{a_i} \Pr\left(s'_i \mid s_i, a_i\right) \Pr\left(a_i \mid s_i\right). \quad (12)$$

With this equation, we only need to calculate the state transition probability $\Pr(s'_i \mid s_i, a_i)$ given a specific action a_i and the probability of selecting this action $\Pr(a_i \mid s_i)$ under current state s_i.

Because the human-driven vehicles' state $s_i = [x_i, I_i]$, the probability $\Pr(s'_i \mid s_i, a_i)$ can be calculated as

$$\Pr\left(s'_i \mid s_i, a_i\right) = \Pr\left(x'_i, I'_i \mid x_i, I_i, a_i\right)$$

$$= \Pr\left(x'_i \mid x_i, I_i, a_i\right) \Pr\left(I'_i \mid x'_i, x_i, I_i, a_i\right). \quad (13)$$

With a certain action a_i, $\Pr(x'_i \mid x_i, I_i, a_i)$ is equal to $\Pr(x'_i \mid x_i, I_{\text{lat},i}, a_i)$. The lateral behavior $I_{\text{lat},i}$ is considered to be a goal-directed driving intention which will not be changed in the driving process. So $\Pr(x'_i \mid x_i, I_{\text{lat},i}, a_i)$ is equal to $\Pr(x'_i \mid x_i, a_i)$ given a reference path corresponding to the intention of $I_{\text{lat},i}$. Using (11), $\Pr(x'_i \mid x_i, a_i)$ can be well solved.

The remaining problem for calculating $\Pr(s'_i \mid s_i, a_i)$ is to deal with $\Pr(I'_i \mid x'_i, x_i, I_i, a_i)$. The lateral intention $I'_{\text{lat},i}$ is assumed stable through the above explanation. And the

longitudinal intention $I'_{\text{lon},i}$ is assumed to be not updated in this process. But it will be updated with new inputs in observation space.

Now $\Pr(s'_i \mid s_i, a_i)$ is well modeled and the remaining problem is to compute the probabilities $\Pr(a_i \mid s_i)$ of human-driven vehicles' future actions:

$$\begin{aligned} \Pr\left(a_i \mid s_i\right) &= \Pr\left(a_i \mid x_i, I_i\right) \\ &= \sum_{x'_{\text{host}}} \Pr\left(a_i \mid x'_{\text{host}}, x_i, I_i\right) \Pr\left(x'_{\text{host}} \mid x_i, I_i\right). \end{aligned} \tag{14}$$

Because x'_{host} is determined by the designed policy, $\Pr(x'_{\text{host}} \mid x_i, I_i)$ could be calculated by (11) given an action a_{host}. The probability $\Pr(a_i \mid x'_{\text{host}}, x_i, I_i)$ means the distribution of human-driven vehicles' actions given the new state x'_{host} of host vehicle, the current state of itself, and its intentions. Instead of building a complex probability model, we designed a deterministic mechanism to calculate the most possible action a_i given x'_{host}, x_i, and I_i.

In this prediction process, the host vehicle is assumed to be maintaining the current actions in the next time step and the action a_i will be leading human-driven vehicle passing through the potential collision area either in advance of host vehicle under the intention I_{NYield} or behind the host vehicle under the intention I_{Yield} to keep a safe distance d_{safe}. In the case with the intention of I_{NYield}, we can calculate the low boundary $a_{i,\text{low}}$ of a_i through the above process and determine the upper one using the largest comfort value $a_{i,\text{comfort}}$. If $a_{i,\text{comfort}} < a_{i,\text{low}}$, $a_{i,\text{low}}$ will be used as the human-driven vehicle's action. If not, we consider the targeted a_i following a normal distribution with mean value μ_{a_i} between $a_{i,\text{low}}$ and $a_{i,\text{comfort}}$. To simplify our model, we use the mean value of these two boundaries to represent human-driven vehicle's action a_i. Similarly, the case with the intention of I_{Yield} can be analyzed in the same process.

After these steps, the transition probability $\Pr(s' \mid s, a)$ is well formulized and the autonomous vehicle could have the ability to understand the future motion of the scenario through this model.

4.6. Observation Model.

The observation model is built to simulate the measurement process. The motion intention is updated in this process. The measurements of human-driven vehicles are modeled with conditional independent assumption. Thus, the observation model can be calculated as

$$\Pr\left(z \mid a, s'\right) = \Pr\left(z_{\text{host}} \mid s'_{\text{host}}\right) \prod_{i=1}^{N} \Pr\left(z_i \mid s'_i\right). \tag{15}$$

The host vehicle's observation function is denoted as

$$\Pr\left(z_{\text{host}} \mid s'_{\text{host}}\right) \sim \mathcal{N}\left(z_{\text{host}} \mid x'_{\text{host}}, \Sigma_{z_{\text{host}}}\right). \tag{16}$$

But in this paper, due to the use of V2V communication sensor, the observation error almost does not affect the decision-making result. The variance matrix is set as zero.

The human-driven vehicle's observation will follow the vehicle's motion intentions. Because we do not consider the observation error, the value in metric state will be the same as the state transition results. But the longitudinal intention of human-driven vehicles in the state space will be updated using the new observations and HMM mentioned in Section 3. The new observation space will be confirmed with the above step.

4.7. Reward Function.

The candidate policies have to satisfy several evaluation criterions. Autonomous vehicles should be driven safely and comfortably. At the same time, they should follow the traffic rules and reach the destination as soon as possible. As a result, we design objective function (17) considering three aspects including safety, time efficiency, and traffic laws, where μ_1, μ_2, and μ_3 are the weight coefficient:

$$\begin{aligned} R\left(s,a\right) &= \mu_1 R_{\text{safety}}\left(s,a\right) + \mu_2 R_{\text{time}}\left(s,a\right) \\ &\quad + \mu_3 R_{\text{law}}\left(s,a\right). \end{aligned} \tag{17}$$

The detailed information will be discussed in the following subsections. In addition, the factor of comfort will be considered and discussed in policy generation part (Section 5.1).

4.7.1. Safety Reward.

The safety reward function $R_{\text{safety}}(s,a)$ is based on the potential conflict status. In our strategy, safety reward is defined as a penalty. If there are no potential conflicts, the safety reward will be set as 0. A large penalty will be assigned due to the risk of collision status.

In an uncontrolled intersection, the four approaching directions are defined as $A_i \in \{1, 2, 3, 4\}$ (Figure 5). The driver's lateral intentions are defined as $I_{\text{lat}} \in \{I_{\text{TR}}, I_{\text{TL}}, I_{\text{GS}}, I_S\}$. So the driving trajectory for each vehicle in the intersection can be generally represented by A_i and $I_{\text{lat},j}$, and we marked it as $T_{A_i, I_{\text{lat},j}}$, $1 \le i \le 4$, $1 \le j \le 4$. The function F is used to judge the potential collision status, which is denoted as

$$F\left(T_x, T_y\right) = \begin{cases} 1, & \text{if potential conflict,} \\ 0, & \text{otherwise,} \end{cases} \tag{18}$$

where T_x and T_y are vehicles' maneuver $T_{A_i, I_{\text{lat},j}}$.

$F(x, y)$ can be calculated through relative direction between two cars, which is shown in Table 1.

The safety reward is based on the following items:

(i) If $F(x, y)$ is equal to 0, then the safety reward is equal to 0 due to the noncollision status.

(ii) If potential collision occurs, there will be a large penalty.

(iii) If $|\text{TTC}_i - \text{TTC}_{\text{host}}| < t_{\text{threshold}}$, there is a penalty depending on $|\text{TTC}_i - \text{TTC}_{\text{host}}|$ and TTC_{host}.

4.7.2. Traffic Law Reward.

Autonomous vehicles should follow traffic laws to interact with human-driven vehicles. Traffic law is modeled as a function $\text{Law}(x, y)$ for each two vehicles x and y

$$\text{Law}\left(T_x, T_y\right) = \begin{cases} 1, & \text{if } x \text{ is prior,} \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$
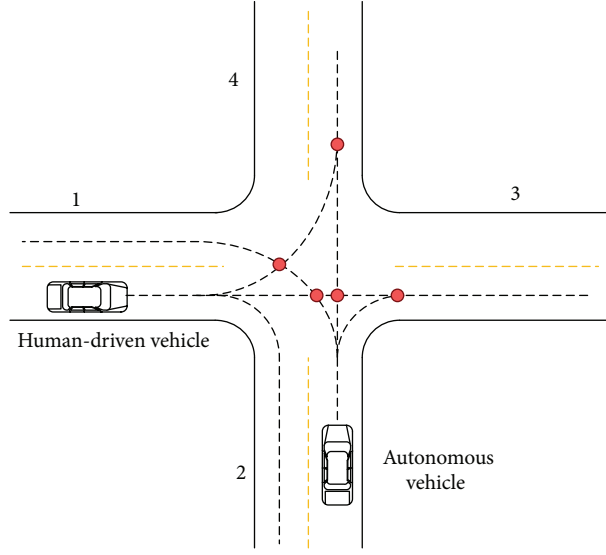
FIGURE 5: One typical scenario for calculating safety reward.

TABLE 1: Safe condition judgments in the intersection.

Human-driven vehicle	Left side				Right side				Opposite side				Same side			
Driving direction	TL	LK	TR	S	TL	LK	TR	S	TL	LK	TR	S	TL	LK	TR	S
Turn left	★	★	○	○	★	★	○	○	★	★	★	○	○	○	○	○
Lane keeping	★	★	○	○	★	★	★	○	★	○	○	○	○	○	○	○
Turn right	○	★	○	○	○	○	○	○	★	○	○	○	○	○	○	○
Stop	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

★ indicates potential collision. ○ indicates no potential collision.

where T_x and T_y are vehicles' maneuver T_{A_i,S_j}. This function Law(T_x, T_y) is formulized as shown in Algorithm 1.

If the behavior will break the law, a large penalty is applied and the behavior of obeying traffic laws will get a zero reward.

4.7.3. Time Reward. The time cost is based on the time to the destination for the targeted vehicles in the intersection area:

$$\text{Cost}_{\text{time}} = \frac{\text{DTG}}{v_{\text{host}}}. \tag{20}$$

DTG is the distance to the driving goal. In addition, we also need to consider the speed limit, which is discussed in policy generation part in Section 5.

5. Approximations on Solving POMDP Problem

Solving POMDP is quite difficult. The complexity of searching total brief space is $\mathcal{O}(|\mathcal{A}|^H |\mathcal{Z}|^H)$ [12], where H is the prediction horizon. In this paper, we model the intention recognition process as a deterministic model and use communication sensors to ignore the perception error, and thus the size of $|\mathcal{Z}|$ is reduced to 1 in the simplified problem. To solve this problem, we first generate the suitable potential policies according to the property of driving tasks and then

select the reasonable total predicting interval time and total horizon. After that, the approximate optimal policy can be calculated through searching all possible policies with maximum total reward. The policy selection process is shown in Algorithm 2 and some detailed explanations are discussed in the subsections.

5.1. Policy Generation. For autonomous driving near intersection, the desired velocity curves need to satisfy several constraints. Firstly, except for emergency braking, the acceleration constraints are applied to ensure comfort. Secondly, the speed limit constraints should be used in this process. We aim to avoid the acceleration commands when autonomous vehicle is reaching maximum speed limit. Thirdly, for the comfort purpose, the acceleration command should not be always changed. In other words, we need to minimize the jerk.

Similar to [11], the candidate policies are divided into three time segments. The first two segments are like "keep constant acceleration/deceleration actions," while keeping constant velocity in the third segment. We use t_1, t_2, and t_3 to represent the time periods of these three segments. To guarantee comfort, the acceleration is limited to the range from -4 m/s^2 to 2 m/s^2 and we discrete acceleration action into a multiple of $[-0.5, 0.5, 0]$. Then, the action space can be represented by a discretizing acceleration set. Then, we can set

$\text{Law}(T_x, T_y) \leftarrow 0$
$t_x \leftarrow d_{x2I}/v_x$
$t_y \leftarrow d_{y2I}/v_y$
if $t_x < t_y - \Delta t$, **then**
    $\text{Law}(T_x, T_y) \leftarrow 1$
else if $|t_x - t_y| < \Delta t$, **then**
    $\text{status} \leftarrow F(T_x, T_y)$
    if status = 1, **then**
        if $I_{\text{lat},x}$ = lanekeeping **and** $I_{\text{lat},y}$ <> lanekeeping, **then**
            $\text{Law}(T_x, T_y) \leftarrow 1$
        else if $I_{\text{lat},x}, I_{\text{lat},y}$ = lanekeeping **and** $A_x - A_y = 1$ or -3, **then**
            $\text{Law}(T_x, T_y) \leftarrow 1$
        else if $I_{\text{lat},x}$ = turnleft **and** $I_{\text{lat},y}$ = turnright, **then**
            $\text{Law}(T_x, T_y) \leftarrow 1$
        end if
    end if
end if
return $\text{Law}(T_x, T_y)$

ALGORITHM 1: Traffic law formulization.

Input:
    Predict horizon H, time step Δt,
    Current states: $s_{\text{host}} = x_{\text{host}}$, $s_{\text{human}} = [x_{\text{human}}, I_{\text{human}}]$
(1) $P \leftarrow \text{genenratepolicyset}()$
(2) **for each** $\pi_k \in P$, **do**
(3)   **for** $i = 1$ **to** $H/\Delta t$, **do**
(4) $a_{\text{host}} \leftarrow \pi_k(i)$
(5) $s'_{\text{host}} \leftarrow \text{updatestate}(s_{\text{host}}, a_{\text{host}})$
(6) $a_{\text{human}} \leftarrow \text{predictactions}(s'_{\text{host}}, s_{\text{human}}, I_{\text{human}})$
(7) $x'_{\text{human}} \leftarrow \text{updatestate}(x_{\text{human}}, a_{\text{human}})$
(8) $I'_{\text{human}} \leftarrow \text{updateintention}(s'_{\text{host}}, x'_{\text{human}})$
(9) $s'_{\text{human}} \leftarrow [x'_{\text{human}}, I'_{\text{human}}]$
(10) $R(i) = \text{calculatereward}(s'_{\text{host}}, s'_{\text{human}})$
(11) $s_{\text{host}} \leftarrow s'_{\text{host}}$
(12) $s_{\text{human}} \leftarrow s'_{\text{human}}$
(13)   **end**
(14) $R_k^{\text{total}} \leftarrow \sum_i \text{sum}(R(i))$
(15) **end**
(16) $k^* \leftarrow \arg\max_k(R_k^{\text{total}})$
(17) $\pi^* \leftarrow \pi_{k^*}$
(18) **return** π^*

ALGORITHM 2: Policy selection process.

the value of t_1, t_2, and t_3 and the prediction period of single step. An example of policy generation is shown in Figure 6.

5.2. Planning Horizon Selection. After building policy generation model, the next problem is to select a suitable planning horizon. Longer horizon can lead to a better solution but consuming more computing resources. However, as our purpose

is to deal with the interaction problem in the uncontrolled intersection, we only need to consider the situation before autonomous vehicle gets through. In our algorithm, we set the prediction horizon as 8 seconds. In addition, in the process of updating the future state of each vehicle using each policy, the car following mode is used after autonomous vehicle passes through the intersection area.
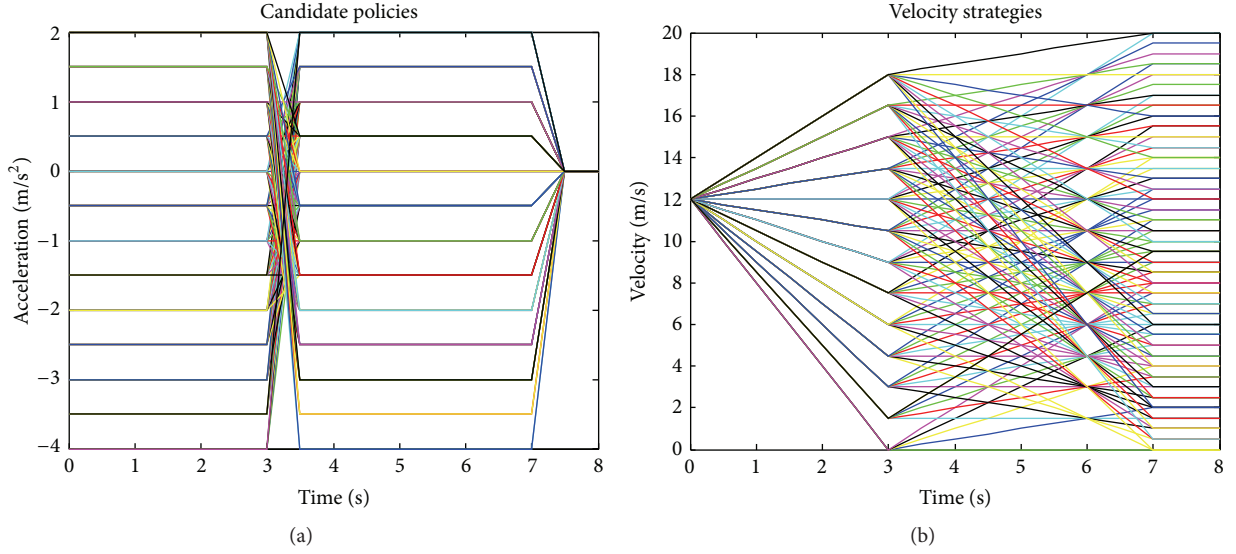
FIGURE 6: An example of policy generation process. (a) is the generated policies and (b) is the corresponding speed profiles. The interval of each prediction step is 0.5 s, current speed is 12 m/s², and the speed limit is 20 m/s². The bold black line is one policy. In the first 3 seconds, autonomous vehicles decelerate in −3.5 m/s², then accelerate at 2 m/s² for 4 seconds, and finally stop in the last one second. In this case, 109 policies were generated, which is suitable for replanning fast.
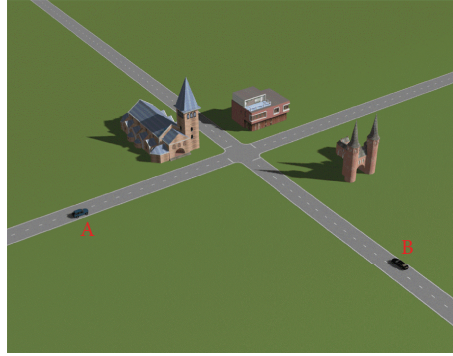
FIGURE 7: Testing scenario. Autonomous vehicle B and human-driven vehicle A are both approaching the uncontrolled intersection. To go across the intersection successfully, autonomous vehicle should interact with human-driven vehicle.

5.3. Time Step Selection. Another problem is the prediction time step. The intention prediction algorithm and the POMDP are computed in each step. If the time step is t_{step}, the total computation times will be H/t_{step}. Thus, smaller time step leads to more computation time. To solve this problem, we use a simple adaptive time step calculation mechanism to give a final value. The time step is selected based on the TTC of autonomous vehicle. If the host vehicle is far away from the intersection, we can use a very large time step. But if the TTC is quite small, the low t_{step} is applied to ensure safety.

6. Experiment and Results

6.1. Settings. In this paper, we evaluate our approach through PreScan 7.1.0 [21], a simulation tool for autonomous driving and connected vehicles. Using this software, we can build the testing scenarios (Figure 7) and add vehicles with dynamic model. In order to get a similar scenario considering social interaction, the driver simulator is added in our experiment (Figure 8). The human-driven vehicle is driven by several people during the experiment and the autonomous vehicle makes decisions based on the human-driven vehicle's driving behavior. The reference trajectory for autonomous vehicle is generated from path planning module and the human-driven vehicle's data (e.g., position, velocity, and heading) are transferred through V2V communication sensor. The decision-making module sends desired velocity command to the PID controlled to follow the reference path. All policies in the experiment part use a planning horizon $H = 8$ s, which is discretized into the time step of 0.5 s.

6.2. Results. It is difficult to compare different approaches in the same scenario because the environment is dynamic and not exactly the same. However, we select two typical situations and special settings to make it possible. The same initial conditions including position, orientation, and velocity

FIGURE 8: Logitech G27 driving simulator.

(a)   (b)
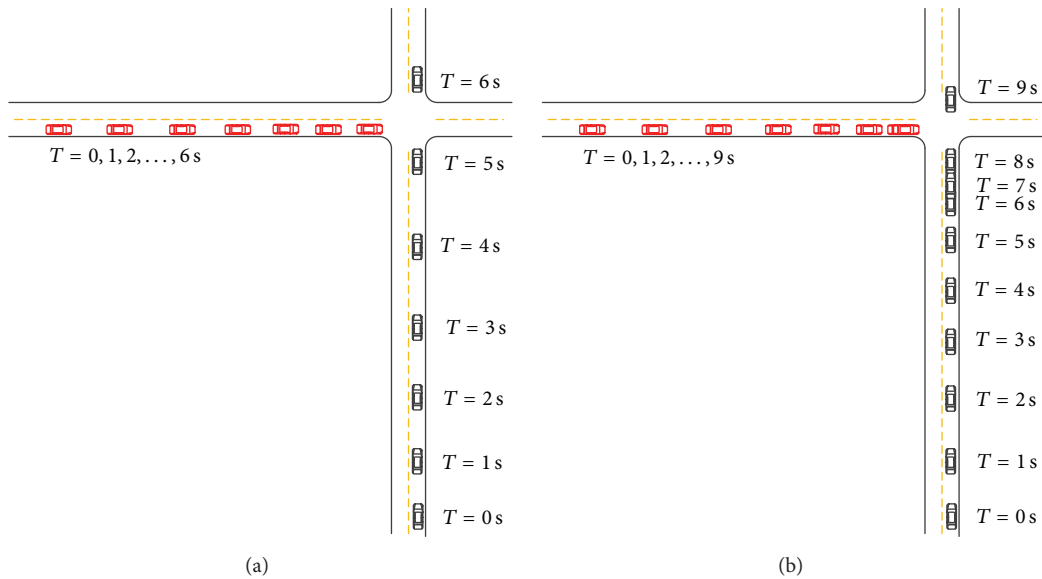
FIGURE 9: The visualized passing sequence. (a) is the result of our approach and (b) represents the result of reactive approach without considering intention. The black vehicle is an autonomous vehicle, while the red car is the human-driven vehicle. Each vehicle represents the position in a specific time T with an interval of 1 second.

for each vehicle are used in different tests. Besides, two typical situations, including human-driven vehicle getting through before or after autonomous vehicle, are compared in this section. With the same initial state, different reactions will occur based on various methods. We compare our approach and reactive-based method [6] in this section. The key difference for these two methods is that our approach considers human-driven vehicle's driving intention.

The first experiment is that human-driven vehicle tries to yield autonomous vehicle in the interaction process. The results are shown in Figures 9 and 10. Firstly, Figure 9 gives us a visual comparison of the different approaches. From almost the same initial state (e.g., position and velocity), our approach could lead to autonomous vehicle passing through the intersection more quickly and reasonable.

Then, let us look at Figure 10 for detailed explanation. In the first 1.2 s in Figures 10(a) and 10(c), autonomous vehicle maintains speed and understands that human-driven vehicle will not perform yielding actions. Then, autonomous vehicle gets yielding intention of human-driven vehicle and understands that human-driven vehicle's lateral intention is to go straight. Based on candidate policies, autonomous vehicle selects acceleration strategy with maximum reward and finally crosses the intersection. In this process, we can obviously find that autonomous vehicle understands human-driven vehicle's yielding intention. Figure 10(c) is an example of understand human-driven vehicle's behavior based on ego vehicle's future actions in a specific time. Our strategy predicts the future actions of human-driven vehicle. Although the velocity curves after 1 s do not correspond, it does not affect the performance of our methods. The reason is that
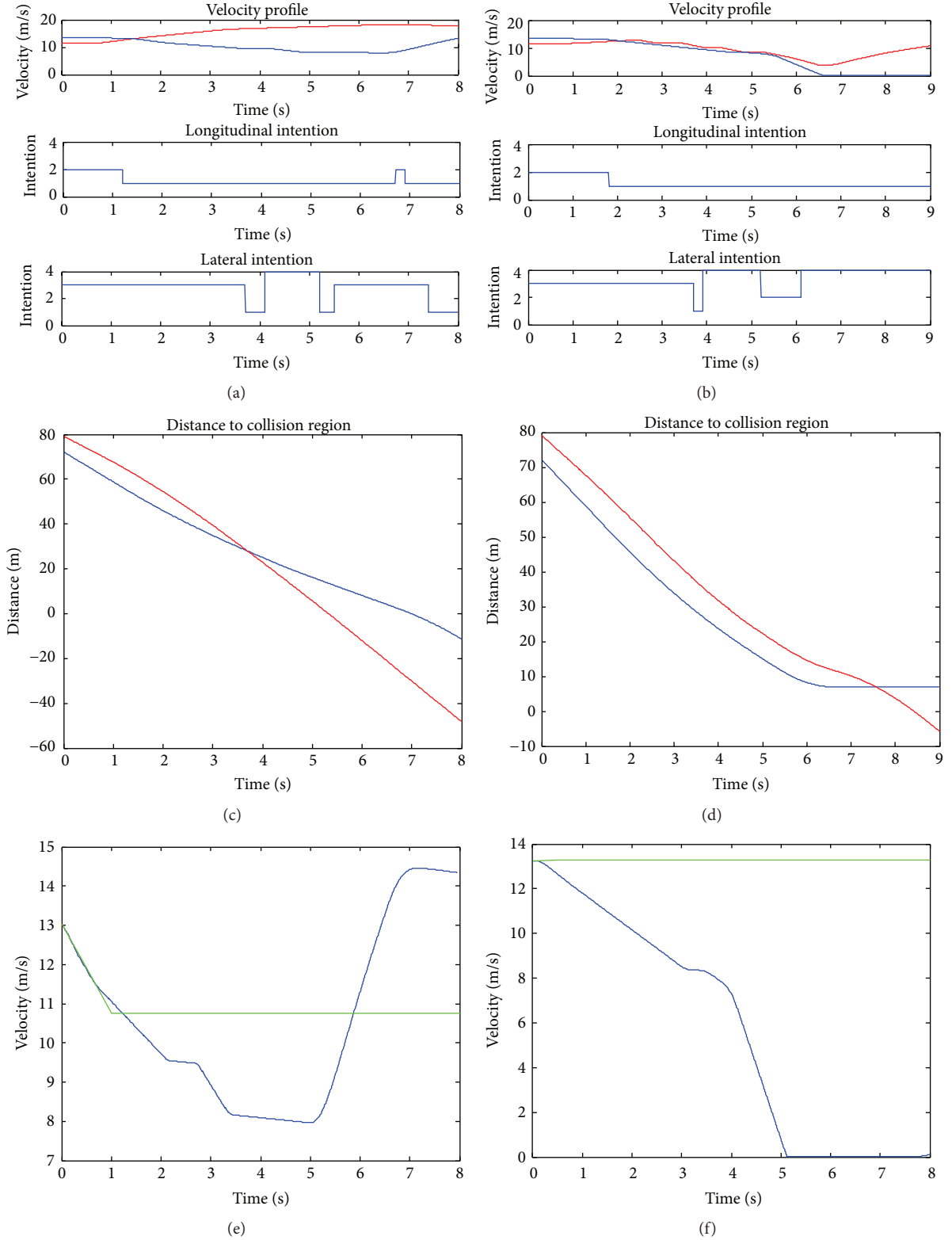
FIGURE 10: Case test 1. In this case, human-driven vehicle passes through intersection after autonomous vehicle. (a), (c), and (e) are the performance of our method, while (b), (d), and (f) are from the strategy without considering the driving intention. (a) and (b) are the velocity profiles and the corresponding driving intention. For longitudinal intention, label 1 means yielding and label 2 means not yielding. In lateral intention, 1 means turning left, 2 means turning right, 3 means going straight, and 4 means stop. The intentions in (b) are not used in that method but for detailed analysis. (c) and (d) are the distance to collision area for autonomous vehicle and human-driven vehicle, respectively. (e) and (f) are the prediction and true motions of human-driven vehicles in time 1.5 s with a prediction length of 8 s. The red curves in these subfigures are from autonomous vehicle while blue lines are from human-driven vehicle. The green lines in (e) and (f) are the prediction velocity curves of human-driven vehicle.

(a)   (b)
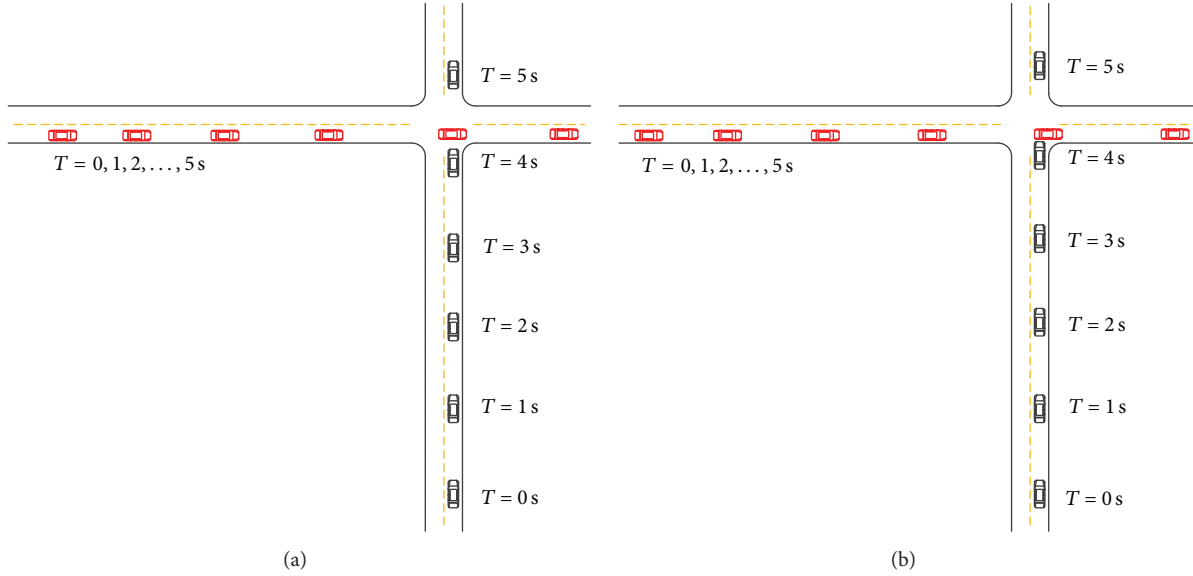
FIGURE 11: The visualized passing sequence for the case of human-driven vehicle first getting through. (a) is the result of our approach and (b) is the reactive-based approach.

we use a deterministic model in the prediction process and the prediction value is inside two boundaries to ensure safety. Besides, the whole actions of autonomous vehicle in this process could also help human-driven vehicle to understand not-yielding intention of autonomous vehicles. In this case, cooperative driving behaviors are performed by both vehicles.

However, if the intention is not considered in this process, we can find the results in Figures 10(b), 10(d), and 10(f). After 2 s in Figure 10(b), while the human-driven vehicle gives a yielding intention, autonomous vehicle could not understand and they find a potential collision based on the constant velocity assumptions. Then, it decreases the speed but the human-driven vehicle also slows down. The puzzled behavior leads both vehicles to slow down near intersection. Finally, human-driven vehicle stops at the stop line and then autonomous vehicle could pass the intersection. In this strategy, the human-driven vehicle's future motion is assumed to be constant (Figure 10(f)). Without understanding of human-driven vehicle's intentions, this strategy can increase congestion problem.

Another experiment is that human-driven vehicle tries to get through the intersection first. The results are shown in Figures 11 and 12. This case is quite typical because many traffic accidents in real world are happening in this situation. In detail, if one vehicle tries to cross an intersection while violating the law, another vehicle will be in great danger if it does not understand its behavior. From the visualized performance in Figure 11, our method is a little more safe than other approaches as there is nearly collision situation in Figure 11(b). In detail, we can see from Figure 12(a) that our strategy could perform deceleration actions after we understand the not-yielding intention in 0.8 s. However, without understanding human-driven vehicle's motion intention, the response time has a 1-second delay which may be quite dangerous.

In addition, it is shown that good performance is in the predictions of human-driven vehicle's future motion in our methods (Figure 12(e)).

The results of these two cases demonstrate that our algorithm could deal with typical scenarios and have better performance than traditional reactive controller. Autonomous vehicle could be driven more safely, fast, and comfortably through our strategy.

7. Conclusion and Future Work

In this paper, we proposed an autonomous driving decision-making algorithm considering human-driven vehicle's uncertain intentions in an uncontrolled intersection. The lateral and longitudinal intentions are recognized by a continuous HMM. Based on HMM and POMDP, we model general decision-making process and then use an approximate approach to solve this complex problem. Finally, we use PreScan software and a driving simulator to emulate social interaction process. The experiment results show that autonomous vehicles with our approach can pass through uncontrolled intersections more safely and efficiently than using the strategy without considering human-driven vehicles' driving intentions.

In the near future, we aim to implement our approach into a real autonomous vehicle and perform real world experiments. In addition, more precious intention recognition algorithm aims to be figured out. Some methods like probabilistic graphic model can be used to get a distribution of each intention. Finally, designing online POMDP planning algorithms is also valuable.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.
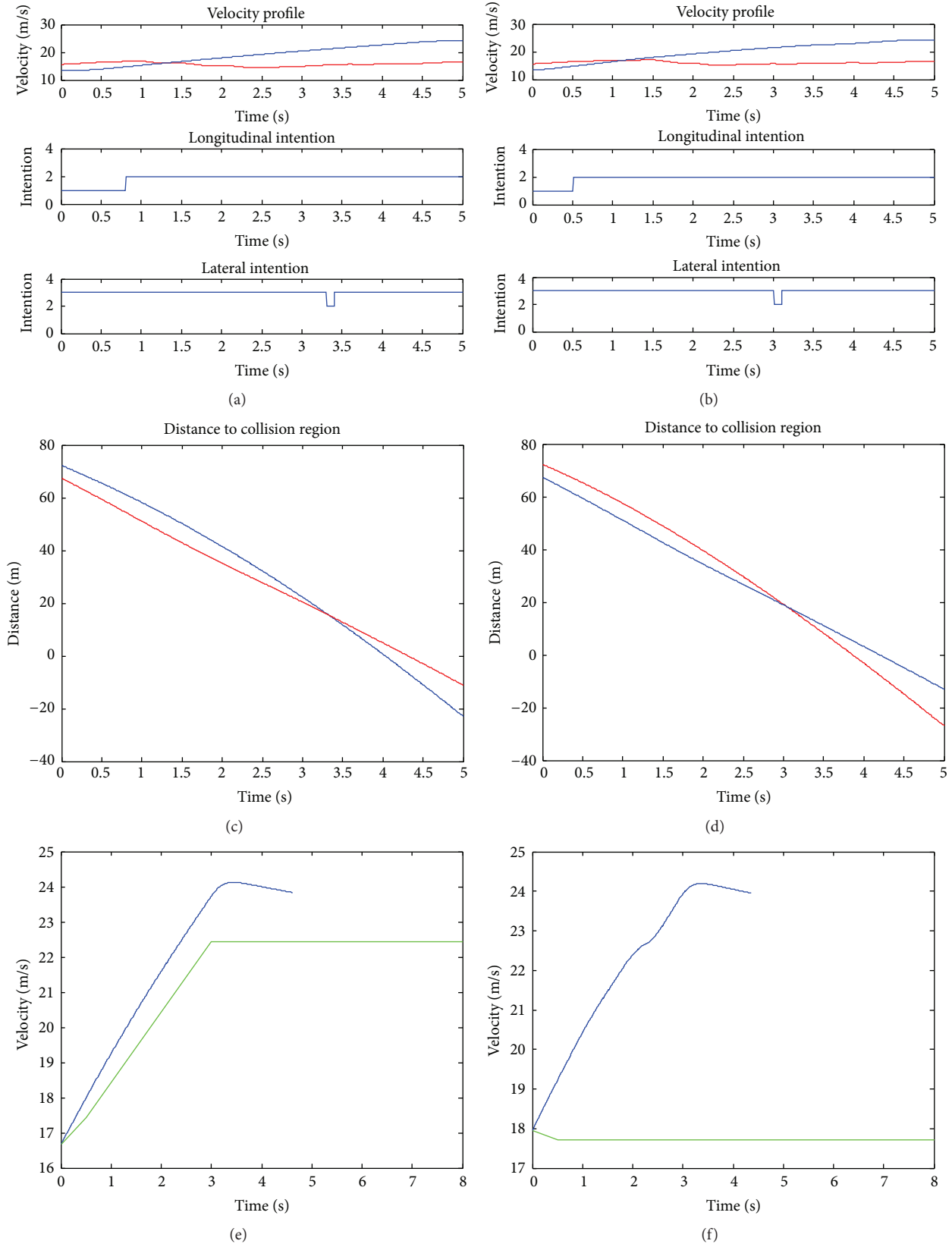
Figure 12: Case test 2. In this case, human-driven vehicle passes through intersection before autonomous vehicle through different strategies. The definition of each subfigure is the same as in Figure 10.

Acknowledgments

References

[1] C. Urmson, J. Anhalt, D. Bagnell et al., "Autonomous driving in urban environments: boss and the urban challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 425–466, 2008.

[2] J. Markoff, "Google cars drive themselves, in traffic," *New York Times*, vol. 9, 2010.

[3] L. R. Rabiner and B.-H. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4–16, 1986.

[4] M. Buehler, K. Iagnemma, and S. Singh, *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*, vol. 56, Springer, 2009.

[5] S. Kammel, J. Ziegler, B. Pitzer et al., "Team AnnieWAY's autonomous system for the 2007 DARPA Urban Challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 615–639, 2008.

[6] C. R. Baker and J. M. Dolan, "Traffic interaction in the urban challenge: putting boss on its best behavior," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '08)*, pp. 1752–1758, IEEE, Nice, France, September 2008.

[7] M. Montemerlo, J. Becker, S. Shat et al., "Junior: the Stanford entry in the urban challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 569–597, 2008.

[8] L. Fletcher, S. Teller, E. Olson et al., "The MIT-Cornell collision and why it happened," *Journal of Field Robotics*, vol. 25, no. 10, pp. 775–807, 2008.

[9] M. Bahram, A. Wolf, M. Aeberhard, and D. Wollherr, "A prediction-based reactive driving strategy for highly automated driving function on freeways," in *Proceedings of the 25th IEEE Intelligent Vehicles Symposium*, pp. 400–406, IEEE, Dearborn, Mich, USA, June 2014.

[10] J. Wei, J. M. Dolan, and B. Litkouhi, "A prediction- and cost function-based algorithm for robust autonomous freeway driving," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '10)*, pp. 512–517, San Diego, Calif, USA, June 2010.

[11] J. Wei, J. M. Dolan, and B. Litkouhi, "Autonomous vehicle social behavior for highway entrance ramp management," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '13)*, pp. 201–207, IEEE, Gold Coast, Australia, June 2013.

[12] H. Bai, S. Cai, N. Ye, D. Hsu, and W. S. Lee, "Intention-aware online POMDP planning for autonomous driving in a crowd," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '15)*, pp. 454–460, IEEE, Seattle, Wash, USA, May 2015.

[13] A. Somani, N. Ye, D. Hsu, and W. S. Lee, "Despot: online pomdp planning with regularization," in *Advances in Neural Information Processing Systems*, pp. 1772–1780, 2013.

[14] S. Brechtel, T. Gindele, and R. Dillmann, "Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs," in *Proceedings of the 17th IEEE International Conference on Intelligent Transportation Systems (ITSC '14)*, pp. 392–399, Qingdao, China, October 2014.

[15] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Mathematics of Operations Research*, vol. 12, no. 3, pp. 441–450, 1987.

[16] O. Madani, S. Hanks, and A. Condon, "On the undecidability of probabilistic planning and related stochastic optimization problems," *Artificial Intelligence*, vol. 147, no. 1-2, pp. 5–34, 2003.

[17] S. Ulbrich and M. Maurer, "Probabilistic online POMDP decision making for lane changes in fully automated driving," in *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC '13)*, pp. 2063–2067, The Hague, The Netherlands, October 2013.

[18] A. G. Cunningham, E. Galceran, R. M. Eustice, and E. Olson, "MPDM: multipolicy decision-making in dynamic, uncertain environments for autonomous driving," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '15)*, pp. 1670–1677, Seattle, Wash, USA, May 2015.

[19] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 1, pp. 72–83, 1995.

[20] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Annals of Mathematical Statistics*, vol. 37, pp. 1554–1563, 1966.

[21] M. Tideman and M. Van Noort, "A simulation tool suite for developing connected vehicle systems," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IEEE IV '13)*, pp. 713–718, Queensland, Australia, June 2013.

Research Article

Evaluation of a Trapezoidal Predictive Controller for a Four-Wire Active Power Filter for Utility Equipment of Metro Railway, Power-Land Substations

Sergio Salas-Duarte,[1] Ismael Araujo-Vargas,[1] Jazmin Ramirez-Hernandez,[1] and Marco Rivera[2]

[1]*Escuela Superior de Ingeniería Mecanica y Eléctrica, Unidad Culhuacan, Instituto Politécnico Nacional, Avenida Santa Ana No. 1000, Col San Francisco Culhuacan, 04430 México, DF, Mexico*
[2]*Universidad de Talca, 2 Norte 685, Talca, Chile*

Correspondence should be addressed to Ismael Araujo-Vargas; iaraujo@ipn.mx

The realization of an improved predictive current controller based on a trapezoidal model is described, and the impact of this technique is assessed on the performance of a 2 kW, 21.6 kHz, four-wire, Active Power Filter for utility equipment of Metro Railway, Power-Land Substations. The operation of the trapezoidal predictive current controller is contrasted with that of a typical predictive control technique, based on a single Euler approximation, which has demonstrated generation of high-quality line currents, each using a 400 V DC link to improve the power quality of an unbalanced nonlinear load of Metro Railway. The results show that the supply current waveforms become virtually sinusoidal waves, reducing the current ripple by 50% and improving its power factor from 0.8 to 0.989 when the active filter is operated with a 1.6 kW load. The principle of operation of the trapezoidal predictive controller is analysed together with a description of its practical development, showing experimental results obtained with a 2 kW prototype.

1. Introduction

The use of Active Power Filters (APFs) in the electrical grid is critical for on-land transportation applications, such as Metropolitan Railway Substations, which reduce the flowing of current harmonics caused by the increased utilization of nonlinear loads, whilst improving the power quality of the supply. APFs are an attractive solution to comply with the national and international power quality standards at every level of the network infrastructure, [1–3], since high-performance switching devices appear available in the market to develop power converters [4]. In addition, the development of fast and versatile microprocessors has facilitated the implementation of nonlinear control techniques, and thereby, APFs are becoming accurate power processors that reshape clean sinusoidal supply currents [5–9].

Four-wire shunt APFs are a commonplace strategy that exhibit attractive characteristics to inject currents and reshape the line currents drawn by unbalanced nonlinear loads, whilst providing a path to cancel the neutral current by using either an additional switching limb or a split DC link [10, 11]. These circuits typically incur in the use of a power theory to calculate the reference currents [12], such that the filter may operate as a current amplifier that injects compensating currents to the grid, causing a complex transistor switching scheme since the generated filter currents must track the references. Predictive control is an attractive method for controlling current waveforms in three-phase converters [6, 7, 13–20], since a piecewise linear model of the converter is used together with a cost function to determine an appropriate converter switching.
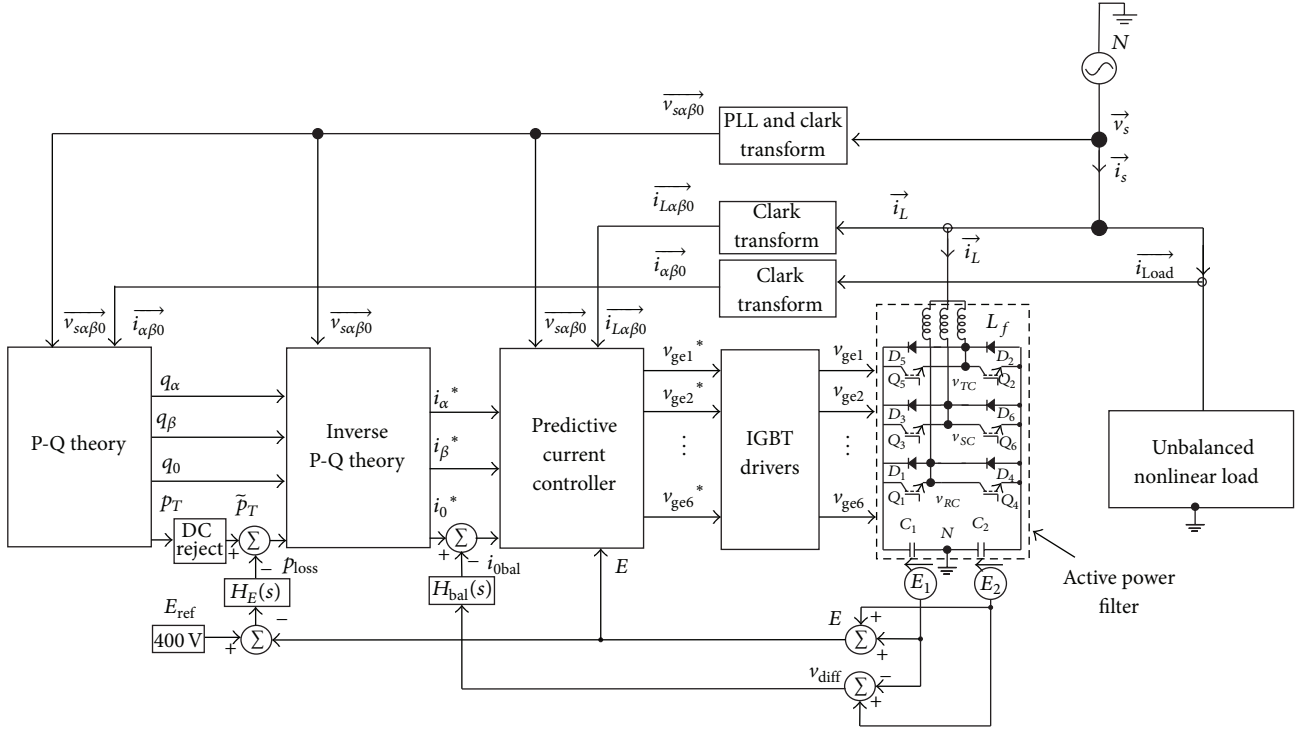
FIGURE 1: Four-wire shunt active filter and its corresponding control block diagram.

This paper presents the realization and experimental verification of a trapezoidal predictive current controller for a four-wire shunt APF that improves the power quality of unbalanced AC loads in contrast to the typical predictive Euler control strategy. The trapezoidal strategy relies its operation on a discrete trapezoidal linear approximation that more accurately determines the switching of the active filter for the one-step ahead current sample, such that three significant advantages are potentially exhibited: first, the trapezoidal predictive controller slightly increments the processing time without affecting the switching of the power converter; second, in contrast to the typical Euler approximation used in other works [6, 7, 13–20], the trapezoidal method generates lower AC current ripple; and third, the convergence time and load operating performance are wider than those obtained using the typical predictive control strategy, which improves the reference current tracking and, therefore, the power quality. Experimental results obtained with a 2 kVA prototype are presented, demonstrating that the trapezoidal predictive control may accurately compensate the currents drawn by an unbalanced nonlinear load under static and dynamic conditions.

2. Four-Wire Shunt Active Filter

2.1. Circuit Description.
The four-wire shunt APF is connected in parallel to the unbalanced nonlinear load as shown at the right-hand side of Figure 1, which consists of a split DC link formed by C_1 and C_2 which refer to the AC supply

neutral node N to provide a path to mitigate a common mode current: a typical three-phase, current-feed active converter, formed by transistors Q_1 to Q_6 and diodes D_1–D_6, and three line filter inductors L_f used to generate the filter current vector, $\overrightarrow{i_L} = \begin{bmatrix} i_{LR} & i_{LS} & i_{LT} \end{bmatrix}^T$, by the difference between the supply and converter voltage vectors $\overrightarrow{v_s} = \begin{bmatrix} v_{RN} & v_{SN} & v_{TN} \end{bmatrix}^T$ and $\overrightarrow{v_C} = \begin{bmatrix} v_{RCN} & v_{SCN} & v_{TCN} \end{bmatrix}^T$, thereby obtaining virtual sinusoidal supply currents $\overrightarrow{i_s} = \begin{bmatrix} i_{SR} & i_{SS} & i_{ST} \end{bmatrix}^T$.

2.2. Principle of Operation of the Active Filter.
The principle of operation of the APF of Figure 1 may be described using the control block diagram presented at the left-hand side of Figure 1. An instantaneous active and reactive power theory, P-Q theory block in Figure 1 [12], is used to obtain an effective calculation of the reference currents that the APF may inject to the supply to instantaneously mitigate the reactive and distorted power components, drawn by the nonlinear load, and balance the active power per phase. The P-Q theory uses the Clarke transformation of the supply voltage and load current as shown in

$$v_{\alpha\beta\theta} = \begin{bmatrix} v_0 \\ v_\alpha \\ v_\beta \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \\ 1 & -\dfrac{1}{2} & -\dfrac{1}{2} \\ 0 & \dfrac{\sqrt{3}}{2} & -\dfrac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} v_{RN} \\ v_{SN} \\ v_{TN} \end{bmatrix},$$

$$i_{\alpha\beta\theta} = \begin{bmatrix} i_0 \\ i_\alpha \\ i_\beta \end{bmatrix} = \sqrt{\frac{2}{3}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} i_{\text{Load}R} \\ i_{\text{Load}S} \\ i_{\text{Load}T} \end{bmatrix} \tag{1}$$

such that the calculation of the active and reactive instantaneous powers in the $\alpha\beta0$ coordinate system is obtained as, respectively, shown in

$$p_T = \overrightarrow{v_{\alpha\beta0}} \cdot \overrightarrow{i_{\alpha\beta0}} = v_\alpha i_\alpha + v_\beta i_\beta + v_0 i_0, \tag{2}$$

$$q = \overrightarrow{v_{\alpha\beta0}} \times \overrightarrow{i_{\alpha\beta0}} = \begin{bmatrix} q_\alpha \\ q_\beta \\ q_\theta \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} v_\beta & v_0 \\ i_\beta & i_0 \end{bmatrix} \\ \begin{bmatrix} v_\theta & v_\alpha \\ i_0 & i_\alpha \end{bmatrix} \\ \begin{bmatrix} v_\alpha & v_\beta \\ i_\alpha & i_\beta \end{bmatrix} \end{bmatrix}, \tag{3}$$

where p_T is the real power or internal product of the voltage and current vectors and q is the imaginary vector power or external product of the voltage and current vectors which is composed of q_α, q_β, and q_0. Since the load uses a fourth conductor, namely, the neutral, which is very common in low-voltage distribution system, the P-Q calculation may include both zero-sequence voltage and current as shown in (2) and (3). Therefore, the instantaneous powers defined above may be combined in a single matrix transformation as shown as follows:

$$\begin{bmatrix} p_T \\ q_\alpha \\ q_\beta \\ q_0 \end{bmatrix} = \begin{bmatrix} v_\alpha & v_\beta & v_0 \\ 0 & -v_0 & v_\beta \\ v_0 & 0 & -v_\alpha \\ -v_\beta & v_\alpha & 0 \end{bmatrix} \begin{bmatrix} i_\alpha \\ i_\beta \\ i_0 \end{bmatrix} \tag{4}$$

which is defined on the $\alpha\beta0$ reference frame. p_T and q are instantaneous power signals that have averaged and oscillatory components that may be used to calculate a reference current vector for the APF control system. The average of p_T, $\overline{p}$, corresponds to the energy per time unity that is transferred from the supply to the load and becomes the power that the system truly uses [8]. In this way, the ideal condition would be to remove the oscillatory portion of the real power p and the imaginary power q of power drawn by the load, such that the calculation of the reference currents for compensating the currents drawn by the load may be given with

$$\begin{bmatrix} i_{\alpha\text{ref}}{}^* \\ i_{\beta\text{ref}}{}^* \\ i_{0\text{ref}}{}^* \end{bmatrix} = \frac{1}{v_{\alpha\beta\theta}^2} \begin{bmatrix} v_\alpha & 0 & v_\theta & -v_\beta \\ v_\beta & -v_\theta & 0 & v_\alpha \\ v_\theta & v_\beta & -v_\alpha & 0 \end{bmatrix} \begin{bmatrix} \widetilde{p} \\ q_\alpha{}^* \\ q_\beta{}^* \\ q_0{}^* \end{bmatrix} \tag{5}$$

which is represented in Figure 1 as the inverse P-Q theory block, which subtracts $\overline{p}$ from p to obtain the oscillatory component of p, $\widetilde{p}$. In this fashion, the reference currents of (5) are used to operate the three-phase converter of Figure 1 as a current amplifier driven by the trapezoidal predictive current controller block shown at the centre of Figure 1.

2.3. DC-Link Voltage Controller. The APF requires a fixed DC-link capacitor voltage E greater than the peak value of the line-to-line supply voltage, for instance, $E = 400$ V when a 220 V, 60 Hz supply is being used. Since the shunt APF topology is identical to that of an active three-phase rectifier [14], the circuit boosts the DC-link voltage using an external voltage control loop that generates a loss power control signal, p_{loss}, which is added to $\widetilde{p}$ to supply energy for the DC-link capacitor and compensate the power losses of the APF circuit. This is shown in the left bottom side of Figure 1, where a linear control loop calculates p_{loss} using the error between the E reference, E_{ref}, and the DC-link voltage E, which is obtained adding the measured DC-link capacitor voltages, and compensating this error with $H_E(s)$.

2.4. DC-Link Capacitor Voltage Balancing Controller. Since the split DC-link node N is used to draw a compensating current for the neutral wire of the supply, the DC-link capacitor voltages may become unbalanced due to the flow of a small DC current. An additional zero-sequence, balancing current, $i_{0\text{bal}}$, is used after the zero-sequence reference current calculation to overcome a voltage unbalance between the capacitors of the split DC link [21]. This is shown at the bottom of Figure 1, where again a linear control loop calculates $i_{0\text{bal}}$ by compensating the error between the DC-link capacitor voltages, v_{dif}, with $H_{\text{bal}}(s)$.

3. Trapezoidal Predictive Current Controller

3.1. Discrete Linear Model of the APF Converter. A space vector AC-side model of the APF three-phase converter is derived calculating the filter inductor voltage vector as shown in

$$\overrightarrow{v_{L\alpha\beta0}} = L\frac{d\overrightarrow{i_{L\alpha\beta0}}}{d_t} = \overrightarrow{v_{s\alpha\beta0}} - \overrightarrow{v_{c\alpha\beta0}} \tag{6}$$

which may be solved to calculate the line current vector $\overrightarrow{i_L}$ as shown as follows:

$$\overrightarrow{i_{L\alpha\beta0}}(t_1) = \overrightarrow{i_{L\alpha\beta0}}(t_1 + kT) + \int_{t_1+kT}^{t_1+(k+1)T} \overrightarrow{v_{L\alpha\beta0}}(t)\,dt. \tag{7}$$

A discrete time model of (7) may be obtained by using the trapezoidal approximation shown in Figure 2, such that (7) becomes

$$\overrightarrow{i_{L\alpha\beta0_{k+1}}}$$
$$\cong \overrightarrow{i_{L\alpha\beta0_k}}$$
$$+ \frac{T_s}{2L}\left[\overrightarrow{v_{s\alpha\beta0_k}} - \overrightarrow{v_{c\alpha\beta0_k}} + \overrightarrow{v_{s\alpha\beta0_{k+1}}} - \overrightarrow{v_{c\alpha\beta0_{k+1}}}\right], \tag{8}$$

$$\overrightarrow{v_{L\alpha\beta0}}(t) = \overrightarrow{v_{s\alpha\beta0}}(t) - \overrightarrow{v_{c\alpha\beta0}}(t)$$
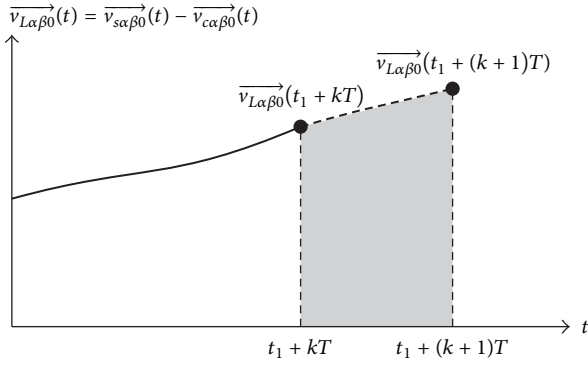
FIGURE 2: Trapezoidal approximation of the volt-seconds integral of (9).

where T is the sampling period that must be small to obtain an accurate model approximation of the system. Since $\overrightarrow{v_{s\alpha\beta0_k}} \approx \overrightarrow{v_{s\alpha\beta0_{k+1}}}$, (8) is rewritten as follows:

$$\overrightarrow{i_{L\alpha\beta0_{k+1}}} \cong \overrightarrow{i_{L\alpha\beta0_k}} + \frac{T_s}{2L}\left[2\overrightarrow{v_{s\alpha\beta0_k}} - \overrightarrow{v_{c\alpha\beta0_k}} - \overrightarrow{v_{c\alpha\beta0_{k+1}}}\right] \quad (9)$$

which may produce eight one-step ahead current vectors, $\overrightarrow{i_{L\alpha\beta0_{k+1}}}^{0}$ to $\overrightarrow{i_{L\alpha\beta0_{k+1}}}^{7}$, since $\overrightarrow{v_{c\alpha\beta0_{k+1}}}$ has six active, $\overrightarrow{v_{c\alpha\beta0}}^{1}$ to $\overrightarrow{v_{c\alpha\beta0}}^{6}$, and two neutral vectors, $\overrightarrow{v_{c\alpha\beta0}}^{0}$ and $\overrightarrow{v_{c\alpha\beta0}}^{7}$, that are listed in Table 1 with respect to their transistor switching states, assuming the common mode voltage due to the AC neutral node connection to the DC link [15]. $\overrightarrow{i_{L\alpha\beta0_{k+1}}}^{0}$ to $\overrightarrow{i_{L\alpha\beta0_{k+1}}}^{7}$ are dispersed around the kth current sample, $\overrightarrow{i_{L\alpha\beta0_k}}$, as shown in $\alpha\beta0$ frame of Figure 3, such that one of these may become near to the reference current sample $\overrightarrow{i_{L\alpha\beta0_k}}^{*}$.

3.2. Cost Function of the Current Controller.

An error current vector, $\overrightarrow{i_{Le\alpha\beta0_k}}$, may be used as a cost function to evaluate which of the transistor switching states causes the nearest one-step ahead current sample to $\overrightarrow{i_{L\alpha\beta0_k}}^{*}$, such that $\overrightarrow{i_{Le\alpha\beta0_k}}$ may be expressed as shown as follows [16]:

$$\overrightarrow{i_{Le\alpha\beta0_k}}^{0-7} = \overrightarrow{i_{L\alpha\beta0_k}}^{*} - \overrightarrow{i_{L\alpha\beta0_{k+1}}}^{0-7}. \quad (10)$$

The size of (10) may be evaluated using the Euclidean norm of $\overrightarrow{i_{Le\alpha\beta0_k}}^{0-7}$, $\|\overrightarrow{i_{Le\alpha\beta0_k}}^{0-7}\|^2$, which is equal to

$$\left\|\overrightarrow{i_{Le\alpha\beta0_k}}^{0-7}\right\|^2 = \left|i_{L\alpha k}^{*} - i_{L\alpha\ k+1}^{0-7}\right|^2 + \left|i_{L\beta k}^{*} - i_{L\beta\ k+1}^{0-7}\right|^2 \\ + \left|i_{L0 k}^{*} - i_{L0\ k+1}^{0-7}\right|^2 \quad (11)$$

such that the minimum $\|\overrightarrow{i_{Le\alpha\beta0_k}}^{0-7}\|^2$ determines the transistor switching state that may be used at the kth instant to produce an appropriate three-phase, filter current tracking with respect to the current reference vector [17].

3.3. Control Algorithm of the Four-Wire APF.

Following the description given above, a flow diagram of the APF control algorithm of Figure 1 is shown in Figure 4. This diagram starts with the parameters initialization of the microcontroller and then enters to an iterative loop control cycle. In this cycle, all the voltage and currents variables are sensed, such as the supply voltage $\overrightarrow{v_s}$, the filter current, $\overrightarrow{i_L}$, the load current, $\overrightarrow{i}_{\text{Load}}$, and the DC-link voltage, E, where the AC inputs are converted to $\alpha\beta0$ plane using (1). Since the APF may operate with a distorted voltage, or high source impedance [22], $\overrightarrow{v_s}$ is processed with a Phase Locked Loop (PLL) to obtain a clean three-phase supply and phase reference. The next process in the algorithm is the calculation of the two external voltage controllers used to maintain charged and balanced DC-link capacitors at a fixed voltage level, which contribute to calculate the reference currents through the inverse P-Q theory, (4) and (5). Once the reference currents are calculated, an "else-if" tree is started to process the trapezoidal predictive current controller with the eight possible transistor state combinations of the APF converter, which uses the discrete current model of (7) and the cost function of (10), such that eight one-step ahead current values are evaluated and then weighted against the current reference vector using (10). Finally, the converter state vector that minimizes the cost function is determined and, thereby, the algorithm applies the selected state vector to the APF converter.

4. Experimental Verification

4.1. Prototype Description.

A 2 kVA, four-wire shunt APF prototype rig was built to evaluate the operation of the APF of Figure 1. Table 2 lists the operating parameters and components of the rig.

A 150 MHz TMS320F28335 Digital Signal Processor (DSP) was used to implement the control strategy of Figures 1 and 4 using a 32-bit data word length for floating point operations ensuring numerical stability. Additional hardware was utilized to interface the DSP with the power converter, such as voltage and current sensors, signal conditioners, IGBT drivers, and fiber optic links. The APF was operated with the aid of a PLL [22] and driven with either the trapezoidal predictive controller (9) or a typical predictive controller that uses the Euler approximation of

$$\overrightarrow{i_{L\alpha\beta0_{k+1}}} \cong \overrightarrow{i_{L\alpha\beta0_k}} + \frac{T_s}{L}\left[\overrightarrow{v_{s\alpha\beta0_k}} - \overrightarrow{v_{c\alpha\beta0_{k+1}}}\right] \quad (12)$$

to experimentally compare the performance.

4.2. Experimental Results.

The 2 kVA APF prototype was verified with the Euler and trapezoidal predictive current controllers and a 127 V, 60 Hz line-to-neutral supply voltage and under three nonlinear load conditions: a 1.6 kW, naturally controlled three-phase rectifier with a LC filter, Figure 5(a); a 0.9 kW, four-wire unbalanced load, Figure 5(b), that consisted of two naturally controlled single-phase rectifiers, both with a LC output filter and each supplied with different single phases, and a resistive load supplied with a single phase; and a 1 kW unbalanced load condition, Figure 5(c), similar to

Table 1: Normalized converter voltage space vectors with respect to the transistor switching states.

| Transistor state combination | | | | | | $\overrightarrow{v_{c\alpha\beta0}}$ | Normalized converter voltages | | |
Q_1	Q_2	Q_3	Q_4	Q_5	Q_6		V_α/E	V_β/E	V_0/E
0	1	0	1	0	1	$\overrightarrow{v_{c\alpha\beta0}}^0$	0	0	$-\dfrac{1}{2}$
1	1	0	0	0	1	$\overrightarrow{v_{c\alpha\beta0}}^1$	$\sqrt{\dfrac{2}{3}}$	0	$-\dfrac{1}{6}$
1	1	1	0	0	0	$\overrightarrow{v_{c\alpha\beta0}}^2$	$\dfrac{\sqrt{6}}{6}$	$\dfrac{1}{\sqrt{2}}$	$\dfrac{1}{6}$
0	1	1	1	0	0	$\overrightarrow{v_{c\alpha\beta0}}^3$	$-\dfrac{\sqrt{6}}{6}$	$\dfrac{1}{\sqrt{2}}$	$-\dfrac{1}{6}$
0	0	1	1	1	0	$\overrightarrow{v_{c\alpha\beta0}}^4$	$-\sqrt{\dfrac{2}{3}}$	0	$\dfrac{1}{6}$
0	0	0	1	1	1	$\overrightarrow{v_{c\alpha\beta0}}^5$	$-\dfrac{\sqrt{6}}{6}$	$-\dfrac{1}{\sqrt{2}}$	$-\dfrac{1}{6}$
1	0	0	0	1	1	$\overrightarrow{v_{c\alpha\beta0}}^6$	$\dfrac{\sqrt{6}}{6}$	$-\dfrac{1}{\sqrt{2}}$	$\dfrac{1}{6}$
1	0	1	0	1	0	$\overrightarrow{v_{c\alpha\beta0}}^7$	0	0	$\dfrac{1}{2}$

Table 2: Operating parameters and components of the four-wire APF prototype.

| Electrical parameters | | |
	Variable	Value
Prototype power rating (laboratory design)		2 kW
Three-phase supply	V_{SN}	127 V, 60 Hz
DC-link voltage reference	E	400 V
Sampling frequency	f_S	21.6 kHz
ADC resolution		12 bits
Prototype components		
Line filters inductors	L_f	10 mH
DC-link capacitors	C	10 mF
Power IGBTs		BSM100GD60DLC, 1200 V, 30 A
IGBT drivers		Infineon 2ED300CL7-s
AC/DC voltage sensors		LEM LV 25-P
AC/DC current sensors		Honeywell CSNA111
DC-link *voltage* controller $H_E(s)$		$\dfrac{3000s + 4.5\text{e}4}{s^2 + 1000 - 6.71\text{e}-8}$
DC-link voltage balance controller $H_{\text{bal}}(s)$		$\dfrac{400s + 10}{s}$

the one used at a power substation of Line B, Metropolitan Railway of Mexico City, for powering electronic utility equipment.

Figure 6(a) shows the experimental supply currents i_{SR}, i_{SS}, and i_{ST} and one supply phase voltage, v_{RN}, obtained with the load condition of Figure 6(a). The experimental line current waveforms were typical of a three-phase, 6-pulse rectifier without the operation of the APF, but, when the APF was turned on using the Euler predictive controller, the supply currents became virtually sinusoidal waveforms, with the supply currents Total Harmonic Distortion (THD) and the total power factor being improved from 29% to 15% and from 0.95 to 0.98, respectively, which confirmed that the APF was properly operating. Two main characteristics were found in the experimental supply current waveforms of Figure 6(a): a 4.2 kHz, high-frequency ripple and a small glitch occurring at every rising and falling slope of the load current waveform.

The first was attributed to the Euler approximation used with the predictive control switching that continuously tracks the reference currents [23], which was confirmed with a dynamic condition of stepping the output filter inductance from 50 mH to 100 mH. Figure 6(b) shows that the operation of the APF with the Euler predictive current control and the P-Q theory is maintained throughout the filter inductance step, since the sinusoidal waveform quality of the supply currents is stable as shown in Figure 6(b), ensuring reliable operation of the APF and, therefore, the predictive controller is likely to be compliant under dynamic conditions, a typical requirement for control techniques; however, the amplitude of the filter current waveforms slightly fell from 2 A to 1.5 A during the transient response, with the supply current THD being barely degraded around 24%. The second characteristic was confirmed by contrasting the measured filter current i_{LR} with its digital reference i_{LR}^*, as shown in the left-hand side
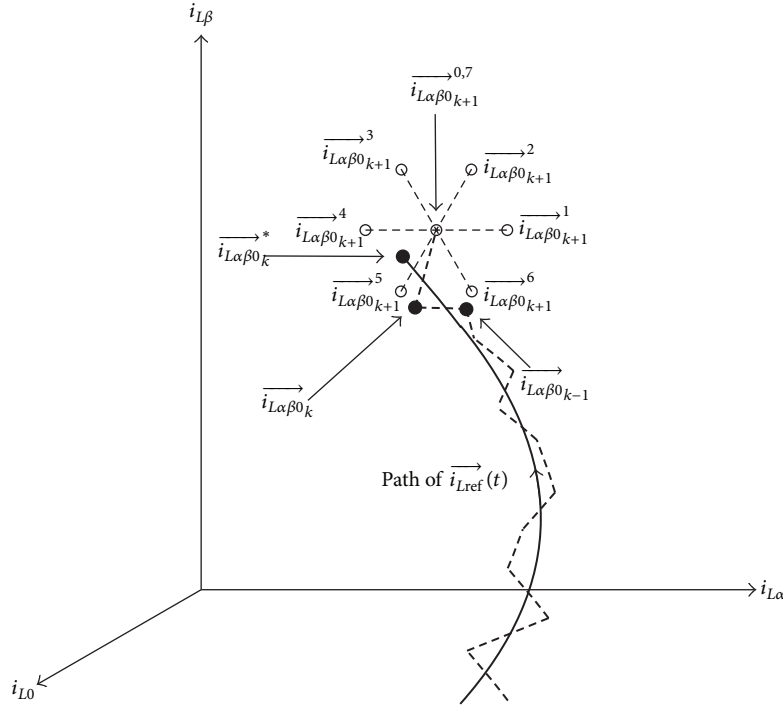
FIGURE 3: One-step ahead current samples, $\overrightarrow{i_{L\alpha\beta0}}_{k+1}^{0}$ to $\overrightarrow{i_{L\alpha\beta0}}_{k+1}^{7}$, around $\overrightarrow{i_{L\alpha\beta0}}_{k}$ and $\overrightarrow{i_{L\alpha\beta0}}_{k}^{*}$ in the $\alpha\beta0$ current frame.

of Figure 6(c), revealing that the predictive current control slightly follows the reference during the high di/dt periods of the load current waveforms due to the simple Euler approximation used in the algorithm, reducing the tracking accuracy of the references, and, therefore, introducing small glitches to the supply current waveforms. This undesired phenomenon was improved by changing the Euler approximation of the predictive controller to the trapezoidal technique as shown in the right-hand side of Figure 6(c), which reveals in its amplification shown in Figure 6(d) that the trapezoidal predictive controller reduces the current ripple amplitude by approximately 50%, increasing its frequency rate from 4.2 kHz to 9.5 kHz and producing a lower supply current THD, 10.2%, and a power factor of 0.989 in contrast to the Euler technique, with the supply current waveforms becoming virtually free of high-frequency glitches and ripple.

Figure 7(a) shows the experimental supply currents i_{SR}, i_{SS}, and i_{ST} and the neutral current i_N obtained with the unbalanced load condition of Figure 5(b) at 0.9 kW. This figure shows that the line currents are typical of an unbalanced nonlinear load before the APF is activated, and, after the APF is on, the supply currents become virtually balanced sinusoidal waveforms of 2.4 A, with the supply current THD and the power factor being improved from an unbalanced 43% to a balanced 25% and from 0.93 to 0.972, respectively, revealing again that the active filter prototype is correctly operating with a four-wire load. In addition, the same figure shows that the power quality of the supply currents becomes much more improved when the predictive controller is changed from the Euler to the trapezoidal technique, with the supply current THD and the power factor becoming 15% and 0.98,

respectively. These experimental current waveforms have again a high-frequency ripple, being 4.2 kHz when the APF is operated with the typical Euler technique and 9.5 kHz with the proposed trapezoidal strategy. In Figure 7(a), the neutral current i_N was virtually mitigated after the APF was activated, becoming more reduced when the APF was driven with the trapezoidal controller. This experiment revealed the effectiveness of the 4-wire, P-Q theory used in this work together with the predictive current control switching.

Figure 7(b) shows the experimental supply currents i_{SR}, i_{SS}, and i_{ST} and the neutral current i_N obtained with the front-end, controlled rectifier drive and monophasic resistive load of Figure 5(c) at 0.9 kW. Before the APF is activated, as shown in Figure 7(b), the experimental supply currents are completely unbalanced, distorted, and phase-displaced due to the biphasic connection of the front-end rectifier of the motor drive and the resistive load connection; however, when the APF is on using, firstly, the Euler predictive technique and then the trapezoidal version, the supply currents become again balanced with virtual sinusoidal waveforms, such that their THD was improved from an unbalanced 40% to a balanced 23% and 18% for the Euler and trapezoidal methods, respectively, and the power factor from 0.8 to 0.97 and 0.98 again for the Euler and trapezoidal methods, respectively. In Figure 7(b), the neutral current i_N was again virtually mitigated after the APF was activated, becoming almost cancelled when the APF used the trapezoidal controller.

A power analyser was used to measure the supply active power, P, the supply apparent power, S, the per-phase supply current THD, I_{RS} THD, I_{SS} THD, and I_{TS} THD, and the total power factor during the experiments described above. The
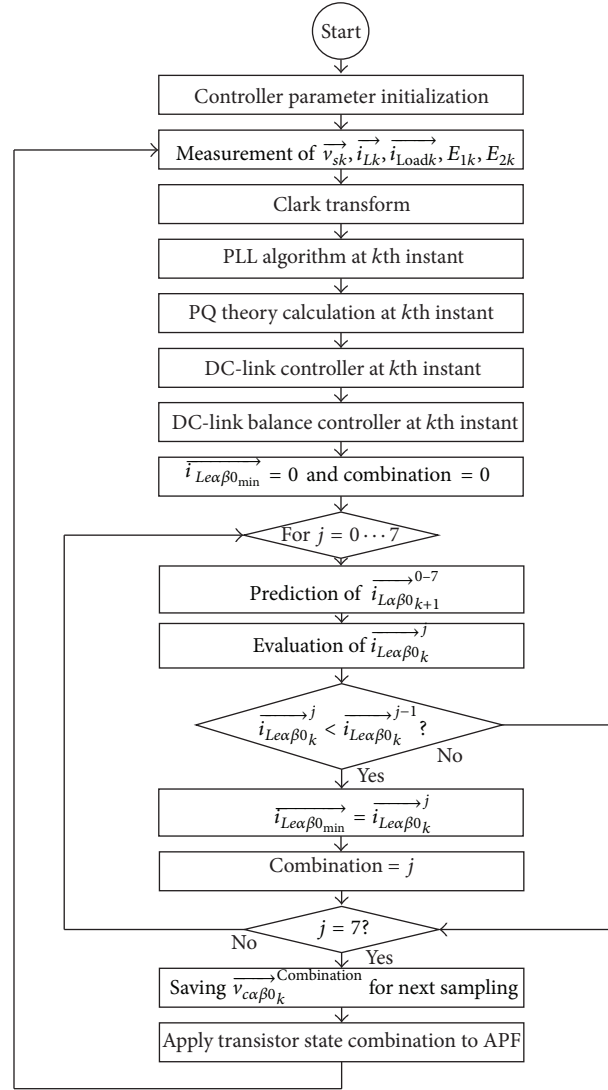
Figure 4: Algorithm flow diagram of the predictive current controller.

results are contrasted in the comparative bar plot of Figure 8 calculated with a 2 kVA APF rating, a 127 V, 60 Hz line-to-neutral supply voltage, and a 400 V APF DC link. Figure 8 shows that P slightly rise by approximately 5%, 100 W, when the APF prototype was used to improve and balance the supply currents among the experiments; the additional power loss occurring in the transistors, filter inductors and DC-link capacitors of the APF converter due to the high-frequency operation. In comparison with the unbalanced load, the current THD reduction is representative when the APF is used together with the trapezoidal predictive controller to compensate the drawn currents of the balanced load as shown in Figure 8, whereas the current THD is slightly improved when the APF is used together with the Euler strategy and the load cases of Figure 5; nevertheless, the drawn current through the neutral wire is noticeably cancelled when the APF is used to correct the power quality of the four-wire AC loads of Figures 5(b) and 5(c). In contrast with the balanced load, the distribution between apparent and active power for the APF

with the unbalanced load becomes equilibrated due to correction of current phase displacement.

Closer inspection of the microprocessor operation revealed that the total period to perform the algorithm of Figure 4 was around 29 μs with the Euler predictive controller, which is well below the sampling period to ensure minimal delay effects of the control system. Unlike the experimental verification of the Euler predictive controller, the experimental verification of the APF with the trapezoidal predictive controller resulted in a slight increment of algorithm processing time of Figure 4, from 29 μs to 30 μs, which was imperceptible during the experimental verification of the APF.

The presented trapezoidal predictive controller is slightly more complex than the typical predictive strategy to perform the current reference tracking and generates a slight increase of power losses, which could make it inadequate for implementation in low-rated rigs; nevertheless, the ripple current reduction, closer current tracking, and power quality

(a)

(b)

(c)
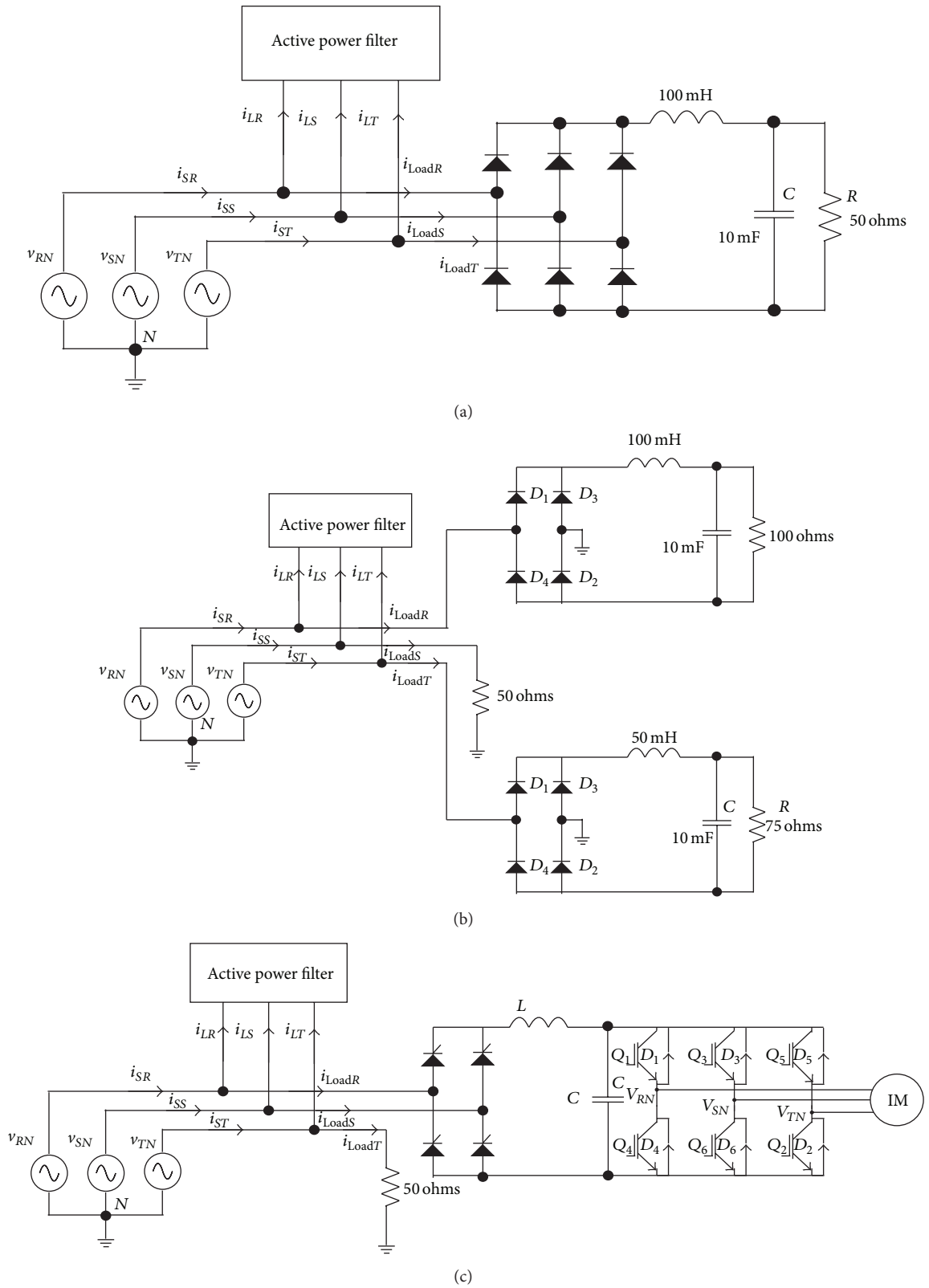
FIGURE 5: Nonlinear loads used to experimentally verify the APF and predictive current controller of Figure 1. (a) Balanced three-wire load, (b) unbalanced four-wire load, and (c) unbalanced load used at a Metro Power Substation.
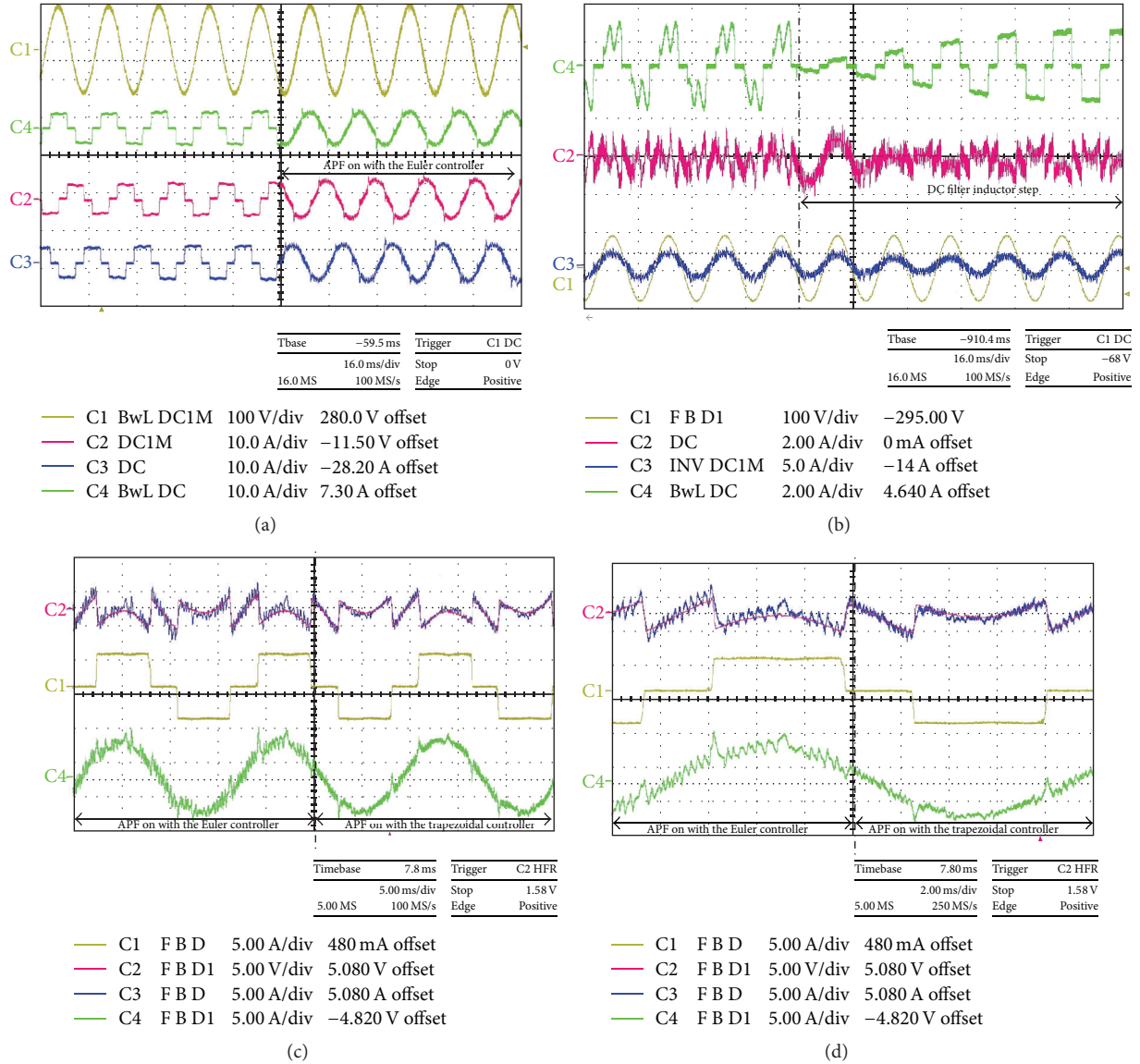
FIGURE 6: Experimental verification of the APF prototype with the 1.6 kW balanced load of (a). (a) Measured waveforms v_{RN} (yellow), i_{SR} (green), i_{SS} (red), and i_{ST} (blue) before and after the activation of the APF. (b) Measured response of v_{RN} (yellow), i_{LR} (red), i_{LoadR} (green), and i_{SR} (yellow) to a filter inductance step from 50 mH to 100 mH. (c) Measured response of i_{LR} (blue) and i_{SR} (green) to a predictive current controller step from Euler to the trapezoidal approximation, which is contrasted against the reference $i_{LR}{}^*$ (red) and load current i_{LoadR}. (d) Time amplification of (c) at the instant of the predictive controller step. 127 V, 60 Hz supply, 400 V APF DC link, and a 21.6 kHz APF sampling frequency.

improvement are important advantages to consider over the traditional predictive controller technique. Furthermore, the digital implementation is acceptable for fast microcontrollers, such as a DSPs and hybrid digital controllers, and will be used once the trapezoidal control strategy is implemented to control other power converter systems, which would be suitable to obtain high power quality results.

5. Conclusions

The utilization of a trapezoidal predictive technique to generate the filter currents of a four-wire, shunt APF allows the

power quality improvement of using unbalanced nonlinear loads for on-land utility applications, such that the supply currents become virtual sinusoidal waves. The latter makes the current control strategy attractive for easy and straight implementation on future power converters that require high-performance power quality; nevertheless, the control technique is suitable for a wide range of power converter applications. In this work, the trapezoidal predictive controller was experimentally verified and evaluated with the four-wire APF under three load conditions; in the first, the load was set up with a three-wire, balanced nonlinear circuit to preliminary check the basic operation of the control
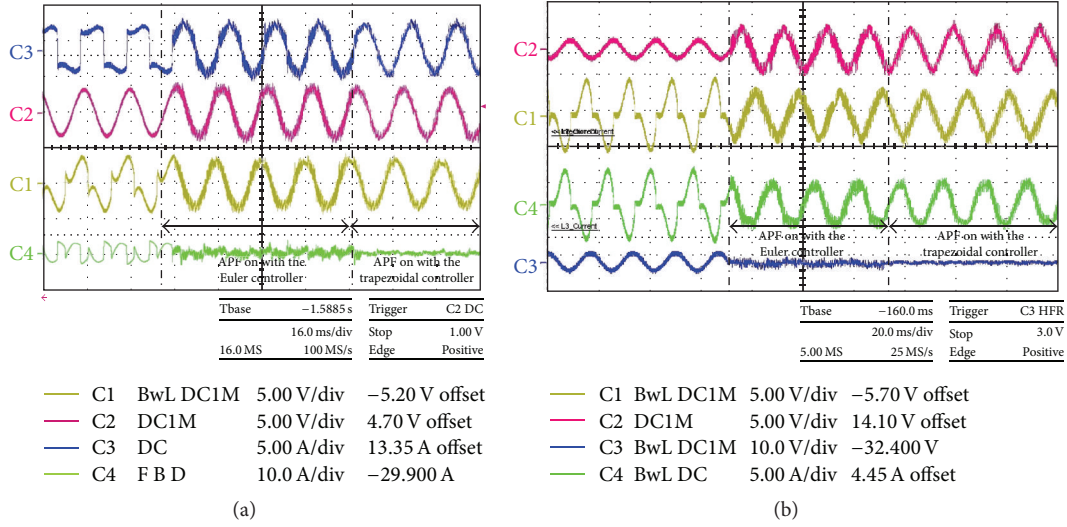
FIGURE 7: (a) Measured supply current waveforms i_{SR} (blue), i_{SS} (yellow), i_{ST} (red), and i_N (green) before and after the activation of the APF with the Euler and Trapezoidal predictive controllers and the load condition of Figure 5(a). (b) Measured supply current waveforms i_{SR} (red), i_{SS} (yellow), i_{ST} (green), and i_N (blue) before and after the activation of the APF with the Euler and trapezoidal predictive controllers and the load condition of Figure 5(b). 127 V, 60 Hz supply, 400 V APF DC link, and 21.6 kHz APF sampling frequency.
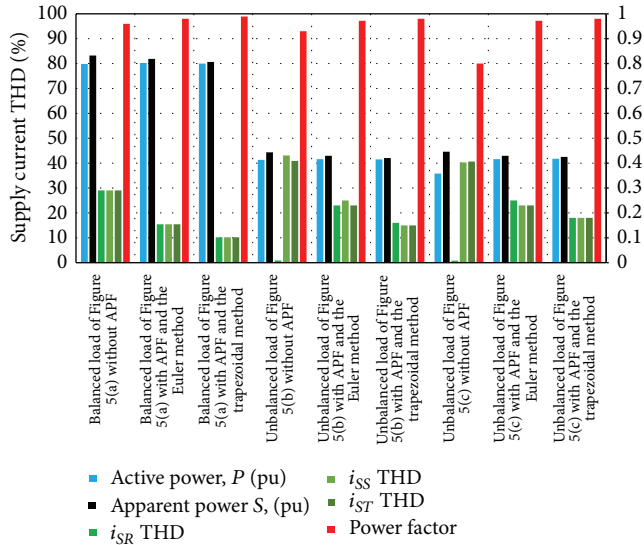
FIGURE 8: Bar plot of the power quality results for the AC load circuits of Figure 5 with the APF prototype operating with the Euler and trapezoidal predictive techniques. 127 V, 60 Hz supply and 2 kVA APF rating.

predictive controller were activated, with the neutral current being mitigated; the current THD was 15% and 18%, respectively, and the power factor was 0.98 in both experiments with a 127 V, 60 Hz three-phase supply voltage. The shape of the supply current waveforms and the power quality were significantly improved in comparison with the original load currents and power quality, with the active power being slightly increased, due to the high-frequency switching losses of the APF power transistors.

The practical realization of the presented trapezoidal predictive controller could consider the use of an extended sampled-data horizon, either forward or backward, to achieve a faster convergence and reduce the current ripple amplitude. This would be convenient for developing power converters with new generation of switching power devices for other applications.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

technique, such that sinusoidal supply current waves were generated and the power quality was improved. A current THD of 10% and a power factor of 0.989 were measured in the first experiment showing a noticeable improvement in contrast to the traditional predictive technique.

In the second and third load conditions, the load was four-wire, unbalanced nonlinear load, with the load currents being much distorted and producing a neutral current path in both load conditions. The supply current waveforms were all improved and balanced when the APF and the trapezoidal

References

[1] IEEE Application Guide for IEEE Standard 1547, "IEEE standard for interconnecting distributed resources with electric power systems," IEEE Standard 1547.2-2008, 2008.

[2] S. W. Mohod and M. V. Aware, "A STATCOM—control scheme for grid connected wind energy system for power quality improvement," *IEEE Systems Journal*, vol. 4, no. 3, pp. 346–352, 2010.

[3] IEEE, "IEEE recommended practices and requirements for harmonic control in electrical power systems," IEEE Standard 519-1992, 1992.

[4] J. D. van Wyk and F. C. Lee, "On a Future for Power Electronics," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 2, pp. 59–72, 2013.

[5] P. Kanjiya, V. Khadkikar, and H. H. Zeineldin, "Optimal control of shunt active power filter to meet IEEE Std. 519 current harmonic constraints under nonideal supply condition," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 2, pp. 724–734, 2015.

[6] P. Acuna, L. Moran, M. Rivera, R. Aguilera, R. Burgos, and V. G. Agelidis, "A single-objective predictive control method for a multivariable single-phase three-level NPC converter-based active power filter," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 7, pp. 4598–4607, 2015.

[7] P. Acuna, L. Moran, M. Rivera, R. Aguilera, R. Burgos, and V. G. Agelidis, "A single-objective predictive control method for a multivariable single-phase three-level NPC converter-based active power filter," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 7, pp. 4598–4607, 2015.

[8] P. Jintakosonwit, H. Fujita, and H. Akagi, "Control and performance of a fully-digital-controlled shunt active filter for installation on a power distribution system," *IEEE Transactions on Power Electronics*, vol. 17, no. 1, pp. 132–140, 2002.

[9] Z. Xiao, X. Deng, R. Yuan, P. Guo, and Q. Chen, "Shunt active power filter with enhanced dynamic performance using novel control strategy," *IET Power Electronics*, vol. 7, no. 12, pp. 3169–3181, 2014.

[10] P. Acuna, L. Moran, M. Rivera, J. Rodriguez, and J. Dixon, "Improved active power filter performance for distribution systems with renewable generation," in *Proceedings of the 38th Annual Conference on IEEE Industrial Electronics Society (IECON '12)*, pp. 1344–1349, Montreal, Canada, October 2012.

[11] P. Acuna, L. Moran, M. Rivera, J. Dixon, and J. Rodriguez, "Improved active power filter performance for renewable power generation systems," *IEEE Transactions on Power Electronics*, vol. 29, no. 2, pp. 687–694, 2014.

[12] H. Akagi, E. H. Watanabe, and M. Aredes, *Instantaneous Power Theory and Applications to Power Conditioning*, IEEE Press Series on Power Engineering, Wiley-IEEE Press, 2007.

[13] R. P. Aguilera and D. E. Quevedo, "Predictive control of power converters: designs with guaranteed performance," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 53–63, 2015.

[14] P. Cortés, J. Rodríguez, P. Antoniewicz, and M. Kazmierkowski, "Direct power control of an AFE using predictive control," *IEEE Transactions on Power Electronics*, vol. 23, no. 5, pp. 2516–2523, 2008.

[15] J. Scoltock, T. Geyer, and U. K. Madawala, "Model predictive direct power control for grid-connected NPC converters," *IEEE Transactions on Industrial Electronics*, vol. 9, no. 62, pp. 5319–5328, 2015.

[16] M. Lopez, J. Rodriguez, C. Silva, and M. Rivera, "Predictive torque control of a multidrive system fed by a dual indirect matrix converter," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 5, pp. 2731–2741, 2015.

[17] A. Bhattacharya and C. Chakraborty, "A shunt active power filter with enhanced performance using ANN-based predictive and adaptive controllers," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 2, pp. 421–428, 2011.

[18] R. Vargas, U. Ammann, B. Hudoffsky, J. Rodriguez, and P. Wheeler, "Predictive torque control of an induction machine fed by a matrix converter with reactive input power control," *IEEE Transactions on Power Electronics*, vol. 25, no. 6, pp. 1426–1438, 2010.

[19] B. S. Riar, T. Geyer, and U. K. Madawala, "Model predictive direct current control of modular multilevel converters: modeling, analysis, and experimental evaluation," *IEEE Transactions on Power Electronics*, vol. 30, no. 1, pp. 431–439, 2015.

[20] L. Tarisciotti, P. Zanchetta, A. Watson, J. C. Clare, M. Degano, and S. Bifaretti, "Modulated model predictive control for a three-phase active rectifier," *IEEE Transactions on Industry Applications*, vol. 51, no. 2, pp. 1610–1620, 2015.

[21] A. Luo, X. Xu, L. Fang, H. Fang, J. Wu, and C. Wu, "Feedback-feedforward PI-type iterative learning control strategy for hybrid active power filter with injection circuit," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11, pp. 3767–3779, 2010.

[22] Y. F. Wang and Y. W. Li, "Three-phase cascaded delayed signal cancellation PLL for fast selective harmonic detection," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 4, pp. 1452–1463, 2013.

[23] M. Rivera, C. Rojas, J. Rodríguez, P. Wheeler, B. Wu, and J. Espinoza, "Predictive current control with input filter resonance mitigation for a direct matrix converter," *IEEE Transactions on Power Electronics*, vol. 26, no. 10, pp. 2794–2803, 2011.

Research Article

Loss Prediction and Thermal Analysis of Surface-Mounted Brushless AC PM Machines for Electric Vehicle Application Considering Driving Duty Cycle

Tianxun Chen,[1] Xiaopeng Wu,[2] Yugang Dong,[2] Chengning Zhang,[2] and Haipeng Liu[2]

[1]*Jinhua Polytechnic, Zhejiang 321007, China*
[2]*Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Xiaopeng Wu; xiaopengwu@hotmail.com

This paper presents a computationally efficient loss prediction procedure and thermal analysis of surface-mounted brushless AC permanent magnet (PM) machine considering the UDDS driving duty cycle by using a lumped parameters' thermal model. The accurate prediction of loss and its variation with load are essential for thermal analysis. Employing finite element analysis (FEA) to determine loss at every load point would be computationally intensive. Here, the finite element analysis and/or experiment based computationally efficient winding copper and iron loss and permanent magnet (PM) power loss models are employed to calculate the electromagnetic loss at every operation point, respectively. Then, the lumped parameter thermal method is used to analyse the thermal behaviour of the driving PM machine. Experiments have been carried out to measure the temperature distribution in a motor prototype. The calculation and experiment results are compared and discussed.

1. Introduction

The massive application of electric vehicle (EV) is a significant way to reduce the emission and to settle the energy crisis [1]. In general, the types of EVs can be categorized as pure electric vehicle (PEV), hybrid electric vehicle (HEV), and fuel cell electric vehicle (FCEV) [1–3]. To all these subcategories, electric machine is always the key traction component, which needs to be well designed and manufactured.

Among all the types of electric machine, brushless AC PM machine is a promising candidate for EV traction motor due to its high power density, high efficiency, fast dynamics, and compactness [4, 5]. However, this kind of machine could easily suffer insulation failure of coils and irreversible demagnetization due to the poor cooling condition on the rotor side and excessive heat generation on the stator side, especially for variable-speed application [6]. Therefore, the accurate temperature prediction for traction brushless AC PM machine is of great importance at the design stage.

In fact, one of the key elements of accurate thermal analysis is the accurate losses calculation. Generally, there are two main sources of loss within an electric machine: mechanical and electromagnetic. Mechanical loss is attributed to the frictional effects within the bearing assembly (bearing loss) and fluid dynamics or aerodynamics effects within the motor body (windage or drag loss) [7]; it can be easily obtained according to the manufacturer's manual. Electromagnetic losses, which effect thermal analysis heavily, are usually associated with active parts of the motor assembly and include the iron, winding, and permanent magnet (PM) loss components [8–10].

For electromagnetic losses prediction, two approaches are widely used: analytical and numerical [11–19]. The analytical approach for iron loss calculation was first developed by Steinmetz [11] and then further modified by Bertotti [12]. The Bertotti formulation which divides the iron loss into three individual parts, that is, eddy current loss, hysteresis loss, and supplementary loss, is commonly used at current electric machine design stage. The analytical winding loss approaches have been well developed to account for the AC effect, for example, the skin effect and the proximity effect [13–15]. The AC equivalent resistance is the commonly used

element to account for the AC effect. However, it is frequency dependent and would change with the change of operation point. Therefore, it would be difficult to calculate the winding losses of all operation points accurately by a single value of AC equivalent resistance. For PM power loss, a variety of analytical techniques have been developed. These are based on simplified assumptions of the field distribution and their use is limited to the selected machine topologies for which the assumptions hold [16, 17].

The numerical approach including time-stepping or frequency domain FEA is a more accurate way to calculate the electromagnetic losses [4, 18, 19]. However, it is time consuming and computationally intensive. Therefore, the numerical approach would not be suitable in the case that a loss map or an efficiency map is required.

Besides the losses prediction approaches, the thermal model is another important element for thermal analysis. Two main models could be found from literatures: FE based thermal model and lumped parameter thermal model [20, 21]. The FE based thermal model is accurate and capable of predicting the hottest pot within a motor, while it is time consuming. The lumped parameter thermal model is quick, while it could be only capable of calculating the mean temperature of each motor component. Thus, the choice of thermal model strongly depends on the design goals.

In addition, there has been increased interest in predicting temperature distribution under the driving duty cycle [4, 22]. A lumped parameter thermal model [20] has been used to calculate the temperature distribution under the Chinese city driving duty cycle in [4]. Good agreement is visible between the analytical and experimental results. However, the procedure presented is time consuming, since the FEA is employed to calculate the loss at each operating point. Some accurate and computationally efficient loss scaling techniques [23–25] have been presented, and some of them have been introduced into the analysis procedure and the equivalent-circuit lumped parameter thermal model has been adopted in the literature [22], while it does not take the PM loss into account, which may be small but can directly heat up magnets.

This paper proposes a computationally efficient loss prediction procedure and a lumped parameter thermal analysis of surface-mounted brushless AC PM machine considering the UDDS driving duty cycle. The machine is applied on a 10-meter motor-direct-driving large coach bus. The FEA based iron loss [23], copper loss [24], and PM loss [25] scaling techniques are utilized to obtain the loss distribution under the driving duty cycle in a timely manner. The equivalent thermal parameter of winding [26] is calculated and introduced into the thermal model. Finally, the temperature of winding, stator, and PM is predicted and is compared with the results from experiment.

The remainder of the paper is organised in the following manner: Section 2 outlines the machine design exemplar and the selected coach bus configuration; Section 3 describes the analysis procedure and the electromagnetic loss scaling technique; Section 4 details thermal modelling; Section 5 describes the experimental setup and results; Section 6 summarizes the research findings.

Table 1: Selected details of driving motor.

Number of poles	8
Number of slots	48
Rated torque	928 Nm
Rated power	350 kW
Motor outer diameter	480 mm
Active length	210 mm
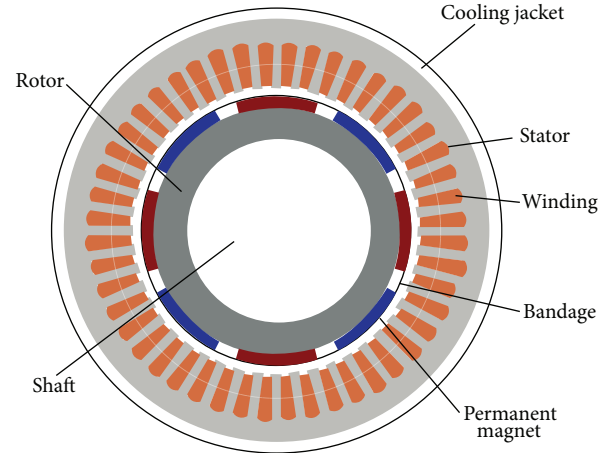
Figure 1: Geometry of analysed PM machine.

2. Study Machine and Bus Model

The analysed motor is a radial-flux, integer-slot, distributed-wound internal-rotor PM machine with water cooling jacket, as shown in Figure 1. Selected details of the driving motor are given in Table 1. And basic traction parameters of selected coach bus are given in Table 2. Please note that the aim of this paper is to accelerate the analysis speed by using computationally efficient loss mapping techniques. Therefore, some of the coach bus traction parameters have been modified to ensure that torque-speed characteristic of analyzed machine can cover the torque-speed requirement.

3. Analysis Procedure

3.1. Outline of Analysis. In order to predict dynamic temperature distribution of the driving motor under the driving duty cycle, an analysis procedure is developed and shown in Figure 2.

The limited number of FEAs is utilized in order to derive the loss functions, which give access to loss predictions over the entire torque-speed envelope quickly. Meanwhile, the UDDS, shown in Figure 3, of EV is adopted. This duty cycle contains 1370 points and each point runs one second. The corresponding operating points of the driving motor including speed and torque are obtained. Then, by the computational loss derivations, the losses at each operating point can be easily achieved. Based on the loss predictions, the dynamic temperature distribution under the driving duty cycle is predicted by using lumped parameter thermal model.

TABLE 2: Traction parameters of coach bus.

Gross vehicle weight m	14000 kg
Final drive ratio i_0	13
Frontal area A	6.6 m^2
Wheel radius R_r	0.46 m
Drivetrain efficiency η	95%
Rolling resistance coefficient μ	0.0075
Aerodynamic drag coefficient C_D	0.7
Vehicle rotary mass coefficient σ	1.2
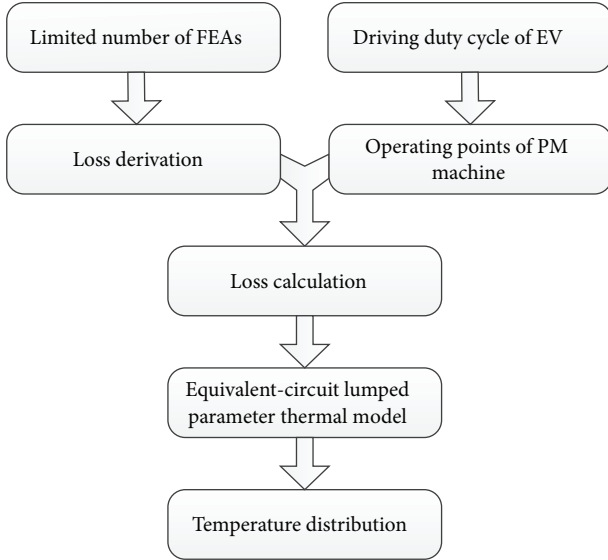
FIGURE 2: Outlines of analysis procedure.

FIGURE 3: The UDDS driving duty cycle.
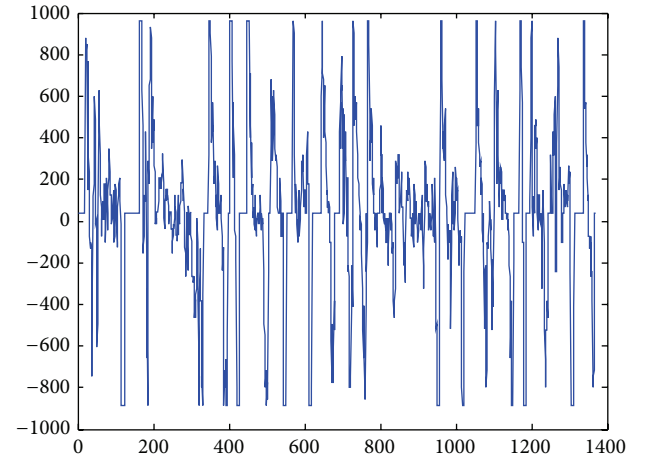
FIGURE 4: Torque requirement of UDDs duty cycle.

3.2. Motor Torque and Speed Calculation.

The corresponding operating points of the driving motor including speed and torque are derived from the following equation:

$$n = \frac{vi_0}{0.377R_r}. \tag{1}$$

Here, the unit of v is km/h, so the vehicle speed in UUDs should be converted to km/h by multiplying 1.61:

$$T = \frac{\left(\mu mg + C_D Av^2/21.15 + \sigma ma\right)R_r}{(i_0\eta)}. \tag{2}$$

The machine torque map under UDDs duty cycle is shown on Figure 4. Here, the generator mode of analysed machine is ignored.

3.3. Loss Derivation

3.3.1. Iron Loss.

The total iron loss occurring in the stator can be viewed as a superposition of two components. One stems from the main magnetizing flux path which flows through the teeth and back iron and couples the coils forming the stator winding. The other occurs during field weakening and is a result of flux emanating from the permanent magnets which flows across the face of the tooth tip and does not couple the stator coils [23]:

$$P_{\text{iron}} = P_1 + P_2, \tag{3}$$

where

$$P_1 = \frac{a_h}{\lambda}V_m + \frac{a_J}{\lambda^2}V_m^2 + \frac{a_{\text{ex}}}{\lambda^{1.5}}V_m^{1.5},$$
$$P_2 = \frac{b_h}{\lambda}V_d + \frac{b_J}{\lambda^2}V_d^2 + \frac{b_{\text{ex}}}{\lambda^{1.5}}V_d^{1.5}. \tag{4}$$

The V_m and V_d used in (4) are determined from the phasor diagram of direct-quadrature d-q axes describing the brushless AC operation of the motor, where

$$\lambda = \frac{E_{\text{ph}}}{f},$$
$$I_q = \frac{T}{k_T},$$

$$V_d = \lambda f \frac{I_d}{I_{\text{SC}}},$$

$$V_m = \lambda f \sqrt{\left(1 - \frac{I_d}{I_{\text{SC}}}\right)^2 + \left(\frac{I_q}{I_{\text{SC}}}\right)^2}. \tag{5}$$

Here, E_{ph} is the phase EMF, T is the torque, I_d and I_q are the magnitudes of the demagnetizing direct axis and quadrature axis components of stator phase current, k_T is the motor torque constant, and I_{SC} is the short-circuit current calculated from FE analysis. The coefficients $a_{h,j,\text{ex}}$ and $b_{h,j,\text{ex}}$ for the hysteresis, Joule eddy-current, and excess losses have been found from curve fitting (4) to the FE open-circuit and short-circuit loss results across the operating frequency f. Note that the iron loss in rotor core is relatively small in this case, so it can be neglected.

3.3.2. Copper Loss. The electrical resistivity of copper increases with operating temperature elevating, which results in simultaneous rise of DC loss component and reduction of the AC loss component caused by skin, proximity, and other eddy current effects [24]. The overall copper power loss and its thermal variation at AC operation depend on the balance between these two loss components [22]. Here, the copper loss scaling approach proposed in [24] has been employed. The end-winding copper loss will be separated from the total copper loss by multiplying a ratio between end-winding length and total winding length.

$$P_{\text{AC}}\big|_T = I^2 R_{\text{DC}}\big|_{T_0} \frac{\left(R_{\text{AC}}/R_{\text{DC}}\big|_{T_0}\right) - 1}{\sqrt{\left(1 + \alpha\left(T - T_0\right)\right)}}$$
$$+ I^2 R_{\text{DC}}\big|_{T_0} \left(1 + \alpha\left(T - T_0\right)\right), \tag{6}$$

where α is the temperature coefficient of resistivity of copper, $3.93e - 3 \text{ K}^{-1}$, and $I^2 R_{\text{DC}}\big|_{T_0}$ is the DC copper loss at reference temperature T_0, which is set at $25°\text{C}$ in this case.

3.3.3. PM Loss. The eddy current loss generated in the magnet array stems from two effects. The first results from the permeance variation caused by stator slotting and the second from the armature reaction field. The complete functional representation of PM loss catering for both constant torque and constant power operating regions can be then written in the following form [25]:

$$P_{\text{PM}} = \left(aI_q^2 + bI_d^2 + cI_d + d\right)\left(\frac{n}{n_W}\right)^2. \tag{7}$$

Here, the coefficients a, b, c, and d can be derived from four individual FEAs accounting for open-circuit, rated current in the quadrature axis, rated current in the direct axis, and reduced current in the direct axis. And n is the rotatory speed; n_W is the reference speed which should be set within the field weakened regime of operation.
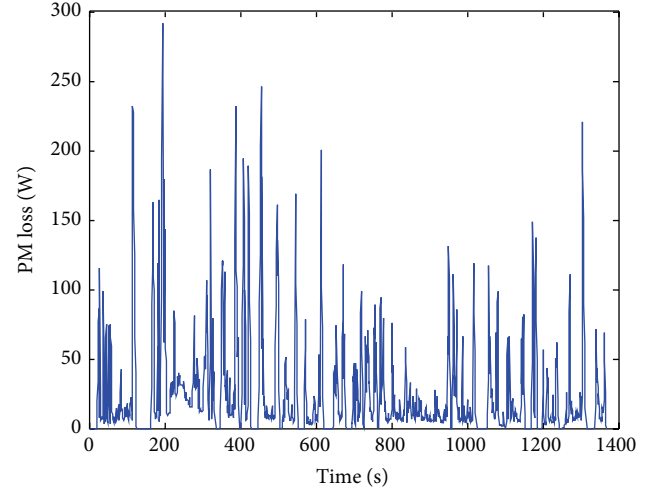
FIGURE 5: PM loss distribution under the UDDS driving duty cycle.
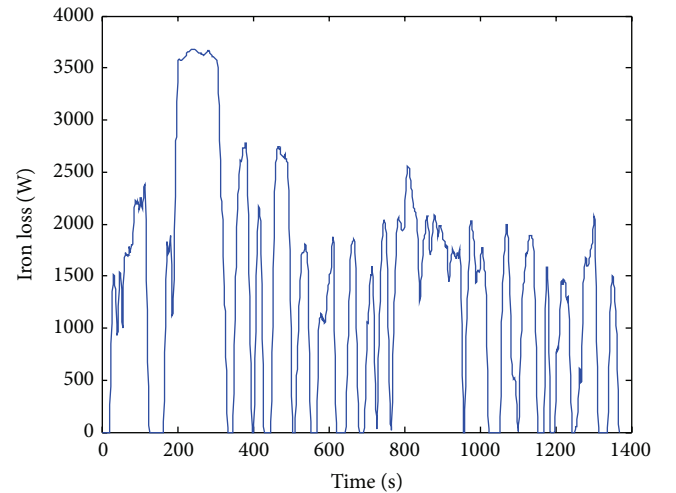
FIGURE 6: Iron loss distribution under the UDDS driving duty cycle.

3.3.4. Other Losses. The bearing and frictional loss is calculated according to the manufacturer's manual.

Figures 5 and 6 show the PM loss and iron loss distribution individually according to (3) and (7) in a single UDDS driving duty cycle. And the mechanical loss distribution is presented in Figure 7. Note that as the copper loss scaling technique requires transient temperature, the copper loss distribution remains unknown at this stage. However, the copper loss at ambient temperature $25°\text{C}$ is easily obtained and would be used as initial condition.

Table 3 illustrates the comparison between the technique proposed in [4] and the presented technique in the paper. As only 6 steps of FEA are utilised in the presented technique, the analysis time of complete duty cycle could be able to be reduced from 2 hours to just several minutes, compared to the previous technique where 1314 FEAs are adopted. In addition, the PM loss is further taken into account in the presented technique, which will result in a more accurate thermal analysis. By using all these means, the analysis time

TABLE 3: Comparison between previous technique in [3] and proposed technique in this paper.

	Technique in [3]	Proposed technique
Numbers of FEA	1314	6
Analysis time of compete duty cycle	More than 2 hours	Around 5 min
Types of electromagnetic loss	Winding loss, iron loss, and mechanical loss	Winding loss, iron loss, PM loss, and mechanical loss
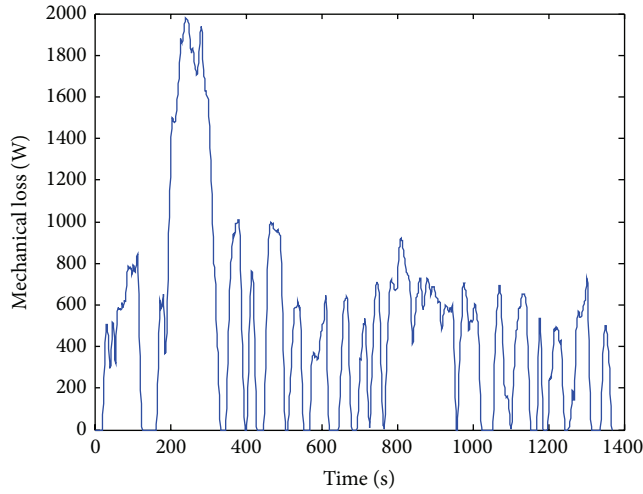
FIGURE 7: Mechanical loss distribution under the UDDS driving duty cycle.

could be reduced substantially, and the results could be more reliable.

4. Thermal Analysis

4.1. Calculation of Thermal Parameters. Figure 5 shows the general thermal network for component based on the assumption that temperature in the PM machine is axially symmetrical. Here, there are three unknown temperatures: the outer surface $T_{r,1}$, the inner surface $T_{r,2}$, and the axial edge T_a. And the four thermal resistances which constitute the network can be calculated by the following equations:

$$R_1 = \frac{1}{2\pi k_r L} \left(1 - \frac{2r_2^2 \ln (r_1/r_2)}{(r_1^2 - r_2^2)} \right),$$

$$R_2 = \frac{1}{2\pi k_r L} \left(\frac{2r_1^2 \ln (r_1/r_2)}{(r_1^2 - r_2^2)} - 1 \right),$$

$$R_3 = \frac{L}{6\pi k_a (r_1^2 - r_2^2)},$$

$$R_m = \frac{-1}{4\pi (r_1^2 - r_2^2) k_r L} \left(r_1^2 + r_2^2 - \frac{4r_1^2 r_2^2 \ln (r_1/r_2)}{(r_1^2 - r_2^2)} \right),$$

(8)
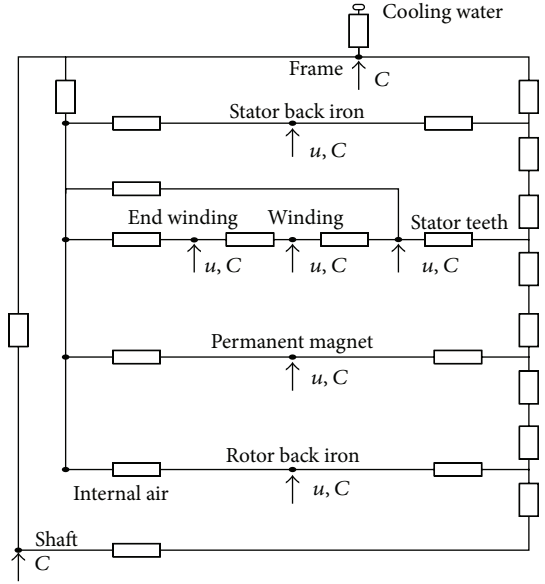
FIGURE 8: Lumped parameter thermal model. C = thermal capacitance, u = heat generation.

where r_1 and r_2 are the outer and inner radius of a general cylindrical component, respectively, and k_a and k_r are the axial and radial thermal conductivities of a general cylindrical component, respectively.

The lump parameter thermal model is shown in Figure 8, which contains nine key components of surface-mounted brushless AC PM machine. Based on the previous assumption, only half of the motor geometry is modelled. Also, it is assumed that heat flows in the radial and axial direction are independent and the heat generation is distributed inside the PM machine.

As a key component of surface-mounted brushless AC machine, the electrical winding is a heterogeneous material consisting of conductors, electrical insulation, impregnation insulation, and air and would easily suffer from severe thermal load condition. A full representation of the individual conductors and surrounding insulation system would add considerable complexity to a thermal model and leads to long solution times which limit the applicability of such models to both transient problems and iterative design/optimization procedures [26]. There are various modelling approaches to capture the thermal parameters of electrical winding [27–30]. This paper uses the methodology presented in [26] to calculate the thermal conductivity and heat capacity:

$$v_c + v_{ci} = \text{PF},$$

$$v_{ii} = 1 - \text{PF},$$

$$v_c = \text{PF} \frac{r_c^2}{(r_c + l_i)^2},$$

$$v_{ci} = \text{PF} \frac{2r_c l_i + l_i^2}{(r_c + l_i)^2},$$

$$k_e = k_a \frac{(1 + v_c) k_c + (1 - v_c) k_a}{(1 - v_c) k_c + (1 + v_c) k_a},$$

$$c_e = \frac{\text{PF} (\rho_c c_c - \rho_p c_p) + \rho_p c_p}{\text{PF} (\rho_c - \rho_p) + \rho_p},$$

$$(9)$$

where v_c, v_{ci}, and v_{ii} are the volume ratio of conductor, conductor insulation, and impregnation insulation, respectively. PF is the packing factor, r_c is the conductor radius, and l_i is the insulation thickness. k_e and k_a are the equivalent thermal conductivity of winding and insulation amalgam, respectively. k_c is the thermal conductivity of conductor. c_e is the equivalent heat capacity, and c_c and c_p are the heat capacity of conductor and impregnation compound. ρ_c and ρ_p are the density of conductor and impregnation compound.

The heat transfer coefficient of the fully developed turbulent flow in the spiral housing water jacket can be calculated according to [31]

$$h = \frac{k_f N_u}{D_h},$$

$$(10)$$

$$N_u = \frac{(f/8)(R_e - 1000) P_r}{1 + 12.7 (f/8)^2 (P_r^{2/3} - 1)},$$

where N_u is the Nusselt number, k_f is the thermal conductivity of fluid, and D_h is the hydraulic diameter. $R_e = \rho v_{av} D_h / \mu$ is the Reynolds number, ρ is the fluid density, v_{av} is the fluid average velocity, μ is the fluid dynamic viscosity, P_r is the Prandtl number, and f is the friction factor for smooth cooling duct.

4.2. Transient State Thermal Model Equation. In order to predict the transient temperature distribution over the entire driving duty cycle, the dynamic thermal equation is utilized according to [31]

$$\frac{dK_i}{dt} = [C_t]^{-1} [P_t] - [C_t]^{-1} ([G_t] + [G_{\text{fluid}}]) [K_i], \quad (11)$$

where C_t is the thermal capacitance matrix, P_t is the heat loss matrix, K_i is the temperature rise matrix, G_{fluid} is the cooling fluid matrix, and G_t is the thermal conductance matrix.

Figure 9 shows the copper distribution individually according to (6) in a single UDDS driving duty cycle. The temperature at every single operating point is used to derive the copper loss. And Figure 10 presents the temperature variations. The highest temperature appears in the stator end winding, which reaches to 93.3°C. Over the entire driving duty cycle, the highest temperature of magnet, stator back iron, stator teeth, and winding is 41.3°C, 57.4°C, 59.1°C, and 88.3°C, respectively, and all of them come from the end of the duty cycle. In this analysis, the ambient temperature is set to 25°C.

5. Experimental Result

The test bench used to measure the temperature of the motor prototype is shown in Figure 11. The temperature measuring
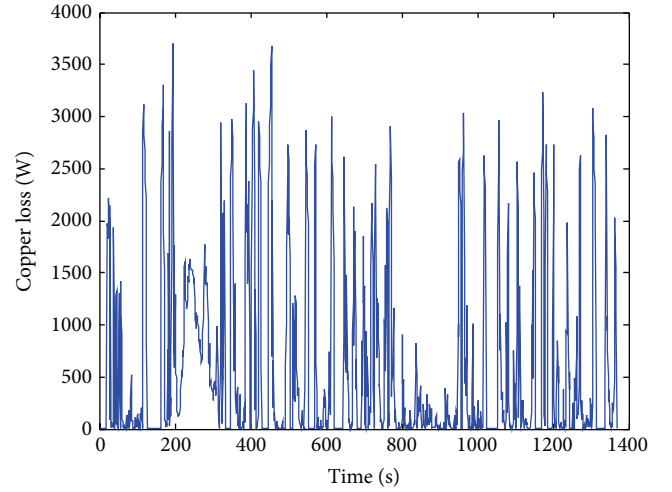
FIGURE 9: Copper loss distribution under the UDDS driving duty cycle.

—— Stator back iron	·—·— End winding
- - - Stator teeth	·—·— Winding
·—·— Magnet	
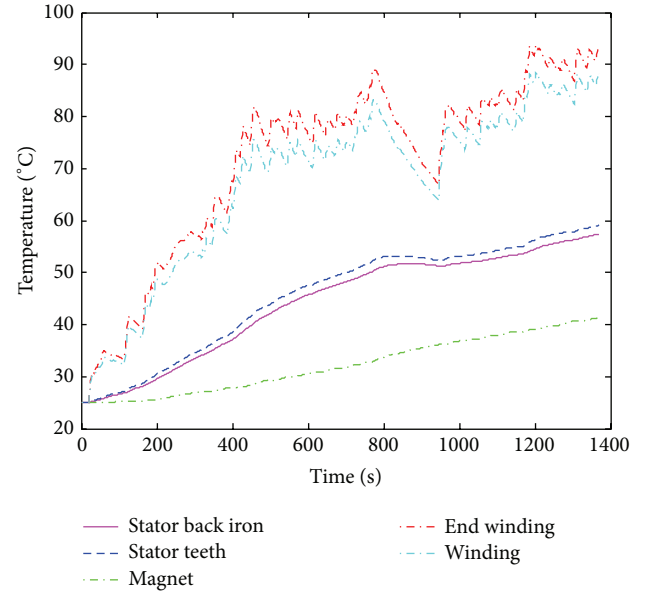
FIGURE 10: Dynamic temperature distribution under the UDDS driving duty cycle.

sensor PT1000 is placed in the end winding, since the end winding suffers from the most severe thermal condition. The ambient temperature is 25°C and the temperature difference between inlet and outlet is 2-3°C.

Both the traction motor and generator are water-cooled and controlled by an integrated controller. The operation points of the traction motor are calculated through (1) and (2); each of them runs a single second. The end-winding temperature measured by the temperature sensor is manually recorded every 20 seconds.

The test and calculated results are shown in Figure 12. Good agreement in trend is visible, while the test results are relatively lower than those derived from calculation. It is important to note that the predicted temperature dropped
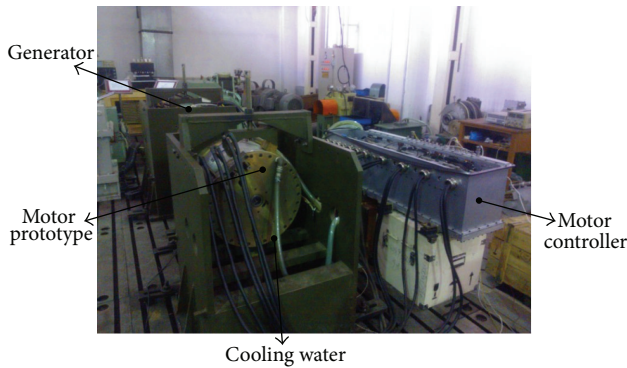
FIGURE 11: Test bench for the temperature measure.
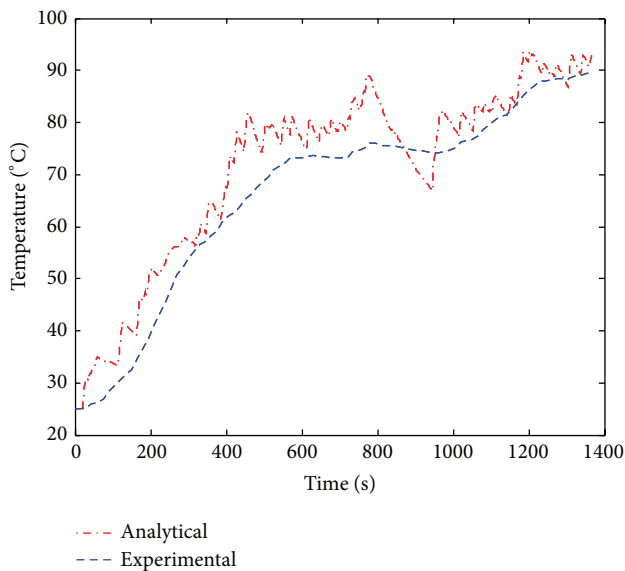
FIGURE 12: Comparison of experimental and analytical temperature distribution of end winding under one driving duty cycle.

rapidly at around 800 s, while the measured temperature remains unchanged. This is caused by time delay of mechanical coupling devices, and this is a drawback of using lumped parameter thermal model. The rest of error component can be attributed to the following issues: (a) simplifications and assumptions used in analysis, (b) loss estimation error comparing to real operating condition, and (c) unknown material thermal property data.

6. Conclusion

A lumped parameter thermal model is used to study the thermal behavior of a surface-mounted brushless AC PM machine for electrical vehicle application under the UDDS driving duty cycle. The equivalent thermal parameter of winding is calculated and introduced into the thermal model. The FEA based iron loss, PM loss, and copper loss scaling techniques are utilized to obtain the loss distribution under the duty cycle in a timely manner. The test bench for temperature measurement has been set up to validate the proposed

analysis method. And the agreement between analysis and experimental results shows this method is convincing.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[2] X. Hu, N. Murgovski, L. Johannesson, and B. Egardt, "Energy efficiency analysis of a series plug-in hybrid electric bus with different energy management strategies and battery sizes," *Applied Energy*, vol. 111, pp. 1001–1009, 2013.

[3] X. Hu, L. Johannesson, N. Murgovski, and B. Egardt, "Longevity-conscious dimensioning and power management of the hybrid energy storage system in a fuel cell hybrid electric bus," *Applied Energy*, vol. 137, pp. 913–924, 2015.

[4] J. Fan, C. Zhang, Z. Wang et al., "Thermal analysis of permanent magnet motor for the electric vehicle application considering driving duty cycle," *IEEE Transactions on Magnetics*, vol. 46, no. 6, pp. 2493–2496, 2010.

[5] K. Chau and W. Li, "Overview of electric machines for electric and hybrid vehicles," *International Journal of Vehicle Design*, vol. 64, no. 1, pp. 46–71, 2014.

[6] J. Pries and H. Hofmann, "Magnetic and thermal scaling of electric machines," *International Journal of Vehicle Design*, vol. 61, no. 1–4, pp. 219–232, 2013.

[7] R. Wrobel, G. Vainel, C. Copeland, T. Duda, D. Staton, and P. Mellor, "Investigation of mechanical loss and heat transfer in an axial-flux PM machine," in *Proceedings of the 5th Annual IEEE Energy Conversion Congress and Exhibition (ECCE '13)*, pp. 4372–4379, September 2013.

[8] D. A. Howey, A. S. Holmes, and K. R. Pullen, "Measurement and CFD prediction of heat transfer in air-cooled disc-type electrical machines," *IEEE Transactions on Industry Applications*, vol. 47, no. 4, pp. 1716–1723, 2011.

[9] D. A. Howey, A. S. Holmes, and K. R. Pullen, "Measurement of stator heat transfer in air-cooled axial flux permanent magnet machines," in *Proceedings of the 35th Annual Conference of the IEEE Industrial Electronics Society (IECON '09)*, pp. 1197–1202, November 2009.

[10] A. C. Malloy, R. F. Martinez-Botas, M. Jaensch, and M. Lamperth, "Measurement of heat generation rate in the permanent magnets of rotating electrical machines," in *Proceedings of the 6th IET International Conference on Power Electronics, Machines and Drives (PEMD '12)*, pp. 1–6, March 2012.

[11] C. P. Steinmetz, "On the law of hysteresis (originally published in 1892)," *Proceedings of the IEEE*, vol. 72, no. 2, pp. 196–221, 1984.

[12] G. Bertotti, "General properties of power losses in soft ferromagnetic materials," *IEEE Transactions on Magnetics*, vol. 24, no. 1, pp. 621–630, 1988.

[13] K. W. E. Cheng and P. D. Evans, "Calculation of winding losses in high-frequency toroidal inductors using single strand

conductors," *IEE Proceedings: Electric Power Applications*, vol. 141, no. 2, pp. 52–62, 1994.

[14] J. Acero, R. Alonso, J. M. Burdío, L. A. Barragán, and D. Puyal, "Frequency-dependent resistance in Litz-wire planar windings for domestic induction heating appliances," *IEEE Transactions on Power Electronics*, vol. 21, no. 4, pp. 856–866, 2006.

[15] A.-T. Phung, G. Meunier, O. Chadebec, X. Margueron, and J.-P. Keradec, "High-frequency proximity losses determination for rectangular cross-section conductors," *IEEE Transactions on Magnetics*, vol. 43, no. 4, pp. 1213–1216, 2007.

[16] K. Atallah, D. Howe, P. H. Mellor, and D. A. Stone, "Rotor loss in permanent-magnet brushless AC machines," *IEEE Transactions on Industry Applications*, vol. 36, no. 6, pp. 1612–1618, 2000.

[17] L. J. Wu, Z. Q. Zhu, D. Staton, M. Popescu, and D. Hawkins, "Analytical modelling and analysis of open-circuit PM power loss in surface-mounted permanent magnet machines," *IEEE Transactions on Magnetics*, vol. 48, no. 3, pp. 1234–1246, 2011.

[18] G. L. Skibinski, B. G. Schram, J. R. Brauer, and Z. Badics, "Finite element prediction of losses and temperatures of laminated composite inductors for AC drives," in *Proceedings of the IEEE International Electric Machines and Drives Conference (IEMDC '03)*, pp. 756–763, June 2003.

[19] Y. Kawase, T. Ota, and H. Fukunaga, "3-D eddy current analysis in permanent magnet of interior permanent magnet motors," *IEEE Transactions on Magnetics*, vol. 36, no. 4, pp. 1863–1866, 2000.

[20] Z. Kolondzovski, A. Belahcen, and A. Arkkio, "Comparative thermal analysis of different rotor types for a high-speed permanent-magnet electrical machine," *IET Electric Power Applications*, vol. 3, no. 4, pp. 279–288, 2009.

[21] P. H. Mellor, D. Roberts, and D. R. Turner, "Lumped parameter thermal model for electrical machines of TEFC design," *IEE Proceedings B: Electric Power Applications*, vol. 138, no. 5, pp. 205–218, 1991.

[22] R. Wrobel, J. Goss, A. Mlot, and P. H. Mellor, "Design considerations of a brushless open-slot radial-flux PM hub motor," *IEEE Transactions on Industry Applications*, vol. 50, no. 3, pp. 1757–1767, 2014.

[23] P. H. Mellor, R. Wrobel, and D. Holliday, "A computationally efficient iron loss model for brushless AC machines that caters for rated flux and field weakened operation," in *Proceedings of the IEEE International Electric Machines and Drives Conference (IEMDC '09)*, pp. 490–494, May 2009.

[24] R. Wrobel, D. E. Salt, A. Griffo, N. Simpson, and P. H. Mellor, "Derivation and scaling of AC copper loss in thermal modeling of electrical machines," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 8, pp. 4412–4420, 2014.

[25] X. Wu, R. Wrobel, P. H. Mellor, and C. Zhang, "A computationally efficient PM power loss derivation for surface-mounted brushless AC PM machines," in *Proceedings of the 21st International Conference on Electrical Machines (ICEM '14)*, pp. 17–23, IEEE, September 2014.

[26] N. Simpson, R. Wrobel, and P. H. Mellor, "Estimation of equivalent thermal parameters of impregnated electrical windings," *IEEE Transactions on Industry Applications*, vol. 49, no. 6, pp. 2505–2515, 2013.

[27] Z. Hashin, "Theory of fiber reinforced materials," NASA Contractor Report, NASA, Washington, DC, USA, 1972.

[28] E. Behrens, "Thermal conductivities of composite materials," *Journal of Composite Materials*, vol. 2, no. 1, pp. 2–17, 1968.

[29] J. Donea, "Thermal conductivities based on variational principles," *Journal of Composite Materials*, vol. 6, no. 2, pp. 262–266, 1972.

[30] R. L. Hamilton and O. K. Crosser, "Thermal conductivity of heterogeneous two-component systems," *Industrial and Engineering Chemistry Fundamentals*, vol. 1, no. 3, pp. 187–191, 1962.

[31] F. P. Incropera and D. P. DeWitt, *Fundamentals of Heat and Mass Transfer*, Wiley, New York, NY, USA, 4th edition, 2001.

Research Article

Genetic Algorithm Based Microscale Vehicle Emissions Modelling

Sicong Zhu,[1] LiSian Tey,[2] and Luis Ferreira[3]

[1]*MOE Key Laboratory for Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, 3 Shangyuan Cun, Haidian District, Beijing 100044, China*
[2]*Faculty of Civil Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[3]*School of Civil Engineering, Faculty of Engineering, Architecture and Information Technology, The University of Queensland, Brisbane St Lucia, QLD 4072, Australia*

Correspondence should be addressed to Sicong Zhu; sczhu@bjtu.edu.cn

There is a need to match emission estimations accuracy with the outputs of transport models. The overall error rate in long-term traffic forecasts resulting from strategic transport models is likely to be significant. Microsimulation models, whilst high-resolution in nature, may have similar measurement errors if they use the outputs of strategic models to obtain traffic demand predictions. At the microlevel, this paper discusses the limitations of existing emissions estimation approaches. Emission models for predicting emission pollutants other than CO_2 are proposed. A genetic algorithm approach is adopted to select the predicting variables for the black box model. The approach is capable of solving combinatorial optimization problems. Overall, the emission prediction results reveal that the proposed new models outperform conventional equations in terms of accuracy and robustness.

1. Introduction

Fuel consumption and emission estimation can be critical for comprehensive transportation planning. In light of the strong relationship between CO_2 emissions and fuel consumption [1] and the high volatility in crude oil prices, this requires more accurate and robust tools to quantify environmental impacts so that project evaluation can adequately address community expectations. Quantitative travel demand and emission models are necessary for the evaluation of future transport/land use options, as well as for the management of existing transport systems. The modelling of emissions is seen as an increasingly important tool in transportation planning and management.

Historically, car-following and traffic flow models have been developed using different theoretical basis. This has given rise to two main kinds of models of traffic dynamics, namely, microscopic representations, based on the description of the individual behavior of each vehicle, and macroscopic representations describing traffic as a continuous flow obeying global rules [2]. Strategic travel demand models tend to be large and regional in nature whereas microsimulation models are used for detailed tactical or operational testing of options. Taking the highest macroscopic level as an example, the total vehicle flow and the average speed over an entire network may be all that is provided [3]. At the lowest level of the hierarchy, high-resolution microscopic transportation models typically produce second-by-second vehicle trajectories (location, speed, acceleration, etc.). Hence, the traffic modelling and emission modelling should match by the accuracy level and aggregation level. For instance, driving cycles used for vehicle emission testing are specified on a second-by-second speed-time profile. Microscopic traffic models should integrate real time emission prediction models, which are able to utilize high-resolution transportation modelling results, therefore generating potentially more precise emission estimations.

Taking the highest macroscopic level as an example, the total vehicle flow and the average speed over an entire network may be all that is provided [4]. Correspondingly,

a number of emission models deliver the predictions based on the aggregated traffic model outputs, for example, the ARTEMIS in Europe and Mobile 6 from US [5, 6]. The common macrolevel modelling approach used to produce a mobile source emission inventory is based on two processing steps. The first step consists of determining a set of emission factors that specifies the rate at which emissions are generated, and the second step is to produce an estimate of vehicle activity. The emission inventory is then calculated by multiplying the results of these two steps together. This methodology has two major shortcomings as follows.

Inaccurate Characterization of Traffic Activities. The overall rate of error in 20-year traffic forecasts resulting from strategic transport models is likely to be large due to the uncertainty attached to input forecasts. The predictive ability of current emission models depends on (a) the errors in the input data used to estimate the amount and characteristics of traffic flow—such data are usually the result of travel demand modelling analysis that is known to have a large degree of uncertainty associated with it—and (b) the adequacy and quality of the information on emission rates that is used in the models.

Emissions Factors May Not Represent Actual Conditions Adequately. The current methods used for determining emission factors are based on average driving characteristics embodied in a predetermined driving cycle. Emission measurements are used as base values to reconstruct statistically the relationship between emission rates and average vehicle speeds. These "averaged speeds" are at variance with the vehicle dynamic operation in microscope.

At the lowest level of the hierarchy, high-resolution microscopic transportation models typically produce second-by-second vehicle trajectories (location, speed, and acceleration). Driving cycles used for vehicle emission testing are also specified on a second-by-second speed-time profile. Microscopic models should be integrated with real time emission prediction models which are able to utilize high-resolution transportation modelling results, thereby generating potentially more precise emission estimates. Several commercial microsimulation traffic packages are widely used to estimate the emissions [7].

There have been a number of modelling approaches on microlevel proposed to estimate future vehicle emissions in conjunction with the outputs of transport models. One such approach is the use of engine power as the main predictive basis. Another is the use of vehicle speed and acceleration as predictive variables. There are three main types of modelling approaches, namely, power-based, speed-based, and hybrid models. Two models from each category, as shown in Table 1, have been analysed [13]. It is found that power-based models have good performance on CO_2 emission and fuel consumption. In contrast, the predicted HC, CO, and NO_x do not match well with the measured results [8]. Moreover, it is difficult to collect some coefficients associated with instantaneous power, such as aerodynamic drag coefficient, vehicle frontal area, and gearing. The results from speed-based models highlight the need to model acceleration,

Table 1: Emission models assessed.

General approach	Model	Reference
Power-based	Commonwealth scientific and industrial research organisation (CRISCO)	[8]
	Comprehensive modal emissions model (CMEM)	[3]
Speed-based	Energy and emissions model (VT)	[9]
	Instantaneous traffic emissions	[10]
Hybrid	Microscale modelling	[11]
	Microscale modelling	[12]

deceleration, and cruising stages of the urban cycle separately. The instantaneous traffic emissions model, a speed-based approach which utilises the microtransportation simulation result as an input, was found to have merit based on the evaluation results. More complex models, whilst theoretically more desirable, may mean additional input measurement errors, such that the overall effect may not yield more accurate estimates [14].

The instantaneous traffic emissions model developed by Int Panis et al. [10] has been adopted by the AIMSUN traffic simulation model [7]. The latter integrates traffic simulation results with emission prediction equations. Emission functions for each vehicle were derived with instantaneous speed and acceleration as parameters using nonlinear multiple regression techniques. The model, shown in (1), was calibrated using data from twenty-five vehicles (six buses, two trucks, and seventeen cars) in Europe. The pollutants modelled are nitrogen oxide (NO_x), hydrocarbon (HC), carbon dioxide (CO_2), and particulate matter (PM):

$$E_n(t) = \max \left[E_0, f_1 + f_2 v_n(t) + f_3 v_n(t)^2 + f_4 a_n(t) \right. \left. + f_5 a_n(t)^2 + f_6 v_n(t) a_n(t) \right], \tag{1}$$

where E_0 is a lower limit of emission (g/s) specified for each vehicle and pollutant type; $v_n(t)$ is instantaneous speed of vehicle n at time t; $a_n(t)$ is acceleration of vehicle n at time t; f_1 to f_6 are emission constants specific for each vehicle and pollutant type determined by the regression analysis.

For certain pollutants, whenever visual inspection of the data plot reveals a clear distinction in the scatter for acceleration and deceleration, the approach suggests that different functional forms should be derived for different driving modes, such as acceleration (with $a_n(t) \geq 0.5$ m/s^2), cruising (with -0.5 m/s$^2 \leq a_n(t) < 0.5$ m/s^2), and deceleration (with $a_n(t) < -0.5$ m/s^2). The modelling results for pollutant emissions such as HC are unsatisfactory, as shown in Section 4.

The current paper aims to develop vehicle emission models for predicting emission pollutants other than CO_2. The models adopt genetic algorithm (GA) for selection of

TABLE 2: Selected average-aged vehicles for model development.

Vehicle[1]	Variant	Odometer (km)	Engine displacement (L)	Mass (kg)
323	Protégé	75,615	1.6	1215
Echo	—	53,859	1.5	921
Vectra	CD	81,666	2.0	1317
323	Protégé	62,229	1.8	1142
Camry	CSI	81,783	2.2	1395
Tarago	GLi	90,749	2.4	1615
Pulsar	LX	65,120	1.6	1067
Commodore	Executive	95,979	3.8	1654

[1] All vehicles are 4-cylinder except the Commodore (V6).

the predicting variables. This paper is structured as follows: Section 2 briefly describes the data sources used and the selection and validation prior to the model development; Section 3 proposes the model development approach; and Section 4 then discusses the findings. Finally, some overall conclusions and limitations are drawn from the results and future research avenues are recommended.

2. Data Sources, Selection, and Validation

The data used for analysis in the current paper were extracted from the Australian national in-service emissions study (NISE2) [15], which was developed using a Composite Urban Emissions Drive Cycle (petrol CUEDC). This cycle represents typical vehicle journeys in a given metropolitan study area, where data were collected in the Brisbane, Sydney, Melbourne, Adelaide, and Perth areas. The emission rates for CO_2, CO, HC, and NO_x of the test-bed vehicles from the NISE2 fleet, which travels on the composite urban driving cycle (CUEDC), were recorded second-by-second in addition to the instantaneous speed. Prior to analysis, the integrity of the emission measurements from NISE2 was reconfirmed and corrected to enhance the reliability of the instantaneous emissions using the method followed by Smit et al. [16]. In addition, cold start affected datasets were filtered prior to analysis by adopting an approach recommended by Favez et al. [17]. Eight average-aged passenger vehicles (which had travelled approximately 50,000–100,000 km each), as listed in Table 2, were shortlisted from the NISE2 database for model development. The vehicles selected are similar to Australian vehicle fleet characteristics (e.g., in mileage and age) [18]. The instantaneous emission observations were averaged over those vehicles.

The AIMSUN emission methodology was adopted for further evaluation using the NISE2 data subset described earlier. Each vehicle was "driven" through the speed profile of the CUEDC drive cycle which was simulated using a purpose-specific program. The averaged CO_2 emission observations from the selected eight vehicles were compared with the predicted second-by-second output from the AIMSUN model. The goodness of fit, R^2, for the two sets of data was 0.80, as shown in Figure 1.
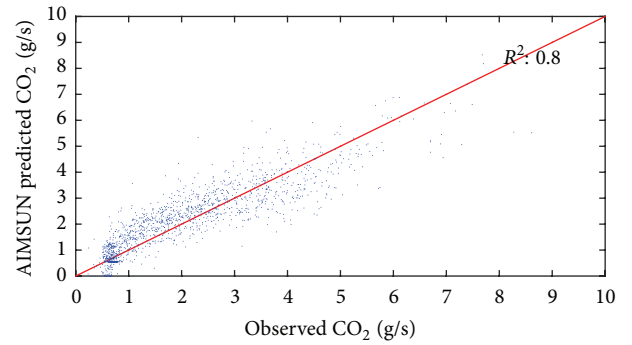
FIGURE 1: CO_2 by AIMSUN model predictions and averaged observed data.

3. Development of Emission Models

Microsimulation based modelling results for pollutants other than CO_2 are usually unsatisfactory [12]. The current paper proposes a methodology to improve those results. The concept is briefly described below.

Current microscopic models use a combination of instantaneous velocity and acceleration to predict various gaseous pollutants including HC and CO. Gaseous pollutants such as HC and CO are primarily formed during in-cylinder combustion processes depending on many factors such as air-fuel ratio, cylinder temperature and pressure, and engine speed [19]. The formation of HC and CO rises in a rich fuel environment. Thus, the fact of high correlation between CO_2 emission rate and fuel consumption was taken into account in the modelling of these gaseous pollutants. In addition, vehicle acceleration or deceleration leads to substantial change in fuel injection per combustion cycle. The change in air-fuel ratio forces the engine to adapt to a new equilibrium and tends to lead to a transient variation in pollutant formation [20]. This effect may be compounded by dynamic effects in the catalyst and exhaust system, such as catalyst malfunctioning, which can cause a sudden increase of the pollutant emissions. For these reasons, modelling of HC and CO, as the products of incomplete combustion fuel, should take the time-lag effect described above into consideration. Hence, several "historical" variables (i.e., variables at previous time steps of

t-1, t-2, t-3, or t-4 seconds) of the time-lag effect are introduced into the models as predicting variables.

A range of variables for the instantaneous and "historical" velocity, acceleration, and CO_2 emission rate were selected and tested:

$$[B_0, \ V_i, \ A_i, \ V_i^2, \ A_i^2, \ V_i A_i, \ V_i^3, \ A_i^3, \ V_i A_i^2, \ V_i^2 A_i,$$
$$V_{i-1}, \ V_{i-2}, \ V_{i-3}, \ V_{i-4}, \ A_{i-1}, \ A_{i-2}, \ A_{i-3}, \ A_{i-4},$$
$$\text{Rate\_CO}_{2i}, \ \text{Rate\_CO}_{2i-1}, \ \text{Rate\_CO}_{2i-2}, \ \text{Rate\_CO}_{2i-3},$$
$$\text{Rate\_CO}_{2i-4}],$$

where B_0 is constant; V_i is velocity at time i; A_i is acceleration at time i; Rate_CO_{2i} is CO_2 emission rate at time i.

A total of 23 candidate predicting variables were identified. An exhaustive enumeration method is not a practical alternative, while stepwise and stagewise regression procedures produce only local optimum solutions [21]. Current highly divided emissions models that are reviewed in Section 1 cannot provide any prototype of variable combination. Compared with other artificial intelligent algorithms, unsupervised heuristic genetic algorithm is independent from any *a priori* knowledge, such as training dataset. In addition, the GA is a stochastic search process that mimics the natural process of "survival of the fittest" through the manipulation of a population of chromosomes [22]. With proper binary encoding, the GA can deliver a fair hybrid emission model which synthesizes variables of different models reviewed in Section 1. To be pointed out, the GA may not deliver an optimized result due to premature convergence and "Hamming cliff" problems [23]. However, it is more difficult, but not impossible, for a GA to become trapped in a local minimum unlike the more conventional gradient methods [24].

The elite individual solution is a balanced one with high accuracy and consisting of strong statistically related variables. The GA programme was compiled using Matlab 8.4. A selective weighted fitness for the GA, (2), was used to implement an automated variables selection procedure to build the calibration models based on least-square regression:

$$\text{Fitness}_i = \frac{1}{\sqrt{\sum_1^N \left(P_i - O_i\right)^2 / N}}, \tag{2}$$

where $P_i = \sum \alpha_i \beta_i \text{Variable}_i$; P_i is estimated instantaneous emission rate at time i; O_i is observed instantaneous emission rate at time i; α_i is binary logic control parameter after regression analysis;

$$\alpha_i = \begin{cases} 1 & \text{if } P \text{ value} < 0.05 \text{ this variable is included} \\ 0 & \text{otherwise;} \end{cases} \tag{3}$$

β_i is coefficient derived from regression analysis; N is number of observations.

P_i is the prediction of emission rate as a product of selected coefficients and the corresponding candidate variables whose P value of the t-statistical analysis is lower than 0.05. The reciprocal relationship of root-mean-square error enables the accurate prediction of solutions with high fitness values.

Table 3: Parameters tuning of genetic operators.

Population size (N)	600
Crossover rate (C)	1
Crossover type	Uniform crossover
Scaling method	Sigma
Sigma scaling coefficient	1
Mutation rate (M)	0.003
Mutation method	Bit string
Selection strategy (S)	Pure selection

Figure 2 shows the flowchart of the GA application. Firstly, chromosomes that represent feasible solutions of a nonlinear optimisation problem are randomly generated to form the initial population following the fitness test (see (2)). Each individual chromosome in binary format symbolises a potential solution to the emission modelling problem. Figure 3 demonstrates an example of an individual chromosome consisting of 23 bits. Each bit represents a corresponding variable in the list of all potential variables. The dichotomous data on each bit are the choice of "included in the model" valued 1 or "not included in the model" valued 0. The size of the solution is therefore 2^{23}. In determining the fitness value, the chromosome tends to take more variables for better prediction accuracy. This may include irrelevant candidate variables and undermine the robustness of the model. In order to avoid this from happening, the calculation of the emission rate prediction, P_i, was limited to those statistically significant variables ($P < 0.05$). In order to achieve acceptable solutions, the calculations of the fitness values and selection of chromosomes continued. This procedure evolves through many generations by a natural genetic process. The genetic process includes three types of operation, namely, crossover, selection, and mutation [25]. It repeats until the number of iterations exceeds a predefined limit. The number of chromosomes in a population and the number of iterations are set to 600 and 100, respectively. The tuning of genetic operator parameters follows the simple genetic algorithm proposed by Mitchell [25]; the details are presented in Table 3.

Figures 4(a) and 4(b) show the maximum and averaged fitness reached over the iterations. The averaged fitness for each generation increases substantially at the beginning stage and reaches stability towards the end. The elite fitness trends upward and fluctuates at the beginning. After 20 generations' evolution, the fitness becomes mature. These results indicate that the chromosome evolvements improve the accuracy of the solution and reach a saturated condition eventually. After 30–40 different runs, the paper shortlists the results and selects the modelling equation based on statistical analysis. Table 4 shows three candidates for acceleration-phase HC emission modelling. Based on modelling fitness and variable significance, this paper selects candidate 1 to be the promising one. As the GA program can be configured to enable the mutation operator only, a random solution is generated by a 600×100 random process. The chosen one is compared with the random solution. However, the randomly selected candidate is hardly a solution, due to variable significance

TABLE 4: HC acceleration phase modelling candidates.

(a) Candidate 1

Modeling R^2	0.6474							
Variable name	$B0$	A	VA	V^3	V^2A	V_{i-1}	CO_2	CO_{2i-3}
P value	0.0151	0.0008	0.00172	$5.52E-05$	0.003202	$1.03E-06$	$7.94E-13$	0.016642

(b) Candidate 2

Modeling R^2	0.6449							
Variable name	$B0$	V	VA	V^3	V^2A	V_{i-1}	A_{i-3}	CO_2
P value	0.0088	0.00264	$5.99E-03$	0.000214	$1.94E-02$	$5.54E-03$	0.035107	$1.19E-10$

(c) Candidate 3

Modeling R^2	0.6493						
Variable name	$B0$	A	V^2	V^3	VA^2	A_{i-4}	CO_2
P value	0.0492	0.01617	$5.2E-08$	$1.57E-07$	0.036635	0.022996	$3.23E-14$
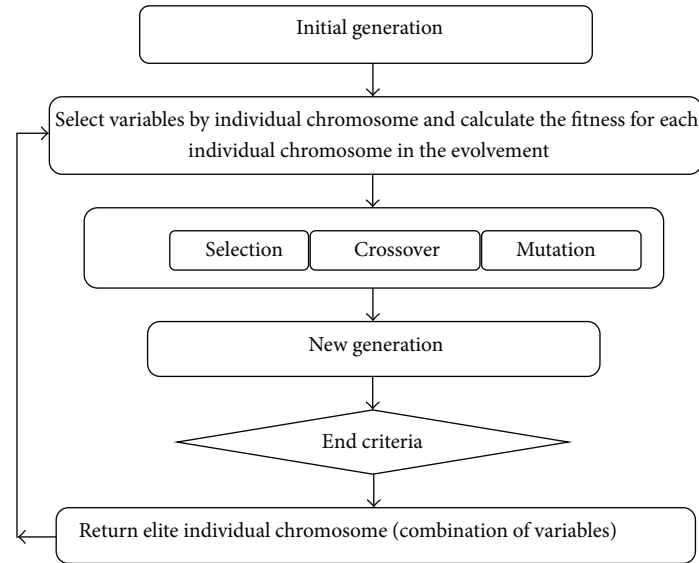
FIGURE 2: Flowchart of genetic algorithm.
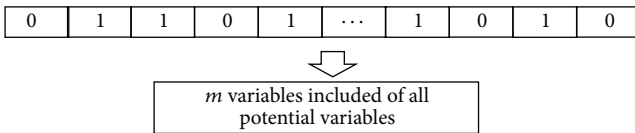
FIGURE 3: An individual chromosome.

constraint (P value). Hence the elite chromosome can be justified as the optimal (or subprime) solution.

4. Results and Main Findings

4.1. General. The proposed GA approach was applied to the HC pollutant for different driving modes, namely, acceleration, cruising, and deceleration. The models for the emission rate, Y, from the ultimate chromosome are presented in (4), (5), and (6), respectively. All the predicting variables selected are statistically significant at $P < 0.05$. Both the instantaneous traffic emissions model (see (1)) and the newly modified equation parameters were calibrated by least-square regression on the same test-bed dataset that was used to develop the new model. To conduct meaningful comparisons of modelling-induced error, the goodness of fit, R^2, for the three driving modes, between the newly modified model and model (1) is shown in Table 5. The proposed new approach significantly improved the modelling results for HC, although the prediction accuracy for the deceleration component was relatively low. The deceleration component accounts for 42% of total driving cycle duration, but the summation of HC deceleration component only takes up 25% of total cycle measurements. As a result, the deceleration component does not play an important role as a percentage of total emissions. Due to the multidimensionality of the solution search space, the GA method is not certain to lead to a global optimum when formulated as an optimization
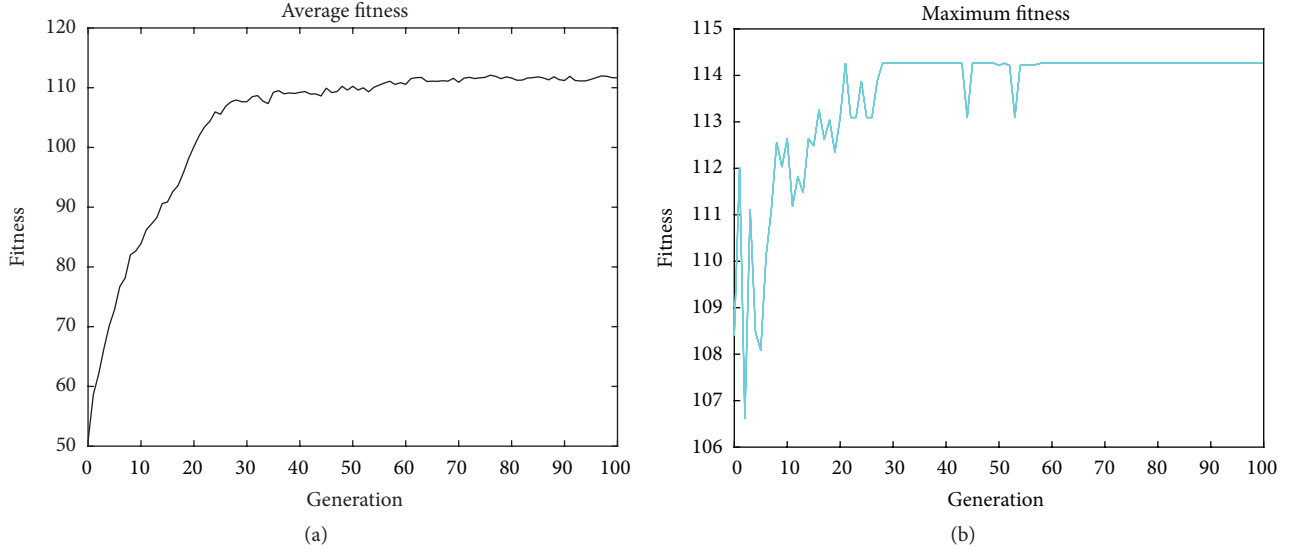
FIGURE 4: (a) Averaged fitness over generations. (b) Maximum averaged fitness over generations.

problem. GA has a tendency to converge towards local optima or even arbitrary points rather than the global optimum of the problem. This means that it is not wise to sacrifice short-term fitness to gain longer-term fitness. To tackle this issue, the proposed GA methodology relies on the fitness function, which enables quantification of individual solution appropriateness in terms of the statistical significance of both model accuracy and shortlisted variables. To test the fitness function, the size of the population was increased to 2,000 in order to enable more potential solutions to search in the multidimensional space, the results being similar to those previously presented. Moreover, taking the HC acceleration component as an example, the new proposed equation R^2 is slightly lower than the theoretical maximum value by which an individual chromosome includes all candidature variables. Hence, the new equation selected by the new algorithm is a close approximation to the global optimum.

Acceleration Component

$$Y = B_0 + B_1 A_i + B_2 V_i A_i + B_3 V_i^3 + B_4 V_i^2 A_i + B_5 V_{i-1}$$
$$+ B_6 \text{Rate\_CO}_{2i} + B_7 \text{Rate\_CO}_{2i-3}. \tag{4}$$

Cruising Component

$$Y = B_0 + B_1 A_i + B_2 V_i^2 + B_3 V_i A_i + B_4 V_i^3 + B_5 V_i^2 A_i$$
$$+ B_6 V_{i-2} + B_7 V_{i-4} + B_8 \text{Rate\_CO}_{2i} \tag{5}$$
$$+ B_9 \text{Rate\_CO}_{2i-1} + B_{10} \text{Rate\_CO}_{2i-3}.$$

Deceleration Component

$$Y = B_0 + B_1 V_i + B_2 V_i^2 + B_3 V_i^3 + B_4 V_i A_i^2 + B_5 A_{i-3}, \tag{6}$$

where Y is emission rate (g/s) for each vehicle and pollutant type; V_i is velocity at time i; A_i is acceleration at time i;

TABLE 5: Results of model development: goodness of fit, R^2.

Driving mode	Overall	Acceleration	Cruising	Deceleration
		R^2		
GA-based model	0.80	0.65	0.75	0.15
Equation (1) model	0.69	0.47	0.60	0.08

HC = hydrocarbon
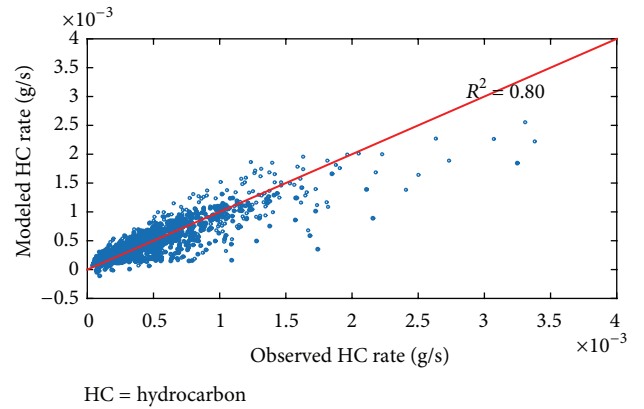
FIGURE 5: New model predicted versus measured HC.

Rate_CO$_{2i}$ is CO_2 emission rate at time i. B_1 to B_n are emission coefficients for each vehicle and pollutant type.

The same experimental dataset was used to test the prediction accuracy of the model. Figure 5 shows the correlation between modelling predictions and HC measurements, and Figure 6 illustrates the corresponding residuals plots.

4.2. Modelling Results and Validations. The proposed new models were validated on different sets of data with various vehicle size and make for different age ranges, namely, new vehicles with mileages from 1,000 to 11,000 km; middle-aged

HC = hydrocarbon
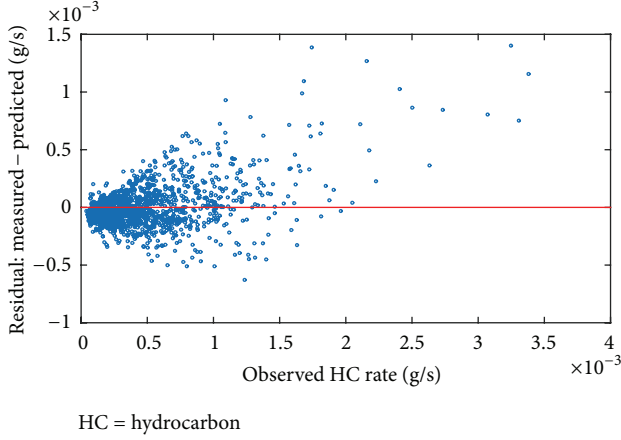
FIGURE 6: New model residuals.

vehicles with mileages from 75,000 to 95,000 km; and older vehicles with mileages from 130,000 to 140,000 km.

Comparisons of the goodness of fit, R^2, of the overall modelling results for the three vehicle age ranges, between the proposed new models (summation of (4) to (6)) and the instantaneous traffic emissions model (1) are shown in Tables 6(a) to 6(c), respectively. Overall, the emission prediction results prove that the proposed new models provide improved results. The new models are more robust and accurate for HC prediction.

Similarly, this methodology is applied to other pollutants including CO and NO$_x$. The new CO emission prediction equations for acceleration, cruising, and deceleration components are shown in (7), (8), and (9), respectively. The overall R^2 for whole test-bed dataset is 0.70.

Acceleration

$$Y = B_0 + B_1 V_i^2 + B_2 V_i^3 + B_3 V_i A_i^2 + B_4 V_i^2 A_i + B_5 V_{i-2}$$
$$+ B_6 \text{Rate\_CO}_{2i} + B_7 \text{Rate\_CO}_{2i-4}. \tag{7}$$

Cruising

$$Y = B_0 + B_1 A_i + B_2 V_i^2 + B_3 V_i A_i + B_4 V_i^3 + B_5 V_i^2 A_i$$
$$+ B_6 V_{i-1} + B_7 A_{i-2} + B_8 \text{Rate\_CO}_{2i} \tag{8}$$
$$+ B_9 \text{Rate\_CO}_{2i-1} + B_{10} \text{Rate\_CO}_{2i-3}.$$

Deceleration

$$Y = B_0 + B_1 V_i^2 A_i + B_2 \text{Rate\_CO}_{2i-4}. \tag{9}$$

New NO$_x$ emission prediction equations for acceleration, cruising, and deceleration components are shown in (10),

(11), and (12), respectively. The overall R^2 for whole test-bed dataset is 0.82:

$$Y = B_0 + B_1 V_i^2 + B_2 V_i^3 + B_3 A_i^3 + B_4 V_i A_i^2 + B_5 V_i^2 A_i$$
$$+ B_6 A_{i-2} + B_7 \text{Rate\_CO}_{2i} + B_8 \text{Rate\_CO}_{2i-2} \tag{10}$$
$$+ B_9 \text{Rate\_CO}_{2i-3}$$

$$Y = B_0 + B_1 V_i + B_2 A_i + B_3 A_i^2 + B_4 V_i A_i + B_5 A_i^3$$
$$+ B_6 V_i A_i^2 + B_7 V_i^2 A_i + B_8 V_{i-4} + B_9 A_{i-4}$$
$$+ B_{10} \text{Rate\_CO}_{2i} + B_{11} \text{Rate\_CO}_{2i-1} \tag{11}$$
$$+ B_{12} \text{Rate\_CO}_{2i-2} + B_{13} \text{Rate\_CO}_{2i-3}$$

$$Y = B_0 + B_1 V_i A_i + B_2 A_{i-4} + B_3 \text{Rate\_CO}_{2i-4}. \tag{12}$$

5. Paralleled Genetic Algorithm

Cantú-Paz [26] classified parallel GAs into three main types:

   (i) Global single-population master-slave GA.

  (ii) Single-population fine-grained GA.

 (iii) Multiple-population coarse grained GA.

Single-population fine-grained GA and multiple-population coarse grained GAs are suitable to tackle dynamic function optimization problems [27]. It has an important role in optimizing complex functions whose optima vary in time (learning-like process). In a master-slave GA there is a single population, but the evaluation of fitness is distributed among several processors. Matlab 8.4 enables the full functionality of the parallel language features by creating a special job on a pool of workers and connecting the pool to the Matlab client [28]. Distributed synchronous GA is based on distribution of workload among processors during the fitness function evaluation phase followed by single central population regeneration. Hence, the massive fitness computations are assigned to workers in order to improve the computation efficiency. When testing on a workstation with i7 CPU (3.6 GHz) and 32 G memory, one generation of genetic algorithm consumes 44.0 seconds without paralleled configuration and 11.8 seconds with 4-worker paralleled configuration.

6. Conclusions

Past research on modelling vehicle emissions other than CO$_2$ reveals relatively weak predicting results. The current paper proposes a GA based methodology to determine the contributing variables for predicting vehicle emissions. This method provides a new approach to the selection of a combination of variables among a large potential set. The applications of the new models show enhanced results for modelling vehicle emissions, supporting the new variable selection methodology using GA. The modified fitness function for the proposed GA demonstrates the ability to establish a balanced multivariate model. In addition, the improved HC prediction results, obtained by introducing "historical" CO$_2$

TABLE 6: (a) Model validation, R^2 (new vehicles). (b) Model validation, R^2 (middle-aged vehicles). (c) Model validation, R^2 (older vehicles).

(a)

	Vehicles							
	323	Echo	Vectra	323	Camry	Tarago	Pulsar	Commodore
GA-based model	0.68	0.59	0.84	0.51	0.82	0.80	0.41	0.84
Equation (1) model	0.61	0.55	0.79	0.44	0.80	0.72	0.38	0.80

(b)

	Vehicles							
	323	Echo	Vectra	323	Camry	Tarago	Pulsar	Commodore
GA-based model	0.68	0.59	0.84	0.51	0.82	0.80	0.41	0.84
Equation (1) model	0.61	0.55	0.79	0.44	0.80	0.72	0.38	0.80

(c)

	Vehicles						
	Festiva	Lancer	Astra	Civic	Astra	Lancer	Pulsar
GA-based model	0.69	0.74	0.61	0.23	0.79	0.85	0.62
Equation (1) model	0.56	0.69	0.53	0.16	0.67	0.82	0.27

emission rates, support the time-lag effect hypothesis. The proposed GA methodology provides a solution for a combinatorial optimization problem, providing high modelling accuracy with statistically significant relationships between the selected predicting variables and the dependant variable.

Future research focus should extend the evaluation of models to include the full set of particles matter (PM). In addition, the accuracy of new models may be compromised when driving with loading or in hilly terrain. Moreover, the developed models produce relatively low accuracy in predicting emissions in deceleration mode. According to Heywood [19], HC, CO, and NO$_x$ emissions depend on the fuel-air equivalence and injection timing. The sharp deceleration (with $a_n(t) < -0.5\,\text{m/s}^2$) results in the misbalance of fuel-air equivalence and variance of injection timing.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] R. Akcelik and M. Besley, "Operating cost, fuel consumption and emission models in aaSIDRA and aaMOTION," in *Proceedings of the 25th Conference of Australian Institutes of Transport Research (CAITR '03)*, University of South Australia, Adelaide, Australia, December 2003.

[2] E. Bourrel and J.-B. Lesort, "Mixing microscopic and macroscopic representations of traffic flow: hybrid model based on Lighthill-Whitham-Richards theory," *Transportation Research Record*, vol. 1852, pp. 193–200, 2003.

[3] M. Barth and G. Scora, *Comprehensive Modal Emissions Model (CMEM), User's Guide, Version 3.01*, University of California, Riverside, Riverside, Calif, USA, 2006.

[4] M. Barth, F. An, J. Norbeck, and M. Ross, "Modal emissions modeling: a physical approach," *Transportation Research Record*, vol. 1520, pp. 81–88, 1996.

[5] M. André and M. Rapone, "Analysis and modelling of the pollutant emissions from European cars regarding the driving characteristics and test cycles," *Atmospheric Environment*, vol. 43, no. 5, pp. 986–995, 2009.

[6] EPA, *User's Guide to MOBILE6.1 and MOBILE6.2; Mobile Source Emission Factor Model, EPA420-R-03-010*, US Environmental Protection Agency, Washington, DC, USA, 2003.

[7] TSS, *Aimsun 6.1 Manual*, Transport Simulation System, Barcelona, Spain, 2010.

[8] D. Y. C. Leung and D. J. Williams, "Modelling of motor vehicle fuel consumption and emissions using a power-based model," *Environmental Monitoring and Assessment*, vol. 65, no. 1-2, pp. 21–29, 2000.

[9] K. Ahn, H. Rakha, A. Trani, and M. Van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *Journal of Transportation Engineering*, vol. 128, no. 2, pp. 182–190, 2002.

[10] L. Int Panis, S. Broekx, and R. Liu, "Modelling instantaneous traffic emission and the influence of traffic speed limits," *Science of the Total Environment*, vol. 371, no. 1–3, pp. 270–285, 2006.

[11] Y. Qi, H. Teng, and L. Yu, "Microscale emission models incorporating acceleration and deceleration," *Journal of Transportation Engineering*, vol. 130, no. 3, pp. 348–359, 2004.

[12] R. Smit and J. McBroom, "Use of microscopic simulation models to predict traffic emissions," *Road and Transport Research*, vol. 18, no. 2, pp. 49–54, 2009.

[13] S. Zhu and L. Ferreira, "Evaluation of vehicle emissions models for micro-simulation modelling: using CO$_2$ as a case study," *Road and Transport Research*, vol. 21, pp. 3–18, 2012.

[14] R. Smit, H. Dia, and L. Morawska, "Road traffic emission and fuel consumption modelling: trends, new developments and future challenges," in *Traffic-Related Air Pollution*, pp. 29–68, Nova Science Publishers, New York, NY, USA, 2009.

[15] Orbital Australia, *Second National In-Service Emissions Study (NISE2): Light Duty Petrol Vehicle Emissions Testing*, Road and Traffic Authority of New South Wales, 2009.

[16] R. Smit, J. Steele, and M. Wilson, *Vehicle Emissions Factor Development—NISE2 Data Verification*, PaeHolmes, Brisbane, Australia, 2010.

[17] J.-Y. Favez, M. Weilenmann, and J. Stilli, "Cold start extra emissions as a function of engine stop time: evolution over the last 10 years," *Atmospheric Environment*, vol. 43, no. 5, pp. 996–1007, 2009.

[18] ABS, *Motor Vehicle Census*, 9309.0, Australian Bureau of Statistics, Canberra, Australia, 2009.

[19] J. B. Heywood, *Internal Combustion Engine Fundamentals*, McGraw-Hill, New York, NY, USA, 1988.

[20] R. J. North, R. B. Noland, W. Y. Ochieng, and J. W. Polak, "Modelling of particulate matter mass emissions from a light-duty diesel vehicle," *Transportation Research Part D: Transport and Environment*, vol. 11, no. 5, pp. 344–357, 2006.

[21] S. Chatterjee, M. Laudato, and L. A. Lynch, "Genetic algorithms and their statistical applications: an introduction," *Computational Statistics and Data Analysis*, vol. 22, no. 6, pp. 633–654, 1996.

[22] I. Vladisavljevic, P. T. Martin, and A. Stevanovic, "Integration of mathematical and physical simulation to calibrate car-following behaviour of unimpaired and impaired drivers," *World Review of Intermodal Transportation Research*, vol. 1, no. 4, pp. 403–418, 2007.

[23] J. J. Grefenstette, "Optimization of control parameters for genetic algorithms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 16, no. 1, pp. 122–128, 1986.

[24] R. Q. Sardiñas, M. R. Santana, and E. A. Brindis, "Genetic algorithm-based multi-objective optimization of cutting parameters in turning processes," *Engineering Applications of Artificial Intelligence*, vol. 19, no. 2, pp. 127–133, 2006.

[25] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, Mass, USA, 1998.

[26] E. Cantú-Paz, "A survey of parallel genetic algorithms," *Calculateurs Paralleles, Reseaux et Systems Repartis*, vol. 10, pp. 141–171, 1998.

[27] G. Luque and E. Alba, *Parallel Genetic Algorithms Theory and Real World Applications*, Springer, New York, NY, USA, 2011.

[28] MathWorks, *MATLAB Documentation*, MathWorks, Natick, Mass, USA, 2012.

Research Article

Online Energy Management of Plug-In Hybrid Electric Vehicles for Prolongation of All-Electric Range Based on Dynamic Programming

Zeyu Chen,[1] Weiguo Liu,[2] Ying Yang,[1] and Weiqiang Chen[1,3]

[1]*School of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China*
[2]*Zhejiang Key Laboratory of Automobile Safety Technology, Hangzhou 310000, China*
[3]*BYD Auto Industry Co., Ltd., Shenzhen 518000, China*

Correspondence should be addressed to Zeyu Chen; chenzy@mail.neu.edu.cn

The employed energy management strategy plays an important role in energy saving performance and exhausted emission reduction of plug-in hybrid electric vehicles (HEVs). An application of dynamic programming for optimization of power allocation is implemented in this paper with certain driving cycle and a limited driving range. Considering the DP algorithm can barely be used in real-time control because of its huge computational task and the dependence on *a priori* driving cycle, several online useful control rules are established based on the offline optimization results of DP. With the above efforts, an online energy management strategy is proposed finally. The presented energy management strategy concerns the prolongation of all-electric driving range as well as the energy saving performance. A simulation study is deployed to evaluate the control performance of the proposed energy management approach. All-electric range of the plug-in HEV can be prolonged by up to 2.86% for a certain driving condition. The energy saving performance is relative to the driving distance. The presented energy management strategy brings a little higher energy cost when driving distance is short, but for a long driving distance, it can reduce the energy consumption by up to 5.77% compared to the traditional CD-CS strategy.

1. Introduction

Circumstance pollution, energy crisis, and global warming aggravated by urban transportation have attracted remarkable attention recently, impelling the rapid development of alternative energy-based vehicles, among which the HEVs have been recognized as a promising type of vehicle due to their potential to enhance the energy economy and reduce exhausted emission [1]. The advantages of HEVs benefit from two onboard energy sources, gasoline and electricity, which enable the power demand to be split between the engine and onboard battery pack. According to the capacity of onboard battery pack and its capability to be recharged directly from power grid, HEVs can be further categorized into normal HEVs and plug-in HEVs. Unlike the traditional normal HEVs that can only be operated at a charge sustaining (CS) manner, plug-in HEVs can also be operated at charge depleting (CD) manner thanks to the possession of onboard large capacity battery pack [2]. Plug-in HEVs can achieve better energy saving performance compared to normal HEVs, but the operation manners and power allocation become more complicated accordingly due to multiple operation manners.

Energy management strategy plays a significant role in the supervisory control of both normal HEVs and plug-in HEVs to reach the optimal power split policy [3–5]. The energy management problem can be normally formulated as an optimization problem, which is to minimize a cost function by determining the rational operation manner and optimal power allocation. To resolve this problem, several advance optimization algorithms have been employed [6–11], the most representative of which is dynamic programming (DP). For example, Zhang and Xiong [12] deployed DP-based adaptive energy management on different driving conditions,

incorporating a fuzzy driving condition recognition algorithm for plug-in HEVs. Pérez and Pilotta [13] implemented DP to resolve the energy management problem of HEVs on the finite time horizon and used the optimization results as a benchmark for other strategies' design and components sizing optimization. The implementation of DP relies on an *a priori* driving cycle. According to the driving cycle it uses, specific or random, DP algorithm can be further divided into two categories: deterministic DP and stochastic DP.

Although the deterministic DP is effective in reaching the global optimality, it can barely be used in real-time control because the specific driving cycle is unknown. A viable method for the deterministic DP to be utilized in the online application is to operate DP offline with some representative driving cycles first and then extract online control policy from the optimization results [14]. The drawback of this method is that the extraction process may lose optimality to some extent. But control performance of the control policy extracted from offline optimization has been recognized as better than that of intuition-based control algorithm [15]. Unlike deterministic DP, stochastic DP is to implement DP algorithm based on prediction of future driving condition, which is obtained by stochastic method using a discrete-time Markov chain [16, 17]. The potential of stochastic DP if useful information of driving cycle is available has been investigated in [18] by dividing the prediction into three levels. Stochastic DP can achieve a near-optimal control performance, and it is time-invariant and capable of being implemented online. Stochastic DP does not rely on a specific known driving cycle as requisite to implement, but it is sensitive to the real-world driving condition as well because the transition probabilities in Markov chain are based on certain collected driving cycles [19]; if the actual driving condition is obviously different from the collected driving data, stochastic DP algorithm cannot always guarantee the validity.

The method that extracts online control policy based on offline DP optimization is employed in this paper. Three contributions have been made. (1) The deterministic DP algorithm is implemented based on certain driving cycle and a limited driving distance. The presented DP algorithm-based optimal strategy is to determine the optimal control policy of power allocation between engine and battery pack in the CS operation manner. (2) An online energy management approach is proposed for the uncertain real-time control application with consideration of the prolongation of all-electric driving range. In this approach, the CD manner and CS manner are switched not only by battery state of charge (SoC) but also by the power request level at each time step. Power allocation in CS manner is determined by the control policy extracted from DP to guarantee the near-optimal energy saving performance. (3) The performance of the presented energy management approach is evaluated by a simulation study and compared to the traditional CD-CS strategy. The correlation between energy saving performance and driving distance is preliminarily disclosed.

The remainder of this paper is organized as follows: configuration of the studied plug-in HEV is introduced and modeled in Section 2; the formulation of energy management problem and dynamic programming algorithm and its
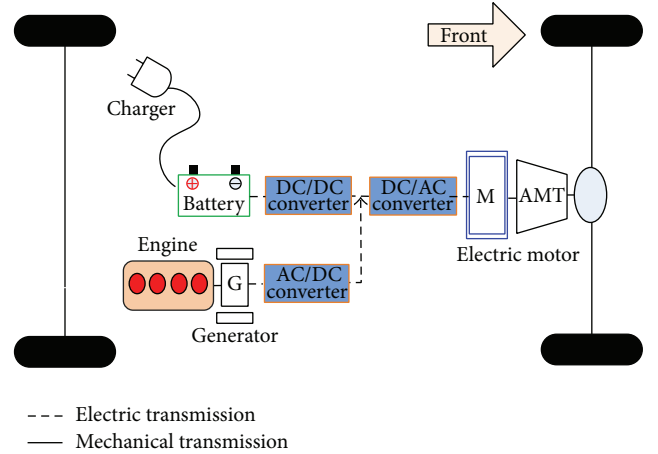
FIGURE 1: Configuration of the plug-in HEV power system.

implementation are described in Section 3. The online energy management strategy against uncertain driving condition is proposed in Section 4. Simulation results and analysis are shown in Section 5 while the conclusions are listed in Section 6.

2. System Model

Configuration of the powertrain in the studied plug-in HEV is shown in Figure 1. An electric motor is coupled through an automatic mechanical transmission (AMT) on the fist axle. A lithium-ion battery pack is used as energy storage system (ESS) while the engine-generator set is used as auxiliary power unit (APU). Parameters of the main components are listed in Table 1.

Dynamics effects of the powertrain components are neglected because they are much faster than energy consumption variation and do not affect the power flow distribution [20]; thus, the power balance relationship is modeled as

$$P_{\text{ESS}}(t) + P_{\text{APU}}(t) = P_{\text{req}}(t), \tag{1}$$

where P_{ESS}, P_{APU} are output power from ESS and APU, respectively, and P_{req} is defined as the power request of driving electric motor, calculated as follows:

$$P_{\text{ESS}}(t) = \begin{cases} P_B(t)\,\eta_{\text{batt}}, & P_B \geq 0 \\ \dfrac{P_B(t)}{\eta_{\text{batt}}}, & P_B < 0, \end{cases}$$

$$P_{\text{APU}}(t) = P_E(t)\,\eta_A, \tag{2}$$

$$P_{\text{req}}(t) = \begin{cases} \dfrac{T_m(t)\,\omega_m(t)}{\eta_{\text{mot}}}, & T_m(t) \geq 0 \\ T_m(t)\,\omega_m(t)\,\eta_{\text{mot}}, & T_m(t) < 0, \end{cases}$$

where $P_B(t)$ is battery power, positive value represents discharging, negative value represents charging, $P_E(t)$ is engine power, T_m is the torque of electric motor, positive value

TABLE 1: Parameters of main components.

Components	Parameters	Value
Engine	Peak power (kW)	75
	Optimal point (kW@r·min^{-1})	30@2200
Electric motor	Peak power (kW)	155
	Maximum torque (Nm)	780
	Maximum speed (r/min)	5500
Battery	Capacity (Ah)	85
	Rated voltage (V)	380

FIGURE 2: Equivalent circuit of generator and rectifier.

FIGURE 3: The universal characteristics curve of engine.

represents driving, negative value represents braking, ω_m is the rotate speed of electric motor, and $\eta_{\text{batt}}(t)$, $\eta_A(t)$, and $\eta_{\text{mot}}(t)$ are the efficiency of battery pack, APU, and electric motor, respectively.

2.1. Battery Model.
Battery pack is modeled by an equivalent circuit analogy. The impact of temperature variation is neglected in this study; then the battery terminal voltage is

$$V_{\text{batt}}(t) = V_{\text{oc}}(\text{SoC}(t)) - R_0(\text{SoC}(t)) \cdot I_{\text{batt}}(t), \quad (3)$$

where V_{oc} and R_0 are the open circuit voltage and internal resistance.

The Ampere-hour counting approach is adopted to calculate SoC:

$$\text{SoC}(t) = \text{SoC}(t_0) - \frac{1}{Q_{\text{nom}}} \int_{t_0}^{t} I_{\text{batt}}(t) \, dt, \quad (4)$$

where Q_{nom} is the nominal capacity of battery.

2.2. Engine-Generator Model.
The electric generator driven through a speed-increase gearbox by the engine is connected with the DC bus by an uncontrolled rectifier. The equivalent electric circuit diagram of electric generator and rectifier is shown in Figure 2, while Figure 3 gives the efficiency map of the engine. Based on the equivalent circuit, the relationship among the engine-generator output current, the DC bus voltage, the torque, and the speed of the electric generator can be obtained as

$$\omega_g(t) = \frac{U_{\text{dc}}(t)}{K_e - I_A(t)K_x},$$

$$T_g(t) = K_e I_A(t) - K_x I_A(t)^2, \quad (5)$$

where $T_g(t)$ and $\omega_g(t)$ are the output torque and speed of the electric generator, respectively, $I_A(t)$ is the current of DC bus, K_e is the equivalent electromotive force coefficient, K_x is the equivalent resistance coefficient, and $U_{\text{dc}}(t)$ is the DC bus voltage.

According to the electric generator torque and the target engine speed, the engine torque is dynamically adjusted by the engine controller. The relationship between the engine torque T_e and the electric generator torque T_g is

$$T_e(t) = \frac{1}{i_z \eta_z} \left(T_g - \left(\frac{J_e}{i_z^2} + J_g \right) \frac{d\omega_g(t)}{dt} \right), \quad (6)$$

where J_e is engine rotational inertia, J_g is the electric generator rotational inertia, and i_z and η_z are the speed-increase gearbox transmission ratio and mechanical efficiency.

2.3. Electric Motor Model.
The electric motor is modeled by using the steady-state experimental data with a dynamic correction. Figure 4 gives the external characteristic and efficiency map of the adopted electric motor. The torque of electric motor is determined by the driver operation, as shown in (7), in which the dynamic response process is equivalent to a first-order delay:

$$T_m(t) = \begin{cases} \dfrac{1}{1+\tau s} T_{m\_\max}(\omega_m(t))(\lambda_d^* - \varepsilon u), & \lambda_d^* \geq 0 \\ \dfrac{1}{1+\tau s} T_{m\_\min}(\omega_m(t))\lambda_d^*, & \lambda_d^* < 0, \end{cases} \quad (7)$$

where τ is the time constant of torque dynamic response, λ_d^* is the operation signal of driver pedals, positive value represents driving pedal, negative value represents braking pedal, $T_{m\_\max}$ and $T_{m\_\min}$ are the upper and lower limit of motor torque under the current motor speed, u is the vehicle velocity, and ε is a small positive constant, used as a velocity feedback gain to improve the control stability.

2.4. Vehicle Model.
Since the electric motor is mechanically connected to the wheel, the motor rotate speed $\omega_m(t)$ can be
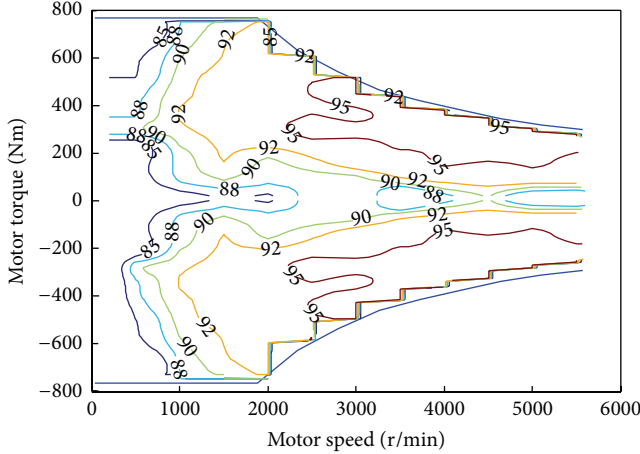
FIGURE 4: Efficiency map of electric motor.

TABLE 2: The values of coefficients in models.

The coefficients	Value or range
λ_d	$[-1, 1]$
ε	0.002
i_z	2
K_e	1.65
K_x	0.00037
τ	0.46
C_d	0.3
A (m^2)	2.2
ψ	0.018

calculated from the vehicle speed. For the energy management problem, the yawing force is neglected, and then the longitudinal dynamics is

$$u(t)$$
$$= \frac{1}{\delta m} \int_{t_0}^{t} \left(\frac{T_m(t) i_T \eta_{\text{mot}}}{r_w} - mg\psi - \frac{1}{2} C_d \rho_a A u(t)^2 \right) dt, \quad (8)$$

where m is vehicle mass, δ is correction coefficient of rotating mass, g is gravitational acceleration, C_d, ρ_a, and A are coefficient of air resistance, air density, and vehicle frontal area, respectively, ψ is coefficient of road resistance, i_T is transmission ratio, and r_w is the radius of wheel.

The values of coefficients in models are shown in Table 2.

3. Optimization Algorithm

3.1. Formulation of Optimal Energy Management Problem.

The purpose of DP algorithm here is to obtain the optimal power allocation between battery pack and engine in CS manner. As mentioned above, plug-in HEVs can be operated both in CD manner and in CS manner. The engine turns off in CD manner and battery pack is used to supply power to electric motor single-handedly so as to achieve a clean and economically driving manner. Obviously, optimization process does not work at this situation because only electricity is consumed. When CS manner is adopted, power demand is split between engine and battery; this is where DP algorithm comes in to play; the optimization objective is denoted as the summation of energy cost from two energy sources:

$$J(u) = m_{\text{equ}} = \int_{t_0}^{t_f} \left(\dot{m}_f(t) + \dot{m}_{\text{batt}}(t) \right) dt, \quad (9)$$

where m_f and m_{batt} are the fuel and electricity consumption, respectively, t_0 represents the time moment that the vehicle switches to CS manner, and t_f is the terminal time of CS manner.

Considering the battery will be unsafe when SoC is quite low, a penalty function is introduced into the "cost function," described as

$$J(u) = m_{\text{equ}} = \int_{t_0}^{t_f} \left(\dot{m}_f(t) + \dot{m}_{\text{batt}}(t) \right) dt + \sigma(\text{SoC}), \quad (10)$$

where $\sigma(\text{SoC})$ is a penalty value when SoC is quite low.

The energy cost rate of oil and electricity can be calculated by

$$\dot{m}_f(t) = f_E \left(\omega_{\text{opt}}(P_E(t)), P_E(t) \right) \cdot P_E(t),$$
$$\dot{m}_{\text{batt}}(t) = Q_{\text{equ}} \cdot P_B(t), \quad (11)$$

where $f_E()$ is the look-up table function from engine efficiency map, ω_{opt} is rotate speed of engine corresponding to optimal efficiency, and Q_{equ} is the equivalent factor to transform the electric energy to fuel consumption.

The control variable is set as $u = [P_E, P_B]^T$; thus, the optimal control problem can be described as to find an optimal control policy u^* to minimize the cost function:

$$J(u^*) \leq J(u) \quad \forall u \in U. \quad (12)$$

To ensure the optimal control policy obtained by DP belongs to feasible solutions, the control variables are subject to some constraints.

3.1.1. Inequality Constraints.

Inequality constraints define the power limitations for the operation characteristics of components in power system, described as

$$0 \leq P_E \leq P_{E\_\text{max}},$$
$$P_{B\_\text{min}} \leq P_B \leq P_{B\_\text{max}}, \quad (13)$$
$$\text{SoC}_{\text{min}} \leq \text{SoC} \leq \text{SoC}_{\text{max}},$$

where P_{E_max} is upper limit of engine power, $P_{\text{batt\_min}}$ and $P_{\text{batt\_max}}$ are upper limit and lower limit of battery power, respectively, and SoC_{min} and SoC_{max} are the upper limit and lower limit of battery SoC, respectively.

3.1.2. Equality Constraints.

The engine power and battery power are subject to the equality constraint described in (1).

The power request and battery SoC are used to represent system state, denoted as $x = [P_{\text{req}}, \text{SoC}]^T$; the equality constraint of initial system state is

$$x(t_0) = x_0, \tag{14}$$

where x_0 is the initial system state when CS manner is switched on.

The system state is uncertain at terminal due to the existence of CD manner. Therefore, the equality constraint of terminal should be eliminated for plug-in HEVs. However, in order to implement the DP algorithm, an equality constraint is put on the terminal state, described as

$$x(t_f) = x_f, \tag{15}$$

where x_f is the system state at the end of CS manner.

3.2. Dynamic Programming. A deterministic DP algorithm is employed to resolve the optimization problem formulated above. DP is a discrete-time global optimal algorithm based on a property that no matter what the previous decisions constitute, the remaining decisions should be an optimal policy. Firstly the previous control problem is rewritten in a discrete-time form for DP application, as shown in Figure 5. At each time step k ($k = n-1, n-2, \ldots, 0$), the function J when moving from the time step k to the end of the optimization horizon is calculated.

The optimization principle is applied from the end step; when $k = n$, there is $J_n(x) = \varphi(x_n)$. Then for each step,

$$J_k(x) = J_{k+1}(x) + \min_{u_k \in U} \left(L(x_k, u_k, t) \right). \tag{16}$$

Starting from the end step to first step, the total cost can be deduced as

$$J_0(\pi) = \varphi(x_n) + \sum_{0}^{n-1} L(x_k, u_k, t), \tag{17}$$

where $\pi = [u_0, u_1, \ldots, u_{n-1}]$ is the discrete control policy.

After the DP operates from the end step to the start point, the optimal control policy π^* can be determined. Since engine power and battery power are subject to an equality constraint, engine power P_E is chosen as the control variable of DP to be discretized at each iterative step, as shown in Figure 3. The terminal state variable is set in advance and then, at each step, SoC_k can be deduced by SoC_{k+1} with the control policy $P_E(k)$:

$$\text{SoC}(k) = \text{SoC}(k+1) + \frac{\left(P_{\text{req}}(k) - P_E(k)\right)\Delta t}{\eta_{\text{batt}} Q_{\text{nom}}}. \tag{18}$$

3.3. Implementation of Optimization Algorithm. Two driving cycles (HWFET and IM240, as shown in Figure 6) are used to implement DP algorithm to resolve the previous optimization problem. As the power allocation is quite clear in CD manner, DP is only employed to obtain the optimal power allocation in CS manner, in which the battery and engine output power together meet the power demand.
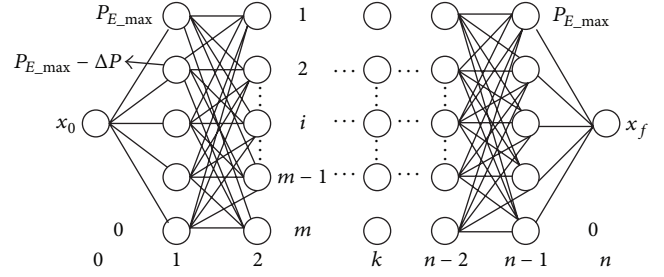
FIGURE 5: Discrete-time state for DP algorithm.
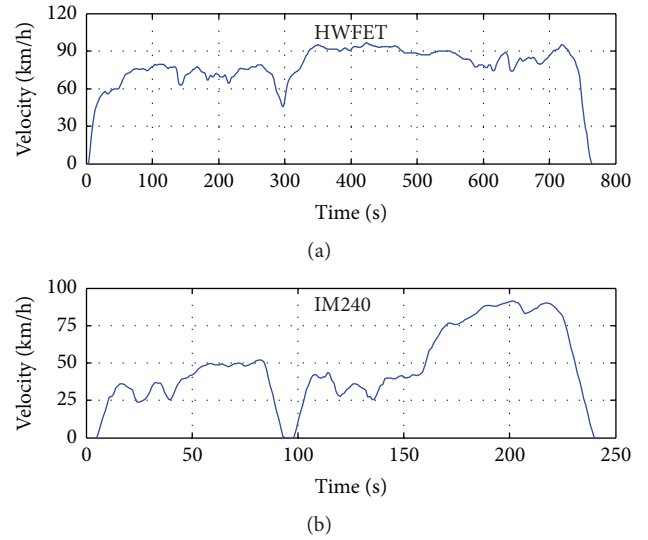
FIGURE 6: Driving cycles for DP implementation.

Figure 7 shows the optimization results of power allocation in two driving cycles. Although DP cannot be directly used in real-time control, it can provide some useful information about the rational power allocation. Two thresholds χ_1 and χ_2 are defined as engine power control instructions with the analysis of DP optimization results, described as follows:

(1) It can be concluded that the engine is turned off by DP when the power request from engine is low. Thereby, χ_1 is chosen as the threshold to determine the lower limit of the engine output power (here, χ_1 is assigned as 30 kW).

(2) Similarly, χ_2 is defined as a threshold corresponding to the upper limits of engine output power in Figures 7(a) and 7(c) (here, χ_2 is assigned as 42 kW). When power request exceeds this threshold, the rest of power demand is provided by the battery.

4. Online Energy Management Strategy

Based on the optimization results of power allocation by DP, some specific control rules will be extracted and further form the online energy management strategy in this section. As mentioned above, the operation manner of plug-in HEVs is
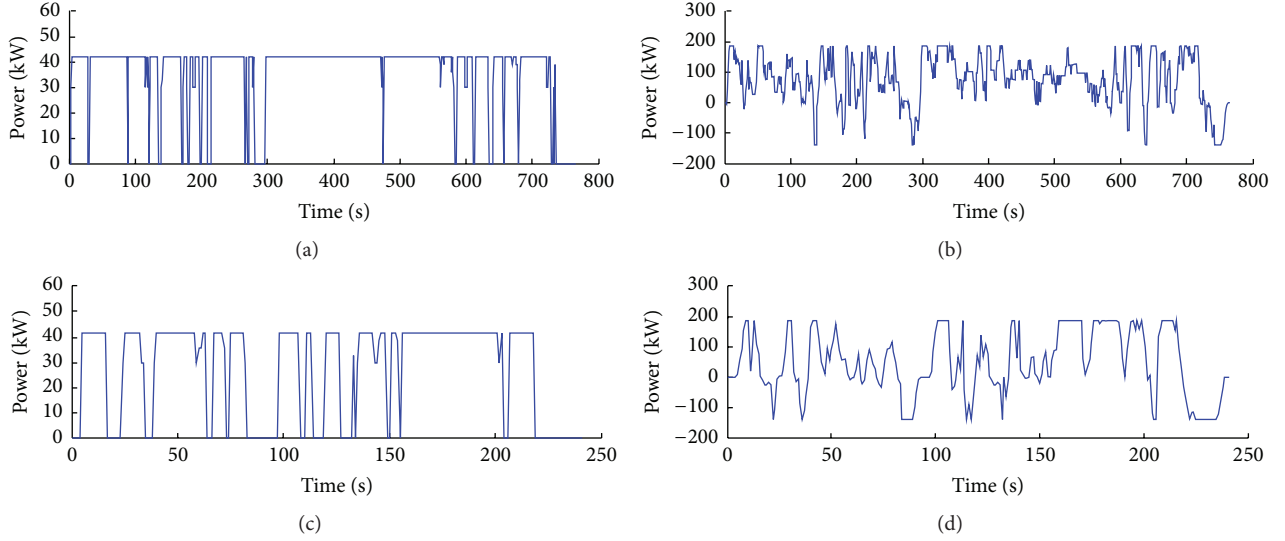
FIGURE 7: Optimization results of DP algorithm from two driving conditions. (a) Engine power in HWFET. (b) Battery power in HWFET. (c) Engine power in IM240. (d) Battery power in IM240.

divided into CD and CS manner. The driving range covered by CD manner is zero emissions, referred to as all-electric range (AER). In existing research results, the most common used online energy management approach is CD-CS strategy [21], which utilizes the CD manner at first and switches to CS manner when a long driving distance is required; the boundary between CD and CS manner is usually controlled according to the threshold of battery SoC. The shortage of CD-CS strategy is that the control policies are independent with power request of driving condition. For example, if the front part of the route has a quite high power request, according to the CD-CS strategy, vehicle will operate at CD manner at first and the electricity will be consumed at a rapid speed, resulting in a very short AER. In this paper, prolongation of AER is considered as a control objective of the online energy management.

The presented energy management strategy contains two layers: the top layer is to determine the operation manner and the bottom layer concerns the specific power allocation based on DP. Figure 8 shows operation manner control policy in the top layer. Unlike the traditional CD-CS strategy that uses a clear boundary to divide the entire trip into CD and CS stages, here CD manner and CS manner are fused together with each other during vehicle utilization. Besides battery SoC, the power request level at each time is a reference of operation manner control as well. The specific control rules in the top layer can be further described as follows:

> *Critical Condition 1: High Power Request.* When P_{req} exceeds a threshold P_M, no matter how high the battery SoC is, the vehicle operation is switched to CS manner, denoted as $CS^{(P)}$, representing the fact that this CS manner is triggered by power request. The engine output power, according to DP, electricity will continue being consumed in this manner due to the high P_{req}, but the rate of electricity consumption is reduced.
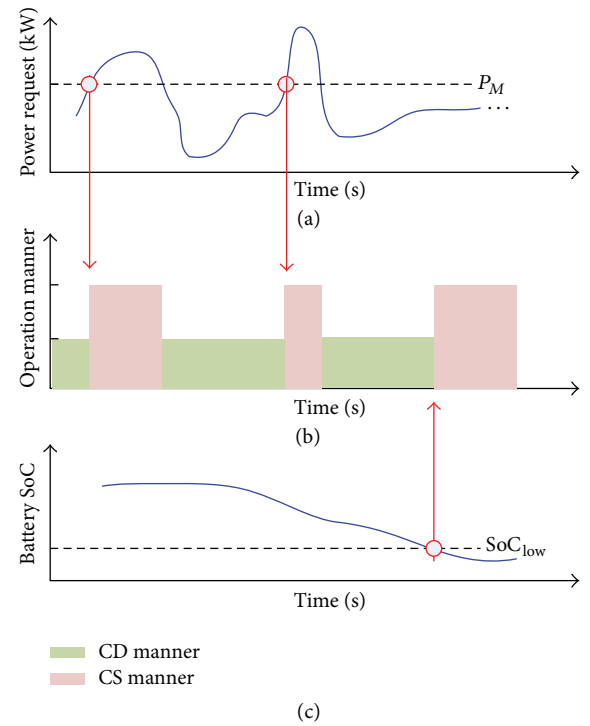
FIGURE 8: Schematic diagram of online operation manner control strategy. (a) Power request. (b) Operation manner: CD or CS. (c) Battery SoC.

> *Critical Condition 2: Low SoC.* CS manner is adopted when SoC reaches the threshold SoC_{Low}, denoted as $CS^{(B)}$, representing the fact that this CS manner is triggered by battery SoC. Different from $CS^{(P)}$, in this manner, electricity is not utilized to supply the power request any longer.
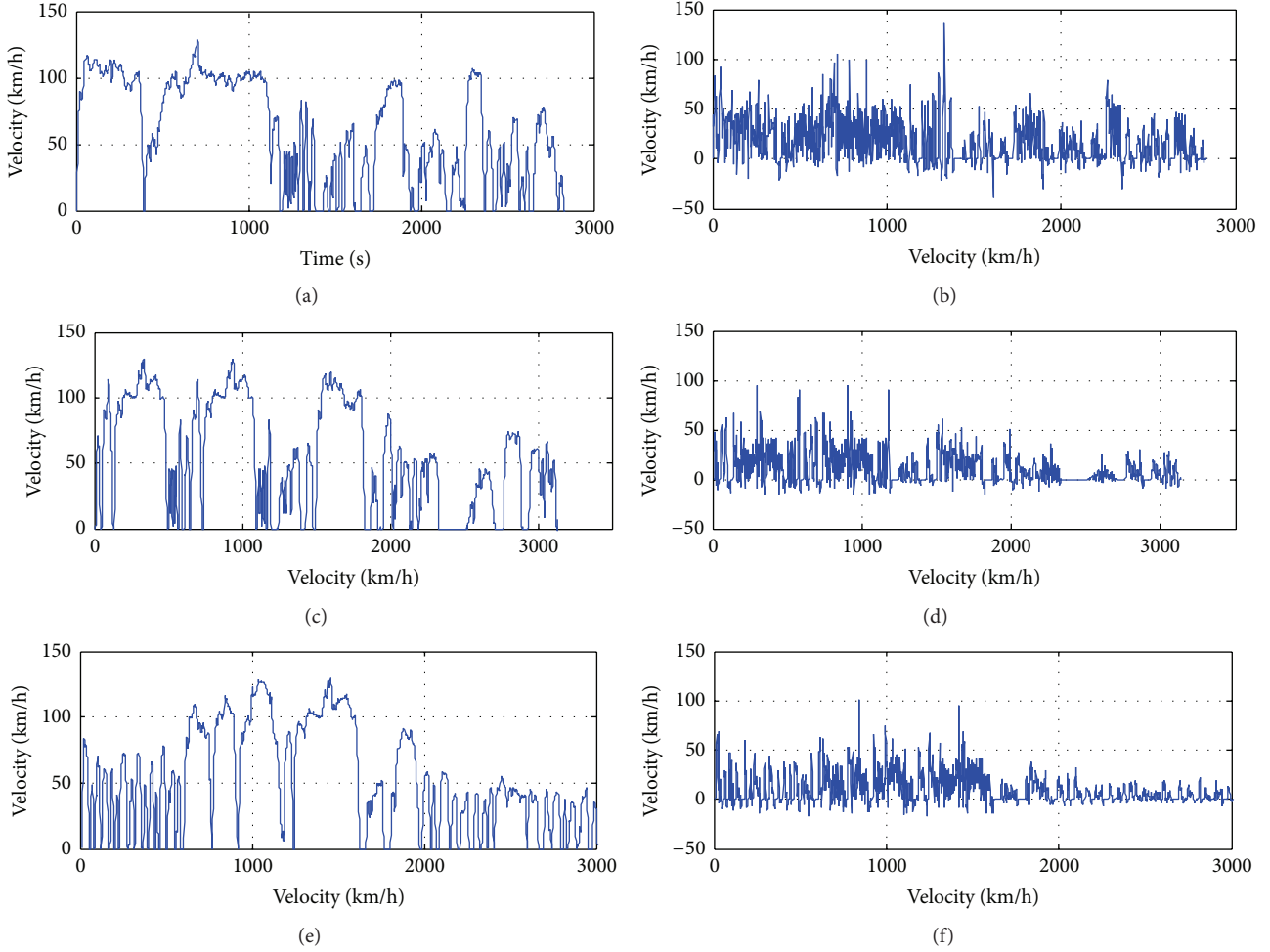
FIGURE 9: The adopted driving cycles in simulation study. (a), (b) Speed profile and power request, respectively, in cycle 1. (c), (d) Speed profile and power request, respectively, in cycle 2. (e), (f) Speed profile and power request, respectively, in cycle 3.

Critical Condition 3: Low Power Request and High SoC. The vehicle is operated at CD manner when P_{req} is below P_M and battery SoC is higher than SoC_{Low}.

In control algorithm of bottom layer, power allocation is based on the optimization result from DP. Two thresholds are used to divide the engine power into three parts according to power request. To achieve the minimum of cost function, DP turns engine off when the power request is low. This control policy reduces the energy consumption but increases the start-stop times of engine. Consequently, the power allocation is modified when power request is low. The power allocation control policy in bottom layer is described as

if $P_{req}(t) > P_M$ and $SoC \geq SoC_{Low}$, calculate:

$$P_E(t) = \begin{cases} \chi_2 & \text{if } P_{req}(t) > \chi_2 \\ P_{req}(t) & \text{if } \chi_1 < P_{req}(t) \leq \chi_2 \\ \chi_1 & \text{if } P_{req}(t) \leq \chi_1 \end{cases}$$

else if $SoC < SoC_{Low}$, calculate:

$$P_E(t)$$
$$= \begin{cases} \min\{P_{E\_max}, P_{req}(t)\} & \text{if } P_{req}(t) > \chi_1 \\ \chi_1 & \text{if } P_{req}(t) \leq \chi_1 \end{cases}$$

else

$$P_E(t) = 0,$$

(19)

where P_{e_max} is the maximum of engine power and χ_1 and χ_2 are two thresholds of engine power which are predetermined by DP.

5. Simulation Results and Discussion

The presented energy management approach is evaluated by a simulation study in this section. Three driving cycles are used as simulation conditions. The speed profiles and power requests of the adopted driving cycles are shown in Figure 9.

(a)

(b)

(c)

(d)

Figure 10: Simulation results of power allocation. (a), (b) Engine power and battery power, respectively, in presented strategy. (c), (d) Engine power and battery power, respectively, in CD-CS strategy.

To better analyze and assess the effectiveness of the presented online energy management approach, the traditional CD-CS strategy is adopted as a benchmark.

The simulation is repeated through the driving cycle until the driving distance exceeds 60 km to make sure that the CS manner could be reached. Results of power allocation of two strategies in the three driving cycles are given in Figures 10–12, respectively. The length of all-electric range is defined as the summation of driving distance covered by CD manner. The engine power indicates the distribution of CD manner in both strategies. In the presented energy management strategy, the CD manner and the $CS^{(P)}$ manner are interlaced while the boundary of CD manner and CS manner is quite clear in CD-CS strategy. This control performance will refrain the battery from the overquick electricity consumption rate when the battery SoC is high. At the last part of simulation, the engine power is enhanced and battery no longer outputs power because the electricity in battery is exhausted and the $CS^{(B)}$ manner is switched on. From the simulation result of battery power, it is clear that the electricity consumption rate in the presented strategy is lower than that in CD-CS

Table 3: The AER results in three cycles.

Cycles	The presented strategy	CD-CS strategy	Prolongation
1	45.76 km	44.61 km	2.58%
2	64.65 km	62.85 km	2.86%
3	65.16 km	64.03 km	1.76%

strategy at most situations and most of the transient large power discharge situations in CD-CS strategy are eliminated in the presented strategy. The lower battery power can result in a more healthy battery working condition and a probably longer AER. The lengths of AER with two strategies have been recorded in three simulation driving cycles, described in Table 3. Compared to traditional CD-CS strategy, the presented strategy can prolong the ARE by up to 2.86% at certain driving condition.

To assess the energy saving performance of the presented strategy with uncertain driving distance, the comparison of energy cost between presented strategy and traditional

Figure 11: Simulation results of power allocation. (a), (b) Engine power and battery power, respectively, in presented strategy. (c), (d) Engine power and battery power, respectively, in CD-CS strategy.

CD-CS strategy is calculated at 20 km, 40 km, and 60 km, respectively. The comparison results in three driving cycles are given in Figure 13. Apparently, the performance of energy saving in the presented strategy is relative to the driving distance. When the trip is quite short, the traditional CD-CS strategy has an obvious advantage. When driving distance is 40 km, energy cost with the presented strategy is lower than that with traditional CD-CS strategy by 5.77% in driving cycle 1. When the driving distance reaches 60 km, the reduction of energy consumption compared to CD-CS strategy is 4.28%, 4.67%, and 4.06%, respectively, in three driving cycles. This result indicates that the presented strategy can achieve a better energy economy when driving distance is quite long, but for a short distance trip, the traditional CD-CS strategy still owns its advantage in terms of energy saving performance.

6. Conclusions

A novel DP-based online energy management approach is proposed for the plug-in HEVs in this paper. The presented approach utilizes control policy extracted from optimization results of DP to reduce the energy consumption. Both the power request and the battery SoC are used to control the operation manners for the prolongation of AER. Three driving cycles are employed to evaluate the presented strategy. Simulation results indicate that the AER in the presented approach can be extended by up to 2.86% compared to CD-CS strategy. The comparison result of energy saving performance between the presented strategy and traditional CD-CS strategy is related to the driving distance. For some certain driving conditions, the reduction of energy consumption can be up to 5.77%.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
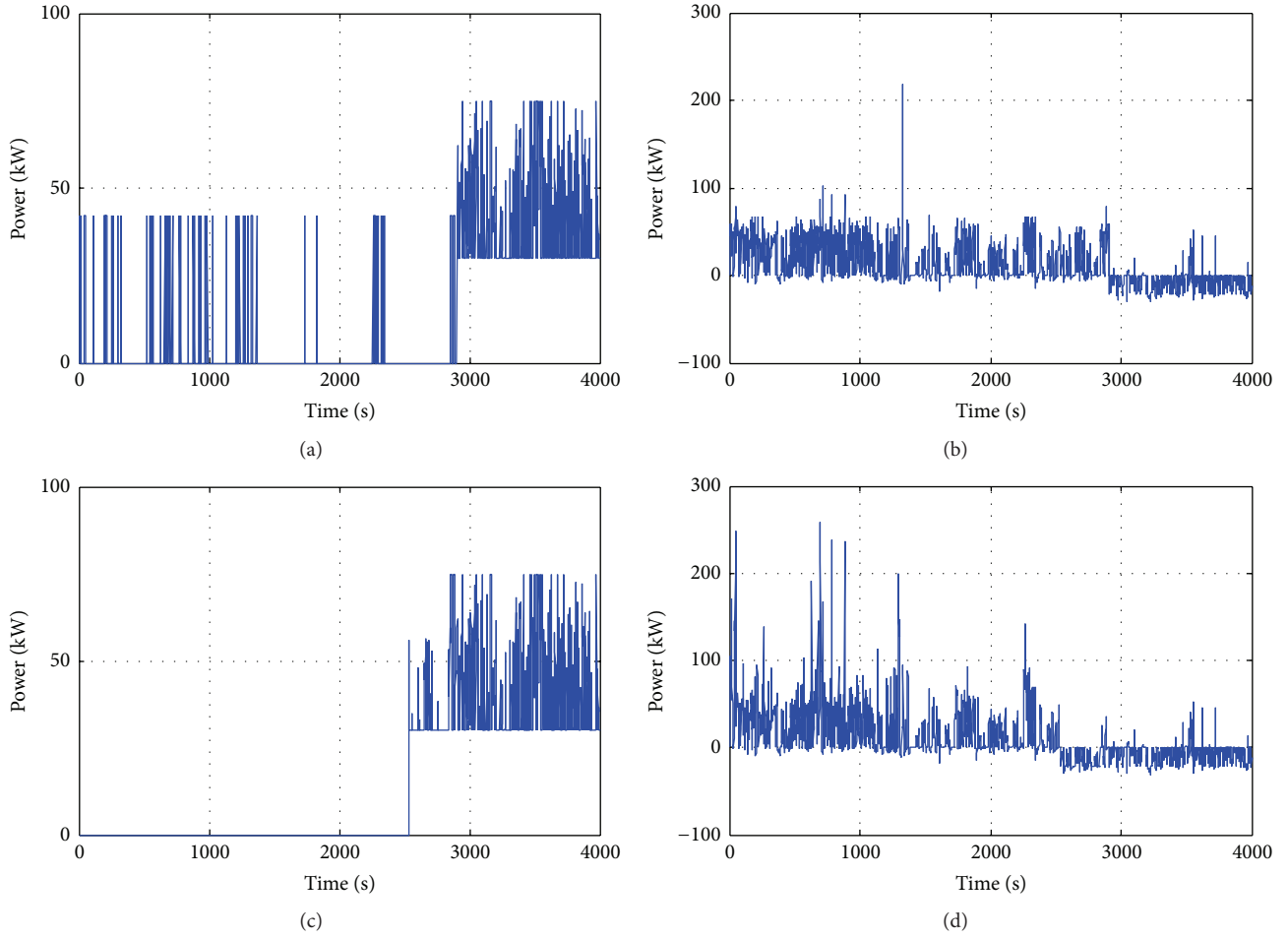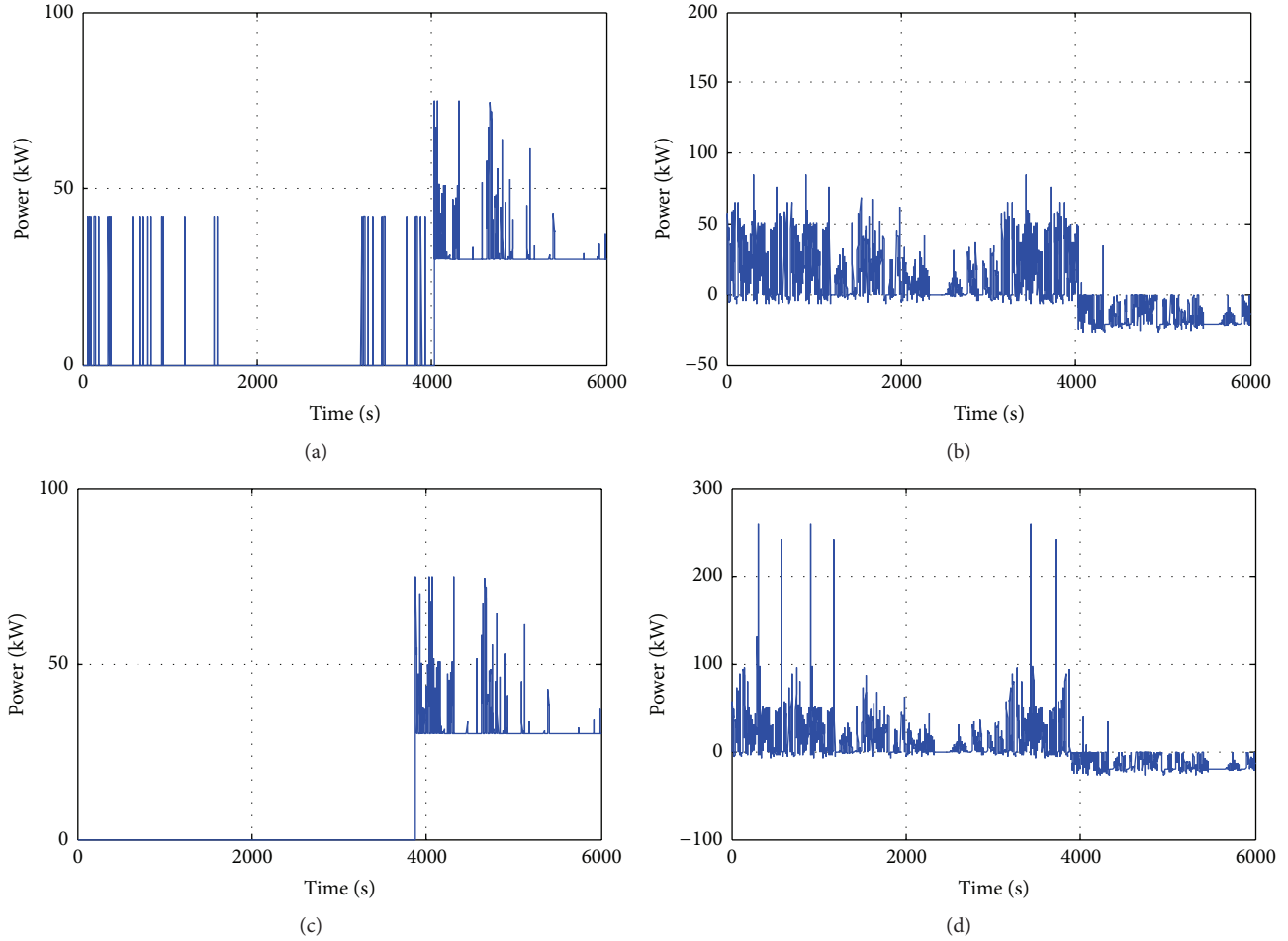
Acknowledgments
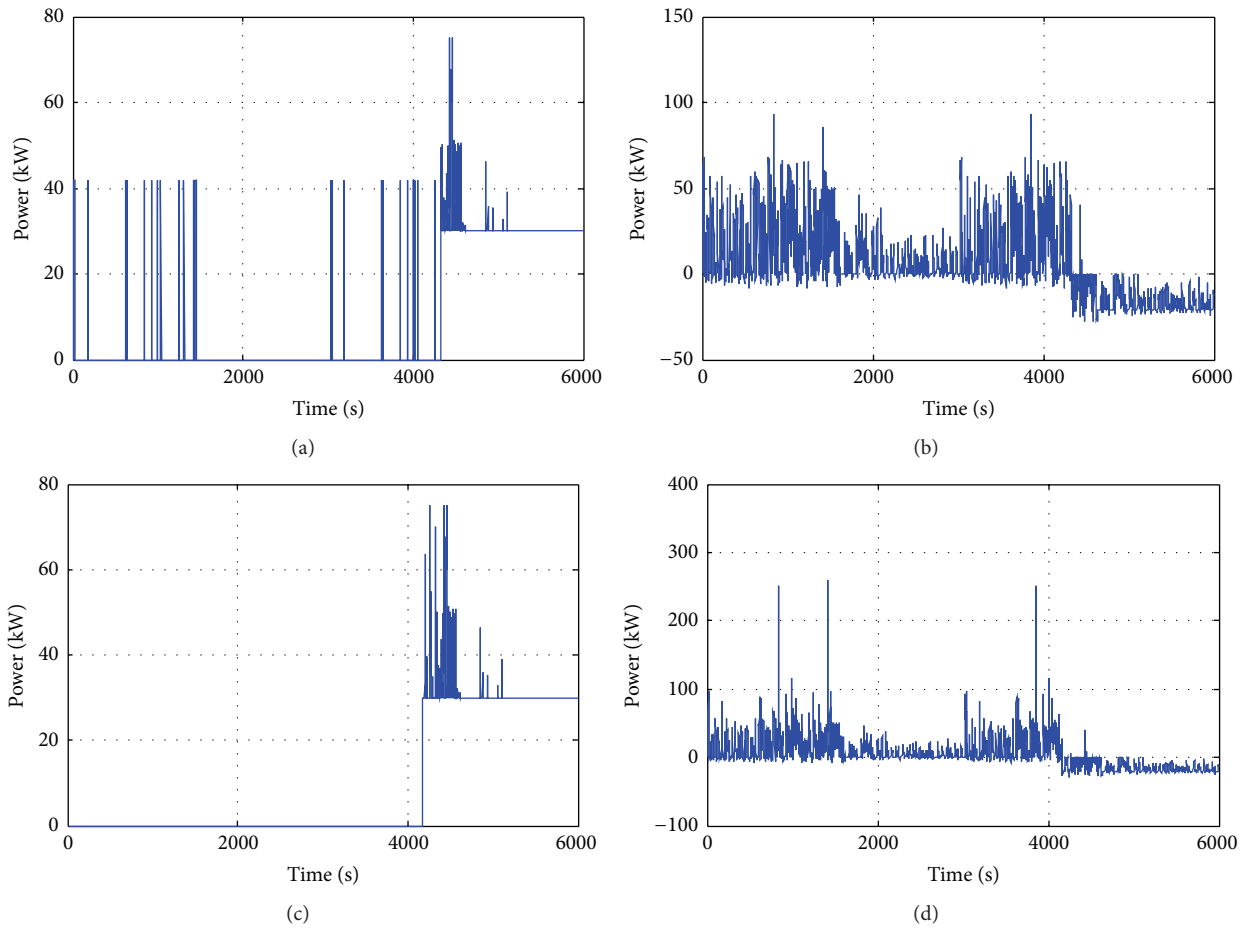
Figure 12: Simulation results of power allocation. (a), (b) Engine power and battery power, respectively, in presented strategy. (c), (d) Engine power and battery power, respectively, in CD-CS strategy.
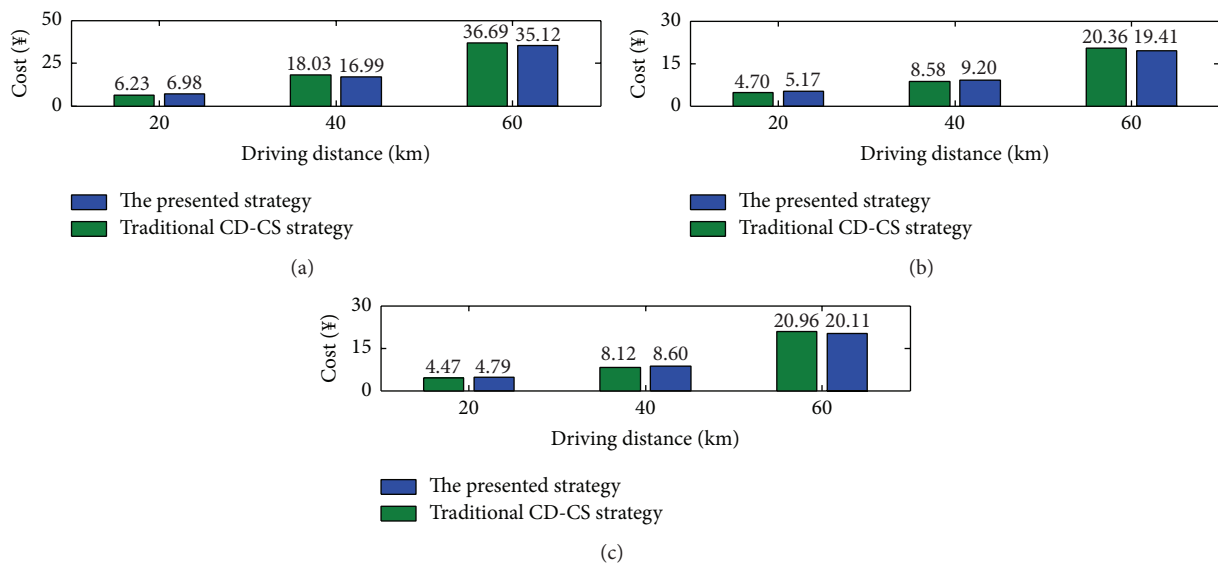
Figure 13: Comparison of energy consumption at different driving distances in three driving cycles. (a) Driving cycle 1; (b) driving cycle 2; (c) driving cycle 3.

References

[1] B. M. Al-Alawi and T. H. Bradley, "Review of hybrid, plug-in hybrid, and electric vehicle market modeling studies," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 190–203, 2013.

[2] N. Rotering and M. Ilic, "Optimal charge control of plug-in hybrid electric vehicles in deregulated electricity markets," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1021–1029, 2011.

[3] A. A. Malikopoulos, "Supervisory power management control algorithms for hybrid electric vehicles: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1869–1885, 2014.

[4] S. G. Wirasingha and A. Emadi, "Classification and review of control strategies for plug-in hybrid electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 111–122, 2011.

[5] H. W. He, R. Xiong, K. Zhao, and Z. Liu, "Energy management strategy research on a hybrid power system by hardware-in-loop experiments," *Applied Energy*, vol. 112, pp. 1311–1317, 2013.

[6] Y.-H. Hung and C.-H. Wu, "An integrated optimization approach for a hybrid energy system in electric vehicles," *Applied Energy*, vol. 98, pp. 479–490, 2012.

[7] X. Hu, L. Johannesson, N. Murgovski, and B. Egardt, "Longevity-conscious dimensioning and power management of the hybrid energy storage system in a fuel cell hybrid electric bus," *Applied Energy*, vol. 137, pp. 913–924, 2015.

[8] Z. Chen, R. Xiong, K. Wang, and B. Jiao, "Optimal energy management strategy of a plug-in hybrid electric vehicle based on a particle swarm optimization algorithm," *Energies*, vol. 8, no. 5, pp. 3661–3678, 2015.

[9] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Optimal dimensioning and power management of a fuel cell/battery hybrid bus via convex programming," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 1, pp. 457–468, 2014.

[10] D. F. Opila, X. Wang, R. McGee, R. B. Gillespie, J. A. Cook, and J. W. Grizzle, "An energy management controller to optimally trade off fuel economy and drivability for hybrid vehicles," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 6, pp. 1490–1505, 2012.

[11] Y. del Valle, G. K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, and R. G. Harley, "Particle swarm optimization: basic concepts, variants and applications in power systems," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 2, pp. 171–195, 2008.

[12] S. Zhang and R. Xiong, "Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming," *Applied Energy*, vol. 155, pp. 68–78, 2015.

[13] L. V. Pérez and E. A. Pilotta, "Optimal power split in a hybrid electric vehicle using direct transcription of an optimal control problem," *Mathematics and Computers in Simulation*, vol. 79, no. 6, pp. 1959–1970, 2009.

[14] D. Kum, H. Peng, and N. K. Bucknor, "Supervisory control of parallel hybrid electric vehicles for fuel and emission reduction," *Journal of Dynamic Systems, Measurement and Control*, vol. 133, no. 6, Article ID 061010, 10 pages, 2011.

[15] L. Wang, Y. Zhang, C. Yin, H. Zhang, and C. Wang, "Hardware-in-the-loop simulation for the design and verification of the control system of a series-parallel hybrid electric city-bus," *Simulation Modelling Practice and Theory*, vol. 25, pp. 148–162, 2012.

[16] S. J. Moura, H. K. Fathy, D. S. Callaway, and J. L. Stein, "A stochastic optimal control approach for power management in plug-in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 3, pp. 545–555, 2011.

[17] E. D. Tate, J. W. Grizzle, and H. Peng, "SP-SDP for fuel consumption and tailpipe emissions minimization in an EVT hybrid," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 673–687, 2010.

[18] L. Johannesson, M. Åsbogård, and B. Egardt, "Assessing the potential of predictive control for hybrid vehicle powertrains using stochastic dynamic programming," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 71–83, 2007.

[19] W. Li, T.-K. Lee, Z. S. Filipi, X. Meng, and C.-N. Zhang, "Development of electric machine duty cycles for parallel hybrid electric Beijing city bus based on Markov chain," *International Journal of Vehicle Design*, vol. 58, no. 2–4, pp. 348–366, 2012.

[20] L. Serrao, S. Onori, and G. Rizzoni, "A comparative analysis of energy management strategies for hybrid electric vehicles," *Journal of Dynamic Systems, Measurement and Control*, vol. 133, no. 3, Article ID 031012, 9 pages, 2011.

[21] X. Hu, N. Murgovski, L. Johannesson, and B. Egardt, "Energy efficiency analysis of a series plug-in hybrid electric bus with different energy management strategies and battery sizes," *Applied Energy*, vol. 111, pp. 1001–1009, 2013.

Research Article

Autonomous Coil Alignment System Using Fuzzy Steering Control for Electric Vehicles with Dynamic Wireless Charging

Karam Hwang,[1] Jaehyoung Park,[1] Dongwook Kim,[1] Hyun Ho Park,[2] Jong Hwa Kwon,[3] Sang Il Kwak,[3] and Seungyoung Ahn[1]

[1] *Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea*
[2] *The University of Suwon, Hwasung 445-743, Republic of Korea*
[3] *Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, Republic of Korea*

Correspondence should be addressed to Seungyoung Ahn; sahn@kaist.ac.kr

An autonomous coil alignment system (ACAS) using fuzzy steering control is proposed for vehicles with dynamic wireless charging. The misalignment between the power receiver coil and power transmitter coil is determined based on the voltage difference between two coils installed on the front-left/front-right of the power receiver coil and is corrected through autonomous steering using fuzzy control. The fuzzy control is chosen over other control methods for implementation in ACAS due to the nonlinear characteristic between voltage difference and lateral misalignment distance, as well as the imprecise and constantly varying voltage readings from sensors. The operational validity and feasibility of the ACAS are verified through simulation, where the vehicle equipped with ACAS is able to align with the power transmitter in the road majority of the time during operation, which also implies achieving better wireless power delivery.

1. Introduction

It is notable that commercialization of electric vehicles (EVs) is becoming more widespread around the world in order to reduce the serious issues related to global warming and energy depletion that are being faced today. However, the EV poses some major flaws which are its high cost and long charging times, which inevitably directs consumers to still resort towards conventional internal combustion engine (ICE) vehicles. Those major flaws are mainly caused due to the drawbacks of current battery technology. To minimize the drawbacks caused from batteries, EVs with dynamic wireless charging systems have been developed. Dynamic wireless charging system allows the vehicle to charge in real time while in motion, which also allows the reduction of the overall battery capacity in the vehicle. This provides the benefit of reducing overall vehicle cost and reduced charging times. Dynamic wireless charging systems are a part

of wireless power transfer (WPT) technology, where power can be transferred from one circuit to another circuit without any physical contact or wiring. WPT technologies are being studied widely around the world as an anticipation to reduce or eliminate any physical wiring elements that restrict power supply to electric or electronic devices [1]. WPT technology has been introduced by N. Tesla in 1914, but it is not until recently that WPT technology has been applied into commercial electronic devices, biomedical products, industrial applications [1–4], and now stationary and dynamic wireless charging vehicles.

A WPT system consists of the power transmitter portion and power receiver portion. The power transmitter portion is composed of a power source and coil, where the power source is generated into an electromagnetic field as it enters through the coil. The power receiver portion, which consists of another coil, will convert the received electromagnetic field into usable energy that can power another source [5–7].

For every WPT system, there is a specific alignment range between the power transmitter and power receiver in which maximum power can be transferred, and this alignment range will vary by system due to its intrinsic parameters such as resistance, inductance, and capacitance [2]. When the range goes past the specific alignment, the power transfer capacity will drop or become near zero in certain cases [7].

In the case of dynamic wireless charging systems for road vehicles, keeping the alignment range becomes more difficult since the vehicle's lateral displacement will continuously change during motion. To achieve maximum power, the driver will have to pay careful attention in keeping the power receiver on the vehicle aligned with the power transmitter, which is installed beneath the center of each road lane. Even when it is assumed that the driver kept the vehicle aligned with the center of the road lane, it does not guarantee that the vehicle's receiver is in perfect alignment with the power transmitter, since the power transmitter is not visible (as it is installed beneath the road). In addition, keeping the vehicle aligned at the center of the road will require a lot of concentration, which can distract the driver to oncoming dangers and can lead to potential car accidents.

To resolve such misalignment issues in dynamic wireless charging systems, many researches have been conducted to increase their alignment range through efficient power receiving modules [6, 8–10]. However, in certain cases, these methods may become difficult or very costly to implement [2]. Therefore in this paper, an autonomous coil alignment system (ACAS) is proposed, which can be implemented for generally all vehicles equipped with dynamic wireless charging systems. The proposed system will detect the misalignment between the power receiver and power transmitter through sensors. When misalignment is detected, the vehicle will adjust itself through autonomous steering control in order to correct the alignment between the vehicle's power receiver and the power transmitter. Through this system, the vehicle will be able to achieve maximum power delivery with minimum driver effort. In addition, since the ACAS operates in a similar manner with the lane keeping assist system (LKAS), the ACAS will be able to keep the vehicle along the center of the road lane, thus providing more safety for the driver as well.

The paper will first describe the principles of WPT and the background research on dynamic wireless charging of vehicles. Then the motivation of the proposed ACAS and its technical operation concept is described, followed by simulation settings and tests to verify its feasibility.

2. Principles of the Wireless Power Transfer (WPT) System

The fundamental circuit of the wireless power transfer system is described as shown in Figure 1. The system is comprised of two RLC circuits: the circuit on the left-side is the source coil circuit (power transmitter) and the circuit on the right side is the load coil circuit (power receiver).
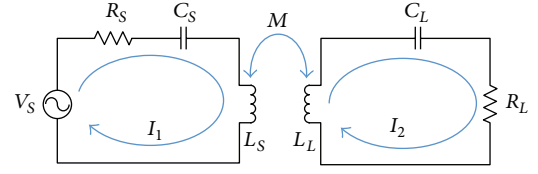
FIGURE 1: Circuit schematic of the source circuit and load circuit in wireless power transfer (WPT).

The equation for source coil and the load coil based on Figure 1 is expressed as

$$\left(R_S + \frac{1}{j\omega C_S} + j\omega L_S\right) I_1 - j\omega M I_2 - V_S = 0,$$

$$\left(R_L + \frac{1}{j\omega C_L} + j\omega L_L\right) I_2 - j\omega M I_1 = 0. \tag{1}$$

In (1), each of R_x, C_x, and L_x represents the resistor, capacitor, and inductor component, while S or L subscript of a certain component represents the source coil or load coil, respectively. ω represents the frequency, and I_1, I_2 represent the current flowing in the source coil and the load coil, respectively, which are expressed as follows:

$$I_1 = \frac{R_L + 1/j\omega C_L + j\omega L_L}{j\omega M} I_2,$$

$$I_2 = \frac{j\omega M}{\left(R_S + 1/j\omega C_S + j\omega L_S\right)\left(R_L + 1/j\omega C_L + j\omega L_L\right) + \omega^2 M^2} \\ \cdot V_S. \tag{2}$$

M shown in (1) to (2) is the mutual inductance that occurs between the source coil and load coil while operating in resonant frequency. This is expressed as

$$M = k\sqrt{L_S L_L}. \tag{3}$$

In (3), k is a value ranging between 0 and 1 and represents the coupling coefficient between the source coil and the load coil. When the value is near 1, it implies that the coupling between the source coil and load coil is very strong and will result in higher mutual inductance. When the value is near 0, it implies that the coupling between the source coil and load coil is very weak and will result in a lower to near nonexistent mutual inductance. In WPT, a larger M value usually facilitates more effective power transfer [11]. Based on (2), the power from the source coil can be determined as follows:

$$P_S = I_1 V_S$$

$$= \frac{R_L + 1/j\omega C_L + j\omega L_L}{\left(R_S + 1/j\omega C_S + j\omega L_S\right)\left(R_L + 1/j\omega C_L + j\omega L_L\right) + \omega^2 M^2} \\ \cdot V_S^2. \tag{4}$$
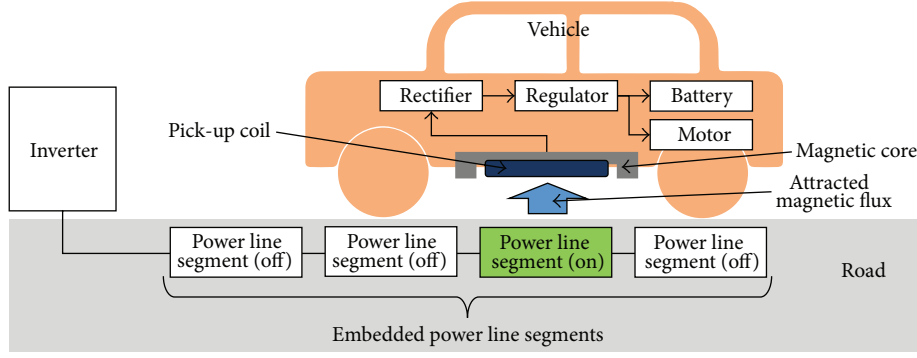
Figure 2: Block diagram of the OLEV system.

Similarly, using (2), the power from the load coil can be determined as

$$P_L = I_2^2 R_L$$

$$= \left| \frac{-\omega^2 M^2}{\{(R_S + 1/j\omega C_S + j\omega L_S)(R_L + 1/j\omega C_L + j\omega L_L) + \omega^2 M^2\}^2} \right| \quad (5)$$

$$\cdot V_S^2 R_L.$$

The equations shown from (1) to (5) are the general principle guidelines used to design a WPT system including vehicles with dynamic wireless charging systems, and the final generated power shown in (5) will power the components in the power receiving portion.

3. Concept on Electric Vehicles (EVs) with Dynamic Wireless Charging

The development on EVs using dynamic wireless charging has been progressing on for the past few decades, and some of the recent advancements (and commercialization) have been made such as the online electric vehicle (OLEV) [11–15]. It has been widely recognized and commercialization of these vehicles is currently progressing in Korea [13]. Figure 2 shows the basic principle of the EV's dynamic wireless charging system (OLEV).

The power transmitter part of the system consists of the inverter and power line, where the inverter provides power to the power lines that are installed beneath the road. The power line is installed as power line segments, where the inverter will only turn on the segment where the vehicle is located in order to mitigate the inefficiency of the power supply [14]. The inverter receives 60 Hz power from the main electrical power plant and converts into a specific constant current at specific resonant frequency. In OLEV systems, the constant current ranges between 200 A to 260 A, and the resonant frequency is set at 20 kHz [11].

The power receiving part of the system is attached on the bottom of the vehicle chassis, which consists of the pick-up coil, magnetic core, rectifier, and regulator. The magnetic core of the receiving system captures the magnetic flux from the power lines and will induce voltage along the power lines. The induced voltage is converted into DC voltage through
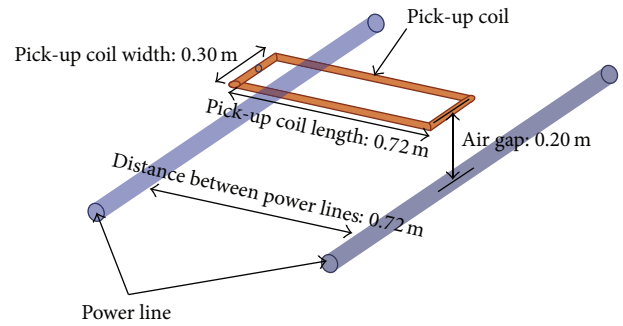
Figure 3: The model and parameters of the pick-up coil and power line used in the Maxwell simulation.

the rectifier and then converted into the desired voltage range through the regulator. A small portion of the received power is used to drive the motor while the rest is used to charge the battery when the vehicle is in motion. When the vehicle is not in motion, all the received power is used to charge the battery. The power receiving part is installed as modules, where each power receiving module is capable of generating 20 kW of power. In case of the OLEV bus, five modules were installed in order to achieve a total of 100 kW target power [11].

4. Motivation for the Coil Alignment System (ACAS)

As mentioned earlier about EVs with dynamic wireless charging, lateral misalignment between the power transmitter and the power receiver will inevitably occur, which will result in reduced power transfer and efficiency.

As described earlier in the principles of WPT, as lateral misalignment increases, it will reduce the mutual inductance, M, due to reduced coupling coefficient, k, which will reduce the output power that can be received from the power receiver. To assess how much power is reduced due to lateral misalignment, a simulation was conducted using ANSYS Maxwell, where a model of the OLEV power line segment and pick-up coil was modeled as shown in Figure 3, using similar dimension parameters used in [11]. The electrical parameters for the power line segment and pick-up coil are listed in Table 1 as well.

TABLE 1: Electrical parameters for power line segment and pick-up coil in Maxwell.

Parameter	Value
Operating frequency	20 kHz
Current fed through power line	200 A
Power line	
Number of turns	8
Inductance	842 uH
Pick-up coil	
Number of turns	50
Inductance	2.71 mH

FIGURE 4: Simulation results showing reduction of power transfer to pick-up coil and its efficiency with increased lateral misalignment.

FIGURE 5: Overall block diagram of the coil alignment system (ACAS).

The induced voltage of the model shown in Figure 3 was observed in ANSYS Maxwell while the misalignment between the power line and pick-up coil was increased from 0 m (meaning that it is at perfect alignment) to 0.6 m. The misalignment was conducted only up to 0.6 m because exceeding this value will imply that the vehicle has crossed to the other lane under the assumption that the vehicle width is 1.8 m and the width of the road lane is 3.0 m. After conducting the simulation, the resultant mutual inductance, inductance of power line and pick-up coil, and the induced voltage on the pick-up coil were implemented into Agilent Advanced Design System (ADS) program, which is an electronic design automation (EDA) simulation tool that analyzes wireless circuit systems. A similar circuit shown in Figure 1 was designed in ADS, and the power generated from the power line and pick-up coil was determined. The power determined from the power line and pick-up based on (4) and (5) was used to determine the power transfer efficiency of the WPT system. It is an important factor in rating the WPT system's performance, and it is determined as follows:

$$\eta = \frac{P_L}{P_S}. \tag{6}$$

The final results were analyzed, and the generated output power from the pick-up coil as well as its efficiency at different lateral alignments is shown in Figure 4.

From the simulation results, the pick-up coil was able to receive 47.83 kW at 70.31% efficiency. But as lateral misalignment increases, the received power and efficiency got reduced significantly, where the output power was at 13.92 kW at 40.57% efficiency even with 20 cm misalignment. At 60 cm, the received power was at 0.85 kW at 4.07% efficiency. Although received power and efficiency can be improved with better pick-up coil design, the lower receiving power and efficiency are inevitable as the pick-up coil moves away from the power line. Therefore, as described in the introduction section, the driver should keep the vehicle aligned with the power lines in order to maximize the power transfer efficiency. However, keeping the vehicle aligned with the power line at all times is near impossible for the driver to conduct, especially since the power line cannot be seen. Therefore, the ACAS system is proposed, which detects the misalignment between the power line and pick-up coil and autonomously steers the vehicle in order to maximize power
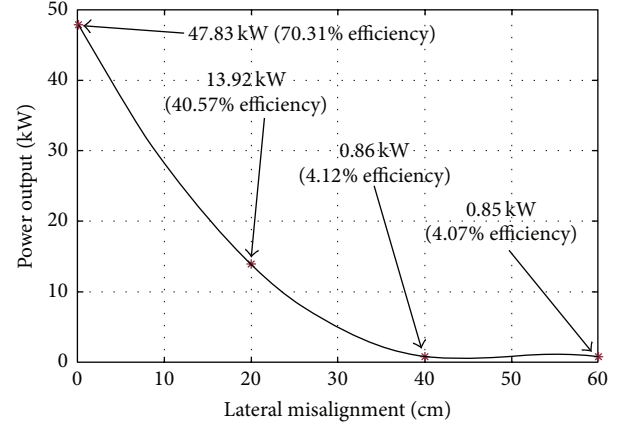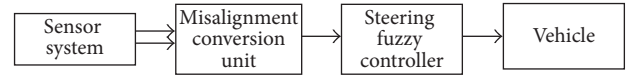
transfer and efficiency and also increase the overall safety and comfort for the driver.

5. Concept of the ACAS

For the ACAS system, the hardware requirements are the electric power steering (EPS) system and sensors. The EPS is typically equipped in modern commercialized vehicles. The sensors are inexpensive and will detect the misalignment between the pick-up coil and power line. In general, the implementation of the ACAS system is inexpensive as it requires minimum hardware modifications, and it is mainly software implementation. The overall block diagram of the ACAS is shown in Figure 5.

There are three subsystems in the ACAS, which consist of the sensor system, misalignment conversion unit, and the steering fuzzy controller. There are two outputs from the sensor system: the output value of the difference between two sensors that are installed on the left-side and right-side of the pick-up coil and the output value of the sensor installed at the pick-up coil regulator's output. The two output values are received by the conversion unit, which will determine the lateral alignment value between the pick-up coil and power line. The alignment value is received by the steering fuzzy controller, where the necessary steering command will be sent to the EPS system of the vehicle.

5.1. Sensor System for the ACAS. The sensor system plays a crucial role in the ACAS system, and its block diagram is shown in Figure 6. There are three main sensor units: left-side voltage sensor unit, right-side voltage sensor unit, and the pick-up voltage sensor unit. The voltage difference between the left-side and right-side voltage sensor unit outputs is
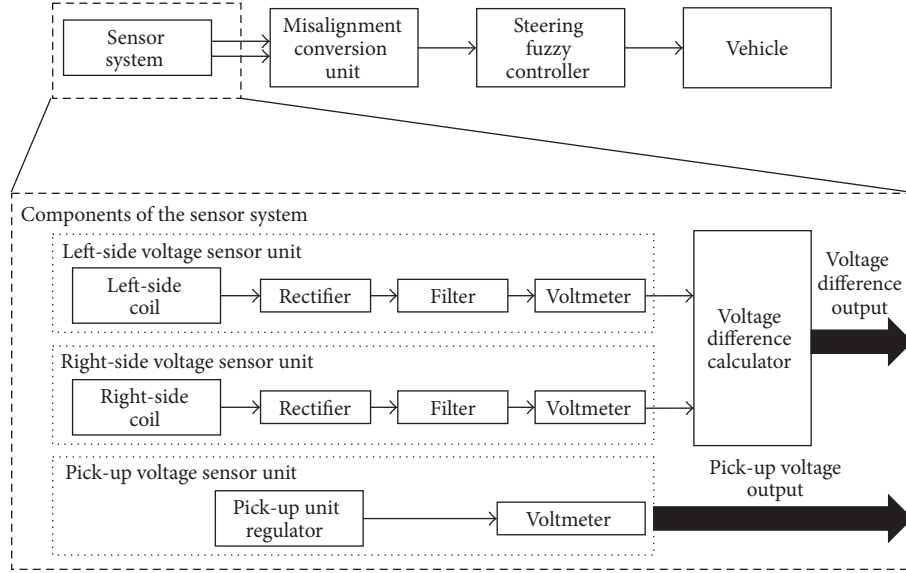
Figure 6: Block diagram of the ACAS sensor system.
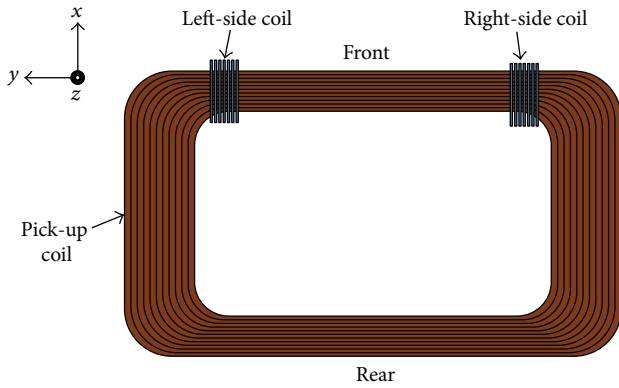
Figure 7: Diagram of left-side coil and right-side coil placement on pick-up coil shown in top view.

calculated by the voltage difference calculator. The calculated output and the output from the pick-up voltage sensor unit are sent to the misalignment conversion unit.

The main component of the pick-up voltage sensor unit is the DC voltmeter, which will measure the voltage output from the pick-up unit regulator. The components in the left-side voltage sensor unit and the right-side voltage sensor unit are the same; it consists of the coil, rectifier, regulator, and voltmeter. The coils described as left-side coil and right-side coil are smaller coils that are wrapped around the pick-up coil, as shown in Figure 7. The two coils are wrapped around the front side of the pick-up coil while keeping enough separation gap between them.

The two coils can also be installed at the rear side of the pick-up coil, but they must be installed along the front or the rear of the pick-up coil because of the magnetic flow direction between the power-line and pick-up coil apparent in the simulation results shown in Figure 8.

Figures 8(a) and 8(b) show the magnetic flow direction in front view and side view, respectively. While significant magnetic flow is visible from the front view, the magnetic flow from the side view is minimal. Even if the required voltage from the two coils is small, the induced voltage from the two coils if installed on the left/right side of the pick-up coil will be near zero volts, which is not desired.

The number of turns and the coil length of the left-side coil and the right-side coil should be kept at minimum, just enough to induce a voltage that can be read by the voltmeter. With bigger turns and bigger length of the two coils, it can disrupt the magnetic flow between the power line and pick-up coil, thus reducing the power transfer efficiency. In this paper, the two coils have been designed to have lengths of 7 cm with 25 turns, which were just enough to induce voltage within 2.5 V while not affecting the power transfer between the power line and pick-up coil. While using the model shown in Figure 3 as a basis, the two coils with the mentioned parameters were implemented and simulation was conducted. The simulation results are shown in Figure 9.

Based on the results shown in Figure 9(a) through Figure 9(d), it shows how the voltage of the left-side coil and right-side coil varies as alignment is increased from 0 cm to 60 cm. However, the output voltage is in AC, where the difference between the two coils cannot be easily identified. Therefore, the two output waveforms are rectified and regulated into DC voltage, which can be read by the DC voltmeter. The resultant DC voltage output is then sent into the voltage difference calculator, where it is calculated as follows:

$$V_d = V_l - V_r. \tag{7}$$

V_l, V_r are the DC voltage readings from the left-side coil voltage sensor unit and right-side coil sensor unit, respectively. The resultant voltage difference between the two coils relative to its alignment is shown in Figure 10.

B (Tesla)

3.7298e − 003
3.4992e − 003
3.2685e − 003
3.0378e − 003
2.8072e − 003
2.5765e − 003
2.3458e − 003
2.1152e − 003
1.8845e − 003
1.6539e − 003
1.4232e − 003
1.1925e − 003
9.6188e − 004
7.3121e − 004
5.0055e − 004
2.6989e − 004
3.9229e − 005

B (Tesla)

3.8137e − 003
3.5776e − 003
3.3415e − 003
3.1054e − 003
2.8693e − 003
2.6332e − 003
2.3971e − 003
2.1610e − 003
1.9249e − 003
1.6887e − 003
1.4526e − 003
1.2165e − 003
9.8043e − 004
7.4433e − 004
5.0822e − 004
2.7212e − 004
3.6012e − 005

(a)   (b)

FIGURE 8: Maxwell simulation results showing the magnetic flow between the power line and pick-up coil from (a) front view and (b) side view.

However, based on the results shown in Figure 10, the nonlinear relationship between lateral misalignment and voltage difference can be observed, where the voltage difference reading can imply left or right misalignment simultaneously. This makes it difficult to determine the exact lateral misalignment location. Therefore, the voltage reading of the pick-up coil is needed in order to determine the exact lateral misalignment location, which is implemented as input for the misalignment conversion unit.

5.2. *Misalignment Conversion Unit for the ACAS.* The misalignment conversion unit shown in Figure 11 will change the nonlinear characteristic between the left-coil/right-coil voltage difference and the lateral misalignment into a more linear characteristic. It consists of the region selector unit, which divides the voltage difference readings into several regions. Each region consists of a model that has the linear

relationship between the voltage difference of the left-coil and right-coil voltage sensor unit, which can be mathematically expressed as follows:

$$m_n = \frac{y_{\max(n)} - y_{\min(n)}}{V_{d\_\max(n)} - V_{d\_\min(n)}}. \tag{8}$$

m_n refers to the slope of the specific linear region which is identified as A, B, or B' and can go up to n regions depending on how many linear regions can be sectioned from the nonlinear voltage difference/lateral misalignment relationship. $y_{\max(n)}$, $y_{\min(n)}$, $V_{d\_\max(n)}$, and $V_{d\_\min(n)}$ refer to the maximum and minimum lateral alignment distance and voltage difference of a specific n region, respectively. The corresponding specific region model is selected by the region selector switch when it meets the specific criteria as follows:

$$m_n = \begin{cases} A, & \text{if } V_L(t) > V_{T1} \\[2mm] B, & \text{if } \left(V_L(t) < V_{T1}\right), \left(V_d(t) > V_d(t-1)\right), \left(V_L(t) > V_L(t-1)\right) \\[2mm] B, & \text{if } \left(V_L(t) < V_{T1}\right), \left(V_d(t) < V_d(t-1)\right), \left(V_L(t) < V_L(t-1)\right) \\[2mm] B', & \text{if } \left(V_L(t) < V_{T1}\right), \left(V_d(t) > V_d(t-1)\right), \left(V_L(t) > V_L(t-1)\right) \\[2mm] B', & \text{if } \left(V_L(t) < V_{T1}\right), \left(V_d(t) < V_d(t-1)\right), \left(V_L(t) < V_L(t-1)\right) \\[2mm] \vdots \\[2mm] n, & \text{if } \left(V_{T1} < V_L(t) < V_{Tn}\right), \left(V_d(t) > V_d(t-1)\right), \left(V_L(t) > V_L(t-1)\right) \\[2mm] n, & \text{if } \left(V_{T1} < V_L(t) < V_{Tn}\right), \left(V_d(t) < V_d(t-1)\right), \left(V_L(t) < V_L(t-1)\right) \\[2mm] n', & \text{if } \left(V_{T1} < V_L(t) < V_{Tn}\right), \left(V_d(t) > V_d(t-1)\right), \left(V_L(t) > V_L(t-1)\right) \\[2mm] n', & \text{if } \left(V_{T1} < V_L(t) < V_{Tn}\right), \left(V_d(t) < V_d(t-1)\right), \left(V_L(t) < V_L(t-1)\right). \end{cases} \tag{9}$$

(a)

(b)

(c)

(d)
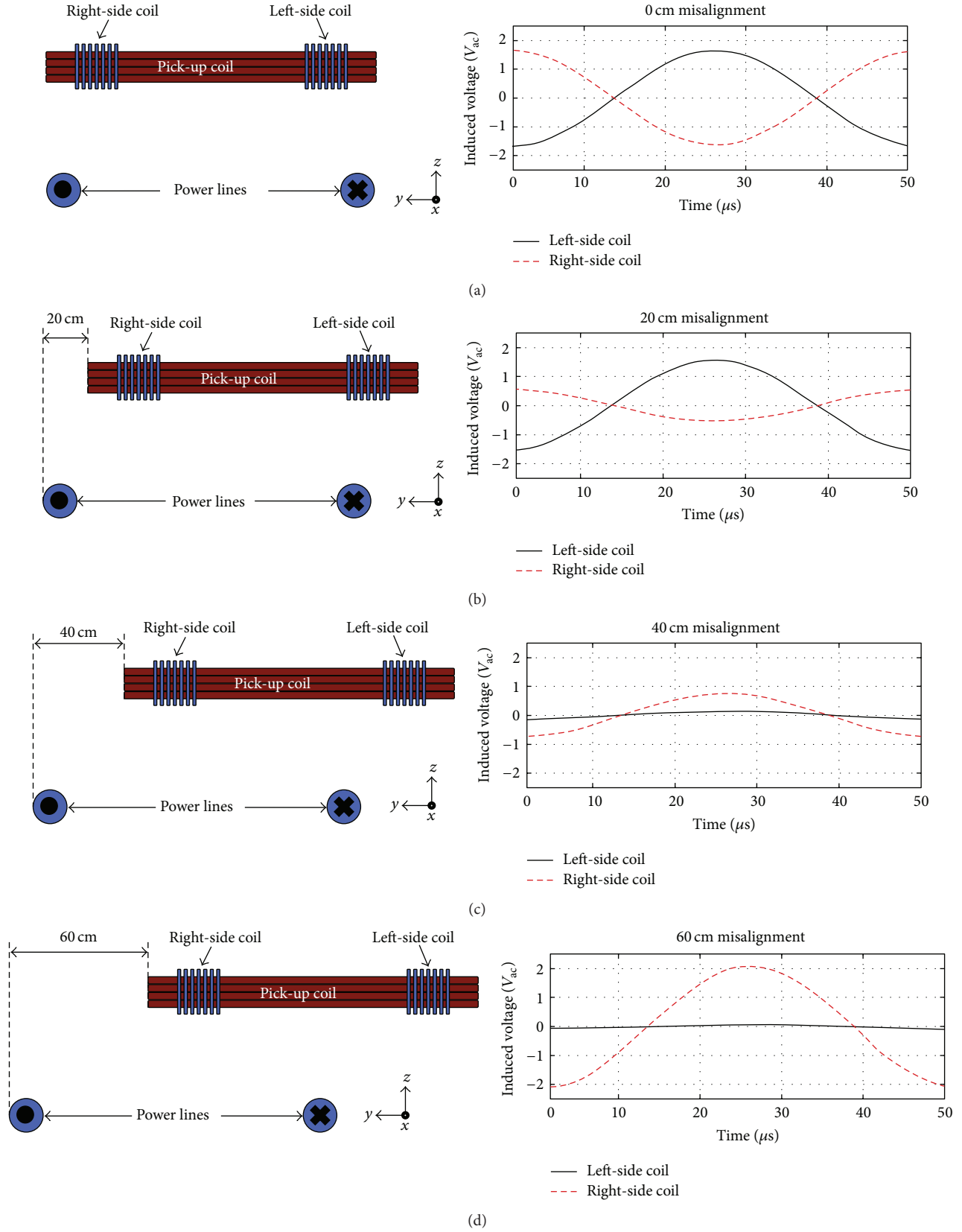
FIGURE 9: Diagram with alignment parameters (drawing not in scale) between pick-up coil and power line and simulation results showing the induced voltage from the left-side coil and right-side coil at (a) 0 cm, (b) 20 cm, (c) 40 cm, and (d) 60 cm misalignment position.
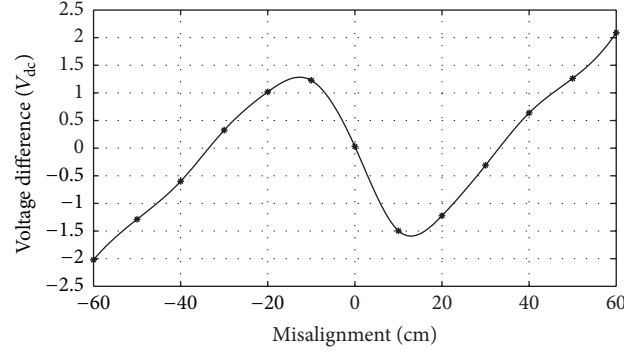
FIGURE 10: Simulation results showing the relationship between voltage difference of left-coil and right-coil and misalignment position.
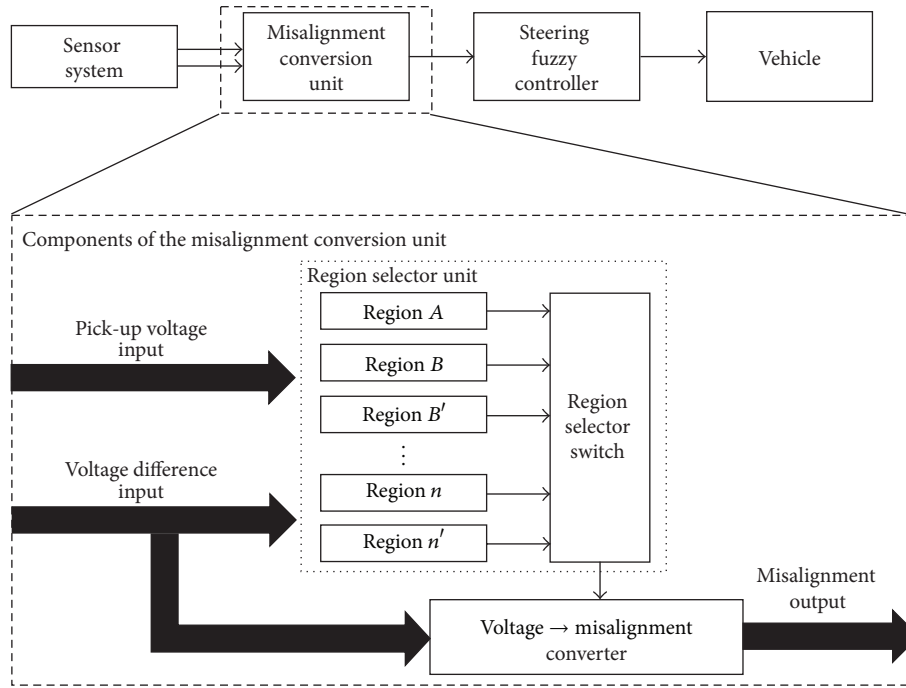
FIGURE 11: Block diagram of the ACAS misalignment conversion unit.

m_n of the corresponding region will be selected depending on the voltage threshold V_{T1}, voltage difference V_d, and the measured voltage from the pick-up voltage sensor unit V_L. In this paper, the number of regions is limited to region A, region B, and region B' as shown in Figure 12. The upper graph in the figure is the same graph shown in Figure 10, and it can be seen that each region consists of a linear relationship characteristic between the voltage difference and the lateral misalignment.

The misalignment range for region A is roughly between ±17 cm, while regions B and B' range between −60 cm and −18 cm and between 60 cm to 18 cm, respectively. If assuming that region A is selected by the region selector unit, it means that V_L is greater than the voltage threshold V_{T1}. If V_L is lower than V_{T1}, the region selector unit will either select region B

or region B'. As shown in the lower graph of Figure 12, V_{T1} is approximately 175 volts, which will toggle from region A to region B or region B's at ±17 cm.

The region selector will determine whether to select region B or region B' based on the previous and current output values of V_L and V_d. As shown in (9), there are two identical conditions that can meet the criteria for each region. If assuming that region B is selected by the region selector unit, it could either mean that the previous value of V_L and V_d is greater or less than its current V_L and V_d values. This implies that the pick-up coil is moving either away or towards the power line in region B, and the condition that is moving towards the power line is always selected. The final selected m_n value from the region selector unit will be converted into estimated lateral displacement location y
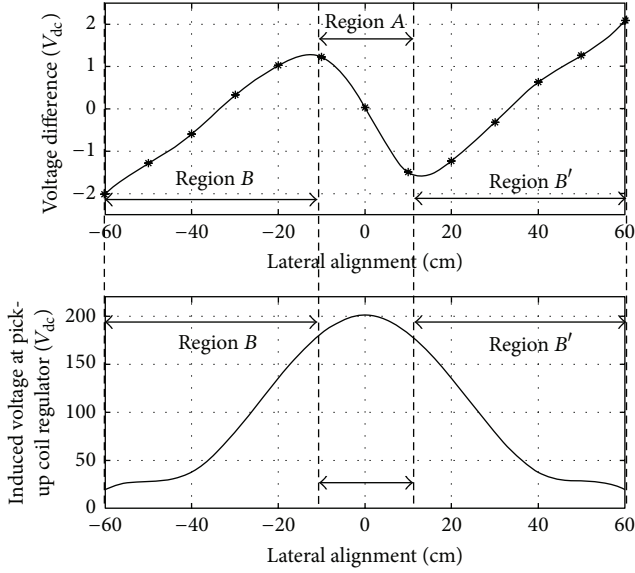
FIGURE 12: Correlation between voltage difference and pick-up coil voltage in order to determine the regions for the ACAS misalignment conversion unit.
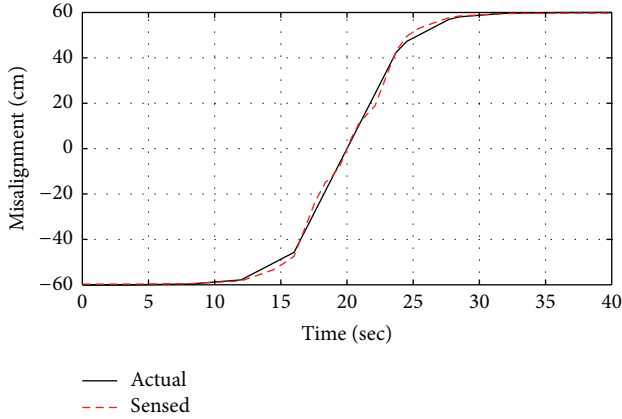
— Actual
--- Sensed

FIGURE 13: Simulation results comparing actual misalignment (desired) and the actual value that have been converted through the ACAS misalignment conversion unit.

with the given voltage difference input in the voltage → misalignment converter as shown in

$$y = m_n V_d. \tag{10}$$

To validate the feasibility of the misalignment conversion unit, a simulation model was designed in SIMULINK that replicates the equations shown in (8) to (10). The simulation begins with misalignment between the power line and the pick-up coil at −60 cm and progresses up until it reaches misalignment of 60 cm. The simulation results are shown in Figure 13. The dashed line is the desired lateral alignment value versus time, and the bold line is the actual converted misalignment value by the misalignment conversion unit. Based on the results, it can be seen that at 0 to 17 sec and 22 to 40 sec range, the converted misalignment value has some

discrepancies with the desired misalignment value, where the maximum error recorded was around 5 cm. However, this is not of great significance as the fuzzy steering controller can compensate for the discrepancies and still achieve high performance to correct detected misalignment.

5.3. Steering Fuzzy Controller for the ACAS. The final subsystem of the ACAS system is the steering fuzzy controller system as shown in Figure 14. Fuzzy control has been used to control the vehicle's steering based on the misalignment reference given by the misalignment conversion unit. Fuzzy control has been used in various applications, such as energy management for hybrid vehicles, parking finding services, vehicle dynamics control, and many other applications [16–19]. This control has also been applied to lane keeping assist systems (LKAS), which has similar resemblances to the fuzzy control specific for the ACAS [20, 21]. In addition, based on the reasons described in [18], fuzzy control method was specifically used for the ACAS application due to the nonlinear characteristics and the imprecise varying variables shown in Figures 12 and 13, respectively.

The received misalignment input is sent into the fuzzification unit, where it is converted into two fuzzy inputs: misalignment and misalignment rate. Converting into a "fuzzy" value means that a "crisp" value, a value that is identified either as TRUE or FALSE, is converted into a value that can be both TRUE and FALSE. The values received as input are converted into a specific membership function (MF). Each MF is a set that contains a certain range of values. To convert into a "fuzzy" value, a triangular MF is used, and it is mathematically expressed as follows:

$$\text{MF (triangle)} = \begin{cases} 0 & \text{if } x \le a_1 \\ \dfrac{x - a_1}{a_2 - a_1} & \text{if } a_1 < x \le a_2 \\ \dfrac{a_3 - x}{a_3 - a_2} & \text{if } a_2 < x \le a_3 \\ 0 & \text{if } a_3 < x. \end{cases} \tag{11}$$

a_1 to a_3 represent the x coordinates for the triangular MF, where a_1, a_2, and a_3 represent the left vertex, center vertex, and right vertex of the triangle, respectively. x is the value received as an input from the fuzzy logic system before it goes through the "fuzzification" unit. However, there are certain situations where the input value is desired more as a "crisp" value than a "fuzzy" value. In this case, a trapezoidal MF is used, which is mathematically expressed as follows:

$$\text{MF (trapezoid)} = \begin{cases} 0 & \text{if } x \le b_1 \\ \dfrac{x - b_1}{b_2 - b_1} & \text{if } b_1 < x \le b_2 \\ 1 & \text{if } b_2 < x \le b_3 \\ \dfrac{b_4 - x}{b_4 - b_3} & \text{if } b_3 < x \le b_4 \\ 0 & \text{if } b_4 < x. \end{cases} \tag{12}$$
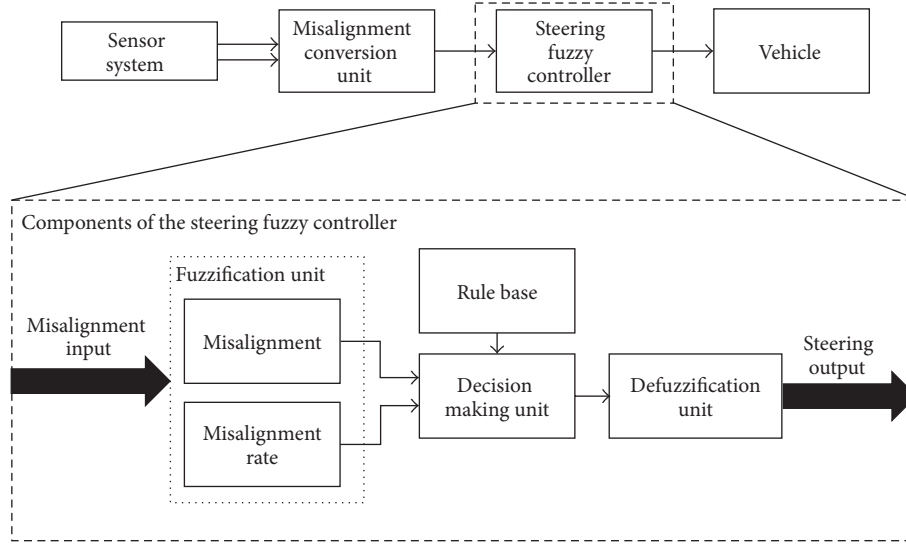
FIGURE 14: Block diagram of the ACAS steering fuzzy controller.
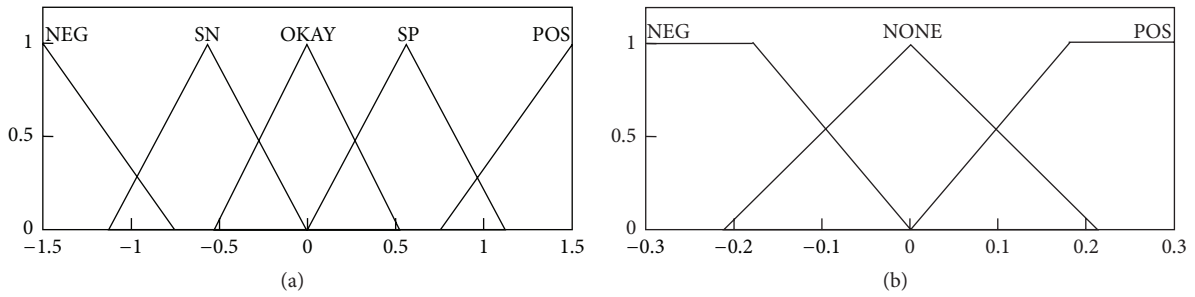
FIGURE 15: Fuzzy input set for (a) lateral displacement and (b) lateral displacement rate.

b_1 to b_4 represent the x coordinates for the trapezoidal MF, where b_1, b_2, b_3, and b_4 represent the left vertex, left-center vertex, right-center vertex, and right vertex of the trapezoid, respectively. Both the triangular and trapezoidal MFs have been used for the ACAS fuzzy logic controller, and the MF set for the lateral misalignment and its rate are shown in Figures 15(a) and 15(b), respectively.

In Figure 15(a), five MFs are defined: NEG (negative), SN (small negative), OKAY, SP (small positive), and POS (positive). Each MF has a specific misalignment range, while the overall set ranges between −1.5 m and 1.5 m, which is the maximum left or right position of one lane, respectively. The SN, OKAY, and SP MFs are put closer together with each other, while the NEG MF and the POS MF are put further away. This is to allow more sensitivity in control within the −1.1 m to 1.1 m range. However, with increased sensitivity, this may increase the oscillation of the system. To avoid this, another fuzzy input, which is the misalignment error rate, is implemented as shown in Figure 15(b). Three MFs are defined: NEG (negative), NONE, and POS (positive). The NEG and POS are trapezoidal MFs while the NONE is a triangular MF. This is to clearly define that the value is negative or positive when the rate is below −0.2 m or above 0.2 m, respectively.
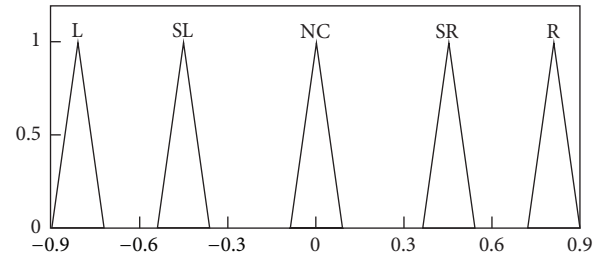
FIGURE 16: Fuzzy output set for the ACAS fuzzy controller.

The fuzzy output block is shown in Figure 16, which will send the final "crisp" value command to the EPS unit of the vehicle. 5 MFs are defined: L (left), SL (small left), NC (no change), SR (small right), and R (right). The MFs range from −0.9 to 0.9, which is the percentage of steering angle. Although it is designed to have maximum steering command, the steering output will not go more than ±10 degrees due to the given design constraints of the two fuzzy inputs and the rule base. This is an ideal case for the vehicle because it
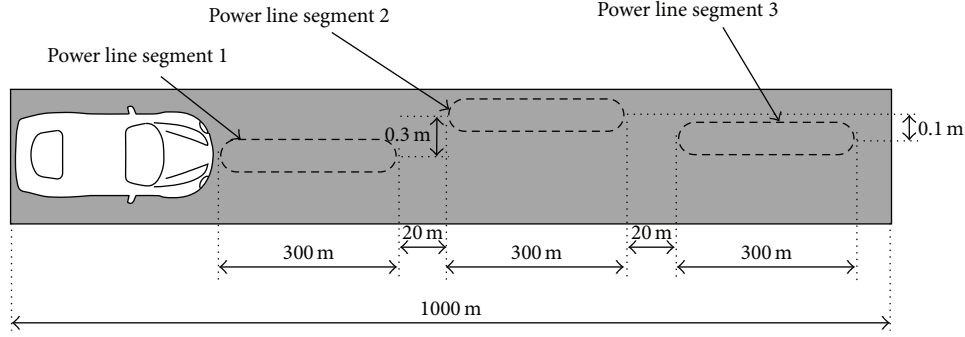
FIGURE 17: Diagram showing the placement of power lines on the road for SIMULINK and CarSim simulation.

will prevent any extreme steering situations, which is highly dangerous during high speed operations.

Rule Base for the Fuzzy Controller

  (1) If "misalignment" is OKAY than "steering" is NC.

  (2) If "misalignment" is POS then "steering" is R.

  (3) If "misalignment" is SP then "steering" is SR.

  (4) If "misalignment" is SN then "steering" is SL.

  (5) If "misalignment" is NEG then "steering" is L.

  (6) If "misalignment" is OKAY and "rate" is POS then "steering" is SL.

  (7) If "misalignment" is OKAY and "rate" is NEG then "steering" is SR.

The rule base shown in the above list is what defines the boundaries of the fuzzy logic controller. It determines the degree of membership of each MF of its fuzzy input sets through the decision-making unit. The degree of membership on both y-axes of the two fuzzy inputs will define how close the relationship is between each MF of its fuzzy set and the rule base. If degree of membership is closer to 1, it means that the relationship is strong, while 0 states the opposite. For example, if looking at the first rule set, it is defined that "misalignment" is OKAY. If the vehicle was positioned at the 0 cm mark, this would mean that the OKAY MF at the misalignment fuzzy input set will have a "1" degree of membership. This degree of membership ranging from 0 to 1 is used to determine the "cut-line" for the output fuzzy set. For example, if the degree of membership was at 0.7, this would mean that the corresponding fuzzy output set will have a "cut-line" at 0.7, where all area above the 0.7 threshold is ignored. From here, the final "crisp" steering output ratio, δ, will be determined in the defuzzification unit using the Center of Gravity (COG) method as shown in (13). The COG finds the center of area which is all output MFs x that are located within the corresponding fuzzy output set with the "cut-line," u_A, within the intervals of a and b. a and b are the corresponding x-axis limits of the fuzzy output set:

$$\delta = \frac{\int_a^b u_A(x)\,x\,dx}{\int_a^b u_A(x)\,x}. \tag{13}$$

TABLE 2: Parameters for vehicle model in CarSim.

Parameter	Value
Model name	C-Class, hatchback
Sprung mass	1274 kg
Length	3.35 m
Wheelbase	2.58 m
Width	1.74 m
Height	1.48 m
Power	125 kW
Tire	205/55 R16

To evaluate the response of the steering fuzzy controller, as well as its ability to keep the vehicle within the power line, a simulation was conducted using SIMULINK and CarSim. To conduct the simulation, a road model similar to the diagram as shown in Figure 17 was designed, where the three power line segments of 300 m length were installed 20 m apart in a 1000 m road lane. The first power line segment was installed at the center of the road, and the second and third power line segments were deviated 0.3 m and 0.1 m from the center of lane, respectively. The vehicle model with the ACAS fuzzy steering controller was also designed as well. The vehicle model name and its parameters used in SIMULINK/CarSim simulation are shown in Table 2. The fuzzy steering controller was tested at four different speeds from 80 km/h up to 140 km/h in 20 km/h increments. This was to observe how well the ACAS fuzzy steering controller maintains its performance as the vehicle's speed was increased.

Figure 18 shows the results of the simulation, where Figure 18(a) shows the lateral versus longitudinal displacement of the vehicle during the simulation. The thickest dash line is the desired position, which is the location of the three power lines. The other lines show the lateral versus longitudinal displacement of the vehicle at 80 km/h, 100 km/h, 120 km/h, and 140 km/h. Based on the result, the fuzzy steering controller was able to maintain its position near the desired lateral alignment position, even with increasing speeds.

Figure 18(b) shows the steering angle position of the fuzzy steering controller for the four different vehicle speeds. It can be seen that as vehicle speed increases, the oscillation

(a)

(b)

(c)
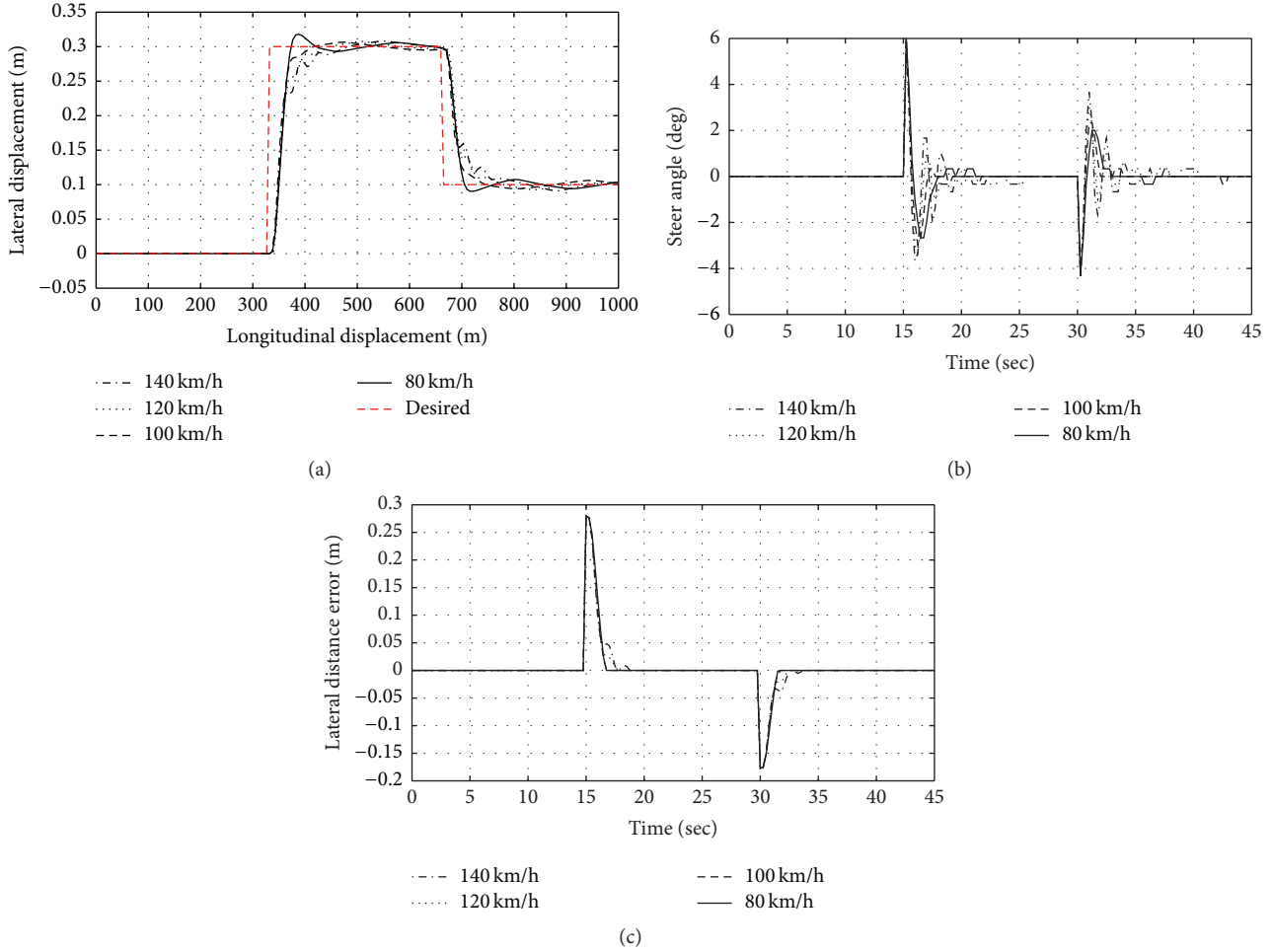
FIGURE 18: Simulation results showing the performance of steering fuzzy controller (a) lateral versus longitudinal displacement, (b) steering angle, and (c) lateral displacement error.

of the steering remained for a longer duration of the time. However, the remaining oscillation of the steering angle was very small, which was less than 1 degree. In addition, as illustrated in Figure 18(b), the maximum steering angle was at 6 degrees for all four vehicle speeds. This states that the vehicle's dynamic was still stable enough or the vehicle would not have been able to maintain its desired path as shown in the results in Figure 18(a). Figure 18(c) shows the lateral displacement error between the desired and actual lateral position. Even with small steering angles, the vehicle was able to position its correct lateral position within 3 to 5 seconds once a significant lateral alignment error was detected, even at higher speeds. Overall, the results shown in Figure 18 validate the performance of the steering fuzzy controller.

6. Comparison Analysis

The operational validity of each subsystem in the ACAS, the sensor system, misalignment conversion unit, and the steering fuzzy controller have been validated through simulation and show that each subsystem has performed as expected.

To view the feasibility of the overall ACAS system, the mitigation of power transfer loss due to misalignment was evaluated. To do so, another simulation was performed using SIMULINK and CarSim. Two vehicles, one vehicle with the ACAS controller and another vehicle without the ACAS controller, were operated on the same road shown in Figure 17. The first vehicle model is identical to the vehicle model used to test the steering fuzzy controller performance. The second vehicle is without the fuzzy controller model; it will only try to maintain the vehicle at the center of the road lane. This simulates a similar situation where the driver will operate the vehicle's steering to maintain the vehicle at the center of the road lane as the driver assumes that the power line is at the center of the road.

Figure 19 shows the results of the simulation, where Figure 19(a) shows the amount of power being transferred to the pick-up coil in the vehicle. At 0 cm misalignment position, the two vehicles receive maximum power at 47.83 kW based on the analysis conducted in Figure 4. However, as the vehicles pass towards power lines 2 and 3, only the vehicle with the ACAS was able to retain most of the maximum power delivery, while the vehicle without ACAS could only

(a)

(b)
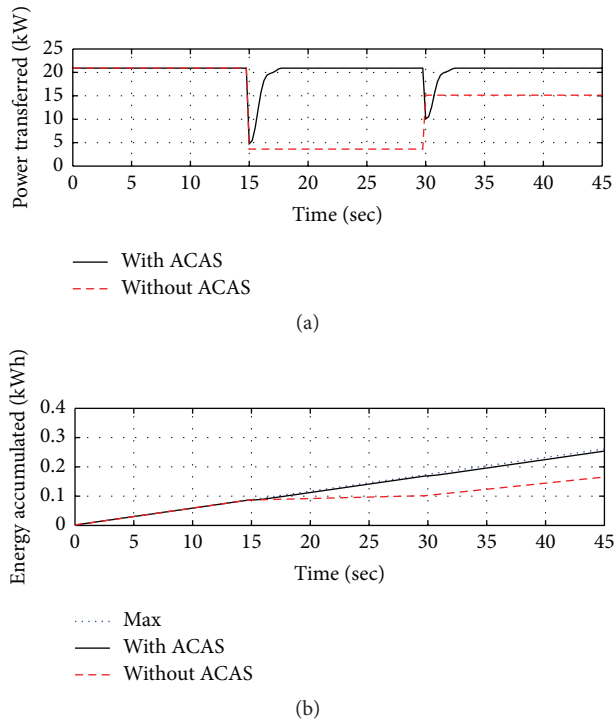
Figure 19: Simulated results showing amount of power transferred to vehicle with/without ACAS (a) and (b) accumulated energy to the vehicle due to power transfer.

receive roughly 4.90 kW and 26.35 kW of power at power lines 2 and 3, respectively.

Figure 19(b) shows the accumulated energy from the transferred power during the simulation time. As shown in the graph, the vehicle with the ACAS was able to accumulate about 578 Wh of energy at the end of the simulation, which is approximately 97% of the total maximum energy at 596 Wh. This maximum energy could have been accumulated only if the vehicle was at perfect alignment with the power line at all times. In case of the vehicle without the ACAS, the vehicle was only able to accumulate 311 Wh of energy at the end of the simulation, which is approximately 52% of the maximum energy that could be accumulated.

7. Conclusion

In this paper, an autonomous coil alignment system (ACAS) for vehicles with dynamic wireless charging is proposed. This system can detect lateral misalignment through three voltage sensors installed near the pick-up coil of the vehicle, and the nonlinear relationship between the voltage difference and misalignment position is converted into a more linear characteristic through the misalignment conversion unit. The lateral misalignment output from the misalignment conversion unit is received by the steering fuzzy controller, where the steering command is given to the electric power steering (EPS) system to control the vehicle's lateral position. The performance and operational feasibility of each subsystem have been verified through various simulations. In

the comparison analysis, it shows that the vehicle equipped with ACAS improves the wireless power transfer efficiency, allowing the vehicle to receive maximum power 97% of the time. With improved pick-up coil designs in electric vehicles with dynamic wireless charging, combination of the ACAS system will certainly provide significant benefits as it will be able to retain near maximum power transfer capacity during operation, thus improving the overall power transfer efficiency for the vehicle.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] S. Nako, K. Okuda, K. Miyashiro, K. Komurasaki, and H. Koizumi, "Wireless power transfer to a microaerial vehicle with a microwave active phased array," *International Journal of Antennas and Propagation*, vol. 2014, Article ID 374543, 5 pages, 2014.

[2] H. Zhou, B. Zhu, W. Hu, Z. Liu, and X. Gao, "Modelling and practical implementation of 2-coil wireless power systems," *Journal of Electrical and Computer Engineering*, vol. 2014, Article ID 906537, 8 pages, 2014.

[3] A. E. Rendon-Nava, J. A. Díaz-Méndez, L. Nino-de-Rivera, W. Calleja-Arriaga, F. Gil-Carrasco, and D. Díaz-Alonso, "Study of the effect of distance and misalignment between magnetically coupled coils for wireless power transfer in intraocular pressure measurement," *The Scientific World Journal*, vol. 2014, Article ID 692434, 11 pages, 2014.

[4] M. A. Adeeb, A. B. Islam, M. R. Haider, F. S. Tulip, M. N. Ericson, and S. K. Islam, "An inductive link-based wireless power transfer system for biomedical applications," *Active and Passive Electronic Components*, vol. 2012, Article ID 879294, 11 pages, 2012.

[5] S. Kim, H.-H. Park, J. Kim, J. Kim, and S. Ahn, "Design and analysis of a resonant reactive shield for a wireless power electric vehicle," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 4, pp. 1057–1066, 2014.

[6] A. Zaheer, H. Hao, G. A. Covic, and D. Kacprzak, "Investigation of multiple decoupled coil primary pad topologies in lumped IPT systems for interoperable electric vehicle charging," *IEEE Transactions on Power Electronics*, vol. 30, no. 4, pp. 1937–1955, 2015.

[7] T. Imura and Y. Hori, "Maximizing air gap and efficiency of magnetic resonant coupling for wireless power transfer using equivalent circuit and Neumann formula," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4746–4752, 2011.

[8] S. Raabe, G. A. J. Elliott, G. A. Covic, and J. T. Boys, "A quadrature pickup for inductive power transfer systems," in *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications (ICIEA '07)*, pp. 68–73, IEEE, May 2007.

[9] M. L. G. Kissin, G. A. Covic, and J. T. Boys, "Steady-state flat-pickup loading effects in polyphase inductive power transfer systems," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 6, pp. 2274–2282, 2011.

[10] G. A. J. Elliott, S. Raabe, G. A. Covic, and J. T. Boys, "Multiphase pickups for large lateral tolerance contactless power-transfer systems," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 5, pp. 1590–1598, 2010.

[11] J. Shin, S. Shin, Y. Kim et al., "Design and implementation of shaped magnetic-resonance-based wireless power transfer system for roadway-powered moving electric vehicles," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 3, pp. 1179–1192, 2014.

[12] S. Ahn, N. Suh, and D.-H. Cho, "Charging up the road," *IEEE Spectrum*, vol. 50, no. 4, pp. 48–54, 2013.

[13] Y. D. Ko and Y. J. Jang, "The optimal system design of the online electric vehicle utilizing wireless power transmission technology," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1255–1265, 2013.

[14] S. Shin, J. Shin, Y. Kim et al., "Hybrid inverter segmentation control for online electric vehicle," in *Proceedings of the IEEE International Electric Vehicle Conference (IEVC '12)*, Greenville, SC, USA, March 2012.

[15] S. Lukic and Z. Pantic, "Cutting the cord: static and dynamic inductive wireless charging of electric vehicles," *IEEE Electrification Magazine*, vol. 1, no. 1, pp. 57–64, 2013.

[16] Z. Chen, J. C. Xia, and B. Irawan, "Development of fuzzy logic forecast models for location-based parking finding services," *Mathematical Problems in Engineering*, vol. 2013, Article ID 473471, 6 pages, 2013.

[17] G. Yin, S. Wang, and X. Jin, "Optimal slip ratio based fuzzy control of acceleration slip regulation for four-wheel independent driving electric vehicles," *Mathematical Problems in Engineering*, vol. 2013, Article ID 410864, 7 pages, 2013.

[18] N. J. Schouten, M. A. Salman, and N. A. Kheir, "Fuzzy logic control for parallel hybrid vehicles," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 3, pp. 460–468, 2002.

[19] J. Chien, J. Lee, and L. Liu, "A fuzzy rules-based driver assistance system," *Mathematical Problems in Engineering*, vol. 2015, Article ID 207675, 14 pages, 2015.

[20] J. E. Naranjo, C. González, R. García, and T. De Pedro, "Lane-change fuzzy control in autonomous vehicles for the overtaking maneuver," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 3, pp. 438–450, 2008.

[21] G. Wang, J. Zhao, X. Zhang, and R. Zhao, "Multi-model fuzzy controller for vehicle lane tracking," in *Proceedings of the IEEE 17th International Conference on Intelligent Transportation Systems (ITSC '14)*, pp. 1341–1346, IEEE, Qingdao, China, October 2014.

Research Article

A Wind Power and Load Prediction Based Frequency Control Approach for Wind-Diesel-Battery Hybrid Power System

Chao Peng,[1] Zhenzhen Zhang,[2] and Jia Wu[1]

[1]*School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2]*College of Electrical & Information Engineering, Southwest University for Nationalities, Chengdu, Sichuan 610041, China*

Correspondence should be addressed to Zhenzhen Zhang; zhangzhenzhen.isit@gmail.com

A frequency control approach based on wind power and load power prediction information is proposed for wind-diesel-battery hybrid power system (WDBHPS). To maintain the frequency stability by wind power and diesel generation as much as possible, a fuzzy control theory based wind and diesel power control module is designed according to wind power and load prediction information. To compensate frequency fluctuation in real time and enhance system disturbance rejection ability, a battery energy storage system real-time control module is designed based on ADRC (active disturbance rejection control). The simulation experiment results demonstrate that the proposed approach has a better disturbance rejection ability and frequency control performance compared with the traditional droop control approach.

1. Introduction

Remote Area Power Supply (RAPS) systems which are used to supply power for rural and remote areas, including island, could not rely on the main grid supply system. The power supply of such areas is always provided with diesel power generation, due to its merits of low installation cost, high reliability, and simple operation. The major drawbacks of this type of power generation is finite fuel, low utilization efficiency, high transportation cost, environmental pollution, and so on [1].

Recently, more and more RAPS systems use renewable power generation, such as wind power, to overcome the draws of diesel power generation. However, the active power output of wind power generation is random and fluctuant, which will seriously affect the stability of RAPS [2]. To solve this problem, wind-diesel-battery hybrid power system (WDBHPS) is introduced, which have become a popular power generation system used in RAPS. It could utilize wind power generation to reduce the use of diesel generators and environmental pollution and utilize battery energy storage system (BESS) to compensate the wind power fluctuation.

Frequency stability reflects the active power balance between the supply and demand, which is important factor in operation security and stability of the wind-diesel-battery hybrid power system. The fluctuation of wind and load would cause system frequency deviation and fluctuation. Thus, frequency control problem is main problem which hinder the development and utilization of wind-diesel-storage hybrid system in practice.

Currently, more and more scholars focus on the research on the frequency control of wind-diesel hybrid power system. Droop control based frequency control approach is the mostly used in wind-diesel hybrid power system in practice [3]. The traditional droop control use a fixed droop coefficient to increase or decrease the active power output of power sources when load demand changes. It could not avoid large deviation of frequency or voltage amplitude for its fixed droop coefficient. Therefore, many scholars studied the enhanced droop control approach. To reduce magnitude of frequency change, an adaptive droop control approach which could adjust the droop coefficient according to the change of active power and load was proposed in [4]. To eliminate the static error of frequency, an integral controller is introduced into frequency droop control in [5]. But its parameter selection is troublesome. In [6], a centralized frequency control strategy is presented, which calculates allocated active power of wind power turbine, diesel generator, and discharging/charging
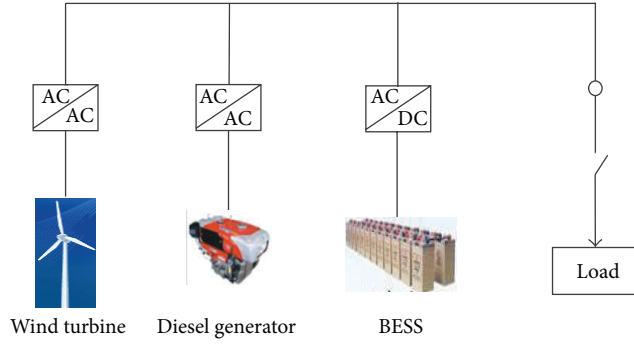
FIGURE 1: Wind-diesel-battery hybrid power system.

FIGURE 2: Control structure of wind power and load prediction based frequency control.

power of BESS according to the deviation of frequency. In [7], a proportional-integral (PI) control based frequency control approach was presented to maintain the balance between power generation and load demand. In [8], a fuzzy logic based supervisory frequency control approach was proposed to maintain the frequency stability of wind-diesel hybrid power system.

Most above frequency control approaches only use BESS to compensate the frequency deviation caused by wind power and load fluctuation and maintain power-load balance by diesel generator. Very few of them consider regulating the active power output of wind turbine to reduce frequency fluctuation. Meanwhile, they did not utilize prediction information of wind power and load to stabilize WDBHPS's frequency. The developments of wind power prediction technology and load prediction technology for microgrid, such as wavelet transformed based wind power prediction approach [9], grey model based wind power prediction model [10], artificial neural network based load prediction approach [11], and bilevel prediction strategy based load prediction approach [12], make their application become possible. If the prediction information could be taken into account in frequency control, the active power output of wind turbine and diesel generator could be regulated to reduce the frequency deviation caused by wind power and load fluctuation.

In this paper, a wind power and load prediction based frequency control approach is proposed for WDBHPS. In this approach, a fuzzy control based wind and diesel power control module is used to reduce frequency deviation by wind turbine and diesel generator according to the wind power and load prediction information. To overcome the draws of traditional droop control, such as difficulty of droop coefficient adjusting, static error of frequency control, and poor disturbance rejection ability, an ADRC based real-time control module of BESS is used to compensate the real-time frequency fluctuation.

2. Wind Power and Load Prediction Based Frequency Control Approach

Wind-diesel-battery hybrid power system is shown in Figure 1. As seen in Figure 1, this system consists of a wind turbine, diesel generator, BESS, AC/AC transformers, and
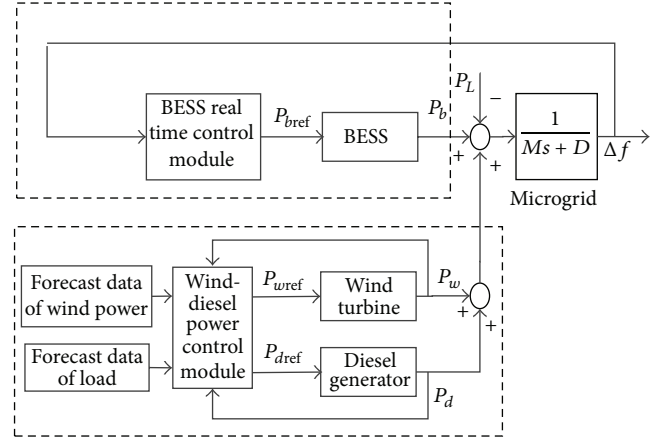
DC/AC convert. The output power of wind turbine and diesel generator and charging/discharging power of BESS are controlled to keep power-load balance and frequency stable in the system.

The control scheme of the proposed control approach is shown in Figure 2, where P_{wref} and P_w are the reference of active power output and actual active power output of wind turbine, respectively, P_{dref} and P_d are the reference of active power output and actual active power output of diesel generator, respectively, and P_{bref} and P_b are the reference of charging/discharging power and actual charging/discharging power of BESS, respectively. P_L is the actual load of system, and Δf is the system frequency deviation.

As seen in Figure 2, the proposed frequency control system mainly consists of two modules: wind-diesel power control module and BESS real-time control module. The wind-diesel power control module regulates the active power output of wind turbine and diesel generator according to wind power and load prediction information. It could regulate the active power output of wind turbine and diesel generator to track the predicted load as close as possible, so as to reduce the frequency deviation to be as small as possible. The wind-diesel power control module is designed based on fuzzy control, which could choose different rules based on wind power and load prediction information. An active disturbance rejection control (ADRC) based BESS real-time control module is designed to compensate the frequency fluctuation in real time, which could use the extended state observer to estimate the disturbance and compensate the real-time frequency fluctuation caused by wind power and load fluctuation.

3. Wind-Diesel Power Control Module

The scheme of wind-diesel power control module is shown in Figure 3. It consists of three parts: interpolation module, power control, and correction module.

3.1. Interpolation Module. The interpolation module determines the predictive value of wind power and load between each two adjacent prediction periods for power control, when
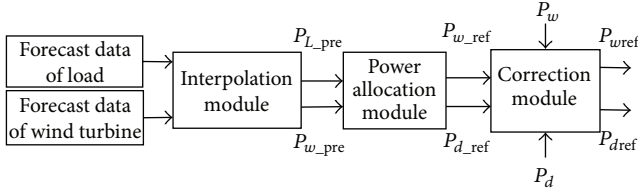
FIGURE 3: The scheme of wind-diesel control module.

FIGURE 4: Control structure of correction module.

prediction period is unmatched with control period. The interpolation module utilizes linear interpolation method to calculate and smooth predictive value at control period. Assumes that the prediction period is T whose unit is s (seconds), the predictive value is y_0 at t_0, and the predictive value is y_1 at $t_0 + T$. The linear interpolation formula is as follows:

$$y(t) = at + b, \tag{1}$$

where $t_0 \le t \le t_0 + T$. Replace t and y in (1) by values of (t_0, y_0) and $(t_0 + T, y_1)$, and coefficients a and b can be obtained:

$$a = \frac{y_1 - y_0}{T},$$
$$b = \frac{(T + t_0)y_0 - t_0 y_1}{T}. \tag{2}$$

3.2. Power Calculation Module. The design objective of power calculation module is to calculate reference power for wind turbine and diesel generation, which could keep the system power balance between generating power and demanding load under the premise of using wind energy as much as possible. Its control strategy is designed as follows:

(1) When $P_{L\_pre} < P_{w\_pre}$, that is, prediction value of load is less than the prediction value of wind power, the load fluctuation would be compensated by wind turbine alone and reference power of diesel generator could be zero. The control of the power value is shown as follows:

$$P_{w\_ref} = P_{L\_pre},$$
$$P_{d\_ref} = 0. \tag{3}$$

(2) When $P_{L\_pre} < P_{w\_pre}$, that is, prediction value of load is equal to or greater than the prediction value of wind power, the reference power of wind power could be set to the prediction value of wind power to maximize the use of wind power and the frequency fluctuation would be compensated by diesel generator. The calculated reference power values for wind turbine and diesel generator are shown as follows:

$$P_{w\_ref} = P_{w\_pre},$$
$$P_{d\_ref} = P_{L\_pre} - P_{w\_ref}. \tag{4}$$

3.3. Correction Module. Correction module adjusts the reference power calculated by power calculation module according to the restraints of active power output change of wind turbine and diesel generator. When wind speed fluctuates sharply, the active power output of wind turbine would fluctuate sharply if wind turbine tracks the calculated reference power in (4) directly. Although diesel generator can be controlled to output any value between zero and the rated power, the rate of change of its power has its constraint.

The objective of correction module is reducing the change of active power output of wind turbine, the operation of diesel generator. The correction module is designed based on fuzzy control, which utilizes the wind power and load prediction information to adjust the active power output of wind turbine and diesel generator.

The fuzzy logic controller for correction module uses two single inputs and two-dimensional output of the structure. Its structure is shown in Figure 4.

The input variables to the fuzzy logic controller are as follows.

The first input variable is the error between reference power and actual active power output of wind turbine ΔP_w, that is, $P_{w\_ref} - P_w$. The second input variable is the error between reference power and actual active power output of diesel generator ΔP_d.

The output variable is adjusting power ΔP which is used to adjust the reference power of wind turbine and diesel generator calculated in power calculation module:

$$P_{wref} = P_{w\_ref} + \Delta P,$$
$$P_{dref} = P_{d\_ref} + \Delta P. \tag{5}$$

Define language variable of ΔP_w as E_1, language variable of ΔP_d as E_2, language variable of ΔP as U, respectively. The fuzzy set of input variables and output variable is {NB, NM, NS, ZO, PS, PM, PB}, which means that the change of power is {negative big, negative middle, negative small, zero, positive small, positive middle, positive big}.

The membership function of the two inputs and output of the fuzzy logic controller adopts the triangle membership function and are shown in Figures 5, 6, and 7. Table 1 shows the fuzzy control rule. The weighted average method is used for defuzzification.

TABLE 1: The rule of fuzzy controller.

U		E_1					
		NB	NM	NS	ZO	PS	PM	PB
	NB	ZO	PS	PS	PM	PM	PB	PB
	NM	ZO	ZO	PS	PS	PM	PM	PB
	NS	NS	ZO	ZO	PS	PS	PM	PM
E_2	ZO	NS	NS	ZO	ZO	ZO	PS	PS
	PS	NM	NM	NS	NS	ZO	ZO	PS
	PM	NB	NM	NM	NS	NS	ZO	ZO
	PB	NB	NB	NM	NM	NS	NS	ZO

FIGURE 6: Membership function of the input E_2.

FIGURE 5: Membership function of the input E_1.
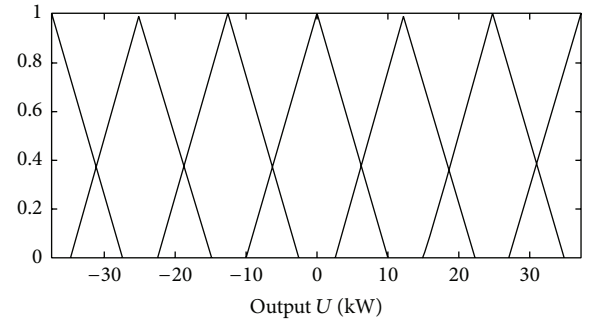
FIGURE 7: Membership function of the output U.

4. Real-Time BESS Frequency Control Module

The design objective of the real-time BESS frequency module is to compensate the frequency fluctuations caused by real-time wind power and load fluctuation and maintain the system frequency in rated frequency range to meet the active power-load balance.

Figure 8 shows the control structure of the real-time BESS frequency control module based on ADRC controller, where Δf_{ref} is reference frequency error, which is always given as zero, that is, the real-time frequency error between rated frequency and actual frequency, ΔP_f is adjusting charging or discharging power of BESS, and Δf is the actual frequency deviation of system. $P_L - P_w - P_d$ is the actual power deviation between load and power generated by wind turbine and diesel. It could be seen as disturbance to system frequency.

In Figure 8, BESS could be seen as a first-order lag loop [13]. Therefore, the whole controlled object can be seen as a concatenation of two first-order loops. A second-order ADRC controller is used for real-time frequency BEES control.

Figure 9 shows the control structure of the second-order ADRC controller, where G_1 is the TD (Tracking Difference), G_2 is the ESO (extended state observer), G_3 is NLSEF (Nonlinear State Error Feedback), and G_0 is the controlled object.

(a) Tracking Difference G_1. The math model of TD is designed as follows:

$$\dot{v}_1 = v_2,$$
$$\dot{v}_2 = fst(v_1, v_2, r, h), \tag{6}$$

FIGURE 8: Control structure of BESS based on ADRC.
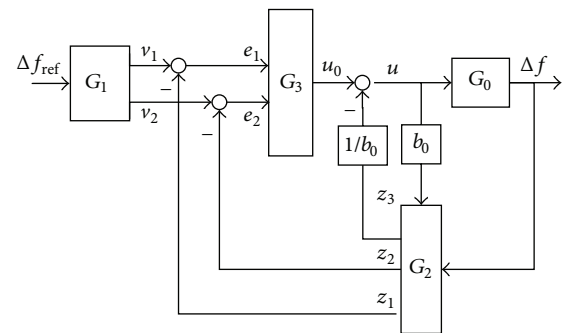
FIGURE 9: Control structure of second-order ADRC.

where function $fst(v_1, v_2, r, h)$ is determined by following equation:

$$\delta = r \cdot h,$$
$$\delta_0 = \delta \cdot h,$$
$$y = v_1 - \Delta f_{ref} + h \cdot v_2,$$

$$a_0 = \sqrt{\delta^2 + 8r|y|},$$

$$a = \begin{cases} x_2 + \dfrac{y}{h} & |y| \leq \delta_0 \\ x_2 + \dfrac{(a_0 - \delta)}{2} \, \text{sign}(y) & |y| > \delta_0, \end{cases}$$

$$\text{fst}(v_1, v_2, r, h) = \begin{cases} -\dfrac{r \cdot a}{\delta} & |a| \leq \delta \\ -r \cdot \text{sign}(a) & |a| > \delta, \end{cases}$$

(7)

where r is the speed factor, h is the filter factor, sign is a sign function, and r and h are adjustable parameters of TD.

(b) Extended State Observer G_2. The math model of ESO is designed as follows:

$$\varepsilon = z_1 - y,$$

$$\dot{z}_1 = z_2 - \beta_{01} \cdot \varepsilon,$$

$$\dot{z}_2 = z_3 - \beta_{02} \cdot \text{fal}(\varepsilon, \alpha_1, \delta_1) + b_0 \cdot u,$$

$$\dot{z}_3 = -\beta_{03} \cdot \text{fal}(\varepsilon, \alpha_2, \delta_1),$$

(8)

where function $\text{fal}(\varepsilon, \alpha, \delta)$ is given as follows:

$$\text{fal}(\varepsilon, \alpha, \delta) = \begin{cases} |\varepsilon|^\alpha \, \text{sign}(\varepsilon) & |\varepsilon| > \delta \geq 0 \\ \dfrac{\varepsilon}{\delta^{1-\alpha}} & |\varepsilon| \leq \delta. \end{cases}$$

(9)

In (8), by choosing appropriate $\{\alpha_1, \alpha_2, \delta_1, \beta_{01}, \beta_{02}, \beta_{03}\}$, z_1, z_2 could be used to estimate the controlled variable y and its differential. z_3 could be used to estimate the disturbance.

(c) Nonlinear State Error Feedback G_3. The math model of NLSEF is designed as follows:

$$e_1 = v_1 - z_1,$$

$$e_2 = v_2 - z_2,$$

$$u_0 = k_1 \cdot \text{fal}(e_1, \alpha_3, \delta_2) + k_2 \text{fal}(e_2, \alpha_4, \delta_2),$$

$$u = u_0 - \dfrac{z_3}{b_0},$$

(10)

where function fal() is showed in (9), z_3 is the extended state variable, and $k_1, k_2, \alpha_3, \alpha_4,$ and δ_2 are the adjusting parameters of NLSEF.

(d) Control Object G_0. The structure of controlled object G_0 is shown in Figure 10, whose control variable is $u = \Delta P_f$, output variable is $y = \Delta f$, and disturbance is $P_L - P_w - P_d$.

The transfer function of control object G_0 could be written as follows:

$$G(s) = -\frac{1}{T_b s (Ms + D)}.$$

(11)
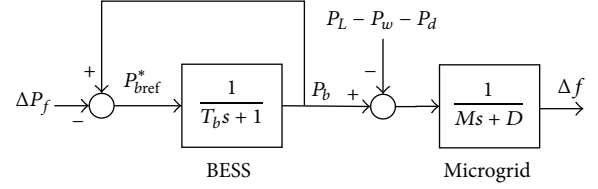
FIGURE 10: Control structure from control variable to output.
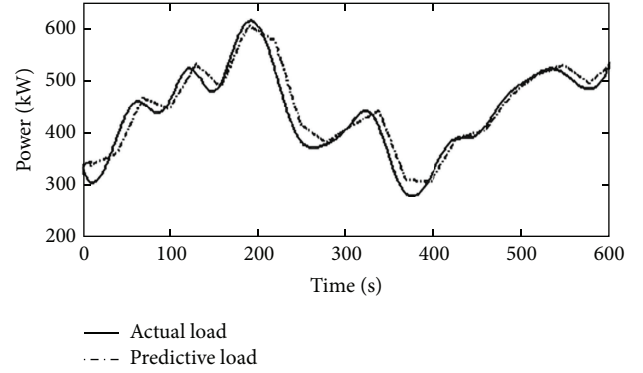
FIGURE 11: Actual load and predictive load.

The corresponding state space expression of G_0 could be written as follows:

$$\dot{x}_1 = \frac{1}{M}(-Dx_1 + x_2),$$

$$\dot{x}_2 = -\frac{1}{T_b}u,$$

(12)

where $x = [x_1, x_2] = [\Delta f, P_b]$ is the state variable, $y = x_1$ is the output, and $u = \Delta P_f$ is the control variable.

5. Simulation and Analysis

In this section, the simulation experiment of frequency control of WPBHPS will be conducted to demonstrate the effectiveness of the proposed approach. A wind-diesel-battery hybrid power system [14], which consists of a 750 kW wind turbine, a 350 kW diesel generator, and a 300 kW BESS with 50 kW·h capacity is used in the simulation experiments.

For comparison, two different frequency control approaches will be used in the simulation experiment: (a) the proposed control approach and (b) the control approach which control the active power output of wind turbine by maximum power point tracking (MPPT) control approach and compensate frequency deviation by diesel generator and BESS only.

A load prediction method based on artificial neural network [11] is used to predict the load. The actual load and predictive load curve is shown in Figure 11.

The actual wind speed data of wind turbines is shown in Figure 12.

A grey model based wind power prediction method [15] is used to predict the available active power of wind turbine. The predictive power of the wind turbine is shown in Figure 13.
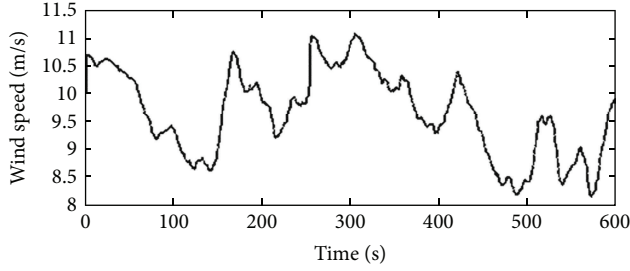
FIGURE 12: Actual wind speed.
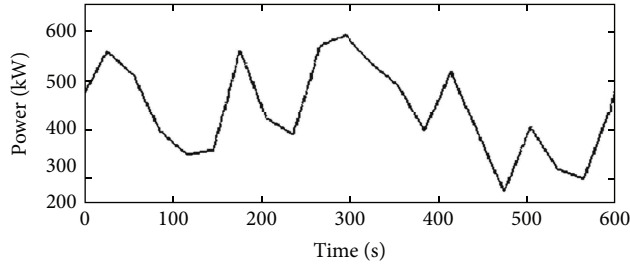
FIGURE 13: Predictive power of wind turbine.

—— Control approach (a)
- - - Control approach (b)
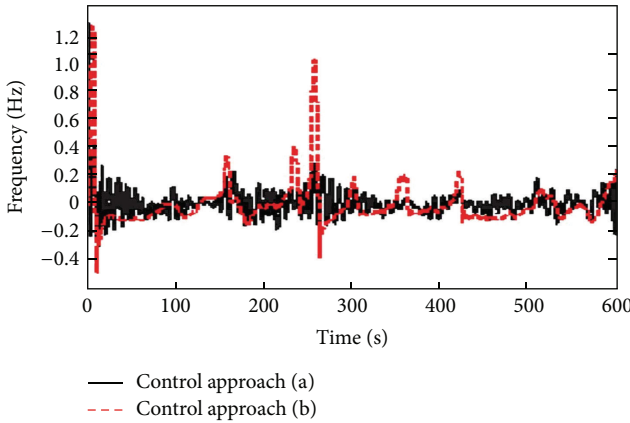
FIGURE 14: Frequency fluctuation of system.

—— Control approach (a)
- - - Control approach (b)

FIGURE 15: Active power output of wind turbine.

—— Control approach (a)
- - - Control approach (b)
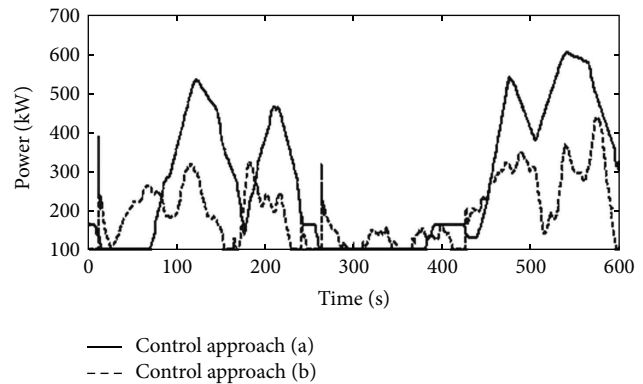
FIGURE 16: Active power output of diesel generator.

—— Control approach (a)
- - - Control approach (b)
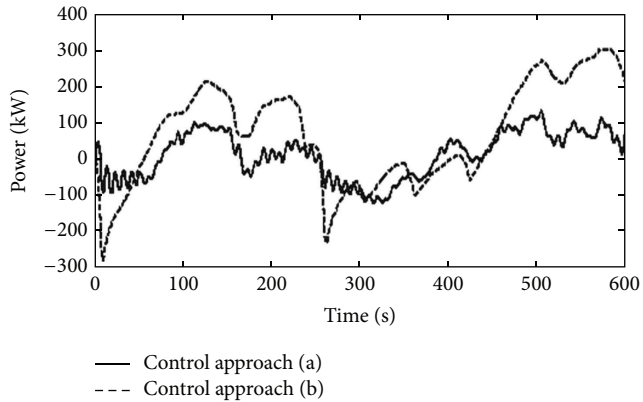
FIGURE 17: Charging/discharging power of BESS.

The results of frequency control simulation experiments are shown in Figure 14. By using control approach (b), the maximum and minimum frequency deviations are 1.2411 Hz and 0.4772 Hz, respectively. The maximum frequency deviation of control approach (a) is 0.4103 Hz, which is 33.5% of that of control approach (b). The minimum frequency deviation of control approach (a) is 0.2920 Hz, which is 61.2% of that of control approach (b). It is obvious that the frequency control performance of control approach (a) is better than approach (b).

The active power output of wind turbine and diesel generator are shown in Figures 15 and 16. As seen in them, the active power output of wind turbine and diesel generator by using control approach (a) is smoother than that by using control approach (b).

The charging/discharging power curve and SOC of BESS are shown in Figures 17 and 18. As seen in Figure 16,

the charging and discharging power of BESS are much smaller by using control approach (a) compared with control approach (b). As seen in Figure 17, the initial SOC of BESS is 0.5. By using control approach (a), the SOC of BESS is around 0.5. By using control approach (b), the SOC of BESS is far below 0.5.

6. Conclusions

This paper proposed a load and wind power prediction based frequency control approach for wind-diesel-battery
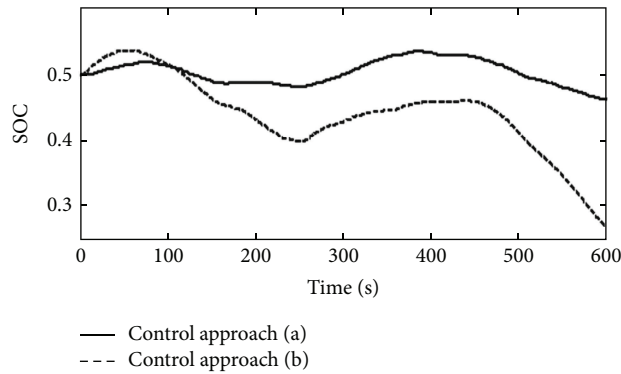
FIGURE 18: SOC of BESS.

hybrid power system. In this control approach, a wind-diesel power control module is designed to regulate the active power output of wind turbine and diesel generator according to the prediction information of load and wind power, which not only reduces frequency deviation caused by fluctuation of load and wind power but also maximizes use of wind power and reduces operation of diesel generator. A real-time BESS frequency control module based on ADRC controller is used to compensate the disturbance caused by load and wind power fluctuation in real time. The simulation results demonstrate that the proposed approach has a better disturbance rejection ability and frequency control performance compared with the traditional control approach without prediction information and traditional droop control approach.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] Z. Chen and Y. Hu, "A hybrid generation system using variable speed wind turbines and diesel units," in *Proceedings of the 29th Annual Conference of the IEEE Industrial Electronics Society (IECON '03)*, vol. 3, pp. 2729–2734, IEEE, Roanoke, Va, USA, November 2003.

[2] M. Imanaka, Y. Onda, J. Baba, T. Yoshihara, and A. Yokoyama, "Feasibility study on compensation of power fluctuation caused by renewable energy source using desalination system in island power system," *Journal of International Council on Electrical Engineering*, vol. 1, no. 3, pp. 345–351, 2014.

[3] P. Basak, S. Chowdhury, S. Halder Nee Dey, and S. P. Chowdhury, "A literature review on integration of distributed energy resources in the perspective of control, protection and stability of microgrid," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 8, pp. 5545–5556, 2012.

[4] Y. W. Zheng, M. Y. Chen, C. Li, R. Xu, and X. Xu, "A microgrid control strategy based on adaptive drooping coefficient adjustment," *Automation of Electric Power Systems*, vol. 37, no. 7, pp. 6–11, 2013.

[5] J. A. Peas Lopes, C. L. Moreira, and A. G. Madureira, "Defining control strategies for microgrids islanded operation," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 916–924, 2006.

[6] A. Madureira, C. Moreira, and J. P. Lopes, "Secondary load-frequency control for micro grids in islanded operation," in *Proceedings of the International Conference on Renewable Energy and Power Quality*, Barcelona, Spain, 2005.

[7] T. Senjyu, T. Nakaji, K. Uezato, and T. Funabashi, "A hybrid power system using alternative energy facilities in isolated island," *IEEE Transactions on Energy Conversion*, vol. 20, no. 2, pp. 406–414, 2005.

[8] L. Leclercq, B. Robyns, and J.-M. Grave, "Control based on fuzzy logic of a flywheel energy storage system associated with wind and diesel generators," *Mathematics and Computers in Simulation*, vol. 63, no. 3–5, pp. 271–280, 2003.

[9] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, "Short-term wind power prediction based on wavelet transform-support vector machine and statistic characteristics analysis," *IEEE Transactions on Industry Applications*, vol. 48, no. 4, pp. 1136–1141, 2012.

[10] T. H. M. El-Fouly, E. F. El-Saadany, and M. M. A. Salama, "Improved grey predictor rolling models for wind power prediction," *IET Generation, Transmission and Distribution*, vol. 1, no. 6, pp. 928–937, 2007.

[11] Z. R. Li, "Short—term load forecasting based on artificial neural network," *Guangxi Electric Power*, vol. 25, no. 4, pp. 7–10, 2002.

[12] N. Amjady, F. Keynia, and H. Zareipour, "Short-term load forecast of microgrids by a new bilevel prediction strategy," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 286–294, 2010.

[13] T. Senjyu, A. Uehara, and A. Yona, "Frequency control by coordination control of wind turbine generator and battery using H_∞ control," in *Proceedings of the Transmission & Distribution Conference & Exposition: Asia and Pacific*, pp. 1–4, Seoul, The Republic of Korea, October 2009.

[14] N. Mendis, K. M. Muttaqi, S. Sayeef, and S. Perera, "A control approach for voltage and frequency regulation of a wind-diesel-battery based hybrid remote area power supply system," in *Proceedings of the 36th Annual Conference of the IEEE Industrial Electronics Society (IECON '10)*, pp. 3054–3060, Glendale, Calif, USA, November 2010.

[15] Z. Z. Zhang, J. X. Zou, and G. Zheng, "Ultra-short term wind power prediction model based on modified grey model method for power control in wind farm," *Wind Engineering*, vol. 35, no. 1, pp. 55–68, 2011.

Research Article

Control Strategy for Power Distribution in Dual Motor Propulsion System for Electric Vehicles

Pedro Daniel Urbina Coronado and Horacio Ahuett-Garza

Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, 64849 Monterrey, NL, Mexico

Correspondence should be addressed to Pedro Daniel Urbina Coronado; pd.urbina.phd.mty@itesm.mx

Electric Vehicles with more than one electric motor can offer advantages in saving energy from the batteries. In order to do that, the control strategy plays an important role in distributing the required torque between the electric motors. A dual motor propulsion system with a differential transmission is simulated in this work. A rule based control strategy for this propulsion system is proposed and analyzed. Two parameters related to the output speed of the transmission and the required torque are used to switch the two modes of operation in which the propulsion system can work under acceleration. The effect of these parameters is presented over the driving cycles of NEDC, UDDS, and NYCC, which are followed using a PID controller. The produced energy losses are calculated as well as an indicator of drivability, which is related to the difference between the desired speed and the actual speed obtained. The results show that less energy losses are present when the vehicle is maintained with one electric motor most of the time, switching only when the extended speed granted by the second motor is required. The propulsion system with the proposed control strategy represents a feasible alternative in the spectrum of sustainable transportation architectures with extending range capabilities.

1. Introduction

Compared with propulsion systems based on the *Internal Combustion Engine* (ICE), the electric propulsion system offers advantage in the efficiency in which the energy is transformed into rotating movement. Typical *Electric Vehicle* (EV) propulsion systems provide efficiencies between 53% and 77%, which are superior when compared to propulsion systems based on ICEs, with efficiencies between 13% and 20% [1]. However, the capacity to store energy in an Electric Vehicle continues to represent a limitation for a more widespread use of EVs. With current technologies, the amount of energy that can be carried lies in a range between 100 and 250 Wh/kg [2, 3], significantly smaller than the amount of energy stored in fossil fuels 15 kWh/kg [4]. This fact, combined with long charging periods and elevated cost, results in EVs with limited driving range that cannot compete with ICE cars.

To address this issue, several research lines have been explored. Lightweight materials [5], new battery technologies [3, 6], and more efficient motors and power electronics [7]

can be counted among the possible improvements. In parallel with those research lines, it has been found that architectures that depart from the traditional electric powertrain can offer a more efficient use of the limited energy stored in the batteries. In such systems, two or more electric motors are used in combination with a planetary gear train, which allows the load to be distributed among the motors. This provides properties of speed ratios that cannot be achieved by conventional transmissions. The control strategy for the power distribution of energy between the motors is a decisive factor in the minimization of power losses of the propulsion system.

In the work of Zhang and coworkers [8], the *dual motor coupling propulsion system* (DMCPS) is presented. The DMCPS can work with one or two electric motors. The shift between one or two motors is produced when the main motor achieves certain speed, which is then maintained constant, and, from that point, the vehicle speed is regulated only by a smaller auxiliary motor. However, when there is a speed differential between the branches of planetary systems, there are also differences in the torque ratios between the power

sources, an effect that was not specified in the cited article. Efficiency analysis and EV driving range simulations have found that efficiency and range are improved if the dual motor type of propulsion system works either with one motor and the other motor locked or with the two motors at the same speed [9]. The operation of the two motors at different speeds can be used as a transition, as we propose in the present work. In the work of Wang and Sun [10], a dual motor propulsion system is also analyzed with an optimization of the components of the powertrain; however details of the control strategy of operation of the two motors were not provided.

The proposed power split control strategy has the capacity to achieve vehicle speeds either working only with one electric motor, working with two electric motors at the same speed, or working with two electric motors at different speeds. The goal is to analyze the input information that the control strategy requires in real time to obtain lower energy losses in the propulsion system during the vehicle acceleration, with the best drivability possible. Due to its simplicity and practicality, a rule based algorithm was chosen to design the control strategy over other popular alternatives like dynamic programming [8, 11] or extremum seeking algorithm [12] whose implementation could be unfeasible or computationally expensive.

Considering the mentioned control strategies and optimizations in powertrains with more than one electric motor, this work aims to contribute with the design of a power split control strategy for a dual motor type of propulsion system for EVs, implementable in real time, which depend on the monitoring of the driver torque request and the vehicle speed. Also, the proposed control strategy was modelled and simulated over driving cycles to analyze the effect of its control parameters on the energy losses and drivability.

The organization of this work is as follows. Section 2 describes the powertrain architecture and Section 3 presents the modelling of its components. Section 4 describes the design of the control strategy. Section 5 provides the results of the proposed control strategy and the effects of the input parameters of the control strategy on energy losses and drivability. Section 6 provides the conclusions of this study.

2. Powertrain Architecture

The powertrain used in this work is presented in the schematic displayed in Figure 1. The first differential (D1) is composed of side gears 1 and 2 (SG1 and SG2), a carrier and planetary gears, and a ring gear (RG) which outputs the torque. The torque is transmitted through a reduction gear and from there to a second differential (D2), which transmits the power to the wheels.

The figure shows that the torque request is sent to the embedded system, which transmits the torque request to the two motors. The control strategy is programmed in the embedded system, which is in charge of the decision of when it is more convenient to use one or both motors according to the conditions of speed and available power. When both motors are used, the torque is transmitted from them to D1. When only one motor is used, one of the motor locks
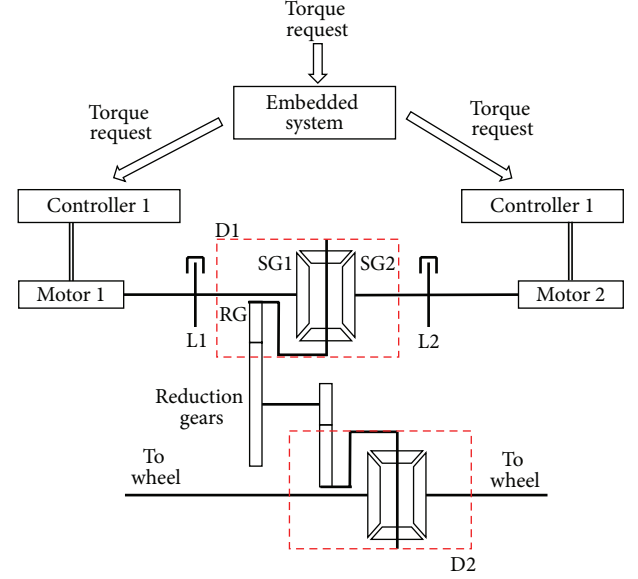
FIGURE 1: Diagram of the propulsion module that shows the internal component of the transmission (differentials and gear reduction) and the motors, controllers, and embedded system.

(labeled in the diagram as L1 or L2) is activated to prevent the rotation of the idle motor. The described powertrain was simulated to account for the power losses incurred in the electric motors. The control strategy objective is to provide the necessary power, considering the torque request and current speed, by choosing the operation mode (one or two motors) that provides less energy losses. The models used for the simulation are described in Section 3.

3. Component Modelling

The EV is modelled using the conventional longitudinal dynamics equations for loads caused by rolling resistance (F_{rr}), aerodynamic drag (F_{ad}), hill climbing (F_{hc}), and inertial forces (F_{la}) in

$$
\begin{aligned}
F_{rr} &= C_{rr} \cdot g \cdot m \cdot \cos\theta, \\
F_{ad} &= \frac{1}{2} \cdot \rho \cdot C_a \cdot A \cdot \left(\frac{dx}{dt}\right)^2, \\
F_{hc} &= g \cdot m \cdot \sin\theta, \\
F_{la} &= m \cdot \frac{d^2x}{dt^2}.
\end{aligned}
\tag{1}
$$

For the rolling resistance force, C_{rr}, g, m, and θ are the rolling resistance coefficient, gravity, vehicle mass, and slope angle, respectively. For the aerodynamic drag, ρ, C_a, A, x, and t are the air density, aerodynamic drag coefficient, frontal area, displacement, and time, respectively. These forces are added to obtain the traction force (F_{tr}) in

$$
F_{tr} = F_{rr} + F_{ad} + F_{hc} + F_{la}.
\tag{2}
$$

TABLE 1: Cases and conditions to obtain the output torque. The columns of L1 and L2 (locks 1 and 2) express whether the lock is activated (1) or deactivated (0).

Operation mode (OM)	Condition				Output torque (T_{D1})
	L1	L2	Driving link	Driven link	
1	0	1	SG1	RG	$T_{EM1} \cdot 2$
2	0	0	SG1, SG2	RG	$T_{EM1} \cdot (2 \cdot \omega_{EM1}/(\omega_{EM1} + \omega_{EM2})) + T_{EM2} \cdot (2 \cdot \omega_{EM2}/(\omega_{EM1} + \omega_{EM2}))$
3	0	0	SG2	SG1	0
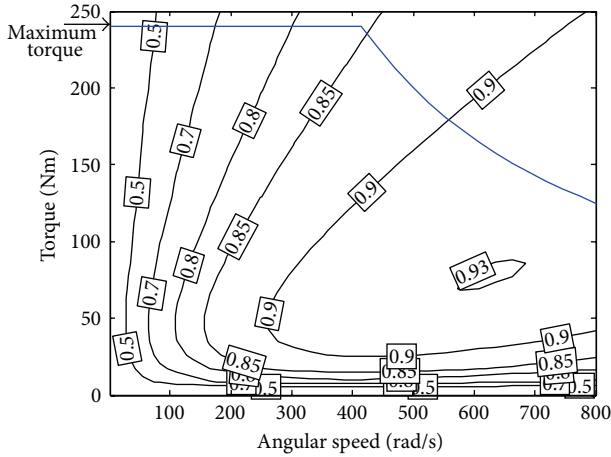
FIGURE 2: Efficiency map used to represent the power losses incurred in the electric motors.

The load torque transmitted to differential 1 (T_{tr}) passing through differential 2 and the gear reduction is presented in (3), where R_w is the wheels radius and G_r is the overall gear ratio:

$$T_{tr} = F_{tr} \cdot R_w \cdot G_r. \tag{3}$$

The model for the electric motor considers the power losses as a function of the torque and speed [13–16]. The efficiency of the motor (η_{EM}) is given by (4) in which T_{EM} and ω_{EM} are the torque and the angular speed of the electric motor. The parameters k_c, k_i, k_ω, and C take the values 0.3, 0.01, 0.000005, and 600, respectively, to simulate a 100 kW induction motor [13]. The efficiency of both electric motors (η_{EM1} and η_{EM2}) in this study is calculated using (4). Figure 2 represents the efficiency map of the motor in its operation range and the line of maximum torque. Consider

$$\eta_{EM}$$
$$= \frac{T_{EM} \cdot \omega_{EM}}{T_{EM} \cdot \omega_{EM} + k_c \cdot T_{EM}^2 + k_i \cdot \omega_{EM} + k_\omega \cdot \omega_{EM}^3 + C}. \tag{4}$$

In this case, the same characteristics for the two electric motors were used. The output angular speed of differential 1 (ω_{D1}) is a function of the speed of electric motors 1 and 2 (ω_{EM1} and ω_{EM2}) as shown in

$$\omega_{D1} = \frac{\omega_{EM1} + \omega_{EM2}}{2}. \tag{5}$$

The output torque of differential 1 (T_{D1}) is a function of the torque of the electric motors (T_{EM1} and T_{EM2}) as shown in Table 1. According to the table, *Operation Mode 1* (OM 1) corresponds to a situation in which lock 2 is actuated and, as a consequence, $\omega_{EM2} = 0$. All the power for the vehicle is provided by electric motor 1. In *Operation Mode 2* (OM 2), both locks are deactivated and the vehicle is powered by both of the electric motors. In *Operation Mode 3* (OM 3), both locks are deactivated but one of the motors is being driven by the other motor, and as a result $T_{D1} = 0$ is obtained. This mode cannot accelerate the vehicle but can be useful in regulating the speed of the electric motors. The three modes are subjected to the condition of (6) which represents the energy balance of the differential:

$$T_{D1} \cdot \omega_{D1} = T_{EM1} \cdot \omega_{EM1} + T_{EM2} \cdot \omega_{EM2}. \tag{6}$$

According to OM 3, a condition in which one of the motors would operate at low speeds (and, e.g., at low efficiencies according to the efficiency map) is avoided.

The diagrams of Figure 3 show the way in which the acceleration is handled in this propulsion system. Figure 3(a) shows the case in which the driver provides a torque request (τ), which then produces a torque request for each motor (τ_{EM1} and τ_{EM2}) and ultimately produces an output angular speed (ω_{D1}) which provides forward movement to the vehicle. In Figure 3(b) the driver is replaced by a PID controller which compares the obtained vehicle speed ($v(t)$) with a set point speed ($v_{SP}(t)$) defined by a driving cycle. The use of a driving cycle provides a predefined standard speed which can be used to produce comparisons in energy losses.

4. Control Strategy for the Propulsion System

The objective of the control strategy is to distribute the torque request signal to each electric motor. The control strategy starts working when a positive torque is requested to the powertrain to produce acceleration in the vehicle (when $\tau > 0$). According to the current speed and the requested torque, a decision is made to calculate the torque request signal for each motor. The inequalities in (7) show the possible values that can be assigned for each torque request signal:

$$0 < \tau < 1,$$
$$0 < \tau_{EM1} < 1, \tag{7}$$
$$0 < \tau_{EM2} < 1.$$

The boundaries of the control strategy consider the limits of the parameters of the electric motors. The electric

(a)

(b)
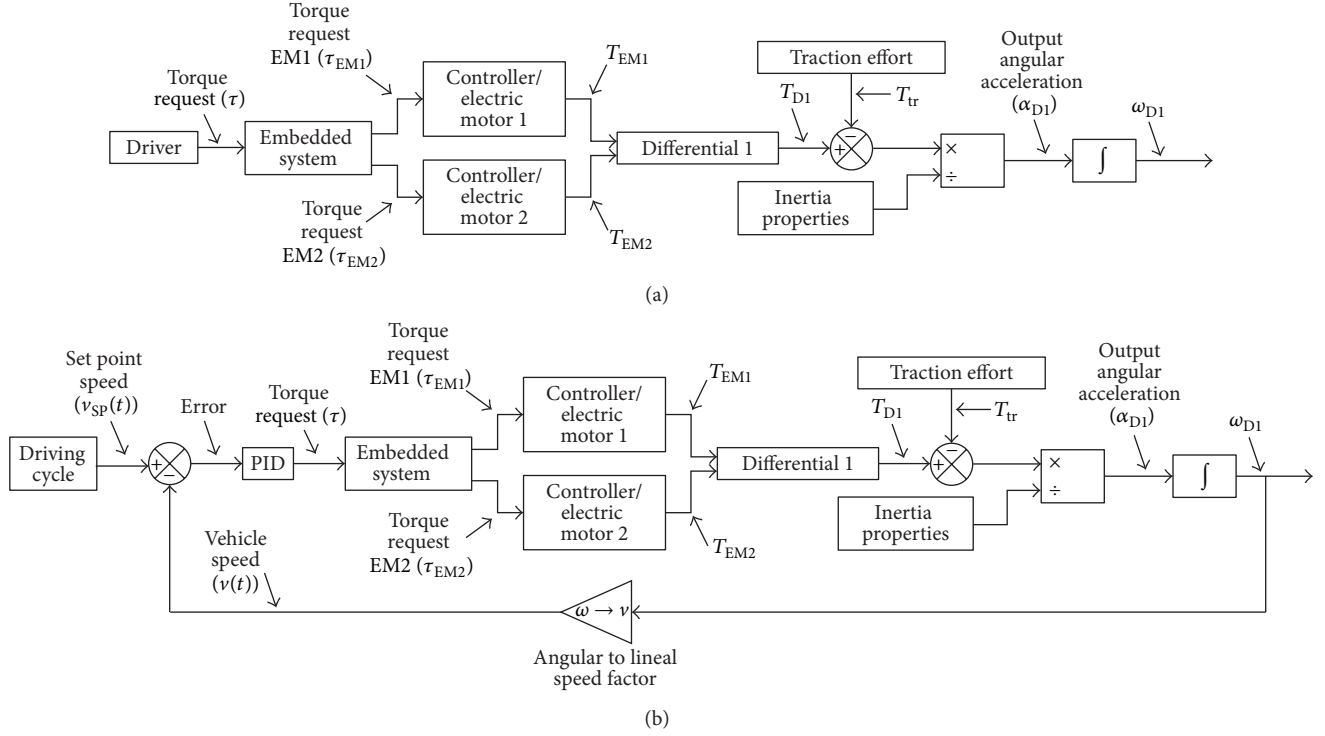
FIGURE 3: Block diagrams of (a) open loop control provided by a driver and (b) closed loop control to follow the speed of a driving cycle.

motor's maximum speed used for this study is 800 rad/s. The maximum achievable torque for the electric motors is shown in Figure 2 in the torque limited zone (0–415 rad/s) and the power limited zone (415 rad/s–800 rad/s).

From (5), the maximum output speed for differential 1 ($\omega_{D1|max}$) can be expressed according to (8) and (9) for OM 1 and OM 2 (previously defined in Table 1):

$$\text{OM 1: } \omega_{D1|max} = \frac{\omega_{EM1|max}}{2}, \tag{8}$$

$$\text{OM 2: } \omega_{D1|max} = \frac{\omega_{EM1|max} + \omega_{EM2|max}}{2}. \tag{9}$$

The implication of (8) and (9) is that OM 2 enables higher speeds (by a factor of two) than what is possible to achieve by using OM 1. The transition speed ($\omega_{1 \to 2}$) is the speed ω_{D1} at which the powertrain changes the operation from one motor (OM 1) to two motors (OM 2). The possible values for $\omega_{1 \to 2}$ are limited by (8) which result in

$$0 \leq \omega_{1 \to 2} \leq \frac{\omega_{EM1|max}}{2}. \tag{10}$$

The transition torque request ($\tau_{1 \to 2}$) is the torque request signal at which the powertrain changes the operation from one motor to two motors. The possible values for $\tau_{1 \to 2}$ can be in the range {0, 1}. The values chosen for $\omega_{1 \to 2}$ and $\tau_{1 \to 2}$ have an impact in the overall efficiency of operation of the powertrain and in the longitudinal dynamics of the vehicle, which is analyzed in the next section.

Figure 4 portrays the control logic in a flow diagram for the transitions between the previously defined operation

TABLE 2: Value of parameters for the simulation of the powertrain.

Symbol	Value
C_{rr}	0.007
ρ	1.2 kg/(m^3)
R_w	0.3 m
Motor maximum torque	240 Nm
m	1700 kg
C_a	0.2
G_r	1/7
Motor maximum speed	800 rad/s
θ	0°
A	2.11 m^2
g	9.81 m/(s^2)
Motor maximum power	100 kW

modes. According to the control logic, OM 3 is used to regulate the speeds of both motors until they are similar within certain range. For that reason OM 3 is used only to transition from OM 1 (motor 1 has speed but motor 2 is at rest) to OM 2 (both motors have speed). According to the control logic, when $\tau_{1 \to 2}$ and $\omega_{1 \to 2}$ are small, there are less chances for the powertrain to work on OM 1. On the contrary, when $\tau_{1 \to 2}$ and $\omega_{1 \to 2}$ are large, the manifestation of OM 3 and OM 2 is delayed.

5. Effect of $\omega_{1 \to 2}$ and $\tau_{1 \to 2}$ on Acceleration and Power Losses

To run simulations on the powertrain, the values of the parameters of the vehicle are shown in Table 2.
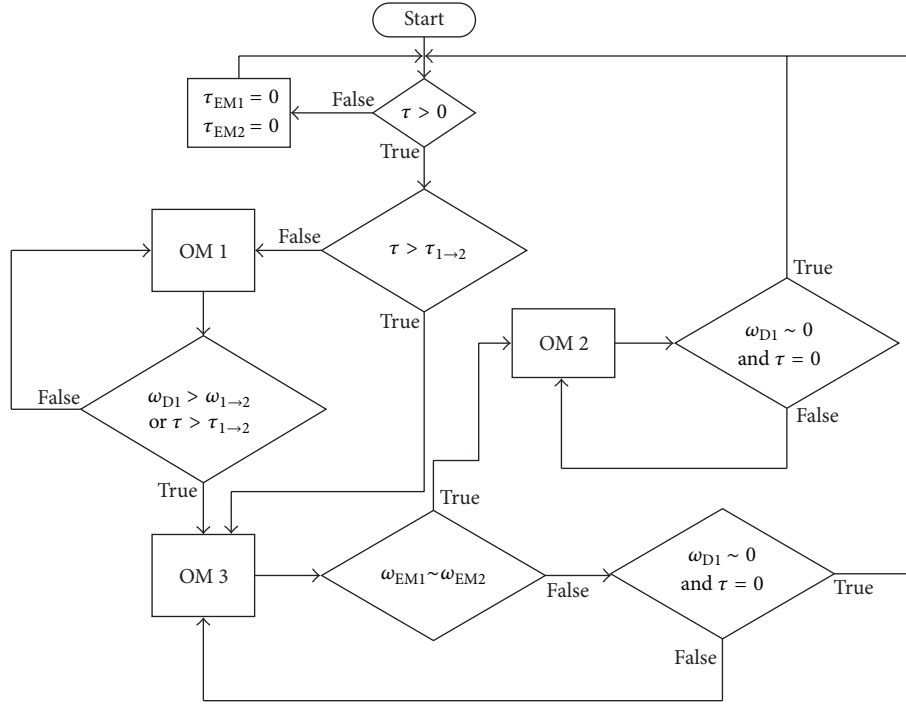
FIGURE 4: Flow diagram of the control strategy programmed in the embedded system.

(a) Vehicle speed

(b) Speed of electric motor 1

(c) Speed of electric motor 2
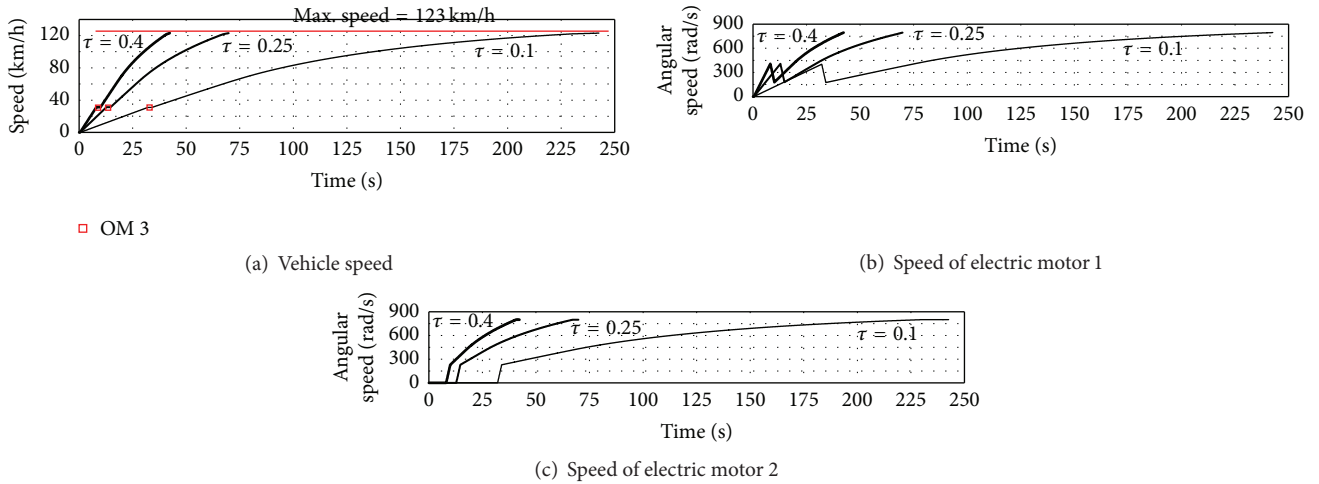
FIGURE 5: Speed of the vehicle (a), electric motor 1 (b), and electric motor 2 (c) for $\tau_{1 \to 2} = 0.5$ and $\omega_{1 \to 2} = 200$ rad/s.

Figure 5(a) shows a simulation of the vehicle speed with $\tau_{1 \to 2} = 0.5$ and $\omega_{1 \to 2} = 200$ rad/s. The three curves shown correspond to the torque requests $\tau = \{0.1, 0.25, 0.4\}$. The powertrain starts accelerating in OM 1, then it transitions through a brief period of time in OM 3 (lasting 1.7 s), and then it follows in OM 2. Figures 5(b) and 5(c) show the rotational speed of electric motors 1 and 2, respectively. The figures show that, in OM 1, only electric motor 1 provides speed (ω_{EM1}) for the powertrain. In OM 3, ω_{EM1} decreases and ω_{EM2} increases until they have the same speed. After that, OM 2 begins to operate.

Figure 6(a) shows the vehicle speed when $\tau = 0.25$ for two different values of $\tau_{1 \to 2}$. When $\tau_{1 \to 2} = 0.15$ the acceleration is handled by the two motors (OM 2). When $\tau_{1 \to 2} = 0.35$ the acceleration is first handled by only one motor (OM 1) and, after the transition speed ($\omega_{1 \to 2} = 200$ rad/s), the powertrain passes through OM 3 to OM 2. According to the figure, when the powertrain is using one motor only, the vehicle achieves slightly higher acceleration; however, during OM 3, the vehicle undergoes a period of coasting before the powertrain (in OM 2) accelerates the vehicle again. Figures 6(b) and 6(c) show ω_{EM1} and ω_{EM2}, respectively. The plots

(a) Vehicle speed

(b) Speed of electric motor 1

(c) Speed of electric motor 2
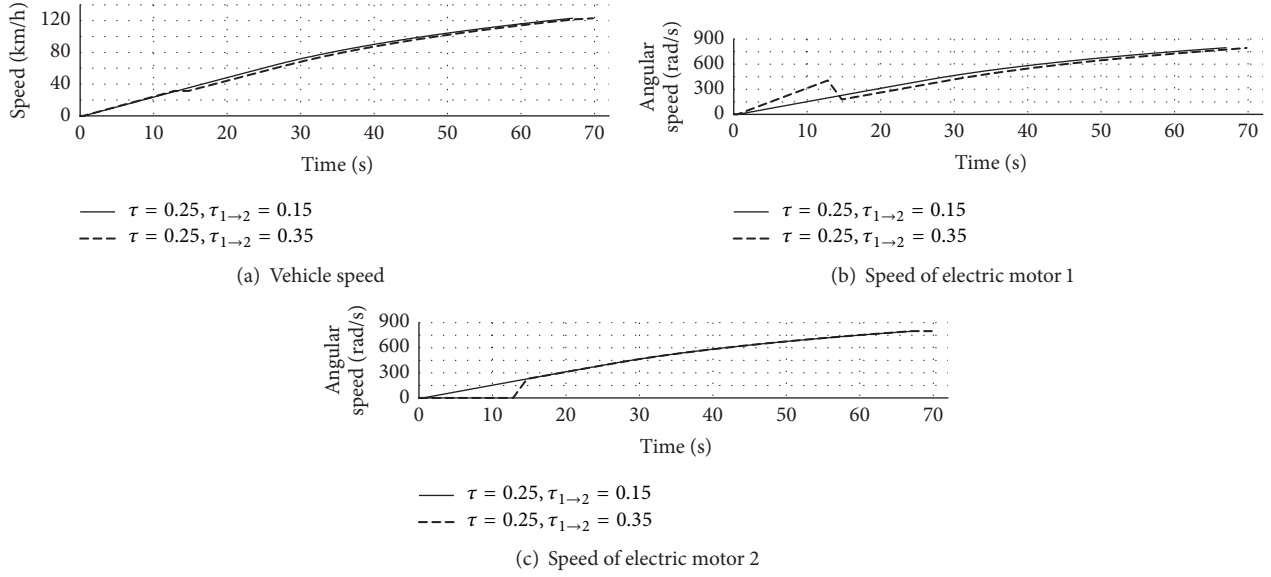
FIGURE 6: Effect of the transition torque request on the vehicle speed (a) and on the rotational speed of electric motors 1 (b) and 2 (c).

(a) Simulated vehicle speed following the NEDC driving cycle

(b) Speed of electric motors
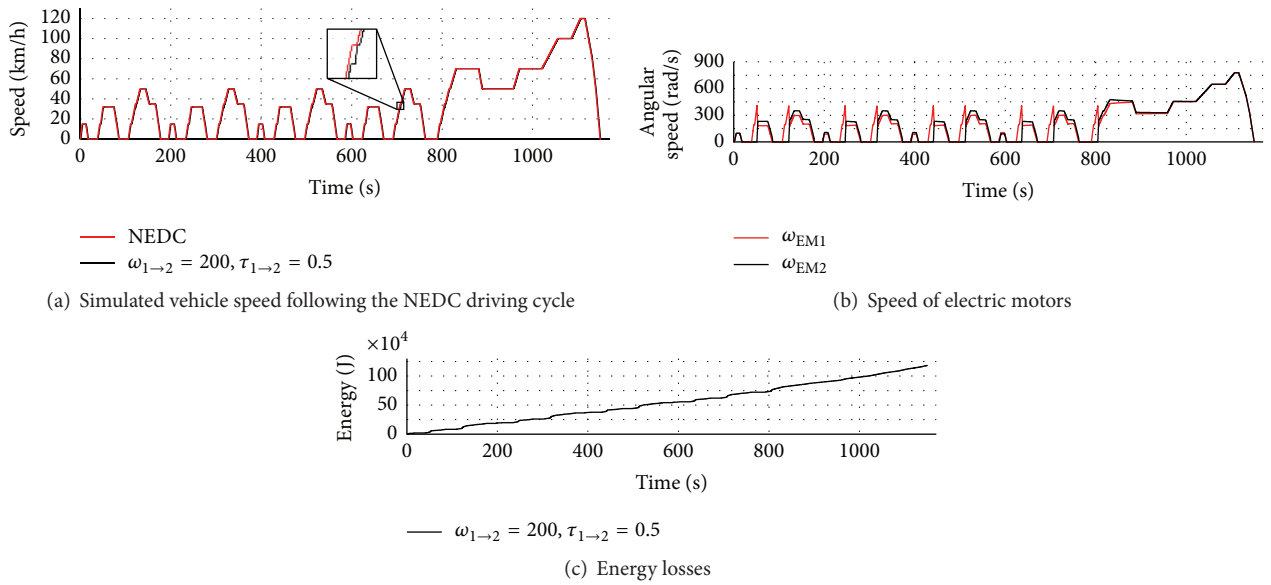
(c) Energy losses

FIGURE 7: Simulated vehicle speed and the speed of the NEDC driving cycle (a) and the rotational speed of the electric motors (b). The energy losses incurred are shown in (c).

corresponding to $\tau_{1 \to 2} = 0.35$ show ω_{EM1} running at almost 400 rad/s just before the speed drops as a consequence of the start of OM 3. At that moment, ω_{EM2} starts to rise until the two speeds are the same. The plots corresponding to $\tau_{1 \to 2} = 0.15$ show the powertrain operating in OM 2 all the time.

According to Figures 5 and 6, OM 3 produced a time in which the vehicle is not accelerated. Drivability, which "describes the driver's complex subjective perception of the interactions between driver and vehicle associated with the longitudinal acceleration aspects" [17], is affected by the implementation of OM 3. Next section continues the analysis of the effects of $\omega_{1 \to 2}$ and $\tau_{1 \to 2}$ on the energy losses and the drivability by using driving cycles.

5.1. Driving Cycle Simulation of the Control Strategy. The torque request (τ) necessary to follow the driving cycle was obtained using a PID. Figure 7(a) illustrates the New European Driving Cycle (NEDC) and the speed followed by the simulated vehicle using $\omega_{1 \to 2}$ and $\tau_{1 \to 2}$ with values of 200 rad/s and 0.5, respectively. The control strategy follows closely the NEDC speed; however when OM 3 is present, an error occurred (the error was magnified in the figure for illustration purposes). The speed error ($e_v(t)$) can be expressed as shown in

$$e_v(t) = v_{\text{SP}}(t) - v(t). \tag{11}$$

(a) Contour plot of energy losses (J) for NEDC

(b) Contour plot of $e_{v|MAX}$(km/h) for NEDC
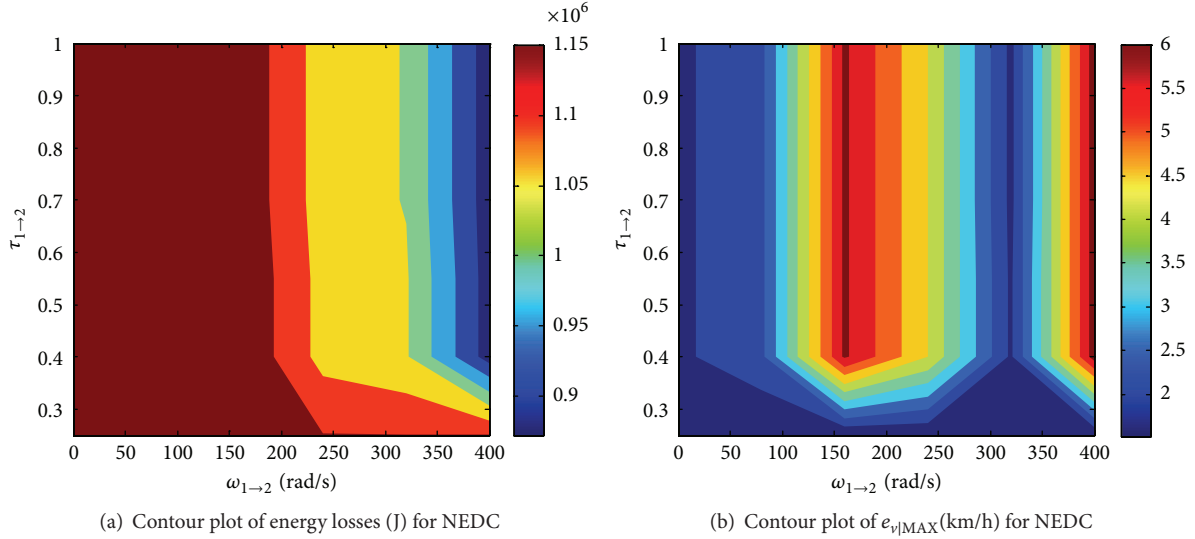
FIGURE 8: Contour plots of energy losses (a) and maximum $e_{v|MAX}$ (b) with $\omega_{1\to2}$ in the x-axis and $\tau_{1\to2}$ in the y-axis for the NEDC.

(a) Vehicle speed following the UDDS

(b) Vehicle speed following the NYCC
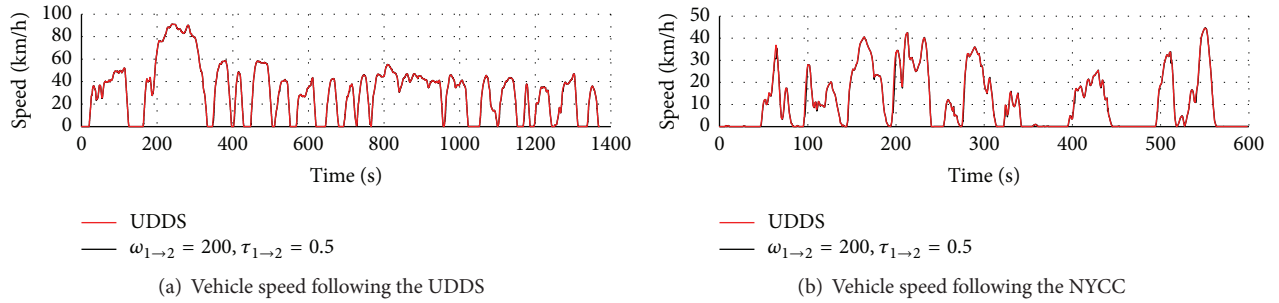
FIGURE 9: PID controller vehicle speed to follow the UDDS (a) and the NYCC (b).

In this research, the maximum $e_v(t)$ produced in the driving cycle ($e_{v|MAX}$) is used as an indicator of drivability. An improved drivability is obtained when $e_{v|MAX}$ is maintained small. Figure 7(b) shows the speed of the electric motors while being controlled to produce the speed of the NEDC. Figure 7(c) shows the accumulated energy losses incurred by the simulated vehicle while following the driving cycle according to the model explained in Section 3.

The effect of $\omega_{1\to2}$ and $\tau_{1\to2}$ on the energy losses while the vehicle follows NEDC is illustrated in Figure 8(a). According to the figure, the energy losses increase when the values of $\omega_{1\to2}$ and $\tau_{1\to2}$ decrease. This situation corresponds to an early switch from OM 1 to OM 2 or when the driving cycle is handled almost completely in OM 2. The maximum energy losses were found when $\omega_{1\to2}$ has a value of 80 rad/s. The maximum energy losses found in this study at the end of the NEDC were of 1,199.753 kJ. The minimum energy losses were found when $\omega_{1\to2}$ has a value of 400 rad/s and $\tau_{1\to2} >$ 0.70. The minimum energy losses found in this study at the end of the NEDC were of 870.823 kJ.

The effect of $\omega_{1\to2}$ and $\tau_{1\to2}$ on $e_{v|MAX}$ while the vehicle follows NEDC is illustrated in Figure 8(b). According to the figure, $e_{v|MAX}$ decreases when the values of $\omega_{1\to2}$ and $\tau_{1\to2}$ decrease. This situation corresponds to a better drivability,

given that the driving cycle is handled almost completely in OM 2 (OM 3 did not appear in the entire driving cycle or appears at small values of speed). The best drivability was found when $\omega_{1\to2}$ has a value of 0 rad/s and $\tau_{1\to2} <$ 0.25, resulting in $e_{v|MAX}$ of 1.52 km/h. The worst drivability was found when $\omega_{1\to2}$ = 400 rad/s and $\tau_{1\to2} >$ 0.4, with a value of $e_{v|MAX}$ = 6.27 km/h. The plot reveals also a second worst case of drivability when $\omega_{1\to2}$ = 160 rad/s and $\tau_{1\to2} >$ 0.25, with a value of $e_{v|MAX}$ = 6.08 km/h.

The Urban Dynamometer Driving Schedule (UDDS) and the New York City Cycle (NYCC) are driving cycles that present higher peak accelerations and lower speeds when compared with the NEDC [18]. The effect of the proposed control strategy under those driving cycles in terms of energy losses and drivability ($e_{v|MAX}$) was also analyzed. The UDDS and the NYCC are shown in Figures 9(a) and 9(b), respectively, along with the speed profile produced by the control strategy applied in the simulated EV. The figure shows that the maximum speed achieved by the vehicle in the UDDS is 91.25 km/h, while for the NYCC it is of 44.6 km/h.

The effect of $\omega_{1\to2}$ and $\tau_{1\to2}$ on UDDS and NYCC is presented in Figure 10. For the UDDS the contour plot for the energy losses (Figure 10(a)) looks similar to the contour plot obtained for the NEDC (Figure 8(a)). The maximum

(a) Contour plot of energy losses (J) for UDDS

(b) Contour plot of $e_{v|MAX}$ (km/h) for UDDS

(c) Contour plot of energy losses (J) for NYCC

(d) Contour plot of $e_{v|MAX}$ (km/h) for NYCC
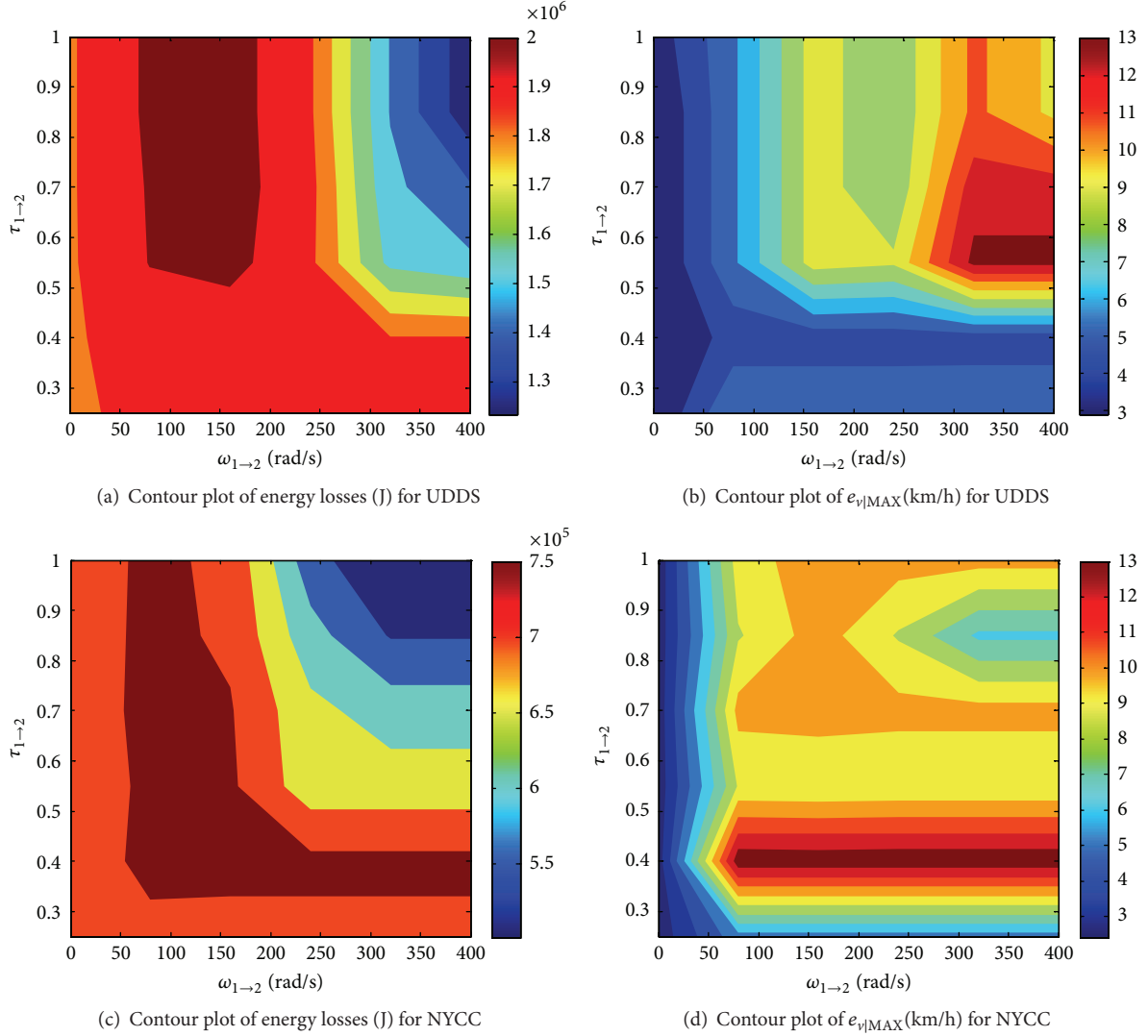
FIGURE 10: Contour plots for energy losses and $e_{v|MAX}$ for the driving cycles of UDDS and NYCC.

energy losses were found when $\omega_{1 \rightarrow 2}$ is less than 240 rad/s (2,043.404 kJ). The minimum energy losses were found when $\omega_{1 \rightarrow 2}$ has a value of 400 rad/s and $\tau_{1 \rightarrow 2} > 0.85$ (1,233.674 kJ). For drivability (Figure 10(b)), the results of the UDDS are similar to those of NEDC. The best drivability was found when $\omega_{1 \rightarrow 2}$ has a value of 0 rad/s and $\tau_{1 \rightarrow 2} < 0.85$ ($e_{v|MAX} = 2.89$ km/h). The worst drivability was found when $\omega_{1 \rightarrow 2}$ has a value of 400 rad/s and $\tau_{1 \rightarrow 2} = 0.55$ ($e_{v|MAX} = 13.29$ km/h).

For the NYCC, according to Figure 10(c), the energy losses follow the same pattern as NEDC and UDDS, increasing in the zone of low $\omega_{1 \rightarrow 2}$ and $\tau_{1 \rightarrow 2}$ values (765.012 kJ) and decreasing when $\omega_{1 \rightarrow 2}$ and $\tau_{1 \rightarrow 2}$ are high (501.146 kJ). For the drivability, the NYYC presented the potential worst case of drivability between the analyzed driving cycles with $e_{v|MAX} = 13.72$ km/h found when $\omega_{1 \rightarrow 2}$ has a value of more than 80 rad/s and $\tau_{1 \rightarrow 2} = 0.4$. The best drivability was found in the same zone as the UDDS and NEDC, when $\omega_{1 \rightarrow 2}$ and $\tau_{1 \rightarrow 2}$ have low values ($e_{v|MAX} = 2.40$ km/h).

5.2. Discussion. The energy losses in all modes were reduced when a high transition speed ($\omega_{1 \rightarrow 2}$) was set in combination with a high transition torque request signal ($\tau_{1 \rightarrow 2}$). This indicates that energy losses were reduced when the driving cycle is handled most of the time in OM 1, with only one electric motor. To handle the whole driving cycle in OM 1 is only possible for the NYCC because of its low maximum speed. For the UDDS and mainly for NEDC, the switching from OM 1 to OM 2 was necessary to achieve the speeds above 61.7 km/h, which correspond to the electric motor maximum speed (of 800 rad/s) in OM 1.

The differences in energy losses can be explained by analyzing the efficiency map of the motors and the zones in which the operation points are located under different values of $\omega_{1 \rightarrow 2}$ and $\tau_{1 \rightarrow 2}$. Figure 11 illustrates this by showing the efficiency map of the motor with the operation points produced superposed, when NEDC is followed. Figure 11(a) uses the parameters $\omega_{1 \rightarrow 2}$ and $\tau_{1 \rightarrow 2}$ with values of 400 rad/s
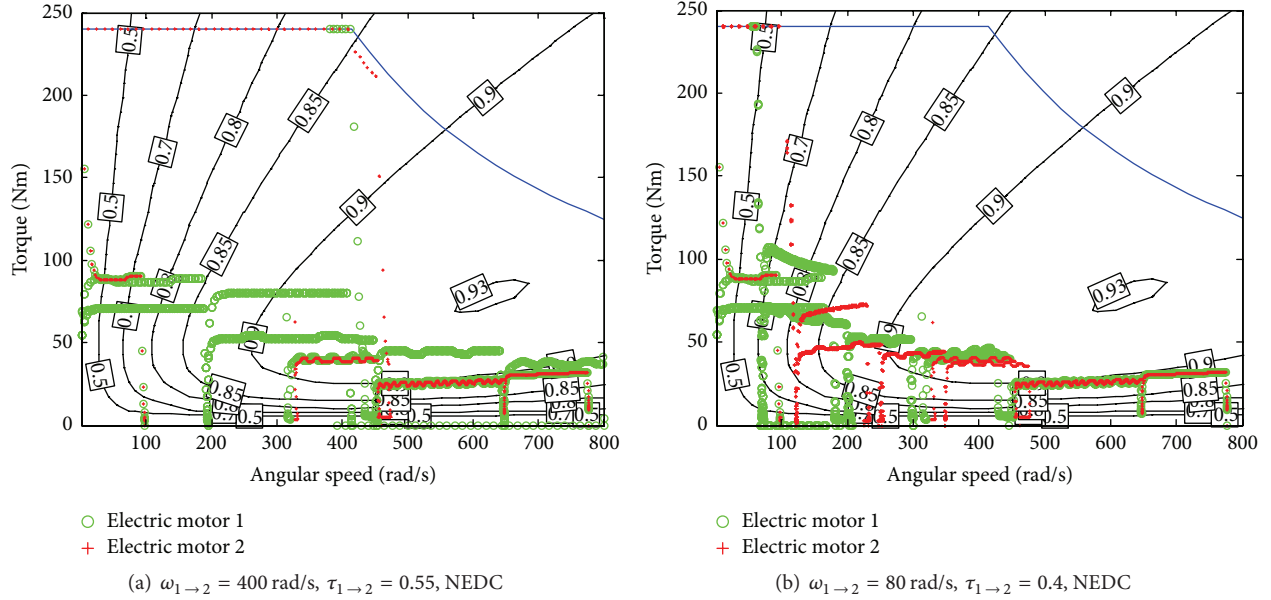
FIGURE 11: Operation points of the two electric motors for the NEDC, superposed in the electric motor efficiency map.

and 0.55, respectively. Figure 11(b) shows the operation points produced when $\omega_{1 \rightarrow 2} = 80$ and $\tau_{1 \rightarrow 2} = 0.4$. Comparing the figures, (a) has more operation points in the high speed zone than (b). The high speed zone in the efficiency map corresponds to the higher efficiency, while zones of low speed present also low efficiencies. This produced a difference in energy losses of 318.617 kJ less between the operation points of (a) with respect to (b). Another observation that can be drawn from Figure 11 is the low utilization of the torque range by the NEDC. Driving cycles with higher accelerations require higher torques. Also, variations in the design of the powertrain can include different transmission gear ratios (which modify the torque requirements in the motors) and different sizes of electric motors to get a better fit of the operation points.

In this work, two electric motors of the same characteristics were used. The effect of having two electric motors with different efficiency maps or size would be that the control strategy would have to account for two types of OM 1, depending on which motor is actuated and which motor is locked. Each motor could offer efficiency advantages in certain conditions of torque and speed, which would add flexibility to the system. On the other hand, two motors of the same characteristics simplify the control strategy and allow fault tolerant traction in the case of failure of any of the electric motors: independently of the faulty motor, any of the motors can continue operating in OM 1 providing the vehicle with the same output torque (however with the limitation in maximum vehicle speed that appears with OM 1).

According to the results, high $\omega_{1 \rightarrow 2}$ and high $\tau_{1 \rightarrow 2}$ combination produces less energy losses but also produces the highest $e_{v|\text{MAX}}$ for the driving cycles NEDC, UDDS, and NYCC.

6. Conclusions

A control strategy for a dual motor propulsion system for EVs is established in this work. The control strategy receives the torque demand and distributes it to the two electric motors to provide the desired acceleration and speed. The objective of the control strategy is to allow the desired acceleration while maintaining low energy losses.

The proposed control strategy is based on rules, with feasibility to be implemented in a real time control system. The control strategy inputs are the current speed of the electric motors and the torque demand signal (τ). The control strategy can be tuned using two parameters: the transition speed ($\omega_{1 \rightarrow 2}$) and the transition torque request signal ($\tau_{1 \rightarrow 2}$). These parameters define when to switch from the use of one motor only (which is identified as OM 1) to two motors (identified as OM 2). The switching produces a transition mode (OM 3), which produces a brief period of coasting of about 1.7 s in this work. This period of coasting can affect drivability by producing a difference between the desired speed and the actual speed. This error was calculated and analyzed.

The effect of $\omega_{1 \rightarrow 2}$ and the transition torque request signal $\tau_{1 \rightarrow 2}$ on energy losses and drivability was studied by running simulations of the speed and energy losses produced by the electric motors in an EV. The simulations made use of three of the most popular driving cycles: NEDC, UDDS, and NYCC. The results show that $\omega_{1 \rightarrow 2}$ and $\tau_{1 \rightarrow 2}$ have a significant effect on the energy losses and drivability. According to the results, the combination of high $\omega_{1 \rightarrow 2}$ and high $\tau_{1 \rightarrow 2}$ provides the best results overall in energy savings in the analyzed driving cycles. The drivability as studied here, using the indicator $e_{v|\text{MAX}}$, is affected by driving cycles with

high accelerations and also when the shifting between modes is produced at higher speeds.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] E. Helmers and P. Marx, "Electric cars: technical characteristics and environmental impacts," *Environmental Sciences Europe*, vol. 24, no. 4, article 14, 2012.

[2] J.-S. Lee, S. T. Kim, R. Cao et al., "Metal-air batteries with high energy density: Li-air versus Zn-air," *Advanced Energy Materials*, vol. 1, no. 1, pp. 34–50, 2011.

[3] A. F. Burke, "Batteries and ultracapacitors for electric, hybrid, and fuel cell vehicles," *Proceedings of the IEEE*, vol. 95, no. 4, pp. 806–820, 2007.

[4] K. Yoshida, S. Tomonari, H. Yoshioka, S. Tanaka, D. Satoh, and M. Esashi, "High energy density miniature electrical and thermal power source using catalytic combustion of butane," in *Proceedings of the 17th IEEE International Conference on Micro Electro Mechanical Systems (MEMS '04)*, pp. 316–321, January 2004.

[5] J. Meschke, M. Kurz, C. Schoenwald, A. Tornow, R. Hackl, and M. Flume, "Advanced lightweight design for electric vehicles— ALIVE," in *Proceedings of the Transport Research Arena on the 5th Conference: Transport Solutions from Research to Deployment (TRA '14)*, Paris, France, April 2014.

[6] A. Khaligh and Z. Li, "Battery, ultracapacitor, fuel cell, and hybrid energy storage systems for electric, hybrid electric, fuel cell, and plug-in hybrid electric vehicles: state of the art," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2806–2814, 2010.

[7] J. De Santiago, H. Bernhoff, B. Ekergård et al., "Electrical motor drivelines in commercial all-electric vehicles: a review," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 475–484, 2012.

[8] S. Zhang, C. Zhang, G. Han, and Q. Wang, "Optimal control strategy design based on dynamic programming for a dual-motor coupling-propulsion system," *The Scientific World Journal*, vol. 2014, Article ID 958239, 9 pages, 2014.

[9] P. D. U. Coronado and H. Ahuett-Garza, "Analysis of energy efficiency and driving range of electric vehicles equipped with a bimotor architecture propulsion system," *International Journal of Electric and Hybrid Vehicles*, vol. 6, no. 2, pp. 152–177, 2014.

[10] Y. Wang and D. Sun, "Powertrain matching and optimization of dual-motor hybrid driving system for electric vehicle based on quantum genetic intelligent algorithm," *Discrete Dynamics in Nature and Society*, vol. 2014, Article ID 956521, 11 pages, 2014.

[11] R. M. Patil, Z. Filipi, and H. K. Fathy, "Comparison of supervisory control strategies for series plug-in hybrid electric vehicle powertrains through dynamic programming," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 2, pp. 502–509, 2014.

[12] E. Dinçmen and B. A. Güvenç, "A control strategy for parallel hybrid electric vehicles based on extremum seeking," *Vehicle System Dynamics*, vol. 50, no. 2, pp. 199–227, 2012.

[13] J. Larminie and J. Lowry, *Electric Vehicle Technology Explained*, John Wiley & Sons, New York, NY, USA, 2003.

[14] A. Elmarakbi and A. Morris, "Modelling and analyzing electric vehicles with geared transmission systems: enhancement of energy consumption and performance," *International Journal of Engineering Research & Technology*, vol. 2, pp. 1215–1254, 2013.

[15] A. Elmarakbi, Q. Ren, R. Trimble, and M. Elkady, "Performance analysis of hybrid and full electrical vehicles equipped with continuously variable transmissions," *Advances in Automobile Engineering*, vol. 2, article 103, 2013.

[16] K. Sedef, A. Maheri, A. Daadbin, and M. Yilmaz, "A comparative study of the performance of DC permanent magnet and AC induction motors in urban electric cars," in *Proceedings of the 2nd International Symposium on Environment Friendly Energies and Applications (EFEA '12)*, pp. 100–105, Newcastle upon Tyne, UK, June 2012.

[17] L. Puhui, X. Tong, and Z. Xuanbai, "Vehicle drivability evaluation and pedal-acceleration response analysis," *International Journal on Advances in Information Sciences and Service Sciences*, vol. 5, no. 10, pp. 506–513, 2013.

[18] B. M. Geller, T. H. Bradley, and B. M. Geller, "Analyzing drive cycles for hybrid electric vehicle simulation and optimization," *Journal of Mechanical Design*, vol. 137, no. 4, Article ID 041401, 14 pages, 2015.

Research Article

A High Power Density Integrated Charger for Electric Vehicles with Active Ripple Compensation

Liwen Pan[1,2] and Chengning Zhang[1]

[1]*School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China*
[2]*Department of Electrical and Electronic Engineering, University of Bristol, Bristol BS8 1TH, UK*

Correspondence should be addressed to Liwen Pan; panliwen2002@163.com

This paper suggests a high power density on-board integrated charger with active ripple compensation circuit for electric vehicles. To obtain a high power density and high efficiency, silicon carbide devices are reported to meet the requirement of high-switching-frequency operation. An integrated bidirectional converter is proposed to function as AC/DC battery charger and to transfer energy between battery pack and motor drive of the traction system. In addition, the conventional H-bridge circuit suffers from ripple power pulsating at second-order line frequency, and a scheme of active ripple compensation circuit has been explored to solve this second-order ripple problem, in which a pair of power switches shared traction mode, a ripple energy storage capacitor, and an energy transfer inductor. Simulation results in MATLAB/Simulink validated the eligibility of the proposed topology. The integrated charger can work as a 70 kW motor drive circuit or a converter with an active ripple compensation circuit for 3 kW charging the battery. The impact of the proposed topology and control strategy on the integrated charger power losses, efficiency, power density, and thermal performance has also been analysed and simulated.

1. Introduction

It is a significant strategy on a global scale to replace fossil fuels vehicles with electric vehicles (EVs) for protecting the environment and achieving energy sustainability [1]. EVs have gained wide attention from the past years as one of the effective solutions for environment deterioration and energy shortages. There are three barrier issues for gaining tremendous acceptance for EVs, which include the high cost and cycle life of batteries, the lock of charging infrastructure, and integrations of chargers. Integrated charger can avoid these problems by integrating with electric drive and battery charger. The main advantages of integrated charging methods are that the weight, volume, and cost are reduced. The configuration of a conventional EV is shown in Figure 1(a). However, the components in the traction circuit, like the inverter, are not used during the battery charging, so it is possible to use it in the charger circuit. The typical structure of an integrated charger is shown in Figure 1(b).

Several organizations such as Society of Automotive Engineers (SAE) have supplied the utility interface for EV conductive charge coupler. As shown in Table 1, most EV charging can be installed at home where the EV can be connected with a convenience household outlet for Level 1 charging. Level 1 charging requires a 120 V or 230 V outlet. Usually single-phase converter is used for that solution. A bidirectional on-board charger needs a highly efficient AC-DC converter that boasts a high power density and fits the limited space and weight requirements.

Battery charger plays an important role in the development of EVs. In most of traditional distributed EVs electrical systems, motor drive circuit and battery charger circuit are separate, so two independent circuits are needed. However, the integrated charger supplies flexibility for layout space, weight, and cost for EVs to obtain high efficiency and higher power density [2–4]. Different types of integrated charger topologies design have been reported and explained in previous papers [5–8]. Similar topologies are introduced in [9–11], without further power losses analysis, efficiency, and thermal stress issues. Besides these, an integrated charger for plug-in electric vehicles based on a special interior permanent magnet motor is introduced. An interior permanent magnet

TABLE 1: Charging power levels (based in part on [37]).

Power level types	Level 1	Level 2	Level 3
Supply voltage	120 V$_{AC}$ (US) 230 V$_{AC}$ (EU)	240 V$_{AC}$ (US) 400 V$_{AC}$ (EU)	208–600 V$_{AC}$ or V$_{DC}$
Power level	≤3.7 kW	3.7–22 kW	>50 kW
Charging time	11–36 hours	1–6 hours	0.2–1 hours
Charger location	On-board 1-phase	On-board 1-phase or 3-phase	Off-board 3-phase
Energy supply interface	Convenience household outlet	Dedicated EV supply equipment	Dedicated EV supply equipment
Battery capacity	16–50 kWh	16–50 kWh	16–50 kWh
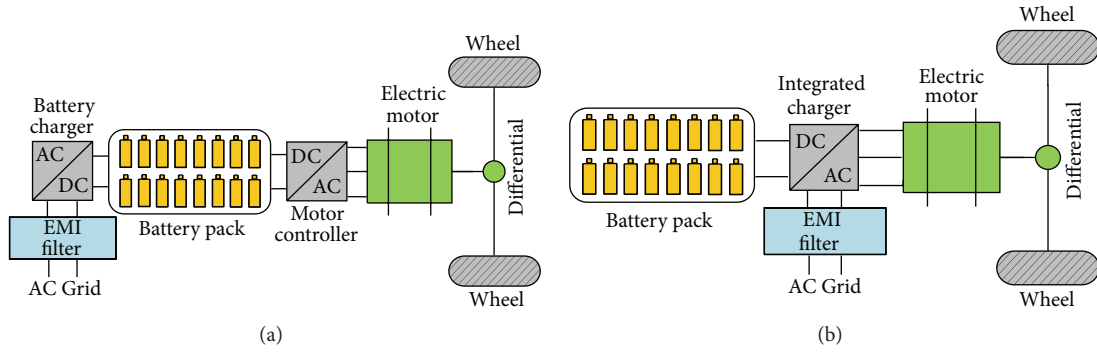
FIGURE 1: Configuration of electric vehicle: (a) configuration of a conventional electric vehicle and (b) configuration of an integrated charger electric vehicle.
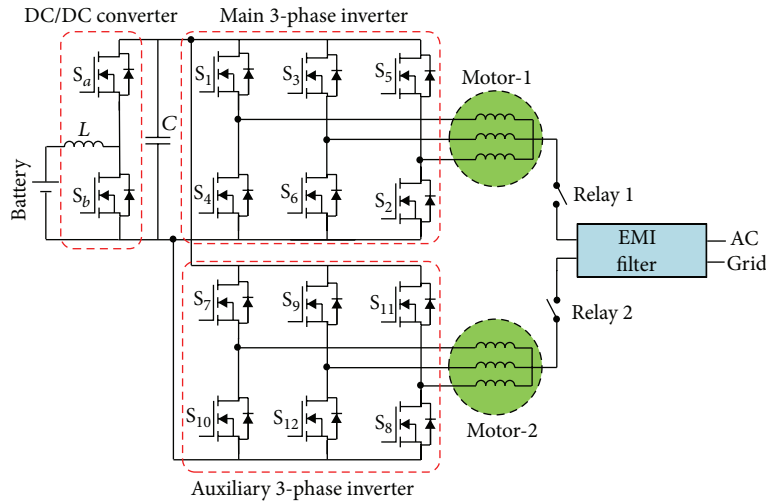
FIGURE 2: Integrated charger with two motors and two inverters.

synchronous is designed with a special winding configuration for traction mode and charging mode [12]. For integrated fast battery charger, a fast on-board battery charger using the motor like filter and the same converter for traction and charging mode is presented as well [13].

The topology as shown in Figure 2 is used for plug-in hybrid vehicles. The two three-phase AC motors are used as inductors for the converter with the neutral points connected to the grid. In traction mode, motor-1 delivers energy as a traction motor while motor-2 is used to charge the battery.

In the charging mode, both of the motors and converters operate as AC-DC boost converter to charge the battery [14]. However, this topology needs two motors or special double winding motor and converters to make the system complex.

A permanent magnet nonisolated integrated charger topology is proposed in Figure 3. The AC three-phase grid is connected to each winding neutral point of the motor. The magnetic motive force is cancelled on the stator. The magnetic decoupling between the stator and the rotor prevent the rotor from vibrating during charging mode. No rotation
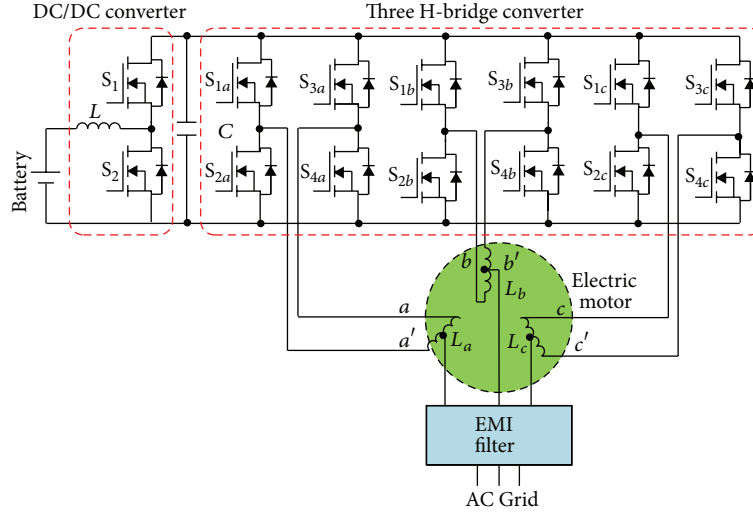
FIGURE 3: Integrated charger topology connecting to PM motor winding as filter.
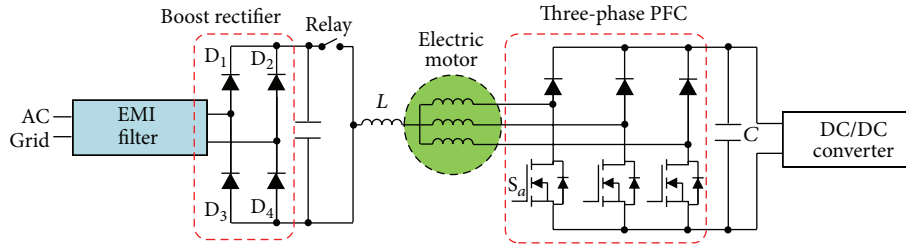
FIGURE 4: Integrated charger with three-phase PFC.

is produced during the charging mode [15]. However, this topology needs to use PMSM windings as coupling filter. It is not proper to use if the motor winding inductance is less than the filter requirement.

An integrated charger is shown in Figure 4. An interior permanent magnet motor drive circuit is operated as a three-phase PFC circuit during charging mode. No additional filter is needed except for electric motor windings. The disadvantage of this topology is a single-phase diode bridge rectifier that is used as the battery charger [16]. Therefore, a large buck capacitor is needed to divert the ripple power from DC-link. The aluminium electrolytic capacitor offers low cost and a high energy density. However, this type of capacitors has a short lifetime which is unacceptable in EVs application. Film capacitor is more reliable than electrolytic capacitor for electric vehicles, which results in low power density. This inevitably leads to an increased system cost and degraded energy efficiency. Various active ripple compensation methods are proposed to absorb the low-frequency ripple energy. An active filter is used to divert the ripple power from the DC-link in existing methods [17].

High efficiency and high power density are expected to achieve for integrated charger design. The converter power density is evaluated by measuring the volume of power modules, cooling system, line inductor, and DC-link capacitor [18] as shown in Figure 5. In addition, the power density is required to constraint on space and weight of the overall
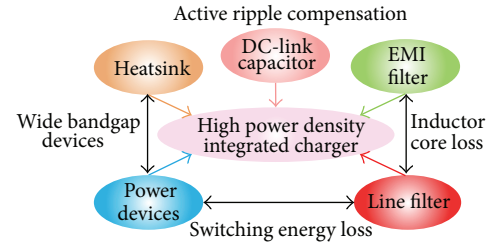
FIGURE 5: High power density converter design.

system. High power density system is considered by the use of new materials switching devices, increase in integrated levels, and design of innovative circuit to reduce the size of DC-link capacitor in this paper.

Conventional converter based on silicon (Si) devices typically operate at lower frequency contrasting with silicon carbide (SiC) MOSFETs. SiC switching devices are developing rapidly in the market in recent years. This supplies opportunities for smaller converter optimization design. High switching frequency can directly affect the size of converter's heatsink, DC-link capacitor, line filter, and EMI filter [19]. In order to optimise the integrated charger efficiency and power density, the SiC switching devices are used as the switches in this paper.

In this paper, a concept of high power density integrated charger for electric vehicles with active ripple compensation
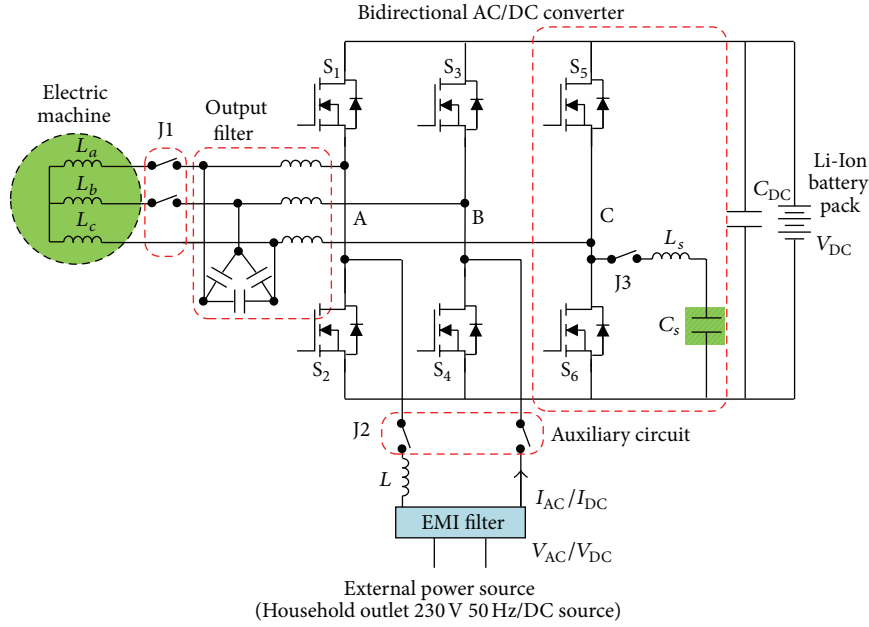
FIGURE 6: Topology of the integrated charger for electric vehicles.

is reported. The paper is organized as follows: the existing onboard battery charger in EV/HEV applications is introduced in Section 1. The topology of a high power density integrated charger with active ripple compensation is explained in Section 2. The design criteria of circuit components are described in Section 3. Section 4 presents an active ripple energy storage control method. Moreover, the main energy storage component inductor method and capacitor method are theoretically analysed and simulated. A capacitor auxiliary energy storage circuit is designed in details. The simulation results in Section 5 show that the design and control strategy meet the demand. The integrated charger system power losses and thermal performance are discussed in Section 5. Finally, the conclusion is given in Section 6.

2. Concept of Integrated Charger with Active Ripple Compensation

The concept of a high power density integrated charger for electric vehicles with active ripple compensation is illustrated in Figure 6. The switches and passive components are shared and multiused for traction mode and charging mode. In traction mode, the grid is not connected to the converter by turning off the relays between grid and converter. The grid current flows in the battery through the H-bridge converter during the battery charging. By this way, the electrical motor cannot rotate, because the relays beside the motor are open in the driving mode [20].

2.1. Topology. A schematic of the proposed integrated charger with active ripple compensation converter is shown in Figure 6. It includes an electric motor, a battery pack, an LC output filter, a DC-link capacitor, and an active ripple

reduction circuit. Relays are added to the circuit to achieve different operations modes [21]. The motor drive system can provide the required drive torque and battery charging in AC power and DC power operational modes. The DC power can be converted into AC power by a three-phase inverter to charge the battery. The H-bridge rectifier can support both AC and DC charging. The power flow between the motor, the battery, and power source is shown in Figure 7.

2.1.1. Motor Drive Modes. In the motor drive mode, relay J1 is switched on for propulsion, and relay J2 and relay J3 are open. The EV is operating at electric propelling mode, and its power flows from the battery pack to the motor. Figure 6 shows the conventional three-phase voltage-source converter structure. The three-phase rectifier circuit consists of three legs A, B, and C. The output of each leg depends on DC-link voltage and the switch status, and the output voltage is independent of the output load current since one of the two switches in a leg is always on at any instant. The DC voltage-source battery feeds the main rectifier circuit, a three-phase bridge. The DC voltage source is battery pack. Six switches are used in the main circuit. Each is composed of a SiC MOSFET and an antiparallel diode to provide bidirectional current flow and unidirectional voltage blocking capability. The converter works as a buck inverter for DC-to-AC power conversion.

2.1.2. AC Power Battery Charging Modes. The battery pack can be charged by the external power supply. When in the charging mode, the drive motor needs to be disengaged for safety purpose. Therefore, relay J1 keeps open to disconnect the three-phase windings in the motor. In addition, relays J2 and J3 are closed for charging. The battery pack can be charged by a low-voltage single-phase grid as a boost rectifier for AC-to-DC power conversion [22].
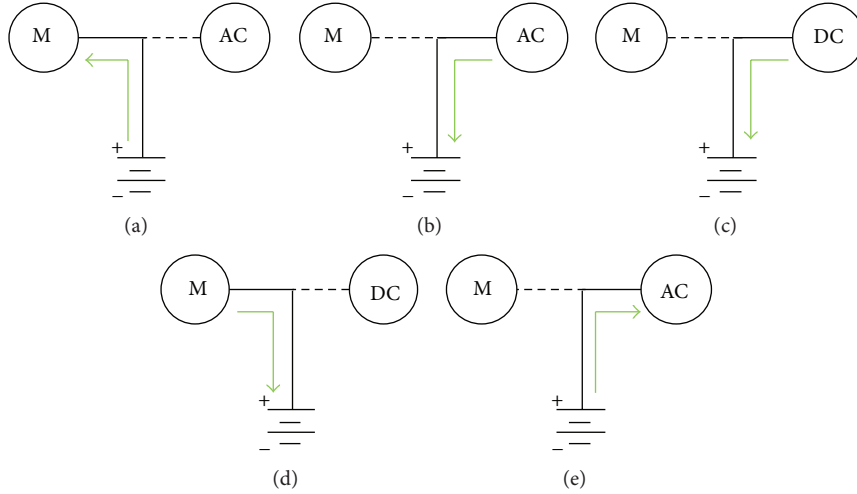
FIGURE 7: Power flow: (a) motor drive mods, (b) AC power charging, (c) DC power charging, (d) regenerative charging, and (e) vehicle-to-grid.
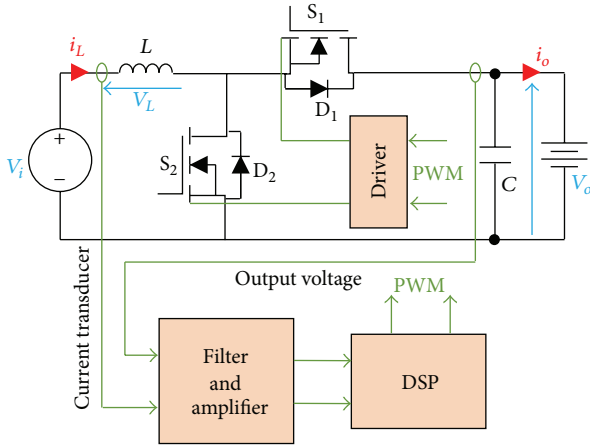
FIGURE 8: Boost DC/DC converter.

FIGURE 9: Active ripple compensation circuit.

2.1.3. DC Power Battery Charging Modes. The battery pack can be charged by low-voltage DC power, as shown in Figure 8. Boost operation from DC input voltage to the high-voltage battery pack of the EV is shown in Figures 10(a) and 10(b).

In this mode, S_2 works for PWM switching, and D_1 provides a free-wheeling path. Other switches and diodes S_1, D_2, S_3, and S_4 maintain the OFF state. The state of charge (SOC) should be regulated by measuring battery voltage and current.

2.1.4. Regenerative Braking. The proposed bidirectional converter is properly combined to select buck-and-boost modes among voltage sources. Regenerative charging uses buck operation from the high-voltage bus to the battery pack.

2.1.5. Vehicle-to-Grid (V2G). V2G is a modified version of EVs for the next generation to spark a revolution in the development of transportation and energy industries [23].

The V2G vehicles have capability of both charging from the grid and discharging to the grid intelligently that utilize bidirectional H-bridge converter properly.

2.1.6. Active Ripple Compensation Circuit. The grid voltage and current are sinusoidal, and twice the line frequency ripple will be generated on the DC-link, which is harmful to both sides of the converter. Especially when a battery connected on the DC-link, the pulsating power will lead to overheat. A relatively large DC-link capacitor is usually used to limit the ripple power, but it results in higher volume, weight, and cost for the integrated charger. To solve this second-order pulsating power, active ripple compensation methods have been explored in the paper.

The active ripple compensation circuit is composed of the active filter switches S_5 and S_6 which are shared with three-phase rectifier for motor drive circuit, an energy store capacitor, and an energy transfer inductor, as shown in Figure 9.

(a)   (b)
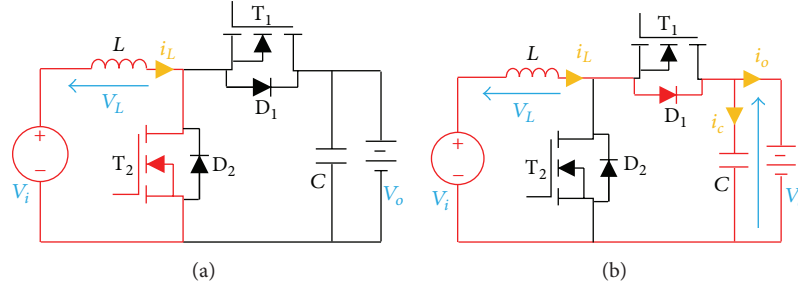
FIGURE 10: Boost converter circuit operating phases: (a) T_2 state-on, D_1 state-off, and (b) D_1 state-on, T_2 state-off.

(a)   (b)   (c)   (d)
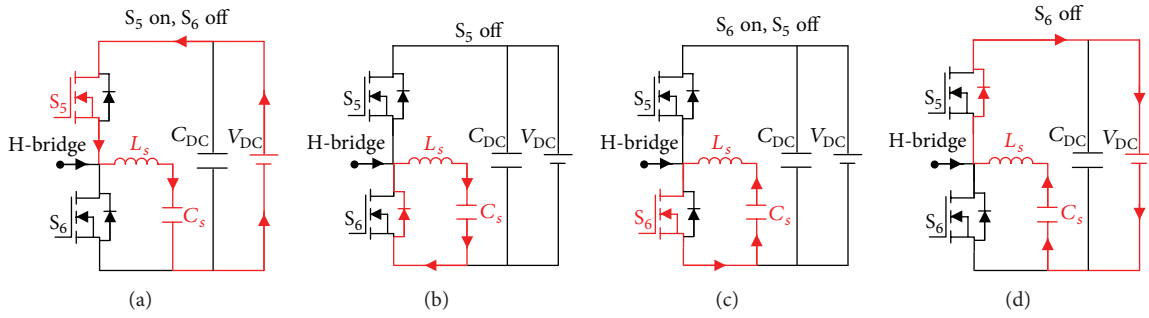
FIGURE 11: Equivalent circuit of active ripple compensation circuit for each stage of operation. ((a), (b)) Buck charging phase and ((c), (d)) boost discharging phase.

The C-phase accomplishes the DC side pulsating power at twice the grid frequency. The effectiveness of the active ripple compensation method is confirmed by simulation results. When switch S_5 turns on, the active ripple compensation circuit works as in buck phase. The capacitor and inductor are charged by DC bus. The inductor will release its energy to capacitor when the S_5 turns off. While switch S_6 is used to control the active ripple compensation circuit works as in boost phase. When the switch S_6 turns on, the capacitor releases its energy to inductor. When the switch S_6 turns off, the DC bus is charged by both of capacitor and inductor. A DC bus capacitor C_{DC} is still needed to filter the high frequency ripple power of the PWM rectifier output. In this case, C_{DC} is smaller than conventional method without the active ripple compensation storage circuit. The equivalent circuit of active ripple compensation circuit for each stage of operation is shown in Figure 11.

2.2. Instantaneous Power Balance Analysis. The grid-side supply voltage u_{AC} and current i_{AC} are assumed to be sinusoidal, the grid power supply can be written in (1), and the power of the input inductor can be expressed as (2):

$$P_{AC} = u_{AC}(t) * i_{AC}(t)$$
$$= \frac{1}{2}U_{AC}I_{AC}\cos\varphi - \frac{1}{2}U_{AC}I_{AC}\cos(2\omega t - \varphi), \tag{1}$$

where U_{AC} and I_{AC} are the peak value of voltage and current, respectively, ω is the angular frequency, and φ is the angle between the grid supply voltage and current:

$$P_L = \omega L I_{AC}^2 \sin(\omega t - \varphi)\cos(\omega t - \varphi). \tag{2}$$

The input power of the single-phase converter after the input inductor can be determined by (1) and (2), where L is the inductance of input filter:

$$P_{in} = P_{AC} - P_L = \frac{1}{2}U_{AC}I_{AC}\cos\varphi$$
$$-\left(\frac{1}{2}U_{AC}I_{AC}\cos(2\omega t - \varphi)\right. \tag{3}$$
$$\left.+\frac{1}{2}\omega L I_{AC}^2 \sin(2\omega t - 2\varphi)\right).$$

As can be seen in (3), a constant power and twice the fundamental frequency 2ω power pulsating consist of the input power of the single-phase battery charging circuit. To ensure a low-frequency ripple in single-phase charger, a large DC capacitor is required which results in low power density. In order to avoid the second-order harmonic and minimize the size of DC capacitor, a DC ripple current reduction method on a single-phase PWM voltage-source rectifier has been investigated in [24]. Several active solutions have been explored to reduce the second-order ripple power, namely, inductive storage method and capacitive storage method. The active ripple compensation charging circuit is shown in Figure 12.
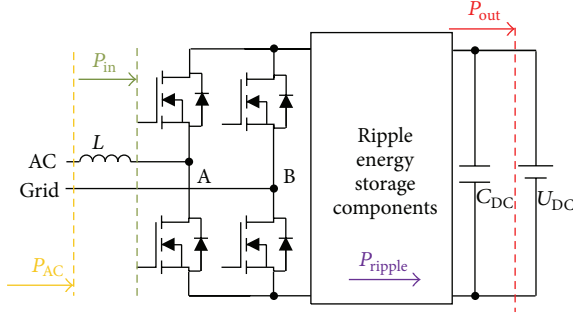
FIGURE 12: Active ripple compensation charging circuit.

A conventional H-bridge rectifier consists of phase-A and phase-B. The AC supply power is P_{AC} which can be expressed by AC source voltage multiplied AC source current. L is the input inductor which is used to keep a constant current flowing to the load throughout the complete cycle of the applied voltage. The input power of the H-bridge rectifier after the input inductor is P_{in}. It is known that there exists second-order harmonic power P_{ripple} on DC-link for H-phase rectifier. The constant power on DC-link can be expressed as P_{out}. Generally speaking, the basic approach behind the ripple reduction circuit involves storage of the ripple power into inductor or capacitor. The comparison between inductive storage method and capacitor method is discussed in Section 3.

3. Design Criteria of Circuit Components

3.1. Selection of Motor Drive Circuit Components. A buck converter with LC filter should be considered in the three-phase inverter to reduce the ripple. An LC filter is critical for AC motor drive application. The capacitors can be configured with a delta connection which has the benefit of smaller short circuit current. Capacitor and choke values are derived to optimize the two-level inverter output performance. The filter is assumed to filter out all the PWM switching harmonics. Damping is required to attenuate the resonance and the output impedance of the filter should be as small as possible across the frequency range. A series resonance occurs at the output filter resonant frequency f_0. Hence, the corner frequency is kept below the switching frequency. The result can be expressed as in

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}}, \tag{4}$$

where f_0 represents the corner frequency, for 10% attenuation of switching frequency harmonics. L and C are the inductance and capacitance, respectively. The key characteristics in the driving mode are shown in Table 2.

3.2. Selection of Battery Charging Circuit Components. A typical H-bridge connected with an active ripple compensation circuit works as battery charging converter. The A-phase and B-phase operate as conventional PWM. A large filter

TABLE 2: Key characteristics in the motor drive mode.

Items	Parameters
Rated output power P_{out}	70 kW
Switching frequency f_{sw}	10 kHz
Modulation index	0.9
Cooling method	Water cooling
DC-link voltage V_{DC}	320 V
Output LC filter inductance L	267 μH
Output LC filter capacitance C	10 μF

TABLE 3: Parameters in battery charging mode.

Items	Parameters
AC voltage source V_{AC} (RMS)	230 V
Line frequency f	50 Hz
Rated output power P_{out}	3 kW
Input inductor L_1	4.35 mH
DC-link capacitance C_{DC}	4665 μF
Switching frequency f_{sw}	20 kHz
DC-link voltage V_{DC}	320 V
Modulation index	0.9
Cooling method	Natural cooling

capacitor is connected on the DC side [25]. The required DC-link capacitance can be expressed as

$$C_{DC} \geq \frac{P_{avg}}{\omega_0 \cdot \Delta V \cdot V_{DC}}, \tag{5}$$

where ΔV is the amplitude of DC-link voltage ripple, for 2% allowable voltage ripple, C_{DC} is the value of DC-link capacitor, V_{DC} is the DC-link voltage, ω_0 is the AC-side circular frequency, and P_{avg} is the average power flowing into the converter. The complete parameters in battery charging mode are listed in Table 3.

3.3. Selection of Ripple Energy Storage Circuit Components. The ripple energy generated by H-bridge can be determined in

$$\begin{aligned}
E_r &= \frac{P_r}{\omega} \\
&= \frac{\sqrt{P_0^2 + \left[\left(2\omega L P_0^2 / U_s^2 \cos^2\varphi\right) - P_0\left(\sin\varphi/\cos\varphi\right)\right]^2}}{\omega}.
\end{aligned} \tag{6}$$

Equation (6) shows the relationship between ripple energy, the angle of grid voltage, the output power, AC supply voltage, the AC supply frequency, and current and input inductor. The parameters of a 3 kW single-phase PWM rectifier are given in Table 3. The relationship between various and the ripple energy is plotted in Figures 13(a) and 13(b).

It can be observed that the phase inductor has an obvious influence on the ripple energy. The higher inductance results in a higher ripple energy. Meanwhile, the supply frequency has an obvious influence on the ripple energy as well.

(a)

(b)
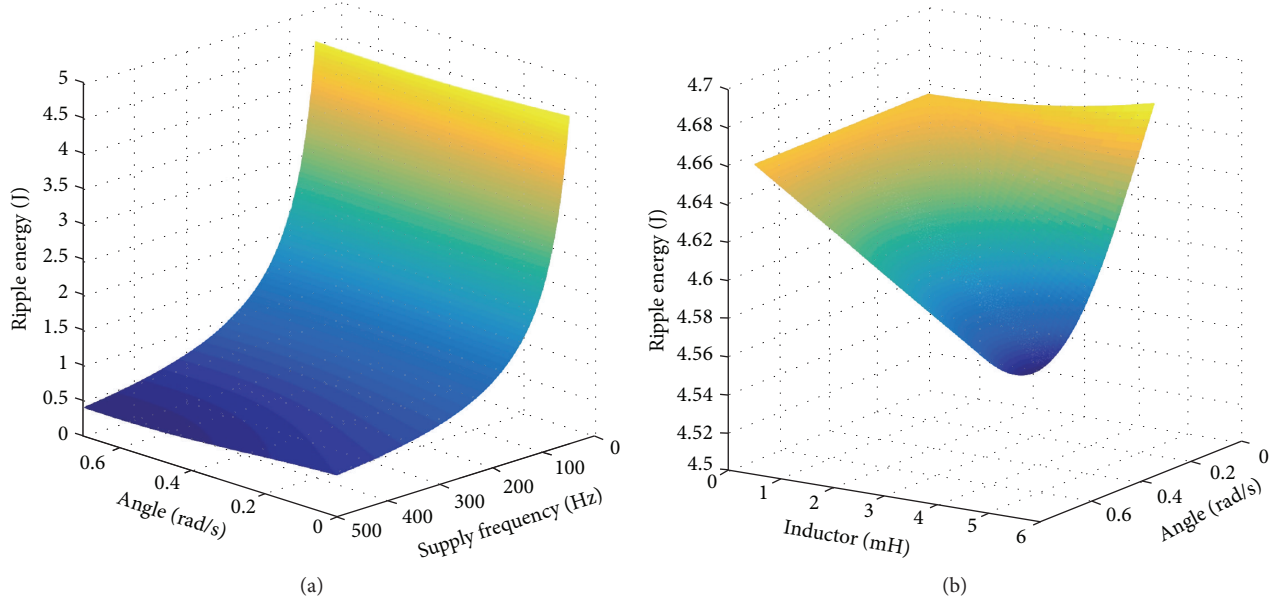
FIGURE 13: Ripple energy: (a) relationship between angle of the supply voltage and current and supply frequency and (b) relationship between angle of the supply voltage and current and inductor.
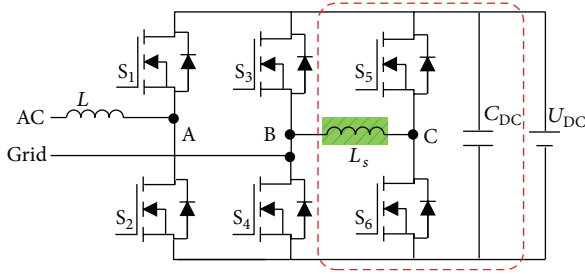
FIGURE 14: Circuit configuration.

FIGURE 15: Compensation region of the inductor.

The lower supply frequency leads to a higher power pulsating. The reason can be expressed in (6). Figure 13(b) shows that there is a minimum of 4.67 J ripple energy for the 3 kW single-phase rectifier system.

3.4. Inductive Storage Method.

Figure 14 shows the circuit configuration of inductive storage method. A conventional single-phase rectifier consists of A-phase leg, B-phase leg, input filter, and DC-link capacitor. An active ripple energy storage circuit consists of C-phase leg and a ripple energy storage inductor. One terminal of the storage inductor is connected to the midpoint of the C-phase leg switches (connection point of S_5 and S_6), while the other terminal is tied to B-phase leg of the H-bridge converter. The compensation inductance L used to store ripple energy should be selected depending on the output power P_{out}. A DC bus capacitor C_{DC} is still necessary to filter the high switching harmonic. The DC ripple energy flows into the inductor. The inductance can be selected in a wide rage when the inductor current is

controlled properly. The maximum inductor current and the minimum inductor current are shown in

$$L_{max} = \frac{2I_{AC}V_{AC}}{I_{min}^2 \omega},$$

$$L_{min} = \frac{\sqrt{2}V_{AC}}{2D\lambda f_{sw}I_{min}},$$

(7)

where L_{max} is maximum inductance, L_{min} is minimum inductance, λ is modulation factor of the single-phase, and V_{AC} and I_{AC} are the grid voltage and grid current, respectively. And D is the ratio between peak ripple current and minimum ripple current. The compensation inductance and current region are shown in Figure 15. Higher switching frequency

(a)

(b)

(c)
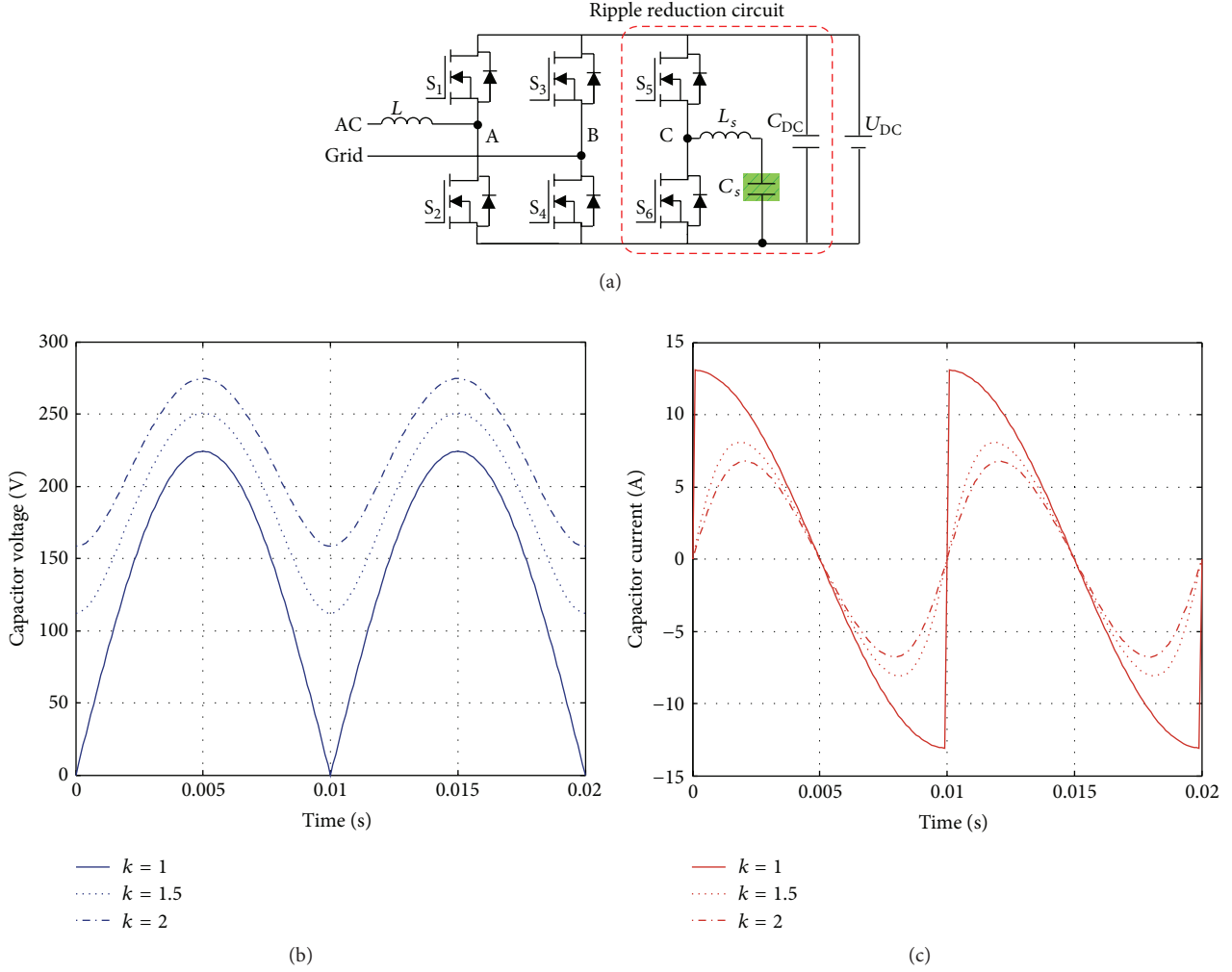
FIGURE 16: Capacitive storage method: (a) circuit configuration of capacitive storage method, (b) capacitive ripple voltage, and (c) capacitive ripple current.

results in less inductance. The trade-off between switching frequency and losses needs to be considered.

3.5. Capacitive Storage Method.

A capacitor is used as the energy storage component is shown in Figure 16(a). One terminal of the storage capacitor is tied to the midpoint of the C-phase leg, while the other terminal is connected to the ground. A ripple energy storage component C_s with an energy transfer element L_s is used as an active ripple compensation circuit. The ripple power can be determined in

$$P_{\text{ripple}} = E_{\text{ripple}} * \omega$$

$$= \sqrt{P_{\text{out}}^2 + \left[\left(\frac{2\omega L P_{\text{out}}^2}{U_s^2 \cos^2 \varphi} \right) - P_{\text{out}} \left(\frac{\sin \varphi}{\cos \varphi} \right) \right]^2}. \tag{8}$$

The ripple energy storage capacitor voltage and current can be expressed as (9); Figures 16(b) and 16(c) show the ripple voltage and ripple current in power pulsating storage capacitor:

$$U_s = \sqrt{\frac{P_{\text{ripple}} (k - \cos (2\omega t))}{C_s \omega}}$$

$$i_s = \frac{P_{\text{ripple}} \sin (2\omega t)}{\sqrt{(P_{\text{ripple}}/C_s \omega) (k - \cos (2\omega t))}}, \tag{9}$$

where the coefficient k is the ripple energy storage margin coefficient defined by the maximum ripple energy and ripple energy stored in the capacitor and C_s is energy storage capacitance. The second-order ripple voltage and capacitive ripple current stored in the capacitor are plotted in Figures 17(a) and 17(b).

The minimum capacitance can be derived as

$$C_s = \frac{2P_{\text{ripple}}}{U_s^2 \omega}. \tag{10}$$
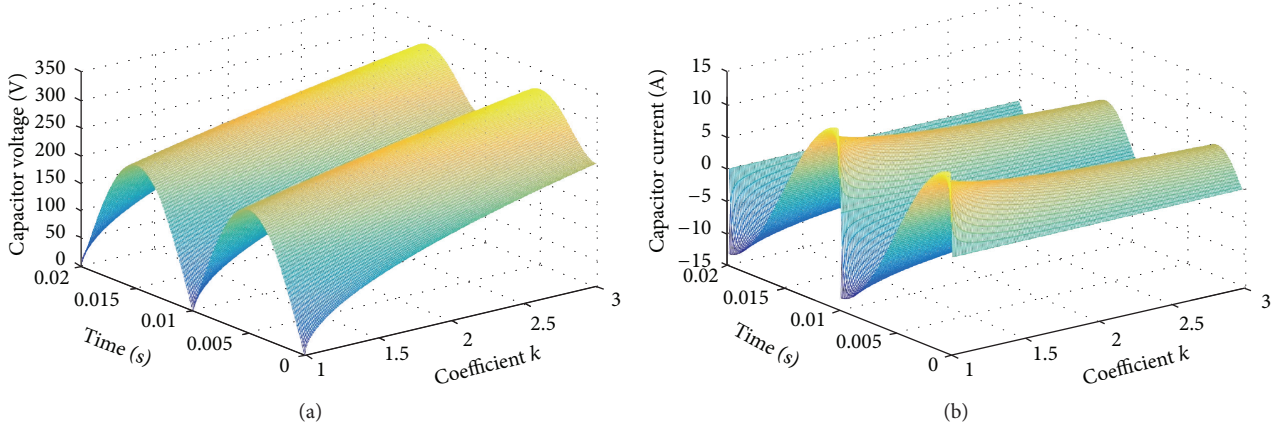
FIGURE 17: Second-order capacitive ripple voltage and capacitive ripple current: (a) capacitive ripple voltage and (b) capacitive ripple current.
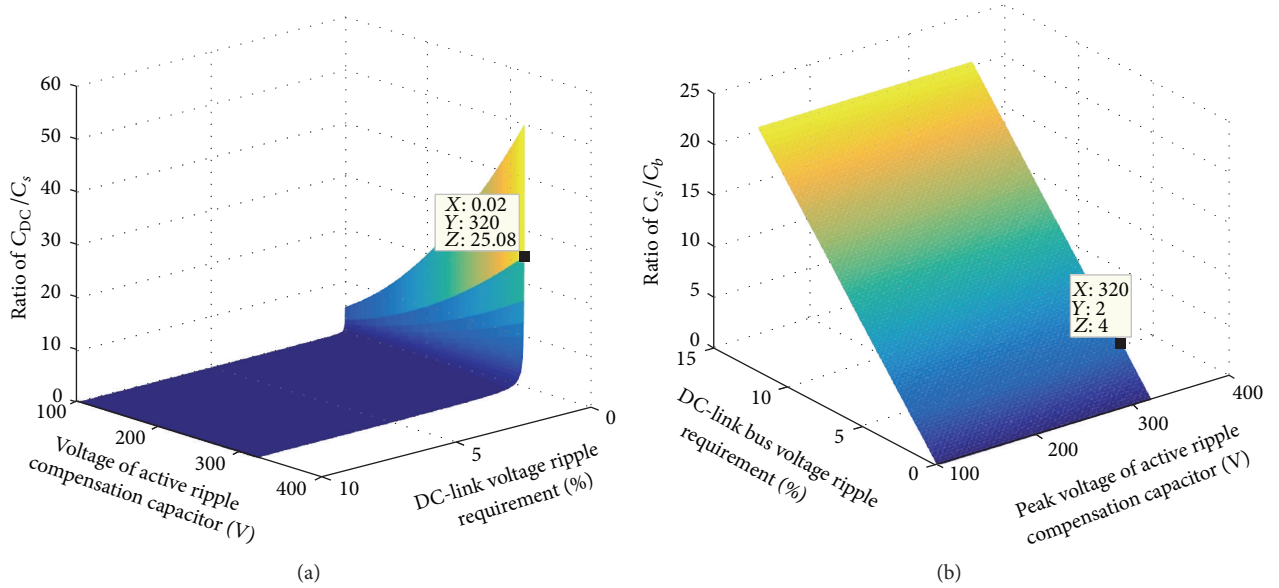
FIGURE 18: Comparison results between the DC-link capacitance and active ripple compensation methods: (a) capacitance comparison result and (b) current rating comparison result.

Equation (5) shows that the conventional design method calculates capacitance C_{DC} needed to filter the second-order harmonic in DC bus, while (10) shows that the active ripple storage capacitance C_s needed to meet the requirement of filtering the second-order ripple power. Figure 18(a) shows the capacitance comparison between the traditional method and the active method. It can be seen the active ripple method will decrease the capacitance 25.08 times compared with the conventional method within the 2% DC-link harmonic requirement. This indicates that the capacitance can decrease from 4.665 mF to 186 μF.

The active ripple compensation circuit works in discontinuous current mode (DCM) to meet the demand of maintaining ripple energy transfer inductance transfer 2ω power pulsating to the ripple energy storage capacitor completely. In DCM the current goes to zero during part of the switching cycle. In order to maintain DCM operation, the ripple

energy transfer inductor selection limit can be expressed as (11):

$$L \leq \frac{T_{sw}}{2 * i_s} \frac{U_d U_s - U_s^2}{U_d}. \tag{11}$$

In addition, the active ripple compensation circuit requires one-phase leg which can share with three-phase inverter motor drive circuit, which has the maximum current rating limitation. The inductor selection limit based on the peak current requirement is expressed as

$$L \geq \frac{2 * \overline{i_s} * T_{sw}}{I_{peak}^2} \frac{U_d U_s - U_s^2}{U_d}. \tag{12}$$

The inductance is selected as 50 μH in the integrated charger system using (11) and (12). Table 4 shows the calculation results of inductive storage method and capacitive

FIGURE 19: Control schematic figure for the circuit working as three-phase inverter.
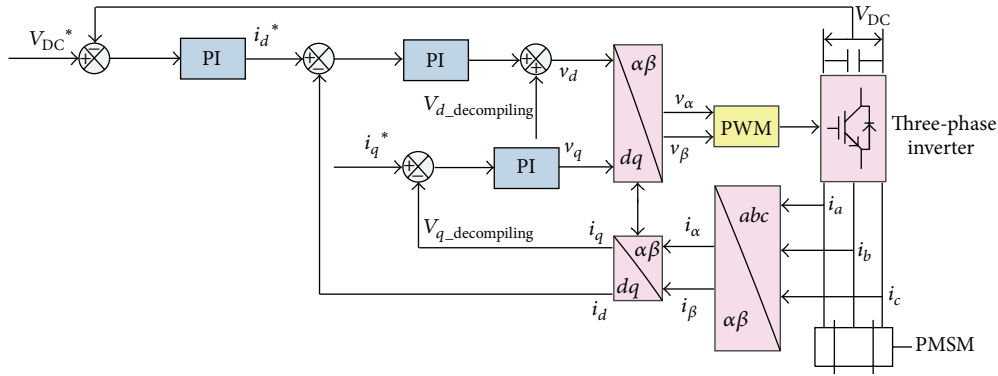
TABLE 4: Comparison of inductive storage method and capacitive storage method.

Method	Capacitance	Inductance	Volume	Weight
Inductive storage method	—	68 mH	2762 cm^3	12.75 kg
Capacitive storage method	298 μF	50 μH	2895 cm^3	4 kg

storage method. Although the volume of the inductive storage devices can be smaller than the capacitive storage method, the weight of the inductor becomes much heavier than the capacitor. This demonstrates that capacitive storage method is superior to the inductive storage method in terms of power density of the charging circuit.

Although inductor is superior to capacitors from the viewpoint of ruggedness and reliability, inductor is inferior from the viewpoint of power density and weight. Power losses of inductor are also much higher than capacitor when working in high switching circuits [26–28]. Therefore, in this paper, an integrated charger topology employs a capacitor as the ripple energy storage element and an inductor as the ripple energy transfer component.

4. Control Design of Integrated Charger with Active Ripple Compensation

4.1. Motor Drive Control. In the motor drive mode, a closed-loop control schematic is illustrated in Figure 19. Both motor speed and current are regulated by applying PWM control to each phase switching device. The conventional PI controller including DC-link voltage control loop and current control loop are applied in three-phase inverter.

4.2. Battery Charging Control. The proposed active ripple energy absorbing method can generate the compensation ripple power successfully according to control capacitor voltage and current based on (9). The waveforms of capacitor reference voltage and reference current are essential issues to be considered besides ripple energy storage components. In Figure 16(b), the sharp turns at the bottom of the waveforms

demonstrate large harmonic which is difficult to track for control strategy. It is possible to reduce the harmonic in the voltage reference and current reference by increasing the energy storage margin coefficient k which is set to 2 in this paper. The ripple current generated by H-bridge rectifier can be expressed as

$$i_{\text{ripple}} = (2D - 1) \, i_{\text{AC}}. \tag{13}$$

The reference compensation current is taken as minus ripple current for the active ripple compensation circuit, where D is duty cycle for H-bridge and i_{AC} is the AC-side current. The active ripple compensation circuit is controlled in bulk type and boost type to charge and discharge the ripple energy storage capacitor. The duty cycle for S_5 and S_6 can be derived as (14) and (15), respectively:

$$D_{\text{charging}} = \sqrt{\frac{2 \cdot i_{\text{ripple}} \cdot f_{\text{sw}}}{(U_d - U_s)/L}}, \tag{14}$$

$$D_{\text{discharging}} = \sqrt{\frac{2 \cdot i_{\text{ripple}} \cdot f_{\text{sw}} \cdot (U_d - U_s)/L}{(U_s/L)^2}}. \tag{15}$$

Based on the above discussion, the control block diagram is depicted in Figure 20. The upper part represents the control schematic for the PWM rectifier battery charging circuit and the lower part is the control strategy for the ripple reduction circuit. For PWM rectifier circuit, the DC-link voltage, DC-link current, and AC-side current are sensed to generate the gate drive signals for the main switches S_1, S_2, S_3, and S_4. In order to improve the performance of DC bus voltage, the AC-side voltage feedforward is adopted. A phase-locked loop (PLL) is constructed to track sinusoidal wave of the AC grid voltage as well. Via the proportional-integral (PI) controller, the command current i_{AC} is obtained to be equal to the command current $i_{\text{AC}}{}^*$. For ripple reduction circuit, the DC-link voltage and ripple energy storage capacitor voltage are sensed to generate the duty cycle for S_5 and S_6. According to the control method for C phase, if the compensation current is positive, the ripple reduction circuit is controlled to absorb the ripple energy from the DC-link charging the ripple compensation capacitor. Similar, if the compensation
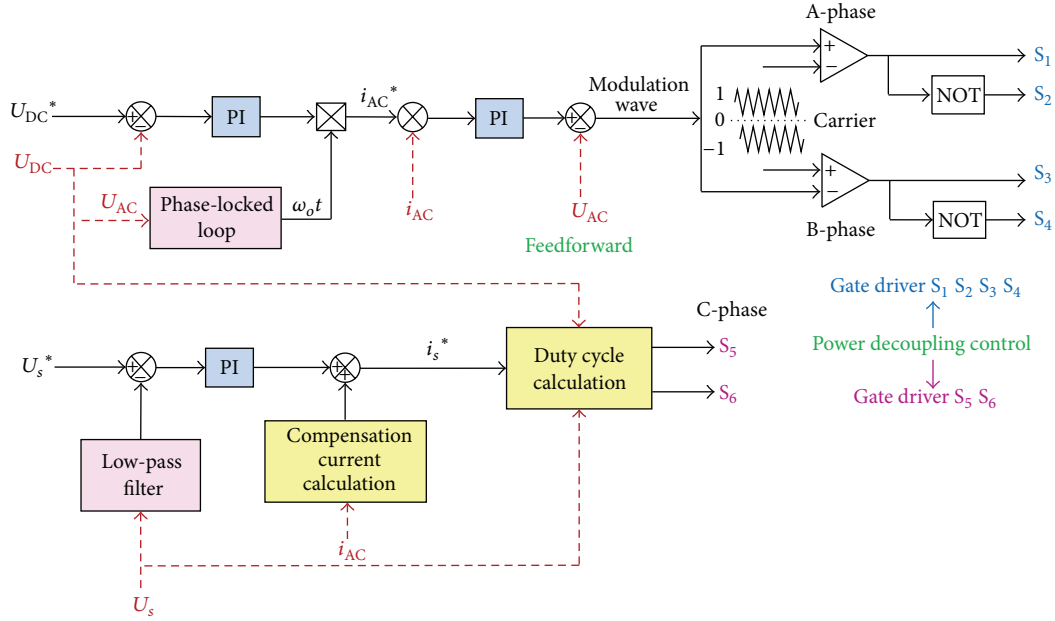
FIGURE 20: Control system for the integrated charger.
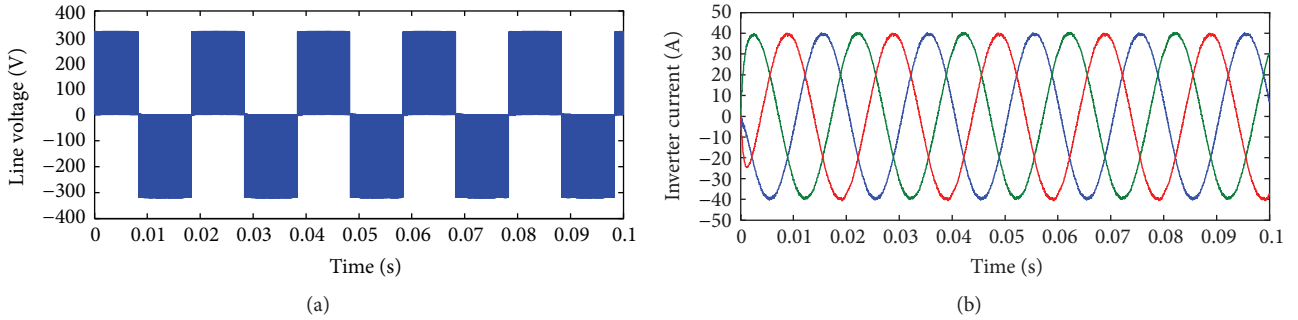
FIGURE 21: Three-phase inverter simulation results: (a) the inverter output line voltage and (b) the inverter current.

current calculation result is negative, the ripple reduction circuit is controlled to release the ripple power into DC-link from ripple energy stored capacitor.

5. Simulation Results

The simulation parameters of the integrated charging converter are summarized in Tables 2 and 3. To evaluate the motor drive mode performance, a three-phase motor drive converter is modeled under PWM control in the simulation. Figure 21 presents the simulation results of motor drive.

Figure 22 presents the simulation results of battery charging without active ripple compensation circuit. Battery charging mode operation has been simulated under two conditions where the AC input voltage is lower than battery voltage, and DC input voltage is lower than battery voltage. When the input voltage is 230 V_{AC}, Figure 22(a) shows the output voltage. Figure 22(b) shows the maximum current ripple in grid side. The boost operation from the low-voltage

bus to the battery has been simulated as shown in Figures 22(c)–22(f) with V_{batt} = 320 V and V_{lv} = 200 V.

Figures 23(a) and 23(b) show the integrated charger with active ripple compendation circuit main switch components duty cycle. It can be seen that the duty cycle for H-bridge PWM rectifier and the active ripple compensation circuit are decoupling. The most of the ripple power is stored in the DC bus capacitor without the active ripple compensation circuit. To meet the requirement of DC bus voltage ripple within 2% limit, a 4665 μF ripple energy storage capacitance is needed. A comparison of the DC bus voltage ripple performance without ripple reduction circuit and with active reduction circuit is provided in Figure 23. The simulation result in Figure 23(c) shows that the DC voltage ripple remains 74.64 V without active ripple compensation. When the C-phase leg is engaged, the 2ω ripple power is absorbed by the capacitor. A 298 μF capacitor and a 400 μF DC-link capacitor are used to be replaced by conventional large DC bus capacitor. The DC bus ripple voltage decreases to 4.6 V as shown in Figure 23(d).
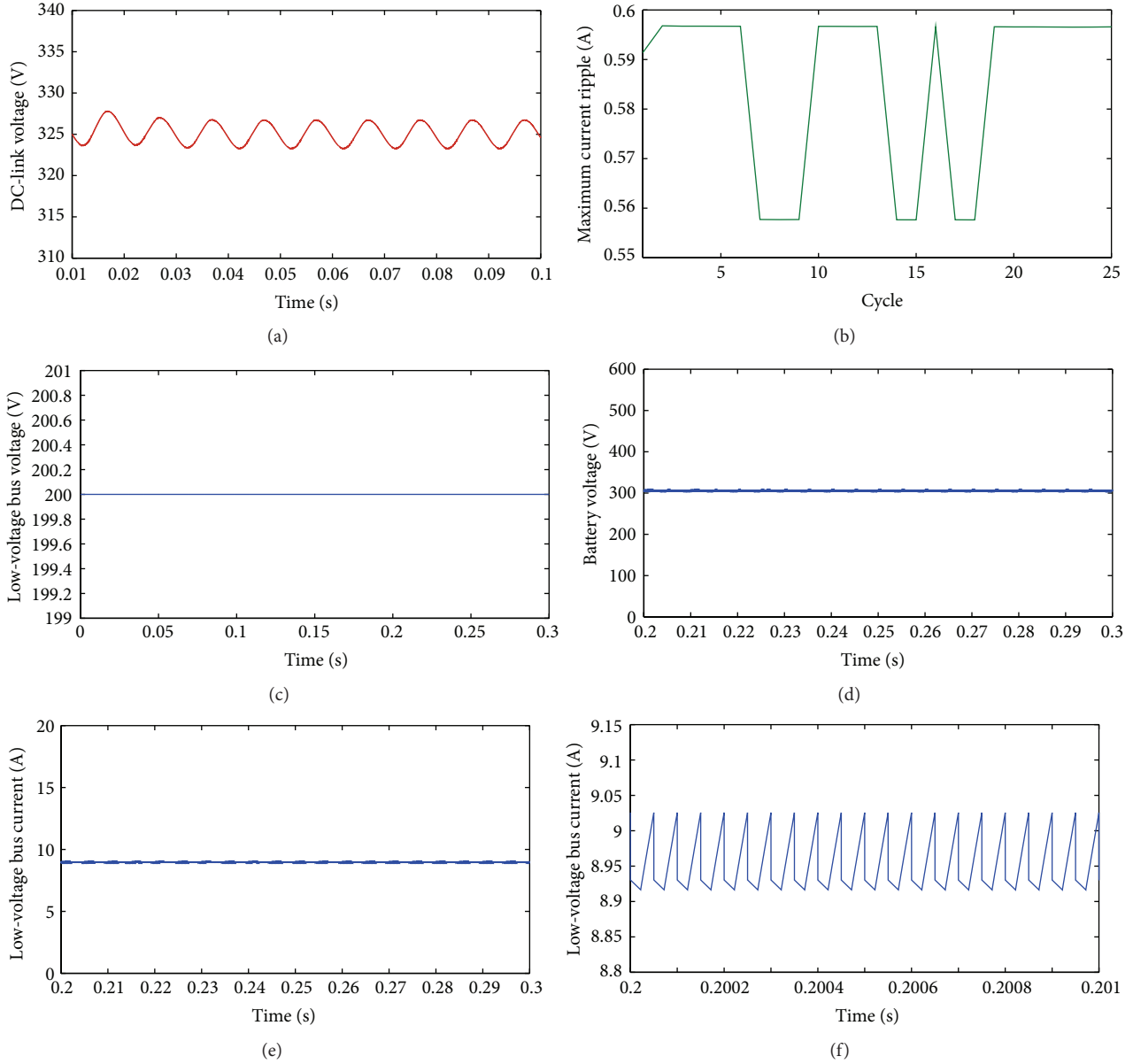
FIGURE 22: Simulation results for the AC/DC battery charging mode: (a) DC-link voltage and (b) AC-side maximum current ripple. Boost-type charging: (c) steady state for V_{lv} = 200 V, (d) steady state for V_{batt} = 320 V, (e) low-voltage bus current, and (f) zoom low-voltage bus current.

From Figures 23(e) and 23(f), it is shown that the DC ripple has been absorbed by the ripple power storage capacitor.

A comparison of the volume of the main components in the conventional method and the active ripple compensation method is illustrated in Figures 24(a) and 24(b). Using the active ripple compensation circuit to store the ripple energy is more effective. Conclusively, the whole integrated charger system volume decreases to 35% compared with the volume of the conventional method.

6. Power Losses and Thermal Analysis

In order to evaluate the thermal performance of the integrated charger with active ripple compensation converter topology, the power losses and device junction temperature are calculated and simulated. SiC-based devices meet the power electronics market demand for high performance 1200 V to 1700 V devices. SiC MOSFETs have unique capabilities such as lower switching loss, higher efficiency, and better temperature performance compared to Si-based devices when operating the same power rating which have essentially reached state-of-the-art limits in performance. In this paper, the performance of SiC MOSFET module from Cree CAS300M12BM2 (1.2 kV, 300 A) will be investigated. Figures 25(a) and 25(b) show the MOSFET and diode average losses of the motor drive mode and battery charging mode, which is derived by analytical models in [29]. Figure 25(c) shows the efficiency variation with various switching frequencies

(a)

(b)

(c)

(d)

(e)

(f)
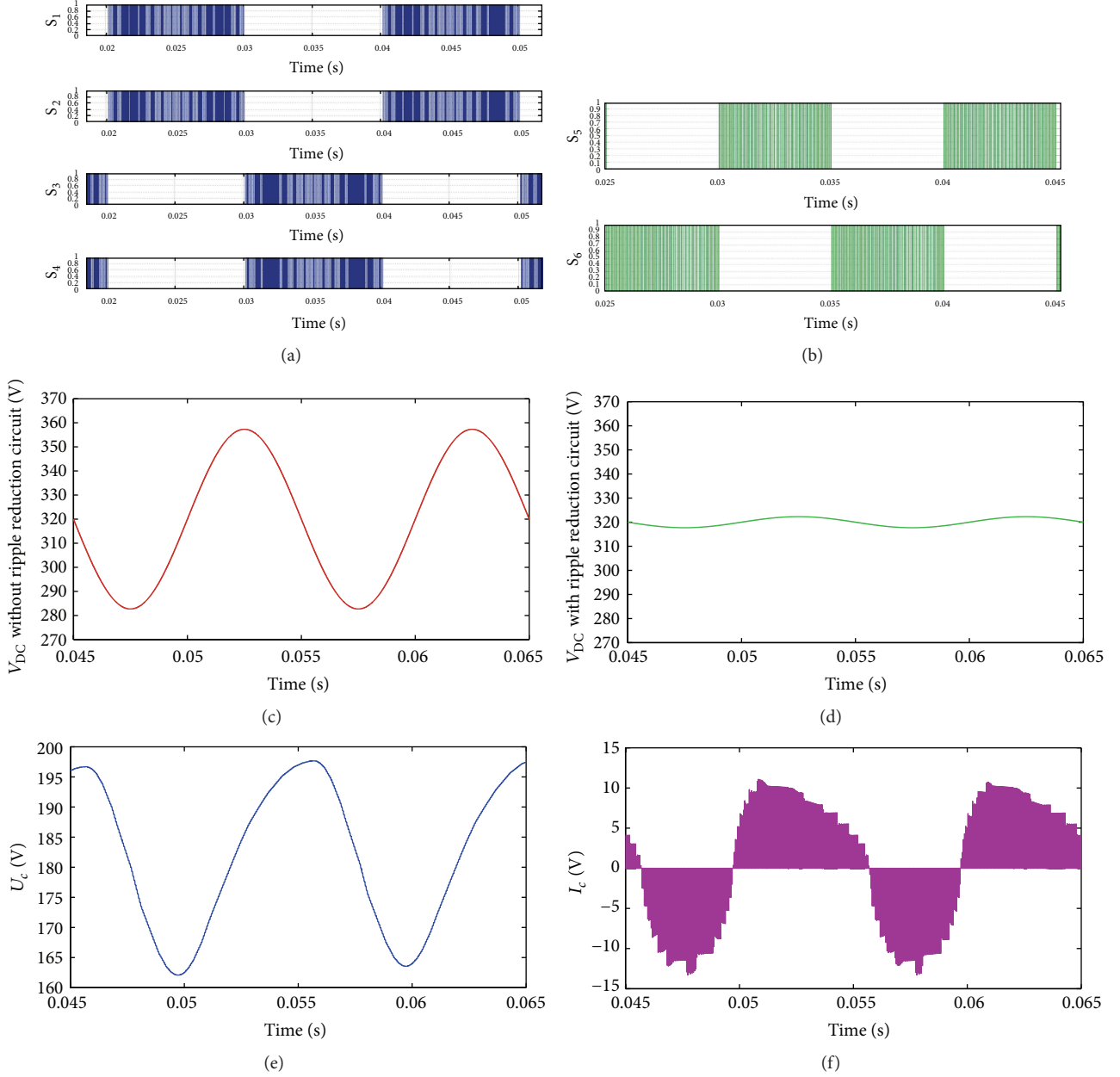
FIGURE 23: Duty cycle for the integrated charger: (a) H-bridge recitifier duty cycle and (b) active ripple compensation circuit duty cycle. Simulation results: (c) DC bus voltage without active ripple compensation circuit, (d) DC bus voltage with active ripple compensation circuit, (e) ripple power storage capacitor voltage, and (f) ripple power storage capacitor current.

for traction and charging operation [30–32]. In comparison to Si IGBT module from SK60GB128 (1.2 kV, 60 A) and SiC MOSFET (CAS300M12BM2) module, as expected, the SiC MOSFET shows a higher efficiency in a wide switching frequency range. With SiC MOSFET lower power losses and operation capability at higher switching frequency, integrated converter can operate up to 73 kHz switching frequency compared to 20 kHz using Si IGBT. The above results show that a clear advantage for SiC MOSFET is the candidate of choice to meet the demand of high power density and high efficiency in power converters application.

It is known that most of the failure mechanism such as bond-wire breakage is related to excessive temperature. Therefore, thermal behaviours management in power converter is essential to increase reliability performance [33]. Real-time junction temperature estimation requires calculation of the instantaneous losses in each MOSFET and diode device of the integrated charger. Therefore, the switching device conduction and switching losses every switching time cycle are calculated instantaneously [34]. The thermal model network includes thermal resistance and capacitance, which can transfer the power losses to the corresponding
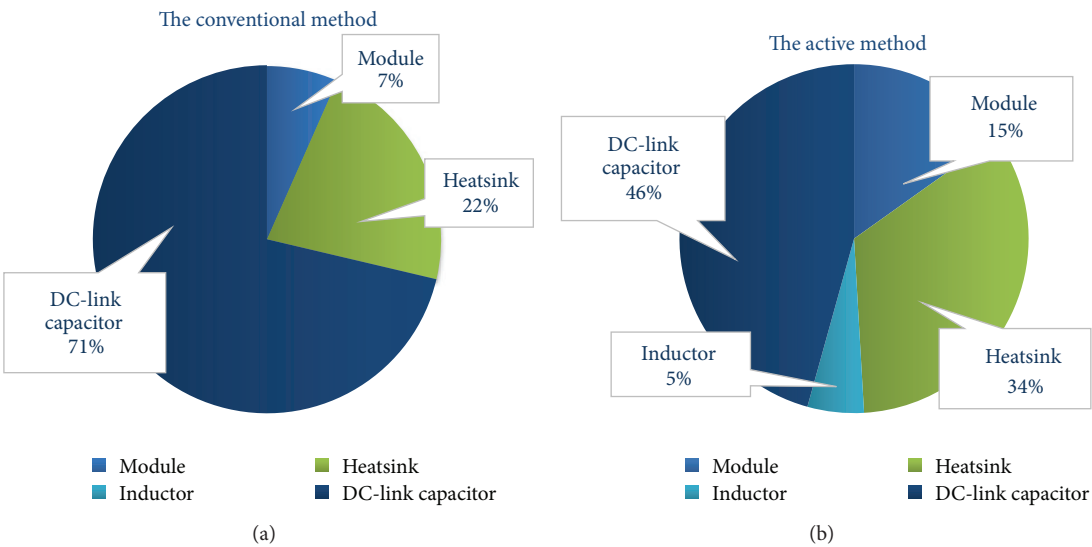
FIGURE 24: Power density: (a) conventional method main components volume comparison and (b) active method main components volume comparison.
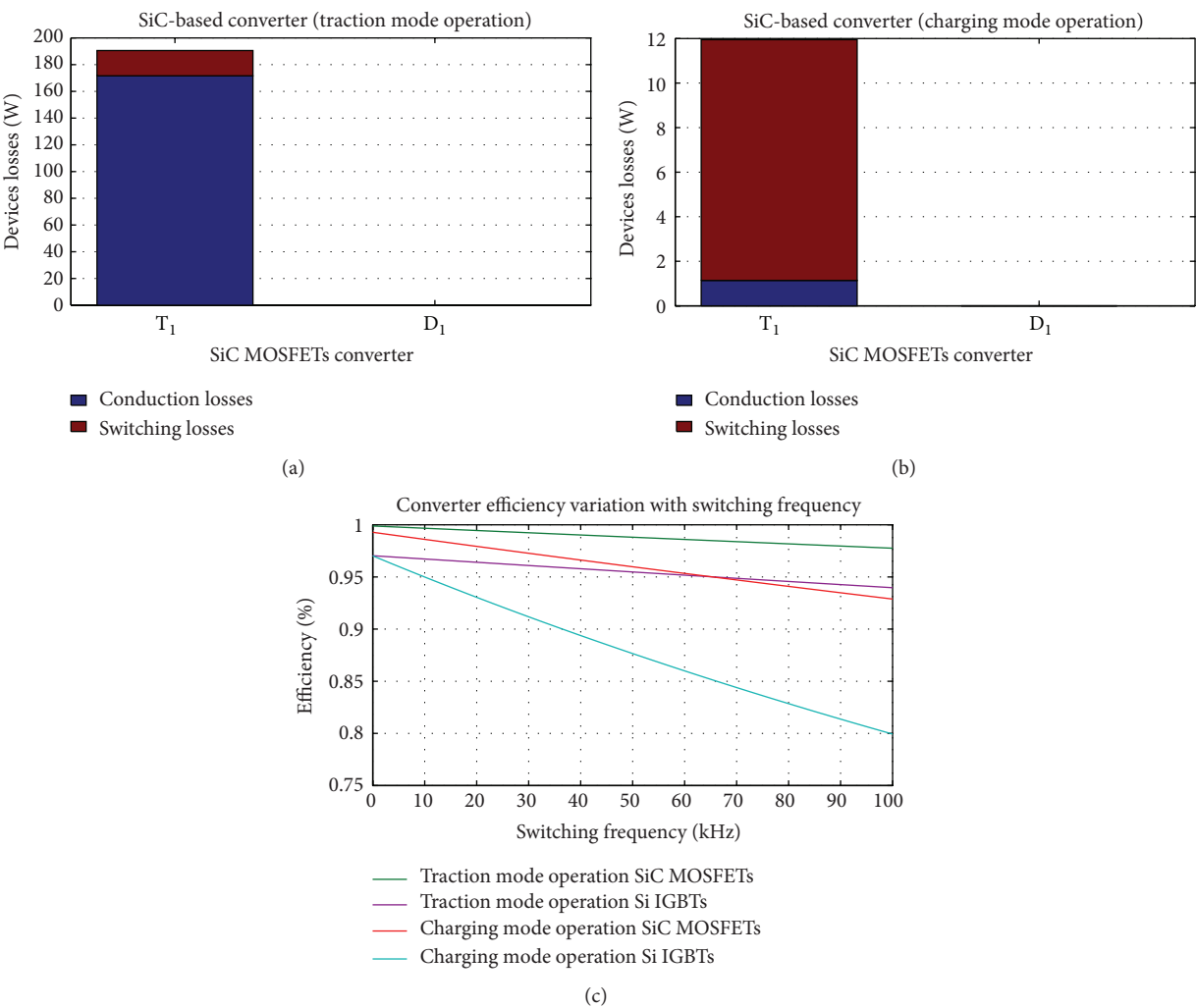
FIGURE 25: Calculation results: (a) traction mode operation, (b) charging mode operation, and (c) converter efficiency variation with switching frequency for SiC MOSFETs and Si IGBTs.
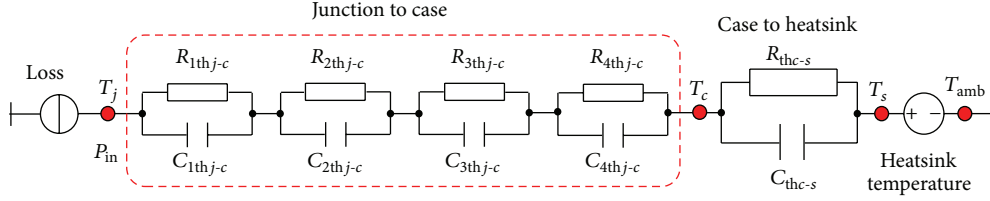
FIGURE 26: Thermal network (Foster model).

(a)   (b)   (c)
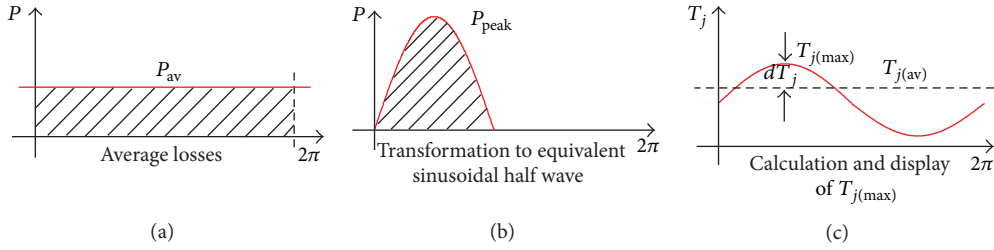
FIGURE 27: Calculation of the junction temperature fluctuation as a function of the converter output frequency. (a) Average losses. (b) Transformation to equivalent sinusoidal half wave. (c) Calculation and display of junction temperature value.

(a)   (b)
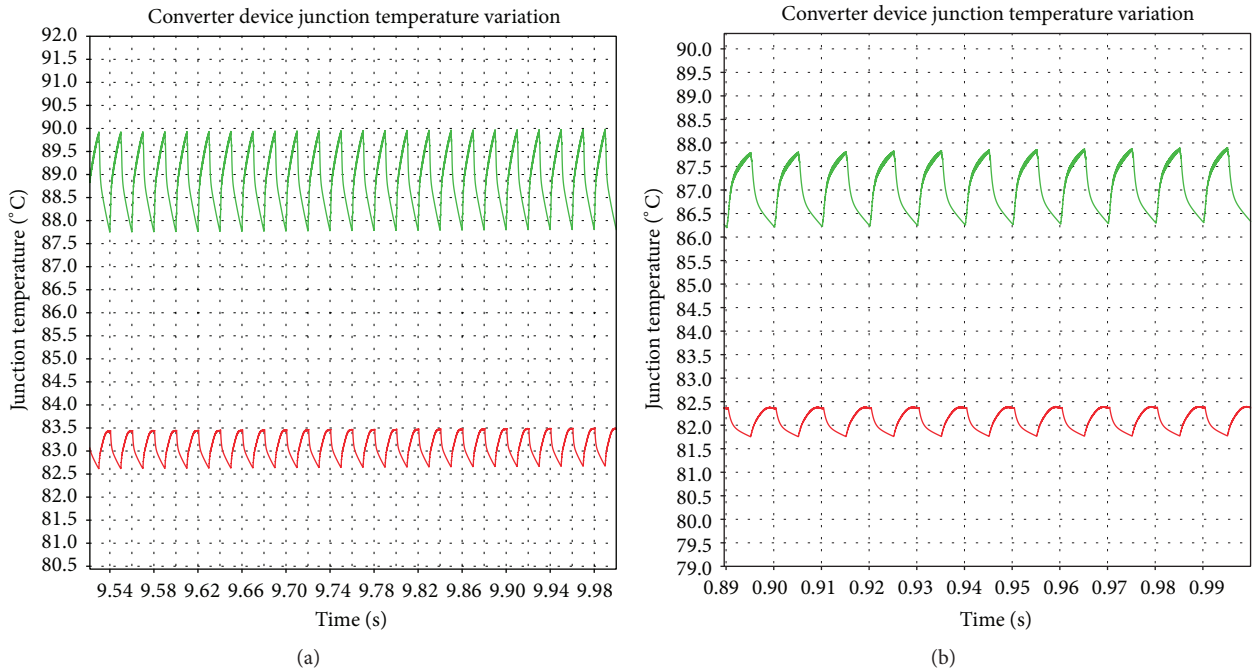
FIGURE 28: MOSFET/diode (battery charging mode) junction temperature variation: (a) single-phase circuit and (b) ripple energy storage circuit.

temperature in power devices. In the simulation, the heatsink temperature is assumed to be fixed at $80°C$ due to its large thermal time constant compared with converter devices [35]. As shown in Figure 26, the Foster thermal network is proposed which is used for junction temperature estimation. The parameters of the thermal network elements are given in the device datasheet [36].

The possible solution for junction temperature variation calculation is shown in Figure 27. A simulation based on transformation average losses to equivalent sinusoidal half

wave method is used to evaluate the junction temperature variation of the battery charging mode. As shown in Figure 28(a), the single-phase SiC MOSFET junction temperature varies between $87.5°C$ and $90°C$. The diode junction temperature varies between $82.5°C$ and $83.5°C$. When the ripple energy storage circuit is applied, the SiC MOSFET temperature varies between $86.2°C$ and $87.7°C$ while the diode temperature varies between $81.8°C$ and $82.5°C$. The junction temperature does not exceed the maximum allowable junction temperature $150°C$.

7. Conclusion

In this paper, a concept of integrated charger with ripple reduction circuit for EV applications is proposed. The integrated converter reduces system cost, increases power density, and may lead to improved efficiency. The proposed integrated converter has been compared with existing topologies, and its advantages have been indicated. Additionally, exact system parameters, control strategy, power losses based on SiC MOSFET devices, efficiency, system power density, and thermal stress are discussed. In order to verify the proposed converter, the functionalities for different operating modes, for example, the boost for charging battery, buck for regenerative braking, and buck for motor drive, have been simulated. According to input/output-voltage-current conditions, the controller chooses the control schemes and proper operating modes. The DC bus low-frequency power pulsating generated from single-phase has been reduced efficiently by active ripple energy compensation circuit. To verity the practicality of the proposed converter for EV applications, an on-board experiment needs to be tested in the future.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] F. Sun and R. Xiong, "A novel dual-scale cell state-of-charge estimation approach for series-connected battery pack used in electric vehicles," *Journal of Power Sources*, vol. 274, pp. 582–594, 2015.

[2] M. Yilmaz and P. T. Krein, "Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles," *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pp. 2151–2169, 2013.

[3] S. Dieckerhoff, S. Bernet, and D. Krug, "Power loss-oriented evaluation of high voltage IGBTs and multilevel converters in transformerless traction applications," *IEEE Transactions on Power Electronics*, vol. 20, no. 6, pp. 1328–1336, 2005.

[4] K. L. Butler, M. Ehsani, and P. Kamath, "A matlab-based modeling and simulation package for electric and hybrid electric vehicle design," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 6, pp. 1770–1778, 1999.

[5] S. Haghbin and M. Alakula, "Electrical apparatus comprising drive system and electrical machine with reconnectable stator winding," International Patent WO/2011/159241, 2011.

[6] A. G. Cocconi, "Combined motor drive and battery recharge system," US Patent 5 341 075, 1994.

[7] AC Propulsion, "AC propulsion EV drive system specifications," Technical Note, AC Propulsion, San Dimas, Calif, USA, 2008.

[8] W. E. Rippel, "Integrated traction inverter and battery charger apparatus," US Patent 4 920 475, 1990.

[9] W. E. Rippel and A. G. Cocconi, "Integrated motor drive and recharge system," US Patent 5 099 186, 1992.

[10] L. De-Sousa and B. Bouchez, "Combined electric device for powering and charging," International Patent WO 2010/057892 A1, 2010.

[11] L. De-Sousa and B. Bouchez, "Method and electric combined device for powering and charging with compensation means," Article ID 057893, International Patent WO 2010/057893 A1, 2010.

[12] S. Haghbin, M. Alakula, K. Khan et al., "An integrated charger for plug-in hybrid electric vehicles based on a special interior permanent magnet motor," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC '10)*, pp. 1–6, IEEE, Lille, France, September 2010.

[13] S. Lacroix, E. Laboure, and M. Hilairet, "An integrated fast battery charger for electric vehicle," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC '10)*, Lille, France, September 2010.

[14] D.-G. Woo, G.-Y. Choe, J.-S. Kim, B.-K. Lee, J. Hur, and G.-B. Kang, "Comparison of integrated battery chargers for plug-in hybrid electric vehicles: topology and control," in *Proceedings of the IEEE International Electric Machines and Drives Conference*, pp. 1294–1299, IEEE, Niagara Falls, Canada, May 2011.

[15] L. De Sousa, B. Silvestre, and B. Bouchez, "A combined multiphase electric drive and fast battery charger for electric vehicles," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC '10)*, pp. 1–6, Lille, France, September 2010.

[16] G. Pellegrino, E. Armando, and P. Guglielmi, "An integral battery charger with power factor correction for electric scooter," *IEEE Transactions on Power Electronics*, vol. 25, no. 3, pp. 751–759, 2010.

[17] H. Li, K. Zhang, H. Zhao, S. F. Fan, and J. Xiong, "Active power decoupling for high-power single-phase PWM rectifiers," *IEEE Transactions on Power Electronics*, vol. 28, no. 3, pp. 1308–1319, 2013.

[18] Y. Hayashi, K. Takao, T. Shimizu, and H. Ohashi, "High power density design methodology," in *Proceedings of the Power Conversion Conference—Nagoya (PCC '07)*, pp. 569–574, IEEE, Nagoya, Japan, April 2007.

[19] I. Laird, X. Yuan, and N. McNeill, "A holistic approach to optimise the power density of a silicon carbide (SiC) MOSFET based three-phase inverter," in *Proceedings of the IEEE 11th International Conference on Power Electronics and Drive Systems (PEDS '15)*, pp. 473–478, Sydney, Australia, June 2015.

[20] M. Yilmaz and P. T. Krein, "Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles," *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pp. 2151–2169, 2013.

[21] Y. Hu, X. Song, W. Cao, and B. Ji, "New SR drive with integrated charging capacity for plug-in hybrid electric vehicles (PHEVs)," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 10, pp. 5722–5731, 2014.

[22] Y. Hu, C. Gan, W. Cao, C. Li, and S. J. Finney, "Split converter-fed SRM drive for flexible charging in EV/HEV applications," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 10, pp. 6085–6095, 2015.

[23] A. Y. Saber and G. K. Venayagamoorthy, "One million plug-in electric vehicles on the road by 2015," in *Proceedings of the 12th International IEEE Conference on Intelligent Transportation*

Systems (ITSC '09), pp. 141–147, IEEE, St. Louis, Mo, USA, October 2009.

[24] T. Shimizu, Y. Jin, and G. Kimura, "DC ripple current reduction on a single-phase PWM voltage-source rectifier," *IEEE Transactions on Industry Applications*, vol. 36, no. 5, pp. 1419–1429, 2000.

[25] X. Yuan and A. Lovett, "DC-link capacitance reduction in a high power medium voltage modular wind power converter," in *Proceedings of the 15th European Conference on Power Electronics and Applications (EPE '13)*, pp. 1–10, Lille, France, September 2013.

[26] R. Wang, F. Wang, D. Boroyevich et al., "A high power density single-phase PWM rectifier with active ripple energy storage," *IEEE Transactions on Power Electronics*, vol. 26, no. 5, pp. 1430–1443, 2011.

[27] H. Li, K. Zhang, H. Zhao, S. Fan, and J. Xiong, "Active power decoupling for high-power single-phase PWM rectifiers," *IEEE Transactions on Power Electronics*, vol. 28, no. 3, pp. 1308–1319, 2013.

[28] W. Qi, H. Wang, X. Tan, G. Wang, and K. D. T. Ngo, "A novel active power decoupling single-phase PWM rectifier topology," in *Proceedings of the 29th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '14)*, pp. 89–95, IEEE, Fort Worth, Tex, USA, March 2014.

[29] X. Yuan, "Analytical averaged loss model of a three-level T-type converter," in *Proceedings of the 7th IET International Conference on Power Electronics, Machines and Drives*, pp. 1–6, April 2014.

[30] J. Liu and K. L. Wong, *Performance Evaluation of Hard-Switching Interleaved DC/DC Boost Converter with New Generation Silicon CarbideMOSFETs*, Cree, Durham, NC, USA.

[31] B. Callanan, "Application considerations for silicon carbide MOSFETs," Application Considerations for SiC MOSFET, January 2011.

[32] R. J. Callanan, A. Agarwal, A. Burk et al., "Recent progress in SiC DMOSFETs and JBS diodes at Cree," in *Proceedings of the 34th Annual Conference of IEEE Industrial Electronics (IECON '08)*, pp. 2885–2890, IEEE, Orlando, Fla, USA, November 2008.

[33] K. Ma, M. Liserre, and F. Blaabjerg, "Lifetime estimation for the power semiconductors considering mission profiles in wind power converter," in *Proceedings of the 5th Annual IEEE Energy Conversion Congress and Exhibition (ECCE '13)*, pp. 2962–2971, Denver, Colo, USA, September 2013.

[34] J. Lemmens, P. Vanassche, and J. Driesen, "Optimal control of traction motor drives under electrothermal constraints," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 2, no. 2, pp. 249–263, 2014.

[35] X. Yuan and A. Lovett, "DC-link capacitance reduction in a high power medium voltage modular wind power converter," in *Proceedings of the 15th European Conference on Power Electronics and Applications (EPE '13)*, pp. 1–10, IEEE, Lille, France, September 2013.

[36] http://www.farnell.com/datasheets/1821845.pdf.

[37] SAE Electric Vehicle and Plug-in Hybrid Electric Vehicle Conductive Charge Coupler, SAE standard J1772, January 2010.

Research Article

An Optimization Scheduling Model for Wind Power and Thermal Power with Energy Storage System considering Carbon Emission Trading

Huan-huan Li, Li-wei Ju, Qing-kun Tan, He Xin, and Zhong-fu Tan

School of Economics and Management, North China Electric Power University, Beijing 102206, China

Correspondence should be addressed to Huan-huan Li; 942893074@qq.com

Wind power has the characteristics of randomness and intermittence, which influences power system safety and stable operation. To alleviate the effect of wind power grid connection and improve power system's wind power consumptive capability, this paper took emission trading and energy storage system into consideration and built an optimization model for thermal-wind power system and energy storage systems collaborative scheduling. A simulation based on 10 thermal units and wind farms with 2800 MW installed capacity verified the correctness of the models put forward by this paper. According to the simulation results, the introduction of carbon emission trading can improve wind power consumptive capability and cut down the average coal consumption per unit of power. The introduction of energy storage system can smooth wind power output curve and suppress power fluctuations. The optimization effects achieve the best when both of carbon emission trading and energy storage system work at the same time.

1. Introduction

Carbon emission trading can promote large-scale wind power development and help power industry achieve energy-saving. Wind power output has the characteristics of randomness and intermittence, which puts impact on power system safety and stable operation and makes consumptive problem become the major factor that hinders large-scale wind power development. Power charging and discharging function of energy storage system can smooth wind power output curve, restrain power fluctuation, and provide backup services for wind power grid integration. Therefore, analysis optimization model for thermal-wind power system and energy storage system collaboration scheduling considering carbon emission trading has important sense in improving power system's wind power consumptive capability.

Literature [1] regarded carbon emission as virtual network flow that attached to the power flow. Based on the analysis results of carbon emission trading and power industry developing trend, a theoretical framework for power system carbon emission was built. Literature [2–4] studied carbon emission right definition problems in cross-regional power trading. An emission right allocation principle was obtained according to the carbon flow tracking mathematical model put forward by those literatures. Literature [5–8] discussed the effects of carbon emission trading. While meeting load demand, CO_2 emission can be effectively controlled when carbon trading and energy storage systems are both considered.

Wind power output is random and intermittent, which makes its output hard to be accurately predicted [9]. This factor makes wind power consumptive problem hard to be solved. The most effective solution is to control wind power output characteristics [10, 11]. Wind-thermal power system and energy storage system collaborative scheduling provides an effective way to solve this problem [12–14]. Literature [15, 16] put forward an operation model for multiple-time-scale hybrid system collaborative scheduling, which can arrange wind power and energy storage online and provide the specific scheduling information for system operators. Literature [17, 18] studied specific energy storage measures and built a planning model for hybrid system joint scheduling model that combined wind power and pumped storage.

The rest of this paper is structured as follows. Section 2 introduces the optimization models of wind-thermal system in different scenarios, including the basic optimization model, the model considering carbon emission trading, the model considering energy storage system, and the model considering both of carbon emission trading and energy storage system. To verify the correctness of the models put forward by this paper and analyze the influence of carbon emission trading and energy storage system on power system's wind power consumptive capability, Section 3 demonstrates a numerical example analysis based on 10 thermal power units and wind farms with 2800 MW installed capacity. Section 4 concludes this paper.

2. Optimization for Wind-Thermal System in Different Conditions

2.1. Optimization Model for Wind-Thermal System Collaborative Scheduling. Wind-thermal system optimization collaborative scheduling is aimed at improving wind power consumptive capability. However, wind power grid connection needs thermal power providing backup service to meet demand load. Therefore, overemphasizing the improvement of wind power consumptive capacity would require higher backup service level, which makes related thermal units start up and shut down more frequently and brings more coal consumption and pollutant emission. To improve wind power consumptive capacity and control thermal units' startup and shutdown, this paper builds a scheduling optimization model of wind power and thermal power. Maximizing the total profit is the optimization objective as follows:

$$\max z_1 = \pi_w + \pi_c, \tag{1}$$

wherein π_w is the total profit of wind farms; π_c is the total profit of thermal power units. π_w and π_c could be, respectively, calculated by

$$\pi_w = p_w \sum_{t=1}^{T} Q_{w,t} \left(1 - \theta_w\right) - OM_w - D_w,$$

$$\pi_c = p_c \sum_{i=1}^{I} \sum_{t=1}^{T} Q_{i,t} \left(1 - \theta_{c,i}\right) - C_{\text{fuel}} - \sum_{i=1}^{I} OM_{c,i} - \sum_{i=1}^{I} D_{c,i}, \tag{2}$$

wherein p_c is the benchmark price of thermal power in power output area; $Q_{i,t}$ is the real-time output of thermal unit i at time t; $\theta_{c,i}$ is power consumption rate of thermal unit i; C_{fuel} is fuel cost; $OM_{c,i}$ is maintenance cost of thermal unit i; $D_{c,i}$ is depreciation cost of thermal unit i.

Fuel cost mainly consists of coal cost and oil cost as follows:

$$C_{\text{fuel}} = \sum_{i=1}^{I} \sum_{t=1}^{T} \left[p_{\text{coal}} u_{i,t} f_i \left(Q_{i,t}\right) + u_{i,t} \left(1 - u_{i,t-1}\right) SU_i \right.$$
$$\left. + u_{i,t-1} \left(1 - u_{i,t}\right) SD_i \right], \tag{3}$$

wherein p_{coal} is the procurement price of standard coal; $Q_{i,t}$ is the real-time output of thermal unit i at time t; $u_{i,t} f_i(Q_{i,t})$

is standard coal consumption of thermal unit i with real-time output $Q_{i,t}$; $u_{i,t}$ is an operation or stop status variable of thermal unit i at time t, if the unit stop $u_{i,t} = 0$ and coal consumption is 0; else $u_{i,t} = 1$ and coal consumption could be calculated by the consumption characteristic function $f_i(\cdot)$ and real-time generation output $Q_{i,t}$. The consumption characteristic function is

$$f_i \left(Q_{i,t}\right) = a_i + b_i Q_{i,t} + c_i Q_{i,t}^2, \tag{4}$$

wherein a_i, b_i, and c_i are parameters of coal consumption function and all greater than 0; $u_{i,t}(1 - u_{i,t-1})SU_i$ is startup cost of thermal power unit i at time t, if and only if $u_{i,t} = 1$ and $u_{i,t-1} = 0$, and $u_{i,t}(1 - u_{i,t-1})SU_i$ does not equal zero; SU_i is the cost of a single startup of thermal unit i, including coal and oil costs; $u_{i,t-1}(1 - u_{i,t})SD_i$ is shutdown cost of thermal unit i at time t, if and only if $u_{i,t-1} = 1$ and $u_{i,t} = 0$, and $u_{i,t-1}(1 - u_{i,t})SD_i$ does not equal zero; SD_i is the cost of a single shutdown of thermal unit i, including coal and oil costs.

The constraints mainly consist of three aspects, namely, demand side, wind power, and thermal output constraints.

(1) Equilibrium Constraint of Power Supply and Demand. Consider

$$\sum_{i=1}^{I} u_{i,t} Q_{i,t} \left(1 - \theta_i\right) + Q_{w,t} \left(1 - \theta_w\right) = \frac{G_t}{(1 - l)}, \tag{5}$$

wherein G_t is the demand load at time t; l is the line losses rate of power system.

(2) Backup Service Constraints. When power system operates, generation side may be inconsistent with demand side. To ensure real-time equilibrium between supply and demand, thermal units should adjust their outputs to coordinate wind power output to meet load demand. The adjustments should meet some constraints, which are depending on thermal units' characteristics:

$$\sum_{i=1}^{I} u_{i,t} \left(Q_{i,t}^{\max} - Q_{i,t}\right) \left(1 - \theta_i\right) \geq R_t^{\text{usr}},$$

$$Q_{i,t}^{\max} = \min \left(u_{i,t-1}\overline{Q_i}, Q_{i,t-1} + \Delta Q_i^+\right) \cdot u_{i,t-1}, \tag{6}$$

$$R_t^{\text{usr}} = \beta_c \sum_{i=1}^{I} Q_{i,t} + \beta_w Q_{w,t}.$$

Equations (6) are upper spinning reserve constraints, wherein $Q_{i,t}^{\max}$ is the maximum possible output of unit i at time t; R_t^{usr} is upper spinning reserve demand, depending on thermal and wind power output in corresponding periods; $\overline{Q_i}$ is the maximum possible output of unit i in unit period, which is determined by installed capacity; ΔQ_i^+ is the upper limit of power climbing speed, namely, the biggest power increment of unit i in unit period; β_c is thermal power units' power

reserve coefficient; β_w is power reserve coefficient of wind turbine. Consider

$$\sum_{i=1}^{I} Q_{i,t} \left(Q_{i,t} - Q_{i,t}^{\min} \right) \left(1 - \theta_i \right) \geq R_t^{\mathrm{dsr}},$$

$$Q_{i,t+1}^{\min} = \max \left(u_{i,t} \underline{Q_i}, Q_{i,t} - \Delta Q_i^- \right) \cdot u_{i,t}, \qquad (7)$$

$$R_t^{\mathrm{dsr}} = \beta_w Q_{w,t}.$$

Equations (7) are lower spinning reserve constraints. $Q_{i,t}^{\min}$ is the minimum output of unit i at time t, which is determined by two factors: one is the minimum output of unit i at the starting state and the other is output-decreasing constraint of unit i in unit period; R_t^{dsr} is the lower spinning reserve demand, depending on wind power output in corresponding period; $\underline{Q_i}$ is the minimum output of unit i at the starting state in unit time, which is the same as the minimum real-time power output; ΔQ_i^- is output-decreasing speed of unit i, namely, the maximum output decrement in unit time.

(3) Thermal Unit Real-Time Output Power Constraint. Real-time output of a thermal power unit is limited by its installed capacity and minimum power output as follows:

$$u_{i,t} \underline{Q_i} \leq Q_{i,t} \leq u_{i,t} \overline{Q_i}. \qquad (8)$$

(4) Output Climbing Speed Constraints. Influenced by technical level, thermal power output changes in a unit period are limited. Real-time output power increment and decrement should meet

$$\Delta Q_i^- \leq Q_{i,t} - Q_{i,t-1} \leq \Delta Q_i^+. \qquad (9)$$

(5) Unit Startup and Shutdown Time Constraints. Frequently startup and shutdown are harmful to unit's performance and cause more fuel cost. Therefore, constraints for the continuous startup and shutdown time are necessary, as shown in the following equations:

$$\left(T_{i,t-1}^{\mathrm{on}} - M_i^{\mathrm{on}} \right) \left(u_{i,t-1} - u_{i,t} \right) \geq 0, \qquad (10)$$

$$\left(T_{i,t-1}^{\mathrm{off}} - M_i^{\mathrm{off}} \right) \left(u_{i,t} - u_{i,t-1} \right) \geq 0. \qquad (11)$$

Equation (10) is the minimum startup time constraint; $T_{i,t-1}^{\mathrm{on}}$ is continuous running time of unit i at time $t-1$; M_i^{on} is the minimum continuous running time. Equation (11) is the minimum downtime constraint; $T_{i,t-1}^{\mathrm{off}}$ is continuous downtime of unit i at moment $t-1$; M_i^{off} is the minimum continuous downtime.

(6) Wind Power Output Constraint. Wind turbine real-time power output is determined by income air velocity:

$$Q_{w,t} \leq \delta_t P_w, \qquad (12)$$

wherein δ_t is the equivalent efficiency of wind farms at time t; P_w is the total installed capacity of wind farms.

2.2. Optimization Model for Wind-Thermal System with Carbon Emission Trading. The introduction of carbon emission trading would redefine thermal power marginal cost, which consists of power generation cost and carbon emission cost. Pollutant emission coefficients of different units are not the same. Therefore, the introduction of carbon emission trading would change the original scheduling plan. To maximize system profit under carbon trading mechanism, this paper built an optimization model with the objective of maximizing the total profit of thermal power and wind power as follows:

$$\max z_2 = \pi_c + \pi_w. \qquad (13)$$

The profit of thermal power π_c can be calculated by

$$\pi_c = p_c \sum_{i=1}^{I} \sum_{t=1}^{T} Q_{i,t} \left(1 - \theta_{c,i} \right) - C_c - \sum_{i=1}^{I} \mathrm{OM}_{c,i} - \sum_{i=1}^{I} D_{c,i}, \qquad (14)$$

wherein p_c is the benchmark price of thermal power in power output area; $\theta_{c,i}$ is self-power-consumption rate of thermal unit i; $\mathrm{OM}_{c,i}$ is the maintenance cost of thermal unit i; $D_{c,i}$ is the depreciation cost of thermal power unit i.

Carbon emission trading mechanism uses emission cost to measure the environmental value of power generation plans. Without carbon emission trading, the variable cost of thermal units mainly consists of coal cost, oil cost, and water cost. But while carbon emission trading is considered, the thermal unit must buy CO_2 emission right when its emission exceeds the initial allocated quota level. Therefore, the variable cost of thermal power could be calculated by

$$C_c = C_{\mathrm{fuel}} + C_{CO_2}, \qquad (15)$$

wherein C_{fuel} is the fuel cost that includes coal cost and oil cost; C_{CO_2} is carbon emission cost, which could be calculated by

$$C_{CO_2} = \left(E_{CO_2} - E_0 \right) p_{CO_2}, \qquad (16)$$

wherein E_{CO_2} is actual carbon emission amount of thermal power; E_0 is the initial allocated carbon emission right, namely, the allowed CO_2 emission amount; p_{CO_2} is carbon emission trading price, determined by the supply and demand relationship in the carbon emission trading market. To simplify the optimization model, this paper hypothesized the price to be a constant value in a short period.

Thermal units' carbon emission amount is related to its power load rate. Generally speaking, carbon emission of unit electricity production can be integrated as a quadratic function as follows:

$$E_i \left(Q_{i,t} \right) = a_{CO_2,i} + b_{CO_2,i} Q_{i,t} + c_{CO_2,i} Q_{i,t}^2, \qquad (17)$$

wherein $a_{CO_2,i}$, $b_{CO_2,i}$, and $c_{CO_2,i}$ are parameters of the carbon emission function.

Total carbon emission of the system can be calculated by

$$E_{CO_2} = \sum_{t=1}^{T} \sum_{i=1}^{I} E_i \left(Q_{i,t} \right). \qquad (18)$$

Other constraints are the same as the basic model in Section 1, from (6) to (12).

2.3. Optimization Model for Wind-Thermal System with Energy Storage System.

While energy storage system is considered, the stakeholders would change from two to three parties. To maximize the total profit, the objective should be changed into (23). Consider

$$\max z_3 = \pi_c + \pi_w + \pi_s, \tag{19}$$

wherein π_s is the profit of energy storage system, determined by charge-discharge price, charge-discharge electric quantity, and fixed cost. Consider

$$\pi_s = p_{s,\text{char}} \sum_{t=1}^{T} Q_{s,t}^+ - p_{s,\text{disc}} \sum_{t=1}^{T} Q_{s,t}^- - F_s, \tag{20}$$

wherein $Q_{s,t}^+$ is energy storage system's charging power at time t; $Q_{s,t}^-$ is energy storage system's discharging power at time t; F_s is the fixed cost of energy storage system; $p_{s,\text{char}}$ and $p_{s,\text{disc}}$ are, respectively, charging and discharging price.

Energy storage system can control power charging and discharging, which make it have the characteristics of both power and load. Wind power output opposite distributed with demand load. In daytime, wind power is not sufficient to satisfy demand load and the energy storage system would act as power to meet the demand load. In night wind power output is far exceeding demand load and energy storage system would act as load to transform the extra wind power into potential energy.

The charge and discharge process of energy storage system is determined by real-time load equilibrium and the charging-discharging capability of the energy storage system. Assuming the power that energy storage stored in system s at time t is $Q_{s,t}$, charge-discharge power balance should meet (21). Consider

$$Q_{s,t} = Q_{s,t-1} + Q_{s,t}^+ - \frac{Q_{s,t}^-}{(1 - \theta_s)}, \tag{21}$$

wherein θ_s is discharging power at time t; θ_s is charge-discharge power loss coefficient that is reflecting the power loss during power transforming.

Charge-discharge capability in unit time is limited by technical level of the energy storage system, as shown in (22) and (23). Consider

$$Q_{s,t}^+ \leq \overline{Q_s}, \tag{22}$$

$$Q_{s,t}^- \leq \overline{Q_s}, \tag{23}$$

wherein $\overline{Q_s}$ is the upper limit of energy storage system charge-discharge power in unit time.

Energy storage system power storage capability also has its upper limit.

Consider

$$Q_{s,t} < Q_s^{\max}, \tag{24}$$

wherein $Q_s^{\max}$ is the maximum storage capacity of storage system.

For the entire study period, the charged power should be equal to the discharged power as follows:

$$\sum_{t=1}^{T} Q_{s,t}^+ (1 - \theta_s) = \sum_{t=1}^{T} Q_{s,t}^-. \tag{25}$$

Based on (25), to ensure energy storage system positive profit, charge and discharge price should meet

$$p_{s,\text{char}} > \frac{p_{s,\text{disc}}}{(1 - \theta_s)}. \tag{26}$$

At any moment power output should equal demand load as follows:

$$\sum_{i=1}^{I} u_{i,t} Q_{i,t} (1 - \theta_i) + Q_{w,t} (1 - \theta_w) + Q_{s,t}^-$$

$$= \frac{G_t}{(1 - l)} + Q_{s,t}^+. \tag{27}$$

Other constraints are the same as the basic model in Section 1, namely, from (6) to (12).

2.4. Optimization Model for Wind-Thermal System with Carbon Emission Trading and Energy Storage System.

The same with Section 3, in this optimization model the optimization objective is still consisting of three parts as follows:

$$\max z_4 = \pi_c + \pi_w + \pi_s, \tag{28}$$

wherein thermal power unit profit π_c is influenced by coal consumption, coal price, carbon emission, and carbon emission price. π_c could be calculated by

$$\pi_c = \left[p_c \sum_{i=1}^{I} \sum_{t=1}^{T} Q_{i,t} (1 - \theta_{c,i}) - C_{\text{fuel}} - C_{\text{CO}_2} \right.$$

$$\left. - \sum_{i=1}^{I} \text{OM}_{c,i} - \sum_{i=1}^{I} D_{c,i} \right], \tag{29}$$

wherein carbon emission cost is determined by the initial allocated carbon emission right and carbon emission price.

Wind power and thermal power real-time output, energy storage system charge-discharge power, and system load should comply with

$$\sum_{i=1}^{I} u_{i,t} Q_{i,t} (1 - \theta_i) + Q_{w,t} (1 - \theta_w) + Q_{s,t}^-$$

$$= \frac{G_t}{(1 - l)} + Q_{s,t}^+. \tag{30}$$

Other constraints are the same as the basic model in Section 1, from (6) to (12).

TABLE 1: Equivalent utilization of wind power units (MW).

Period	Load	Utilization ratio	Period	Load	Utilization rate	Period	Load	Utilization rate
1	1100	0.33	9	2300	0.28	17	1700	0.32
2	1200	0.55	10	2500	0.11	18	1900	0.29
3	1400	0.68	11	2600	0.26	19	2100	0.17
4	1600	0.76	12	2500	0.23	20	2500	0.13
5	1700	0.67	13	2400	0.12	21	2300	0.23
6	1900	0.51	14	2300	0.20	22	1900	0.38
7	2000	0.36	15	2100	0.09	23	1500	0.33
8	2100	0.32	16	1800	0.21	24	1300	0.38

TABLE 2: Dispatching optimization result of power system under different scenarios.

Scenario	Wind power			Thermal power			Profit (10^4 Yuan)
	Output (MW·h)	Grid accessed rate (%)	Abandoned rate (%)	Output (MW·h)	Grid accessed rate (%)	Coal consumption (kg/MW·h)	
1	18407.1	35.1	16.9	35274.8	64.9	343.5	327.8
2	18413.6	35.1	16.9	35294.8	64.9	346.8	295.9
3	18896.9	36.0	14.7	34772.9	64.0	344.4	305.6

3. Numerical Example Analysis

3.1. Basic Data. This paper did a simulation based on 10 thermal power units and wind power farms with 2800 MW installed capacity. Thermal power units' operating parameters are referred to in literature [19]. A typical day's system load and wind load output data is shown in Table 1. Assume wind power tariff to be 540 Yuan/MW·h, maintenance and depreciation costs to be 600 million, and thermal power tariff to be 380 Yuan/MW·h, equivalent to 800 Yuan/t of standard coal price.

3.2. Numerical Example Results. With the optimization objective of maximizing total profit, this paper solved the scheduling optimization model of wind and thermal power with or without carbon trading and energy storage system by the mean of GAMS.

3.2.1. Carbon Emission Trading's Impact on Wind Power Consumption. To study different carbon emission prices' impact on wind power consumption, three carbon emission mechanisms scenarios are set. Carbon emission trading is not considered in scenario 1; namely, the carbon trading is not levied. In other two parts the part where carbon emission quantity exceeds initial carbon emission right should levy emission fee. The fee is 80 Yuan/t in scenario 2 and 100 Yuan/t in scenario 3.

In scenario 1, total carbon emission is 29079.7 t. Assume the allocated carbon emission right is 98% of the total emission; the initial carbon emission right that thermal power gains would be 28498.1 t. Wind power consumptive optimization results under different carbon emission mechanisms are listed in Table 2.

According to Table 2, price increasing of carbon emission trading would increase wind power output and decrease abandoned wind. In scenario 1, wind power generation
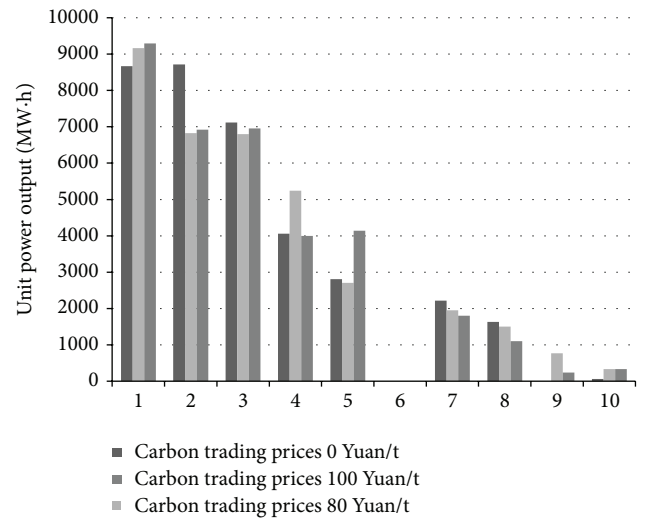
FIGURE 1: Thermal power output with different carbon emission trading prices.

is 18407.1 MW·h. When carbon emission trading price is 80 Yuan/t, wind power output is increased to 18413.6 MW·h. When carbon emission trading price is 100 Yuan/t, wind power output is 18896.9 MW·h and abandoned wind rate would decrease to 14.7%. Thermal power outputs with different carbon emission trading prices are shown in Figure 1.

The introduction of carbon emission trading mechanism makes thermal power market structure change with the margin output cost. For example, units 2# and 3# are with big carbon emission coefficients; the increasing of carbon trading price would decrease their output. Contrarily, unit 5# has small carbon emission coefficients; then the increasing of carbon emission trading would increase its power output. To meet system supply and demand balance constraints,

TABLE 3: Dispatching optimization result of power system under different scenarios.

Scenario	Wind power			Thermal power			Profit (10^4 Yuan)
	Generation (MW·h)	Electricity grid accessed rate (%)	Wind abandon rate (%)	Generation	Electricity grid accessed rate (%)	Coal consumption (kg/MW·h)	
1	18407.1	35.1	16.9	35274.8	64.9	343.5	327.8
2	18542.1	35.3	16.3	35237.2	64.7	344.3	301.1
3	18620.6	35.4	15.9	35252.5	64.6	344.6	290.0

TABLE 4: Dispatching optimization result of power system under different scenarios.

Scenario	Wind power			Thermal power			Profit (10^4 Yuan)
	Generation (MW·h)	Electricity grid accessed rate (%)	Wind abandon rate (%)	Generation (MW·h)	Electricity grid accessed rate (%)	Coal consumption (kg/MW·h)	
1	18407.1	35.1	16.9	35274.8	64.9	343.5	327.8
2	18620.6	35.4	15.9	35252.5	64.6	344.6	290.0
3	18896.9	36.0	14.7	34772.9	64.0	344.4	305.6
4	18963.2	36.1	14.4	34837.5	63.9	342.6	296.4

unit backup constraints, and unit output constraints, thermal power generation structure does not show obvious change regulation.

3.2.2. Energy Storage System's Impact on Wind Power Consumption.

To study energy storage system impact on wind consumption, this paper sets three scenarios according to the energy storage capacity. And optimization results are shown in Table 3.

With the access and scale-expansion of energy storage system, abandoned wind rate showed a downward trend and unit utilization efficiency increased gradually. Without energy storage system, abandoned wind rate is 16.9%. With 20 MW energy storage system connected to the system, abandoned wind rate decreased to 16.3% and electric quantity increased by 135.0 MW. With 40 MW energy storage system connected to the system, abandoned wind rate decreased to 15.9%, and electric quantity increased by 213.5 MW.

When decreasing the abandoned wind rate, thermal power output decreased and its net coal consumption rate increased to some extent. Without energy storage system, unit output coal consumption quantity is 343.5 kg/MW·h. And when the connected energy storage system is with 20 MW power storage capacity, the unit coal consumption quantity changes to be 344.3 kg/MW·h and 344.6 kg/MW·h, separately increased by 0.8 kg/MW·h and 1.1 kg/MW·h.

From the aspect of system profit, system profit decreased with the increasing power storage capacity, which is because of energy storage system's high investment cost and lack of commercial promotion in large scale. From the aspect of policy, China's policies gradually concentrate on large-scale energy storage system development but still lack industrial planning, industrial standards and financial subsidy, and other substantive supports. From the aspect of economic benefits, only pumped storage power plants can gain good economic benefit; other storage techniques are constrained by the high investment cost and unsound energy storage electricity price mechanism.

The development of China's large-scale energy storage system means both opportunities and challenges. Currently, challenges that are brought by price mechanism and investment cost are much more than opportunities. But in the long run, with the establishment of price mechanism and mature energy storage technology, China's large-scale energy storage system has a huge potential market.

3.2.3. Carbon Trading and Energy Storage System's Impact on Wind Consumption.

To compare wind power consumption of different combinations, 4 scenarios were set according to wind consumption combination with assistance of generation side. Scenario 1 is wind and thermal power joint scheduling optimization. Scenario 2 is wind and thermal power, energy storage system integrated scheduling optimization. Scenario 3 is wind and thermal power integrated scheduling optimization under carbon trading mechanism. Scenario 4 is wind power, thermal power, and energy storage system joint scheduling optimization with carbon trading mechanism. This paper uses GAMS to optimize. The optimization results are listed in Table 4.

From the aspect of wind power output, when there are only thermal power and wind power joint scheduled, the abandoned wind rate is 16.9%. With the introduction of carbon emission trading or energy storage system, abandoned wind rate decreased. And when both of carbon emission trading and energy storage system are considered, abandoned wind rate decreased to 14.4%, which achieves the minimum in 4 scenarios. From the aspect of thermal power, its power output and grid accessed rate decreased from scenario 1 to scenario 4.

From the aspect of system total profits, due to the high fixed cost of energy storage system, the profits will be higher without energy storage system. For energy storage system, its real-time charge and discharge power in scenario 4 and system power storage are shown in Figure 2. In scenario 4, energy storage system total charged power quantity is 488.1 MW·h, discharged power quantity is 346.9 MW·h, and
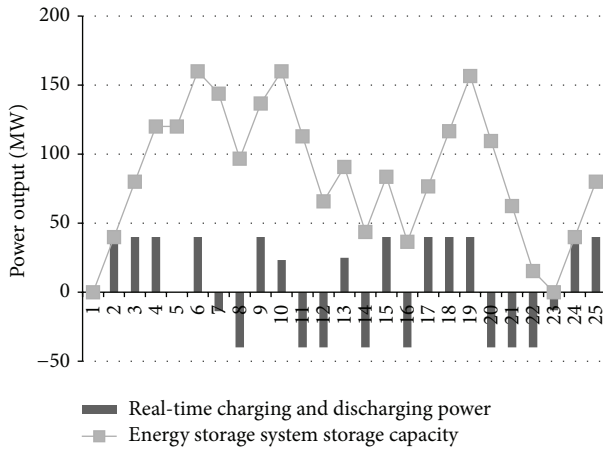
Figure 2: Charge and discharge optimization result of energy storage system.
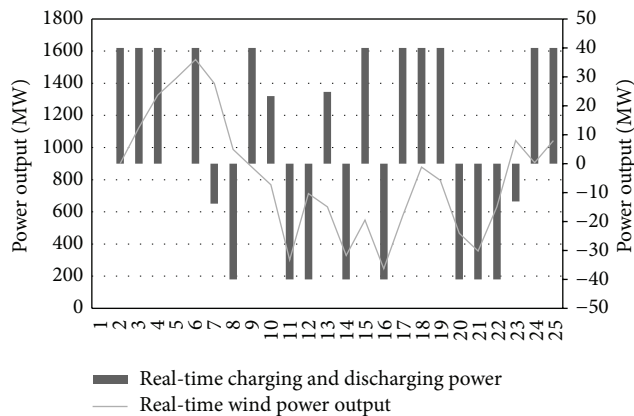
Figure 3: Comparison of wind power output and charge-discharge power of energy storage system.

final power storage is 80 MW·h. According to (22), total profit of energy storage system is −348000 Yuan. Wherein profit in charge-discharge process is 12000 Yuan, fixed cost is 360000 Yuan.

For energy storage system, it should release all the power in the final time to gain more economic benefits by selling the stored power. However, to reduce wind power output fluctuation's impact on system, charge and discharge decisions are determined by wind power output, which can reduce the pressure on thermal power peak shaving.

Figure 3 shows the change curves of real-time wind power output and energy storage charge-discharge power. In Figure 3, most of the charging time is in the wind power output-increasing period and most of the discharging time is in the wind power output-decreasing period.

From the aspect of system carbon emission level, thermal power's carbon emission is 28765.3 t in scenario 3, which is 267.2 t more than the initial allocated quota, and 26700 Yuan should be levied as the carbon emission cost. Thermal power's carbon emission is 28685.4 t in scenario 4, which is 187.3 t,

and is more than the initial allocation quota, and 187000 Yuan should be levied as the carbon emission cost.

Based on the above analysis, the introduction of carbon emission trading and energy storage systems can improve wind power consumptive capacity, improve wind power generation efficiency, and reduce thermal power output as well as coal consumption. However, due to the high fixed costs of energy storage system, power profits will be reduced by the access of energy storage system. For the examples in this paper, total charge quantity of the energy storage system is 488.1 MW·h, and total profit is −348000 Yuan.

4. Conclusion

To promote large-scale wind power grid connection and achieve energy-saving, this paper introduced carbon emission trading, which can bring economic benefits for wind power. To alleviate randomness and intermittence of wind power output and its impact on wind power consumptive capacity, this paper introduced energy storage system to provide backup services for wind power and built wind power energy storage collaborative scheduling optimization model with carbon emission trading and made a numerical example; the conclusions are as follows:

(1) Carbon emission trading can bring economic benefits for wind power and transform its cleaning feature into economic value, improve wind power grid connection, and reduce average coal consumption if there is power generation. The introduction of energy storage system can smooth wind power output, suppress fluctuation, and provide backup services for wind power that accessed the grid; electric quantity of wind power paralleling in the grid increased with the increasing capacity of energy storage system access.

(2) Energy storage system and carbon emission trading's introduction can achieve security and stability while running to maximize wind power capacity and increase economic benefits of wind power. However, due to the high fixed cost of energy storage systems, the above measures would reduce system total profit. Therefore, to maximize wind power utilization, related subsidies for the energy storage system need to be formulated.

Conflict of Interests

The authors declare that there is no conflict of interests.

Authors' Contribution

Huan-huan Li, Qing-kun Tan, and Li-wei Ju have built the demand response model, energy storage model, and power generation scheduling optimization model and He Xin and Zhong-fu Tan have improved the models and built the scenario simulation and reduction strategy and done the simulation analysis.

Acknowledgments

References

[1] J. Hetzer, D. C. Yu, and K. Bhattarai, "An economic dispatch model incorporating wind power," *IEEE Transactions on Energy Conversion*, vol. 23, no. 2, pp. 603–611, 2008.

[2] M. Grubb, T. Jamasb, and M. G. Pollitt, *Delivering a Low-Carbon Electricity System*, Crockett Book Company, 2008.

[3] IPCC, "Climate change 2007: the physical science basis," in *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK, 2007.

[4] X. Yao, H. Zhou, A. Zhang, and A. Li, "Regional energy efficiency, carbon emission performance and technology gaps in China: a meta-frontier non-radial directional distance function analysis," *Energy Policy*, vol. 84, pp. 142–154, 2015.

[5] L. Liu, C. Chen, Y. Zhao, and E. Zhao, "China's carbon-emissions trading: overview, challenges and future," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 254–266, 2015.

[6] A. M. Oestreich and I. Tsiakas, "Carbon emissions and stock returns: evidence from the EU emissions trading scheme," *Journal of Banking & Finance*, vol. 58, pp. 294–308, 2015.

[7] Z. Liao, X. Zhu, and J. Shi, "Case study on initial allocation of Shanghai carbon emission trading based on Shapley value," *Journal of Cleaner Production*, vol. 103, pp. 338–344, 2015.

[8] L. Tang, J. Wu, L. Yu, and Q. Bao, "Carbon emissions trading scheme exploration in China: a multi-agent-based model," *Energy Policy*, vol. 81, pp. 152–169, 2015.

[9] Y. Liu, L. Ren, Y. Li, and X.-G. Zhao, "The industrial performance of wind power industry in China," *Renewable and Sustainable Energy Reviews*, vol. 43, pp. 644–655, 2014.

[10] L. Baringo and A. J. Conejo, "Wind power investment within a market environment," *Applied Energy*, vol. 88, no. 9, pp. 3239–3247, 2011.

[11] Y. Zhou, L. Wang, and J. D. McCalley, "Designing effective and efficient incentive policies for renewable energy in generation expansion planning," *Applied Energy*, vol. 88, no. 6, pp. 2201–2209, 2011.

[12] C. Luo and B.-T. Ooi, "Frequency deviation of thermal power plants due to wind farms," *IEEE Transactions on Energy Conversion*, vol. 21, no. 3, pp. 708–716, 2006.

[13] H. Daneshi, A. K. Srivastava, and A. Daneshi, "Generation scheduling with integration of wind power and compressed air energy storage," in *Proceedings of the IEEE PES Transmission and Distribution Conference and Exposition*, pp. 1–6, IEEE, New Orleans, La, USA, April 2010.

[14] P. Zhao, J. Wang, and Y. Dai, "Capacity allocation of a hybrid energy storage system for power system peak shaving at high wind power penetration level," *Renewable Energy*, vol. 75, pp. 541–549, 2015.

[15] M. Moca, C. Litan, G. C. Silaghi, and G. Fedak, "Multi-criteria and satisfaction oriented scheduling for hybrid distributed computing infrastructures," *Future Generation Computer Systems*, 2015.

[16] S. Wen, G. Guo, and W. S. Wong, "Hybrid event-time-triggered networked control systems: scheduling-event-control co-design," *Information Sciences*, vol. 305, pp. 269–284, 2015.

[17] M. Dicorato, G. Forte, M. Pisani, and M. Trovato, "Planning and operating combined wind-storage system in electricity market," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 209–217, 2012.

[18] J. Garcia-Gonzalez, R. M. R. dela Muela, L. M. Santos, and A. M. Gonzalez, "Stochastic joint optimization of wind generation and pumped-storage units in an electricity market," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 460–468, 2008.

[19] D. Anyou, *Research on the Combined Approach and Optimization Model of Wind Power in China*, North China Electric Power University, 2014.

Research Article

Dynamic Modeling and Control Strategy Optimization for a Hybrid Electric Tracked Vehicle

Hong Wang,[1] Qiang Song,[1] Shengbo Wang,[2] and Pu Zeng[3]

[1]*Beijing Co-Innovation Center of Electric Vehicles, National Engineering Laboratory for Electric Vehicles,*
 Beijing Institute of Technology, No. 5 South Zhongguancun Street, Haidian District, Beijing 100081, China
[2]*Construction Machinery R&D Center, Intelligence Research Department, Shantui Construction Machinery Co., Ltd.,*
 No. 58 Highway G327, Jining, Shandong 272073, China
[3]*Shanghai Volkswagen, Ningbo 315336, China*

Correspondence should be addressed to Qiang Song; songqiang@bit.edu.cn

A new hybrid electric tracked bulldozer composed of an engine generator, two driving motors, and an ultracapacitor is put forward, which can provide high efficiencies and less fuel consumption comparing with traditional ones. This paper first presents the terramechanics of this hybrid electric tracked bulldozer. The driving dynamics for this tracked bulldozer is then analyzed. After that, based on analyzing the working characteristics of the engine, generator, and driving motors, the power train system model and control strategy optimization is established by using MATLAB/Simulink and OPTIMUS software. Simulation is performed under a representative working condition, and the results demonstrate that fuel economy of the HETV can be significantly improved.

1. Introduction

As a type of construction vehicle, the quantity of bulldozers is increasing significantly with tremendous social development, usually causing energy unsustainability and poor air quality. Additionally, the energy efficiency of the conventional bulldozer is only 20%. The electrification of the bulldozer as construction machinery is a good way to decrease air pollution and oil shortages. For these reasons, developing an electric bulldozer has significant effects on energy savings and emission reductions [1]. Studies have been performed to address these energy issues, notably regarding the utilization of electricity as a viable replacement of oil. Hybrid power systems are more suitable than traditional power trains for the reduction of fuel consumption and emissions. In the United States, Caterpillar produced the first D7E hybrid electric drive tracked bulldozer in March 2008. Compared with traditional models, CO and NO_x emissions were reduced by approximately 10% and 20%, respectively. The D7E can improve fuel economy by 25% [2].

Hybrid electric tracked bulldozer is a complex system due to its two power sources: engine-generator sets and ultracapacitor. The working speeds of the engine and charge-discharge properties of the ultracapacitor impact the fuel economy of the bulldozer together. A reasonable energy distribution control strategy could achieve the minimum fuel consumption. How to coordinate the power flow between engine-generator sets, ultracapacitor, and drive motor effectively to achieve the minimum fuel consumption is a complex design optimization problem.

In recent years, scholars have proposed a variety of different power allocation control strategies for electric vehicles [3–8], which can be roughly classified into two categories: thermostat type and power follow type [9–12]. The working principle of the thermostat type control strategy is as follows: engine outputs a set constant power when the SOC of battery is lower than the set minimum value; engine does not work when SOC is higher than the set maximum value. At this time, the battery needs to meet the transient high power discharge requirements that could do harm to the discharge

efficiency and the lifetime of the battery. Power follow control strategy requires the engine-generator sets output power to always follow the load demand power of the vehicle, causing engine speed frequently fluctuate and affecting the engine efficiency and emission properties. Xiong et al. [13] adopted the thermostat control strategy for PHEV, taking battery and ultracapacitor as the auxiliary power source. This control method improved the discharge efficiency and the lifetime of the auxiliary power source. However, it is complex, rarely used in the construction machinery vehicles, like bulldozer. Kwon et al. [14] adopted power follow control strategy for a series hybrid wheel excavator, taking ultracapacitor as the auxiliary power source which improved the fuel economy by 24%. However, this control method limited the improvement of the fuel economy as engine was controlled to work at one single point.

In this paper, one power follow control strategy is put forward to combine the working features of the proposed series hybrid electric tracked bulldozers [15, 16]. Bulldozers do not require high speed and acceleration performance. But they demand power when low-speed operation fluctuates remarkably. Considering working features of the bulldozer, ultracapacitor is taken as the auxiliary power source due to its high power density and short-term high power and current output capabilities. By analyzing the engine-generator working point and the SOC state of the ultracapacitor, a reasonable power control strategy is proposed to improve the fuel economy of the hybrid electric tracked bulldozer. Test bench experiment was performed to collect actual test bench data to correlate and validate the proposed control strategy for this hybrid electric tracked bulldozer.

Parameter optimization for the power follow control strategy is also researched because of the importance of accurate match of control strategy parameters for fuel economy improving [17, 18]. Commonly, genetic algorithm is famous for its global optimization and parallel computing capabilities [19–21] and has been widely used in multiple areas, including parameter estimation in system identification, optimization, and neural network training [22]. A genetic algorithm is established to solve the parameter optimization problem in the control strategy in this study.

The organization of this paper is as follows. In Section 2, the new configuration of the hybrid electric tracked bulldozer is described and the detailed terramechanics of the tracked bulldozer is researched to provide the theoretical basis for the modeling of the dynamics in Section 3. The power control strategy is described in Section 4. The optimization problem, procedure, and results of the control strategy are introduced in Section 5. Finally, the conclusions are presented in Section 6.

2. Terramechanics of the Tracked Bulldozer

2.1. Configuration of the Hybrid Electric Tracked Bulldozer.
The schematic of this hybrid electric tracked bulldozer (using one traditional bulldozer as the prototype) is given in Figure 1. This hybrid electric tracked bulldozer maintains the vehicle body, hydraulic and operation system of the traditional bulldozer, just changing the transmission form.

The paths for the electric power and mechanical power are tandem in the series configurations. This construction simplifies the propulsion and transmission system and provides greater flexibility in the power train system.

This hybrid electric tracked bulldozer uses an integrated controller to control two motors on both sides of the bulldozer independently and to transfer electric energy from the DC BUS into mechanical energy to drive the bulldozer.

The terramechanics of this hybrid electric tracked bulldozer is the theoretical basis for further analysis of the driving dynamics. The working states of the tracked bulldozer can be divided into the following six stages: soil-cutting, soil-transportation, returning, climbing, turning, and accelerating. The external travel resistance, operating resistance, driving force, and track slide curve under the working states are researched in the following context.

2.2. External Travel Resistance, F_E.
The resistance caused by the vertical deformation of the soil under the anterior track of the vehicle when driving is called external travel resistance. It mainly results from the energy consumption of soil compaction and the effects of bulldozing resistance. F_E can be written as the following [23]:

$$F_E = F_c + F_b,$$
$$F_c = \frac{2b}{(n+1)\,k^{1/n}}\left(\frac{G}{2bL}\right)^{(n+1)/n}, \tag{1}$$
$$F_b = \gamma Z^2 b k_\gamma + 2bZc k_{\text{pc}},$$

where

$$k_\gamma = \left(\frac{2N_\gamma}{\tan\psi} + 1\right)\cos^2\psi,$$
$$k_{\text{pc}} = \left(N_c - \tan\psi\right)\cos^2\psi, \tag{2}$$

where F_c is compaction resistance (N); F_b is bulldozing resistance (N); b is track width (m); G is vehicle weight (N); L is track length (m); c is soil cohesion coefficient (KPa); ψ is soil internal friction angle (°); n is soil deformation index; k is soil deformation modulus (KN/m^{n+2}); Z is track amount of sinkage (m); γ is unit weight (N/m^3); and N_γ and N_c are the soil Terzaghi coefficients of bearing capacity [24].

2.3. Operating Resistance F_T.
Consider

$$F_T = F_1 + F_2 + F_3 + F_4,$$
$$F_1 = 10^6 B_1 h_p k_b,$$
$$F_2 = G_t \mu_1 \cos\alpha = \frac{V\gamma\mu_1 \cos\theta}{k_s},$$
$$F_3 = 10^6 B_1 X \mu_2 k_y, \tag{3}$$
$$F_4 = G_t \mu_2 \cos\delta^2 \cos\theta,$$
$$V = \frac{B_1\left(H - h_p\right)^2 k_m}{2\tan\alpha_0},$$
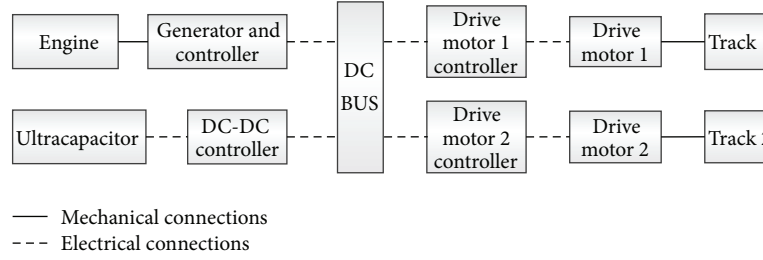
FIGURE 1: Configuration of hybrid electric tracked bulldozer.

where F_1 is soil-cutting resistance (N); F_2 is the pushing resistance of the mound before the blade (N); F_3 is frictional resistance between the ground and blade (N); F_4 is the horizontal component of the frictional resistance when the soil rises along the blade (N); k_b is cutting resistance per unit area (MPa); B_1 is blade width (m); h_p is average cutting depth (m); G_t is the gravity of the mound in front of the bulldozing plate; μ_1 is the friction coefficient between soil particles; μ_2 is the friction coefficient between the soil and blade; θ is slope (°); V is the volume of the mound in front of the bulldozing plate; k_s is the loose degree coefficient of the soil; k_m is the fullness degree coefficient of the soil; H is blade height (m); α_0 is the natural slope angle of the soil (°); k_y is cutting resistance per unit area after the blade is pressed into the soil (MPa); X is the length of the worn blade contacting the ground (m); and δ is the cutting angle of the blade (°) [1].

2.4. Driving Force, F. The soil driving force is partly consumed by overcoming external travel resistance and bulldozer operation, acceleration, climbing, or load traction. The relationship between maximum driving force (adhesion force) $F_{\max}$ and slide ratio i is given as the following [25]:

$$
F_{\max} = 2bLc \left(1 + \frac{2h}{b} \right)
$$
$$
+ G \tan \varphi \left\{ 1 + 0.64 \left[\frac{h}{b} \text{arc} \cot \left(\frac{h}{b} \right) \right] \right\}. \tag{4}
$$

Thus, driving force F can be written as the following:

$$
F = F_{\max} \left[1 - \frac{K}{iL} \left(1 - e^{-iL/K} \right) \right], \tag{5}
$$

where h is the grouser height (m) and K is the soil horizontal shear deformation modulus (m):

$$
i = 1 - \frac{v}{r_0 w_0} = 1 - \frac{v}{v_T} = \frac{v_T - v}{v_T} = \frac{v_j}{v_T}, \tag{6}
$$

where v is the actual speed of vehicle (m/s); w_0 is the angular velocity of the sprocket wheel (rad/s); r_0 is the pitch radius of the sprocket wheel (m); v_T is the theoretical speed of vehicle; v_j is the track slip velocity relative to the ground (rad/s).

2.5. Track Slide Curve. Based on the parameters of the bulldozer and sandy loam shown in Table 1, the track slide curve of this hybrid electric tracked bulldozer on sandy loam is shown in Figure 2.

TABLE 1: Parameters of the bulldozer and soil.

Name	Value	Unit
Vehicle weight (G)	280	KN
Track width (b)	0.61	m
Track length (L)	3.05	m
Grouser height (h)	0.07	m
Soil cohesion coefficient (c)	13.79	KPa
Soil internal friction angle (ψ)	28	Degree
Soil deformation index (n)	0.3	Null
Soil deformation modulus (k)	146	KN/m^{n+2}
Unit weight (γ)	17700	N/m^3
Soil Terzaghi coefficients of bearing capacity (N_c)	10.8	Null
Soil Terzaghi coefficients of bearing capacity (N_r)	3.8	Null
Soil horizontal shear deformation modulus (K)	0.02	m

From Figure 2, the driving force reaches a maximum of 213 KN when the track slips completely (slide ratio $i = 1$, meaning that the track slip velocity relative to the ground is equal to the theoretical speed of vehicle), and the difference between the driving force and external travel resistance is used for the operation, acceleration, climbing, and load traction of the bulldozer.

3. Dynamic Modeling of the Hybrid Electric Tracked Bulldozer

The series hybrid power system is composed of a diesel engine, permanent magnet generator, motor drive system, and tracks, as shown in Figure 1. The hybrid electric bulldozer uses the integrated controller to control the motors on both sides of the bulldozer independently and to transfer the electric energy from the generator and auxiliary power supply (also from DC BUS) into mechanical energy to drive the bulldozer [26]. Based on the working principle analysis and test bench experiment of various parts of the power train system, a mathematical model of the hybrid power system was established.

3.1. Engine-Generator Model. To ensure the accuracy of the modeling, test bench experimental data are used to establish
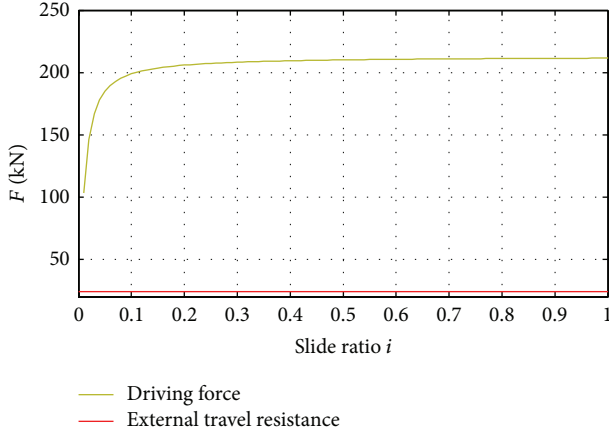
FIGURE 2: Track slide curve of the tracked bulldozer on sandy loam.
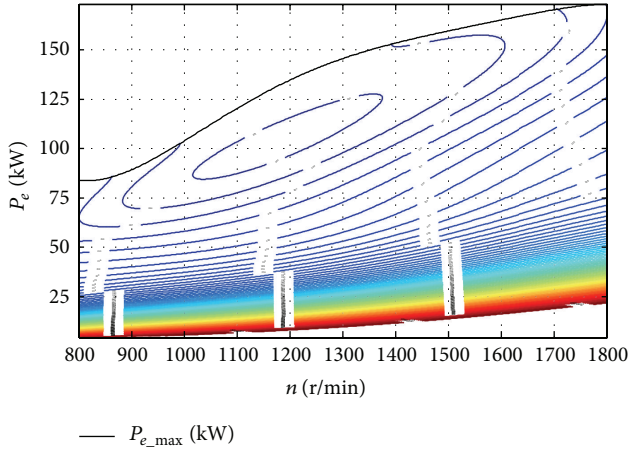
FIGURE 3: Diesel engine universal characteristic curve.

FIGURE 4: Motor drive system efficiency.

where s is model order number and A_k is the polynomial coefficient $k = (j^2 + j + 2 * i)/2$.

The generator provides current to DC BUS under generating mode; the kinetic equation for the generator is given as the following:

$$J_g \frac{dn_g}{dt} = T_g - T_e^D, \tag{9}$$

where T_g is generator shaft torque; n_g is generator speed; and J_g is the rotary inertia of the generator.

The relationship between generator shaft torque, speed and output DC voltage, and current of the controller side is given as the following:

$$\frac{U_g \cdot I_g}{1000 \cdot \eta_g} = \frac{T_g \cdot n_g}{9549}, \tag{10}$$

where U_g is output DC voltage from the generator; I_g is output DC current; η_g is power generation efficiency.

3.2. Motor Model. A permanent magnet motor is adopted in this hybrid bulldozer with a torque of 500/800 N·m, a power of 75/105 kW, a rated speed of 1430 rpm, and a maximum speed of 6000 rpm. In this power train system, the two motors follow the instructions of the control unit for torque output. The focus of the motor model is to establish the relationship between the output torque, speed of the motor and the input DC voltage, and current of the motor controller. Therefore, the motor model is obtained based on the test bench data of the motor drive system. Figure 4 shows the efficiency map of the system.

Considering the response time of the motor drive system, a first-order link is added between the target torque and the actual output torque:

the diesel engine universal characteristic curve, as shown in Figure 3.

The engine fuel injection amount is determined by throttle position and engine speed. Engine output torque is determined by the following:

$$T_e^D = T_e - J_e \frac{dw_e}{dt}, \tag{7}$$

$$T_e = \alpha * T_{e\_max} (n_e),$$

where T_e^D is dynamic engine output torque; T_e is steady-state engine flywheel output torque; α is throttle position; $T_{e\_max}(n_e)$ is maximum torque at the engine speed n_e; and J_e is the rotary inertia of the rotating parts in the engine.

The engine fuel consumption b_e (g/kWh) is a function of T_e and n_e:

$$b_e = f (T_e, n_e) = \sum_{j=0}^{s} \sum_{i=0}^{j} A_k * T_e^i * n_e^{j-1} \tag{8}$$

$$(i, j = 0, 1, 2, \ldots, s),$$

$$T_m = \begin{cases} \dfrac{T_{ref}}{\tau s + 1} & T_{ref} \le T_{max} (n) \\[3mm] \dfrac{T_{max} (n)}{\tau s + 1} & T_{ref} > T_{max} (n), \end{cases} \tag{11}$$

where T_{ref} is target torque; T_m is output torque; n is motor speed; $T_{max}(n)$ is maximum torque at speed n; and τ is response time.
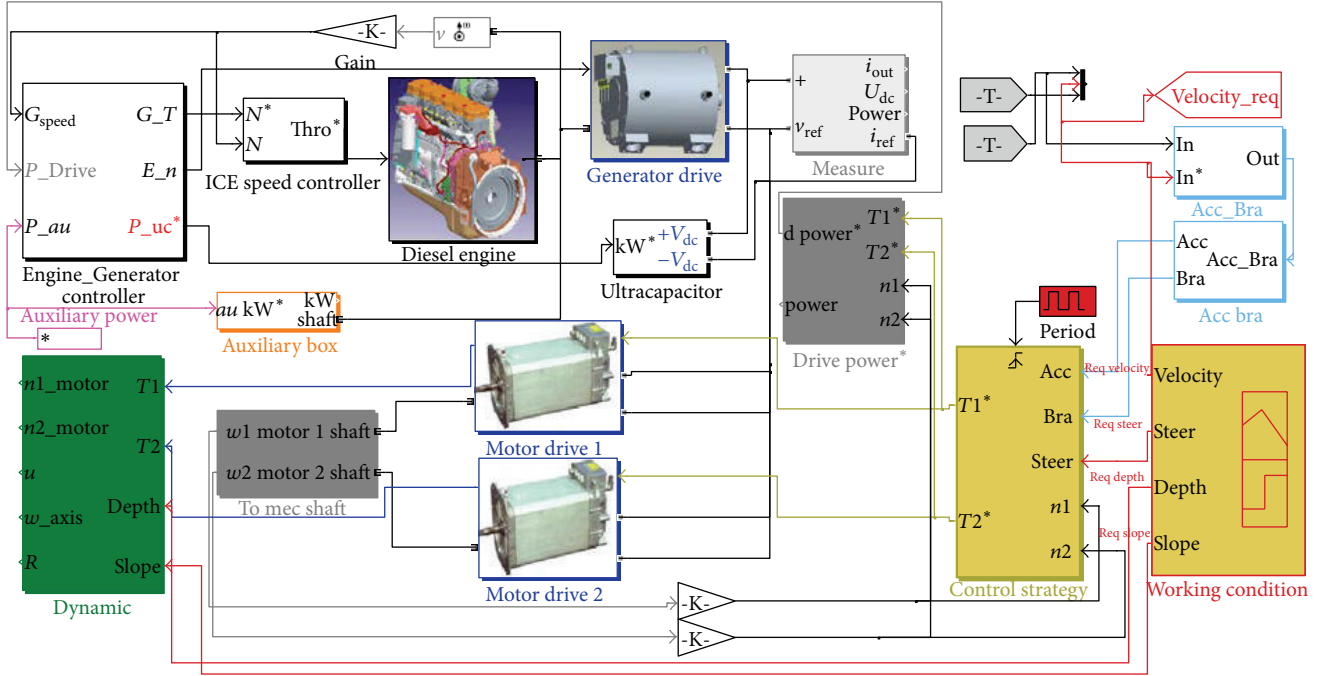
FIGURE 5: The power train system model of the hybrid electric tracked bulldozer.

The dynamic equation for the motor is as follows:

$$J_m \frac{dn}{dt} \frac{2\pi}{60} = T_m - T_{\text{load}}, \tag{12}$$

where J_m is motor rotational inertia and T_{load} is load torque.

The relationship between the input DC voltage, current of the motor controller and the shaft output torque, and speed is as follows:

$$\frac{U \cdot I \cdot \eta_d}{1000} = \frac{T_m \cdot n}{9549} \quad (T_m > 0),$$

$$\frac{U \cdot I}{1000 \cdot \eta_b} = \frac{T_m \cdot n}{9549} \quad (T_m < 0), \tag{13}$$

where U is the input DC voltage; I is the input DC current; η_d is motor efficiency when driving; and η_b is motor efficiency when braking.

3.3. Driving Dynamic Model. Based on the study of the terramechanics of this hybrid electric tracked bulldozer, a driving dynamic model is established as follows:

$$(T_1 + T_2) \frac{i_0 \eta}{r} - R_L - R_R = m\dot{v},$$

$$(T_1 - T_2) \frac{i_0 \eta}{r} \frac{B}{2} + (R_L - R_R) \frac{B}{2} - T_r = I\dot{w},$$

$$T_r = \begin{cases} 0, & w = 0, \\ \dfrac{\mu lmg}{4} \left[1 - \left(\dfrac{2\lambda_1}{l} \right)^2 \right], & w \neq 0, \end{cases} \tag{14}$$

where $T_{1,2}$ is motor output torque; i_0 is the gear ratio from the motor to the driving wheel; η is the efficiency from the motor shaft to the track; r is driving wheel radius; R_L and R_R are left and right side track resistance, respectively; m is vehicle mass; v is the vehicle drive speed along the longitudinal direction; T_r is steering resistance torque; I is vehicle rotational inertia; w is vehicle angular velocity; μ is steering resistance coefficient; l is vehicle length; and λ_1 is the offset of the track contact with the ground alone longitudinal direction.

The power train system simulation model of the hybrid electric tracked bulldozer including the driver model, working condition model, engine-generator model, motor drive system model, vehicle dynamics model, and control strategy in MATLAB/Simulink is shown in Figure 5, where (1)~(3) and (14) are used in the module "dynamic"; (7)~(8) are used in the module "diesel engine"; (9)~(10) are used in the module "generator drive"; (11)~(13) are used in the module "motor drive 1" and "motor drive 2."

3.4. Experiment Validity Analysis of the Dynamic Model of the Hybrid Bulldozer. To verify the accuracy of the simulation model of the hybrid electric tracked bulldozer, the parameters and the real working conditions (as shown in Figure 6) were substituted into this simulation model to compare the simulation results with the real vehicle experimental data.

In Figure 6, V (km/h) is the bulldozer velocity; depth (m) is soil-cut depth; and slope (°) is bulldozer gradeability. The working stages are described as follows: 1~4 s traveling stage; 4~16 s soil-cutting stage; 16~31 s soil-transportation stage; 31~33 s unloading soil stage; and 33~50 s no-load stage.

Figures 7 and 8 show the simulation result and real bulldozer's working velocities and drive forces of single track
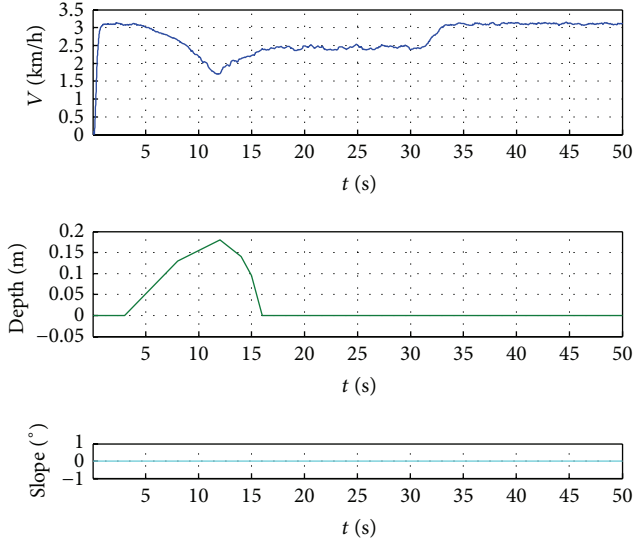
FIGURE 6: The real working condition of the bulldozer.
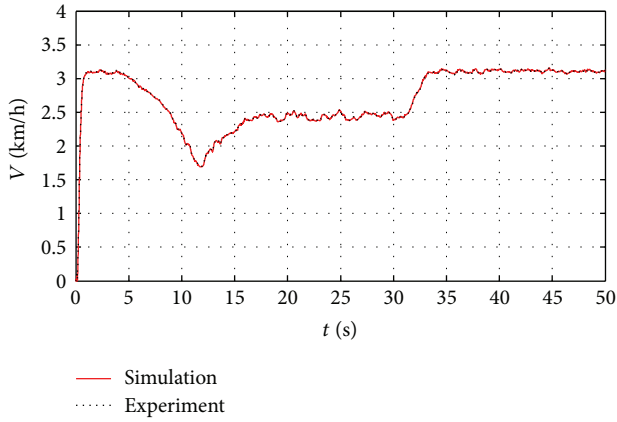
FIGURE 7: Bulldozer velocities in the simulation and experiment.

FIGURE 8: Single track drive forces in the simulation and experiment.
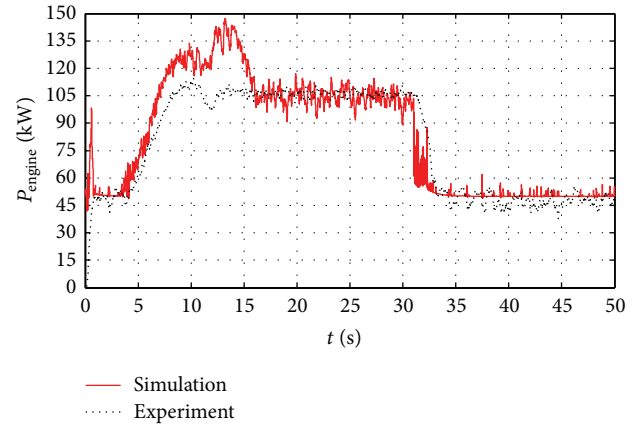
FIGURE 9: Engine output power in the simulation and experiment.

under the working condition shown in Figure 6. Figure 9 shows the engine output power in the simulation and experiment. Figure 10 shows the generator and motor efficiency under the working condition.

As we can see from Figures 7 to 8, bulldozer velocities and single track drive forces in the simulation and experimental data are identical well. As can be seen from data, the variance's relative error of the single track drive forces in the simulation and experimental data is 1.45%.

In Figure 9, the engine output power in the simulation and experimental data are identical well-expected 4–16 s soil-cutting stage. The engine output power in simulation is higher about 15~25 kW than that in experiment in 4–16 s soil-cutting stage. This is because the efficiency of the power train system is different in simulation (hybrid) and experiment (traditional).The experiment transmission efficiency is set approximately 0.9, which is higher than that of simulation transmission. Figure 10 shows that the simulation efficiency is below 0.90 for motor in 4–16 s soil-cutting stage. So the engine must provide higher power in simulation than that in

experiment in order to satisfy the same demand power of the bulldozer.

Under the state above, the simulation model (Figure 5) is valid and can be used for the simulation research of control strategy for the hybrid electric tracked bulldozer.

4. Control Strategy

4.1. Power Allocation Strategy. One power follow control strategy is put forward to combine the working features of the proposed series hybrid electric tracked bulldozers, which should be able to coordinate the power supply and need relationship among the engine generator, ultracapacitor, and the motors. The engine output power should follow the load demand power of the bulldozer, and the ultracapacitor should supply the power shortage caused by the excessive load demand power as the auxiliary power source. The SOC of the ultracapacitor and load power requirement determines the working point of the engine generator. The power allocation strategy is shown in Table 2.

In Table 2, P^* is the target demand mechanical power; P_e is engine output power; P_{e_max} is engine maximum output power; P_{e_min} is engine minimum output power; P_{DC} is DC
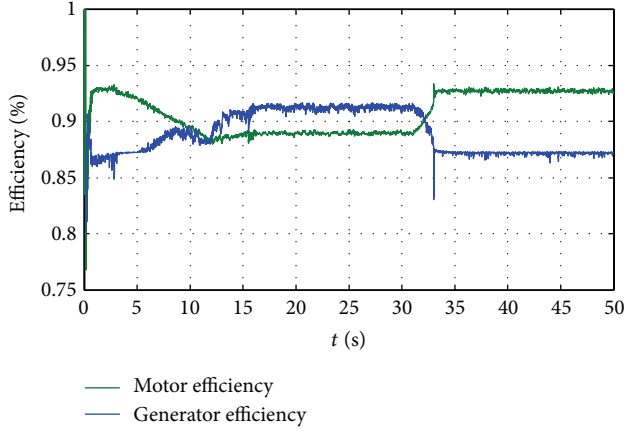
FIGURE 10: Generator and motor efficiency.

TABLE 2: Power allocation strategy.

Judgment	State of the ultracapacitor	Power supply
$P^* < P_{e\_max}$ SOC < SOC_{max}	Charging	$P_g = \eta_1 * P_e$ $P_{uc} = P_{DC} - P_g$
$P^* < P_{e\_max}$ SOC $\geq \mathrm{SOC}_{max}$	Not working	$P_g = \eta_1 * P_e$ $P_{uc} = 0$
$P^* > P_{e\_max}$ SOC > SOC_{min}	Discharging	$P_g = \eta_1 * P_{e\_max}$ $P_{uc} = P_{DC} - P_g$
$P^* > P_{e\_max}$ SOC $\leq \mathrm{SOC}_{min}$	Not working	$P_g = \eta_1 * P_{e\_max}$ $P_{uc} = 0$

BUS demand electric power; P_g is generator output power; P_{uc} is ultracapacitor power; η_1 is power efficiency of the generator; SOC_{max} and SOC_{min} are ultracapacitor maximum and minimum state of charge, respectively.

P^* is target demand mechanical power, which can be given as

$$P^* = \frac{P^*_{\text{Track}}}{\eta_{E\text{-}T}} = \frac{F_{\text{Track}} * r}{\eta_{E\text{-}T}} = \frac{(F_E + F_T) * r}{\eta_{E\text{-}T}}, \qquad (15)$$

where P^*_{Track} is the demand power of the track; $\eta_{T\text{-}E}$ is the transmission efficiency from engine to track; F_{Track} is the off-road motion resistance; F_E is external travel resistance calculated by (1)~(2); F_T is operating resistance calculated by (3).

As can be seen from (15) and (1)~(3), we can conclude that the average cutting depth h_p and the volume of the mound in front of the bulldozing plate V are linear to P^*; P^* is a quadratic function of the track amount of sinkage Z. These primary parameters h_p, V, and Z impact the energy management at the above level.

The flowchart for the power allocation strategy is shown in Figure 11. The driver's intention is taken as the target required power P^*, and the vehicle management system (VMS) calculates the drive motor required power P^*_m according to the working conditions of all components and the vehicle. Simultaneously, the VMS calculates the engine target
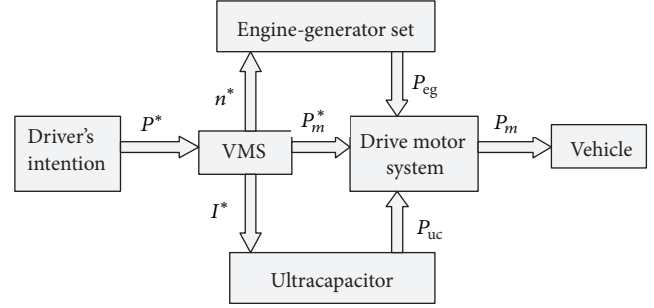
FIGURE 11: The flowchart for the power allocation strategy.

TABLE 3: Parameters of the control strategy.

Parameters	Instructions
SOC_{max}	Maximum SOC
SOC_{min}	Minimum SOC
N_1	Engine working speed 1
N_2	Engine working speed 2
N_3	Engine working speed 3

speed n^* and the ultracapacitor target current I^*. The engine-generator set makes the output power P_{eg}, whereas the ultracapacitor makes the output power P_{uc} according to a data table lookup. The output power of the engine-generator set and ultracapacitor is then sent to the drive motors through DC BUS.

In this control strategy, the parameters must be properly adjusted and optimized to reduce fuel consumption as much as possible in order to satisfy the operational requirements. The parameters of the control strategy are shown in Table 3.

4.2. Engine Control Strategy. An engine multipoint speed switching control strategy is adopted here, and engine working speed is determined by the bulldozer load demand power. The engine operates at low speeds when the load demand power is low and operates at higher speeds when the load demand power increases. Figure 12 shows the schematic of the engine multipoint speed switching control strategy.

In Figure 12, x-axis represents engine speed and y-axis represents load demand power. Load demand power is divided into three areas: low, medium, and high. The engine speed in each area is fixed as N_1, N_2, and N_3. When the load demand power is in low load area, engine speed will be fixed at N_1; as load demand power increases to medium load area, engine speed will be fixed at N_2; as load demand power increases to high load area, engine speed will be fixed at N_3. Power hysteresis band is set between each adjacent load power area to avoid the engine speed frequent switching.

4.3. Real Experimental Testing of Control Strategy. Test bench experiment was performed to collect actual test bench data to correlate and validate the proposed power follow control strategy for this hybrid electric tracked bulldozer. The experimental bench consists of engine, generator, drive
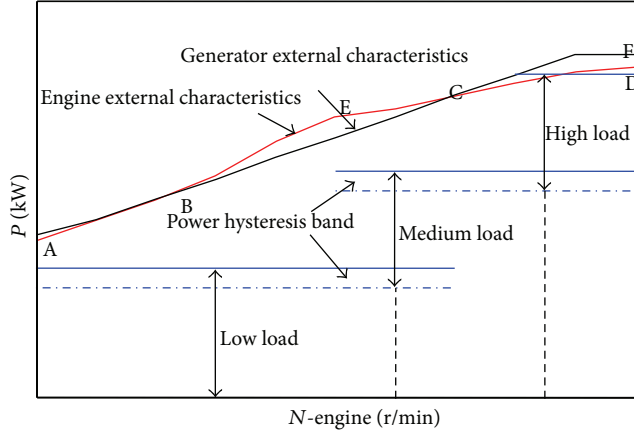
FIGURE 12: Schematic of the engine multipoint speed switching control strategy.

TABLE 4: Basic parameters of the power train system.

Name	Value	Unit
Maximum power of the engine (P_{e_max})	175	kW
Rated power of the motor (P_{m_rated})	75	kW
Maximum power of the motor (P_{m_max})	105	kW
Rated power of the generator (P_{g_rated})	175	kW
Maximum power of the generator (P_{g_max})	180	kW

motor system, auxiliary power source, dynamometer, and the vehicle management system, as shown in Figure 13(a).

This test bench adopts power follow control strategy which was written in the vehicle management system to coordinate the power supply among the engine generator and the motors. Dynamometer control interface shown in Figure 13(b) controls the working speed of the dynamometer and drive motor, and the drive motor controller controls the load torque of the drive motor. In the process of test, the load torque of the drive motor was increased up from 30 N·m to 100 N·m suddenly and then down to 30 N·m suddenly to simulate the load sudden changes process of the bulldozer, as shown in Figure 14. Load torque is linear related to the accelerator pedal opening degree.

As we can see from the test bench result shown in Figure 15, the engine-generator output power follows the load demand power well under the working condition shown in Figure 14. Under the state above, the proposed power follow control strategy is effective and can be used for the control strategy optimization for the hybrid electric tracked bulldozer.

5. Control Strategy Optimization

5.1. Optimization Model. To solve the problem of the parameter optimization of the power follow control strategy, a mathematical model of the entire power train system including the fuel consumption model is established based on the dynamic modeling in Section 3 and the control strategy in Section 4.

The important parameters of the power train system of the bulldozer are given in Table 4.

The fuel consumption model is obtained through the fuel consumption MAP graph shown in Figure 16, where b_e (g/Kwh) is fuel consumption rate. Minimizing fuel consumption is the optimization goal, and the working condition has no effect on the control strategy put forward here. Therefore, one representative bulldozer working condition is adopted here, as shown in Figure 17.

In this optimization problem, the object function is engine fuel consumption under the representative working condition. Engine minimum fuel consumption can be obtained by integration [27]:

$$\min B = \int_{t_0}^{t_1} B\left(n, P_e, t\right) dt, \tag{16}$$

where B is engine fuel consumption; n is engine speed at time t; P_e is engine output power at time t; and t_0 and t_1 are the start and end time of the bulldozer working condition, respectively.

In this problem, the SOC of the ultracapacitor at the start and end times should remain identical. Therefore, the sum of the ultracapacitor output and input energies is zero:

$$\int_{t_0}^{t_1} P_{uc}\left(t\right) dt = 0, \tag{17}$$

$$SOC_{min} \leq SOC \leq SOC_{max}.$$

The other constraint conditions are the following:

$$P_{e\_min} \leq P_e \leq P_{e\_max},$$
$$n_{e\_min} \leq n_e \leq n_{e\_max}, \tag{18}$$
$$I_{uc} \leq I_{uc\_max},$$

where n_{e_max} and n_{e_min} are the maximum and minimum speeds of the engine, respectively, and I_{uc_max} is the maximum charging/discharging current of the ultracapacitor.

5.2. Control Strategy Optimization. Based on the characteristics of the series hybrid tracked bulldozer, a genetic algorithm (GA) is adopted to solve the parameter optimization problem in the control strategy. The engine working speeds N_1, N_2, and N_3 at different power levels are set as the objective optimized parameters. The initial values of N_1, N_2, and N_3 are determined by engineering experience. Because GA cannot address the parameters directly, the parameters must be converted into a chromosome composed of genes with a certain structure. Therefore, the chromosome here is X_i (N_1, N_2, N_3). The GA operations are as follows: selection, crossover, and mutation.

A design of experiment (DOE) is used to simplify the calculation. As a design space exploration technique, DOE is used for preliminary design space exploration to reflect the relationship between the design variables and objective function in fewer numbers of trials. According to the given optimization variables, 16 test sample points are constructed

(a)

(b)
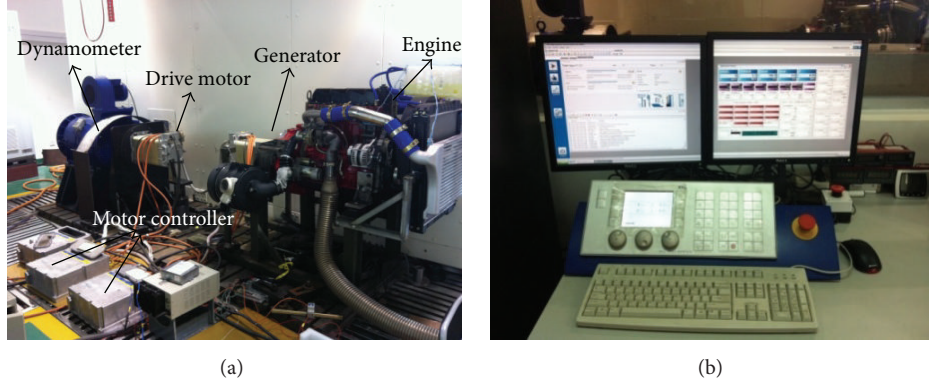
FIGURE 13: Hardware setup for experiments. (a) Structure of the hybrid bulldozer test bench; (b) dynamometer control interface.

(a) Accelerator pedal opening degree

(b) Load torque of the drive motor
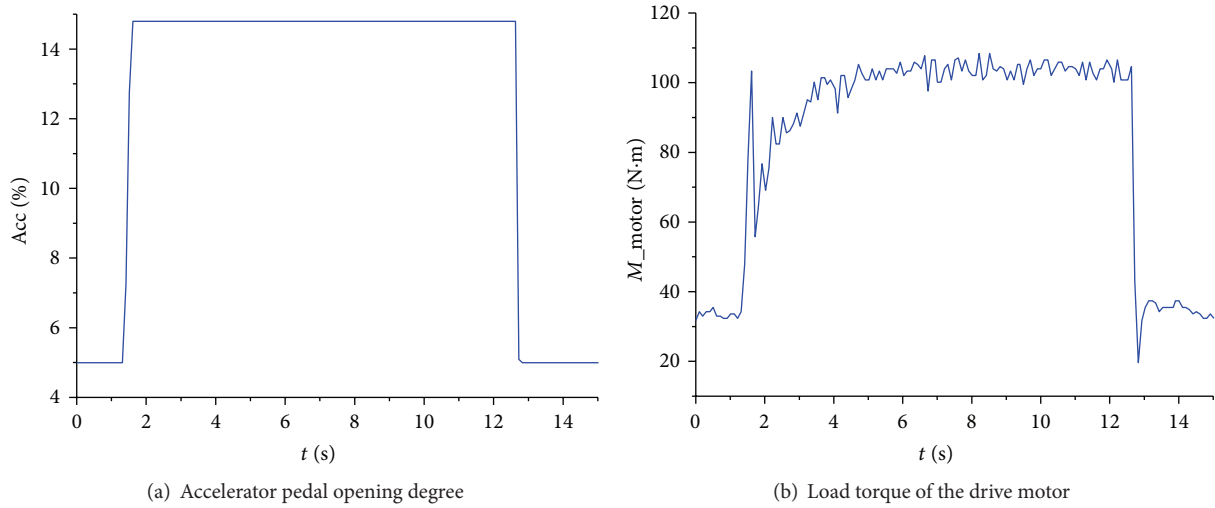
FIGURE 14: The test bench working condition.

using an orthogonal design method. The response surface is constructed on the basis of the DOE and can replace the original model with certain accuracy. On the basis of the DOE, the GA is then applied to optimize the parameters in the solution space to determine the optimal solution [28, 29]. The GA work flow is then established (Figure 18) by setting the minimum fuel consumption as the target function under typical working conditions.

In this flowchart, $f(x)$ and $g_j(x)$ are the objective function and constraint conditions given in (16)~(18). The main optimization steps are as follows: (1) initialize the first-generation chromosomes randomly, X_i, $i = 1$; (2) simulate the control strategy parameters represented by each chromosome using the vehicle simulation model for one complete simulation and determine the fitness of each individual according to the predetermined fitness function; (3) select a new generation of chromosomes, X_i', according to the fitness values (the selected probability is greater when the fitness is larger); (4) crossover and mutate, to generate the new chromosomes, X_i, $i = i + 1$; (5) return to (2); and (6) halt the process when satisfying the stopping condition.

TABLE 5: Optimization results for the parameters.

Parameters	Before optimization	After optimization
N_1 (rpm)	1000	1102
N_2 (rpm)	1400	1395
N_3 (rpm)	1650	1621
B (g)	4200	3917

5.3. Optimization Results. Figure 19 shows the convergence process of the control strategy parameters and the objective function under the combined simulation. N_1, N_2, and N_3 are the engine working speeds at the engine power ranges of 0–60 kW, 60–120 kW, and 120–175 kW, respectively. With the proceeding of the optimization process, N_1, N_2, and N_3 are constantly adjusted in the engine working speed range and gradually converge to stable values.

After 16 optimization calculations, through the optimization in software environment of OPTIMUS, one solution in the solution space is finally found as the optimal (Table 5).
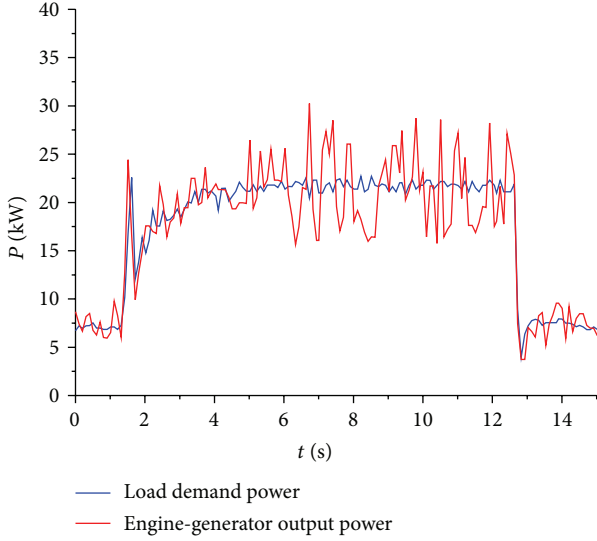
FIGURE 15: Engine-generator output power and load demand power.

FIGURE 17: Typical working conditions of the bulldozer.

FIGURE 16: Fuel consumption MAP graph.

$$\text{fitness}(x) = \begin{cases} f(x)g_j(x) \ge 0 & , \nabla j = 1, 2, \ldots, m \\ f_{\text{worst}} + \sum_{j=1}^{m} \langle g_j(x) \rangle, \text{other} \end{cases}$$
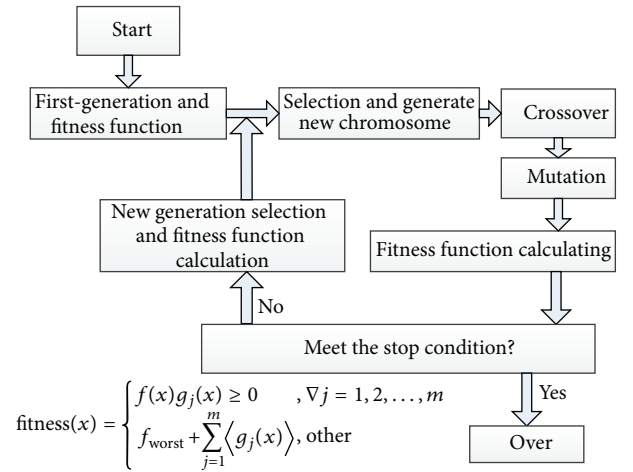
FIGURE 18: The GA flowchart of the control strategy optimization.

Figure 20 shows the right and left track velocity of the hybrid bulldozer under the typical working condition (Figure 17). As shown in Figure 21, the adjustment of the engine speed is great after optimization, especially in low load power and high load power areas. Engine speed was adjusted from low efficiency to high efficiency working point to improve the fuel economy of the bulldozer as shown in Figure 22.

The adjustment of the engine speed impacts the power allocation in a certain degree. As can be seen from Figure 21, in the high load power area, the engine maximum output power decreased due to the decreasing of the engine speed after adjustment. Under the same working condition, the ultracapacitor power supplement will be increased when the demand power is very large. Discharge current will increase to reduce the life of the ultracapacitor to some extent.

6. Conclusions

Based on the study of the terramechanics of the tracked bulldozer and test bench experimental data, a dynamic model of the hybrid electric tracked bulldozer power train system was established, combining the operation principle of each part of

the system. To verify the accuracy of the dynamic simulation model, the parameters and the real working conditions of the prototype traditional bulldozer were substituted into the simulation model, and the simulation results were compared with the experimental data to analyze the accuracy of the simulation model. A power follow strategy is proposed, combining the working features of the series hybrid electric tracked bulldozer. Test bench experiment was performed to collect actual test bench data to correlate and validate the proposed power follow control strategy for this hybrid electric tracked bulldozer. Based on the dynamic model of this hybrid electric tracked bulldozer, a genetic algorithm is proposed to solve the control strategy parameter optimization problem. Based on the optimized control strategy parameters, the simulation is performed to compare engine fuel consumption before and after control strategy optimization. By analyzing the optimization results, the fuel consumption of the bulldozer after optimization is reduced by approximately 6.74% compared
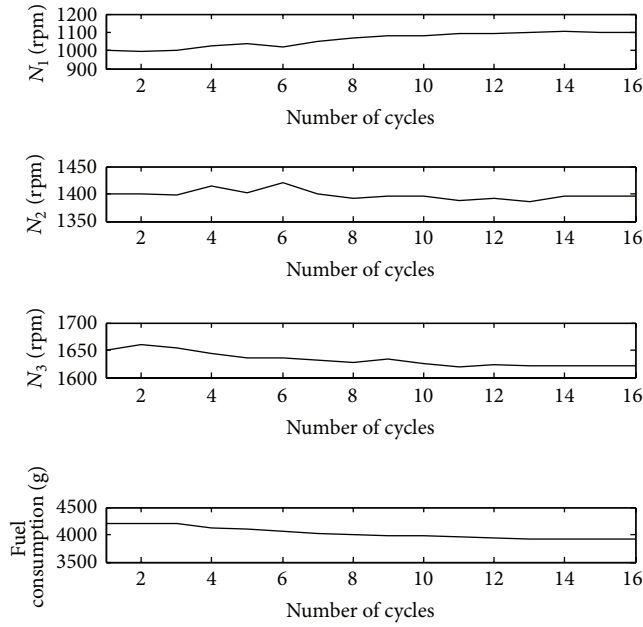
FIGURE 19: The convergence process of the objective function.
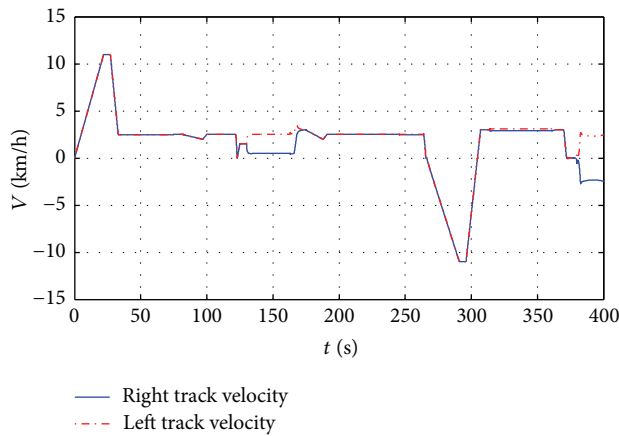
FIGURE 20: The right and left track velocities of the bulldozer.
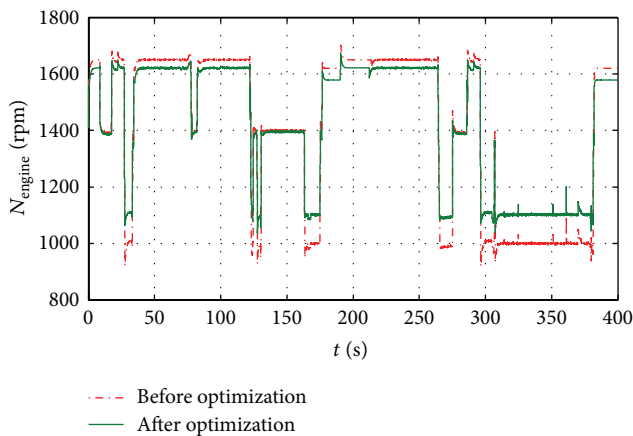
FIGURE 21: Comparison of engine speed.
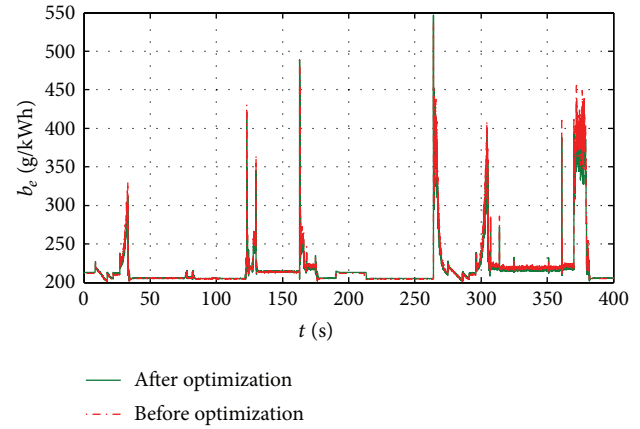
FIGURE 22: Comparison of engine fuel consumption.

with the former condition. This result verifies the validity of the method proposed in this paper which can reduce the difficulty of the design and optimize the control strategy.

Conflict of Interests

The authors declare no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] H. Wang and F. C. Sun, "Dynamic modeling and simulation on a hybrid power system for dualmotor—drive electric tracked bulldozer," *Applied Mechanics and Materials*, vol. 494-495, pp. 229–233, 2014.

[2] Q. Song and H. Wang, "Parameters matching for dual-motor-drive electric bulldozer," *Journal of Beijing Institute of Technology*, vol. 20, pp. 169–170, 2011.

[3] J. Wang, Q. N. Wang, P. Y. Wang, J. N. Wang, and N. W. Zou, "Hybrid electric vehicle modeling accuracy verification and global optimal control algorithm research," *International Journal of Automotive Technology*, vol. 16, no. 3, pp. 513–524, 2015.

[4] C. Sun, X. Hu, S. J. Moura, and F. Sun, "Velocity predictors for predictive energy management in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1197–1204, 2015.

[5] R. Ghorbani, E. Bibeau, P. Zanetel, and A. Karlis, "Modeling and simulation of a series parallel hybrid electric vehicle using REVS," in *Proceedings of the American Control Conference (ACC '07)*, pp. 4413–4418, IEEE, New York, NY, USA, July 2007.

[6] P. Caratozzolo and M. Canseco, "Design and control of the propulsion system of a series hybrid electric vehicle," in *Proceedings of the Electronics, Robotics and Automotive Mechanics Conference (CERMA '06)*, pp. 270–273, Mexico, Russia, September 2006.

[7] Z. Y. Song, H. Hofmann, J. Q. Li, J. Hou, X. B. Han, and M. G. Ouyang, "Energy management strategies comparison for

electric vehicles with hybrid energy storage system," *Applied Energy*, vol. 134, pp. 321–331, 2014.

[8] X. Hu, L. Johannesson, N. Murgovski, and B. Egardt, "Longevity-conscious dimensioning and power management of the hybrid energy storage system in a fuel cell hybrid electric bus," *Applied Energy*, vol. 137, pp. 913–924, 2015.

[9] X. Y. Li, X. M. Yu, J. Li, and Z. X. Wu, "Control strategy for series hybrid-power vehicle," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 2, pp. 122–126, 2005.

[10] E. Mehrdad, Y. Gao, and A. Emadi, *Modem Electric, Hybrid Electric, and Fuel Cell Vehicles: Fundamentals, Theory, and Design*, CRC Press, New York, NY, USA, 2010.

[11] X. Hu, N. Murgovski, L. Johannesson, and B. Egardt, "Energy efficiency analysis of a series plug-in hybrid electric bus with different energy management strategies and battery sizes," *Applied Energy*, vol. 111, pp. 1001–1009, 2013.

[12] C. Sun, S. J. Moura, X. Hu, J. K. Hedrick, and F. Sun, "Dynamic traffic feedback data enabled energy management in plug-in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1075–1086, 2015.

[13] R. Xiong, H. He, Y. Wang, and X. Zhang, "Study on ultracapacitor-battery hybrid power system for PHEV applications," *High Technology Letters*, vol. 16, no. 1, pp. 496–500, 2010.

[14] T.-S. Kwon, S.-W. Lee, S.-K. Sul et al., "Power control algorithm for hybrid excavator with supercapacitor," *IEEE Transactions on Industry Applications*, vol. 46, no. 4, pp. 1447–1455, 2010.

[15] G. G. Wang, I. Horowitz, S. H. Wang, and C. W. Chen, "Control design for a tracked vehicle with implicit nonlinearities using quantitative feedback theory," in *Proceedings of the 27th IEEE Conference on Decision and Control*, pp. 2416–2418, IEEE, Austin, Tex, USA, December 1988.

[16] Y. H. Li, X. M. Lu, and N. C. Kar, "Rule-based control strategy with novel parameters optimization using NSGA-II for powersplit PHEV operation cost minimization," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 7, pp. 3051–3061, 2014.

[17] S. Zhang, C. N. Zhang, R. Xiong, and W. Zhou, "Study on the optimal charging strategy for lithium-ion batteries used in electric vehicles," *Energies*, vol. 7, no. 10, pp. 6783–6797, 2014.

[18] J. K. Peng, H. W. He, and N. L. Feng, "Simulation research on an electric vehicle chassis system based on a collaborative control system," *Energies*, vol. 6, no. 1, pp. 312–328, 2014.

[19] C. M. Qi and P. Li, "An exponential entropy-based hybrid ant colony algorithm for vehicle routing optimization," *Applied Mathematics & Information Sciences*, vol. 8, no. 6, pp. 3167–3173, 2014.

[20] K. B. Wipke, M. R. Cuddy, and S. D. Burch, "ADVISOR 2.1: a user-friendly advanced powertrain simulation using a combined backward/forward approach," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 6, pp. 1751–1761, 1999.

[21] L. Guzzella and A. Amstutz, "CAE tools for quasi-static modeling and optimization of hybrid powertrains," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 6, pp. 1762–1769, 1999.

[22] J. H. Pu, C. L. Yin, and J. W. Zhang, "Application of genetic algorithm in optimization of control strategy for hybrid electric vehicles," *China Mechanical Engineering*, vol. 16, Article ID 649650, 2005.

[23] M. G. Bekker, *Off-Road Locomotion*, The University of Michigan Press, Ann Arbor, Mich, USA, 1960.

[24] C. G. Zhang, K. H. Zeng, and J. H. Zhao, "Analytical solutions of critical load and Terzaghi's ultimate bearing capacity for unsaturated soil," *Journal of Tongji University (Natural Science)*, vol. 38, no. 12, pp. 1736–1739, 2010.

[25] G. Bekker M, *Theory of Land Locomotion*, The University of Michigan Press, Ann Arbor, Mich, USA, 1956.

[26] Q. Song, P. Zeng, and H. Wang, "Study on the power flowing control strategy of a series hybrid tracked bulldozer at the typical working condition," *Journal of Mechanical Engineering*, vol. 50, no. 2014, pp. 136–143, 2014.

[27] R. C. Wang, R. He, J. B. Yu, and C. H. Hu, "Parameter optimization of a PSHEV based on genetic algorithm," *China Mechanical Engineering*, vol. 24, no. 18, pp. 2544–2549, 2013.

[28] A. Piccolo, L. Ippolito, V. Zo Galdi, and A. Vaccaro, "Optimisation of energy flow management in hybrid electric vehicles via genetic algorithms," in *Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, pp. 434–439, Como, Italy, July 2001.

[29] B. Wang, Z. P. Yang, F. Lin, and W. Zhao, "An improved genetic algorithm for optimal stationary energy storage system locating and sizing," *Energies*, vol. 7, no. 10, pp. 6434–6458, 2014.

Research Article

Dynamic Model of Kaplan Turbine Regulating System Suitable for Power System Analysis

Jie Zhao, Li Wang, Dichen Liu, Jun Wang, Yu Zhao, Tian Liu, and Haoyu Wang

School of Electrical Engineering, Wuhan University, Wuhan, Hubei 430072, China

Correspondence should be addressed to Li Wang; 2009302540275@whu.edu.cn

Accurate modeling of Kaplan turbine regulating system is of great significance for grid security and stability analysis. In this paper, Kaplan turbine regulating system model is divided into the governor system model, the blade control system model, and the turbine and water diversion system model. The Kaplan turbine has its particularity, and the on-cam relationship between the wicket gate opening and the runner blade angle under a certain water head on the whole range was obtained by high-order curve fitting method. Progressively the linearized Kaplan turbine model, improved ideal Kaplan turbine model, and nonlinear Kaplan turbine model were developed. The nonlinear Kaplan turbine model considered the correction function of the blade angle on the turbine power, thereby improving the model simulation accuracy. The model parameters were calculated or obtained by the improved particle swarm optimization (IPSO) algorithm. For the blade control system model, the default blade servomotor time constant given by value of one simplified the modeling and experimental work. Further studies combined with measured test data verified the established model accuracy and laid a foundation for further research into the influence of Kaplan turbine connecting to the grid.

1. Introduction

Hydraulic turbine generator units are increasingly being relied upon to meet a variety of control requirements as the size and complexity of interconnected systems increase. To establish models reflecting the actual characteristics of the prime mover and its governor is an important foundation work for the accuracy of power grid security and stability analysis [1, 2]. It is necessary to adjust both the wicket gate opening and the runner blade angle under different water head by the governor to keep efficient operation of a Kaplan turbine, which is one of the common forms of turbines in hydropower systems [3, 4]. There is an urgent need to establish Kaplan turbine regulating system dynamic model suitable for power system analysis based on measured experimental data to obtain the parameters [5].

There have been many studies on the modeling and verification of Francis turbine [6–10]. The common Francis turbine model contains the single regulator system model, the wicket gate control system model, and the turbine and water diversion system model. The ideal turbine model or linearized turbine model reflects the basic characteristics of the prime mover, but the ideal turbine model reflects unit features at rated operating point and the linearized turbine model is suitable for small power fluctuations [8, 9].

The operation of Kaplan turbine involves control of the wicket gates and the runner blades position to regulate the water flow into the turbine [2, 11]. Due to the on-cam relationship between the wicket gate opening and the runner blade angle, the impact of blade angle on Kaplan turbine transient simulation accuracy should not be ignored [12, 13]. The linear fitting [14], polynomial interpolation and fitting [15], or intelligent algorithm [16] was used to obtain the on-cam relationship expression. However, the linear fitting obtained via limited points [14] to an extent may fail to reflect the function completely. Different Kaplan turbine and water diversion system models based on different functions of wicket gate opening multiplied by function of blade angle as inputs of nonlinear turbine model were compared [17]. The parameter identification of a turbine and water

diversion system model may become difficult with high-order fitting function of the wicket gate opening or the blade angle. As a result, the applicability of the established model declines.

The Kaplan turbine regulating system model contains the mathematical models of the governor, the dual-regulated vane control system and blade control system, and the turbine and water diversion system. In this paper, the Kaplan turbine regulating system dynamic model and parameters were optimized suitable for power system analysis, experimental measurement, and precision requirements. The main contribution lies in the blade control system model simplification, the on-cam relationship obtainment, the derived five-parameter linearized Kaplan turbine model, parameter obtainment method, and different Kaplan turbine regulating system dynamic models comparison in gird connected system. The structure is organized as follows. In Section 2, the Kaplan turbine regulating system model is presented. Section 3 contains the development of three kinds of Kaplan turbine and water diversion system model (i.e., the linearized Kaplan turbine model, the improved ideal turbine model, and the nonlinear Kaplan turbine model). The established model is verified by comparison with the measured data in Section 4. The results and discussion are presented in Section 4 as well. Section 5 presents the conclusions drawn thereof.

2. Kaplan Turbine Regulating System Mathematical Model

Figure 1 shows the frame diagram of hydroelectric power with double-regulated turbine [18]. The vane control system is based on the same principle of Francis turbine. The on-cam relationship presents the runner blade angle changes with the wicket gate opening and the blade is kept in a certain best angle fitting with the wicket gate opening to improve the power generation efficiency.

ω and ω_{ref} are the unit frequency and given frequency value, respectively, and P_{ref} and P_e are the given power and turbine load, respectively. y_{ref}, y, and y_r are the given wicket gate opening, the actual wicket gate opening, and the blade opening, respectively. The blade opening is transferred from the blade angle with the maximum value no more than one. q is the turbine working flow, h is the turbine working head, and P_m is the mechanical power output.

2.1. The Regulating System Model. The Kaplan turbine governor system is consistent with that of the general Francis turbine. Figure 2 shows the digital governor model. Also, Figure 3 shows the vane control system model considering the speed limit and the amplitude limit [18]. K_P, K_I, and K_D are the proportional, integral, and differential coefficients of the governor, respectively. K_{sp}, K_{si}, and K_{sd} are the proportional, integral, and differential coefficients of the integrated amplifier module, respectively. e_p is the power deviation magnification. T_y is the main servomotor time constant. VEL$_{\text{open}}$ and VEL$_{\text{close}}$ are the open and close servomotor speed limits, respectively. $y_{g\text{MAX}}$ and $y_{g\text{MIN}}$ are the limiting values of the wicket gate opening.

2.2. Blade Control System Model. The principle of the blade control system is similar to that of the vane control system, assuming that the on-cam relationship between the wicket gate opening and the runner blade angle is neglected [19]. The blade control valve adopts a proportional valve to handle the blade servomotor. Figure 4 shows the blade control system model under a certain water head. T_{yr} is the blade servomotor time constant. VEL$_{\text{open1}}$ and VEL$_{\text{close1}}$ are the open and close servomotor speed limits, respectively. $y_{r\text{MAX}}$ and $y_{r\text{MIN}}$ are the limiting values of the blade opening.

The function between the blade opening, the gate opening, and the water head is expressed as

$$y_r = f(y, H). \tag{1}$$

Under a certain water head, the function of the wicket gate opening and the runner blade opening is given by

$$y_r = a_5 y^5 + a_4 y^4 + a_3 y^3 + a_2 y^2 + a_1 y + a_0, \tag{2}$$

where a_5, a_4, a_3, a_2, a_1, and a_0 are the fitting coefficients.

3. Kaplan Turbine and Water Diversion System Model

The linearized Kaplan turbine model, improved ideal turbine model, and nonlinear Kaplan turbine model were developed. The five-parameter linearized Kaplan turbine model was derived for parameters obtainment convenience. The improved ideal turbine model and nonlinear turbine model were demonstrated based on the principle of the turbine that the blade opening affected the Kaplan turbine power like the linearized turbine model did.

3.1. Model One: Linearized Kaplan Turbine Model. The dynamic characteristic of hydraulic turbine is associated with the dynamic characteristic of the water flow in the pipe. The turbine characteristics in the vicinity of the operating point are approximately linear. Compared with the Francis turbine [9], the torque and water flow of Kaplan turbine can be expressed as

$$m_t = e_x x + e_y y + e_r y_r + e_h h,$$
$$q = e_{qx} x + e_{qy} y + e_{qr} y_r + e_{qh} h, \tag{3}$$

where e_x, e_y, e_r, and e_h are the transfer coefficients of turbine torque to the rotational speed, wicket gate opening, blade opening, and water head, respectively. e_{qx}, e_{qy}, e_{qr}, and e_{qh} are the transfer coefficients of the water flow to the rotational speed, wicket gate opening, blade opening, and water head, respectively.

For the turbine water diversion system, considering the wall elasticity of water flow and water diversion system and ignoring the friction between them, the mathematical model of water diversion system can be obtained as [6]

$$G_h(s) = \frac{2T_w}{T_r} \frac{1 - e^{T_r s}}{1 + e^{T_r s}}, \tag{4}$$
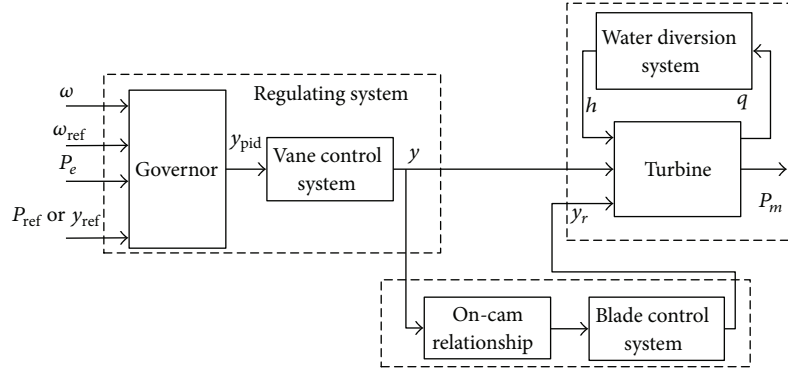
Figure 1: Kaplan turbine regulating system frame diagram.
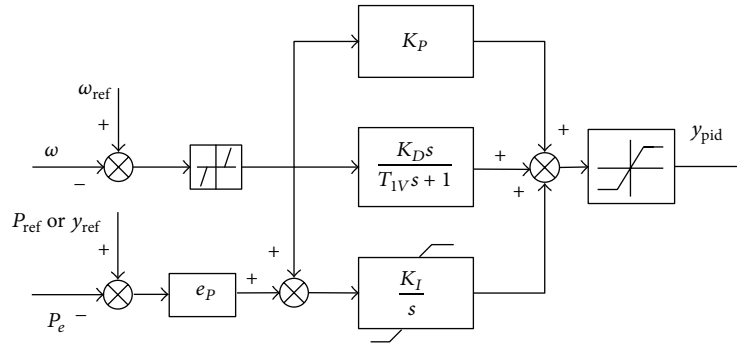
Figure 2: Kaplan turbine governor model.
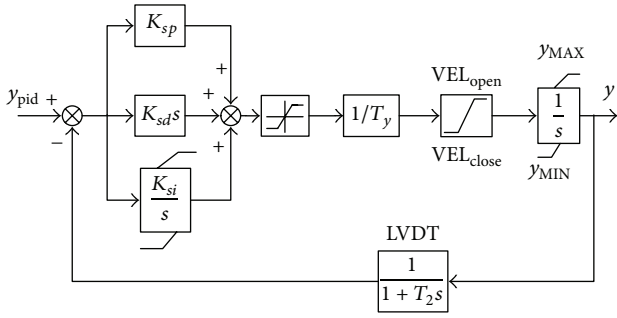
Figure 3: Kaplan turbine vane control system model.
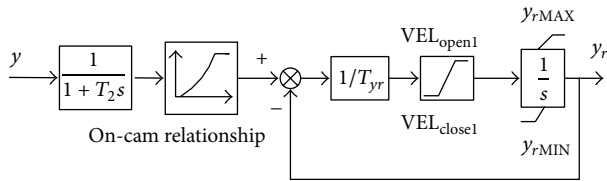
Figure 4: Blade control system model.

where T_w is the water inertia time constant and T_r is the water hammer phase length.

If the high level terms in the Taylor expression of (4) are ignored, then the transfer function of the simplified elastic model is given as

$$G_h(s) = -\frac{T_w s \left(1 + (1/24) T_r^2 s^2\right)}{1 + (1/8) T_r^2 s^2}. \tag{5}$$

Furthermore, if the flexibility of the water flow and the water diversion pipe is ignored, the transfer function of the rigid water hammer model is expressed as

$$G_h(s) = -T_w s. \tag{6}$$

The speed influence on the turbine power in (3) is neglected due to the little unit speed change under the grid connected condition. Figure 5 shows the linearized Kaplan turbine model and its deformed five-parameter linearized turbine model with rigid water hammer model, while a, b, c, d, and f are constants above zero.

3.2. Model Two: Improved Ideal Turbine Model. Assume that the transfer coefficients in (3) and the turbine efficiency are certain under various working conditions and the wicket gate opening varies linearly. Ignoring the unit speed influence on turbine power, the ideal turbine model of the Francis turbine at rated operating point is described as [8]

$$G(s) = \frac{1 - T_w s}{1 + 0.5 T_w s}. \tag{7}$$

(a)

(b)

FIGURE 5: Linearized Kaplan turbine model. (a) Linearized Kaplan turbine model. (b) Five-parameter linearized turbine model.
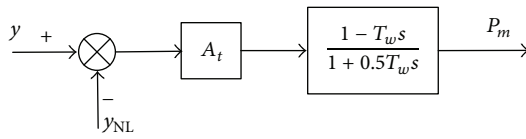
FIGURE 6: Improved ideal turbine model.

Figure 6 shows the improved ideal turbine considering the effect of the wicket gate and the blade on Kaplan turbine power. y_{NL} is the equivalent no-load opening. A_t is the proportional coefficient. The blade effect on the Kaplan turbine power is reflected by the calculation of A_t and y_{NL}, which are identified by an optimization algorithm in this paper.

3.3. Model Three: Nonlinear Kaplan Turbine Model. For the Francis turbine, considering the relationship between the turbine flow, wicket gate opening, and the water head as a valve [19], the nonlinear turbine model is described by

$$
\begin{aligned}
q &= y\sqrt{h}, \\
P_m &= k_p\left(q - q_0\right)h, \\
\dot{q} &= \frac{1}{G_h\left(s\right)}\left(h - h_0\right),
\end{aligned}
\tag{8}
$$

where q is the water flow in the diversion pipeline, $\dot{q}$ is the derivative of q, h_0 is the static head (its per unit value is 1), and q_0 is no-load flow. k_p is proportional coefficient.

The nonlinear Kaplan turbine model is established based on the following four aspects:

(a) Under the normal grid connected operation, the unit speed is generally in the vicinity of rated speed, so the unit speed impact on the turbine model is ignored.

(b) The hydraulic turbine with the water diversion pipe length less than 800 m can adopt the rigid water hammer model to simulate the dynamic process of the water diversion pipe.

(c) Consider the nonlinear relationship between the wicket gate opening and the turbine mechanical power.

(d) The effect of the blade opening on the Kaplan turbine is viewed as increasing flow amount.

The variable G is introduced to represent the equivalent opening for the combined effect of the wicket gate opening and the blade opening given by

$$
G = f\left(y\right) \cdot f\left(y_r\right),
\tag{9}
$$

where $f(y)$ and $f(y_r)$ are the function of y and y_r, respectively.

Different to the Kaplan turbine and water diversion system models based on concrete different functions of y and y_r [17], function $f(y)$ is represented by cubic polynomial fitting and function $f(y_r)$ is represented by linear fitting given by

$$
\begin{aligned}
f\left(y\right) &= b_3 y^3 + b_2 y^2 + b_1 y + b_0, \\
f\left(y_r\right) &= y_r \cdot B_{\mathrm{flow}} + 1,
\end{aligned}
\tag{10}
$$

where b_3, b_2, b_1, and b_0 are the fitting coefficients. B_{flow} is the coefficient representing the blade angle impact on the turbine water flow. Herein the effect of the blade opening on the turbine is considered as incremental flow referring to the principle of Kaplan turbine and the parameters in (10) are easy to obtain for engineering application and model validation, which reflects the nonlinear characteristics and working principle of Kaplan turbine.

There is a time delay between the wicket gate movement and the consequent blade angle movement in the case studied. Figure 7 shows the nonlinear Kaplan turbine model with rigid water hammer model considering the above four aspects combined with (8)–(10).

4. Results and Discussion

The testing of the Kaplan turbine regulating system model is based on the Kaplan turbine at Tukahe hydropower plant in Yunnan province of West China. The hydropower station has three units with one unit capacity of 55 MW, normal reservoir storage level of 368 m, and rated head of 25 m.

Different fitting functions of on-cam relationship were compared based on the whole range measured under given wicket gate opening disturbances while the gate-runner relationship was determined from the turbine characteristics (TC) data in [17]. The blade control effect is mainly determined by the on-cam relationship, speed limits, and blade
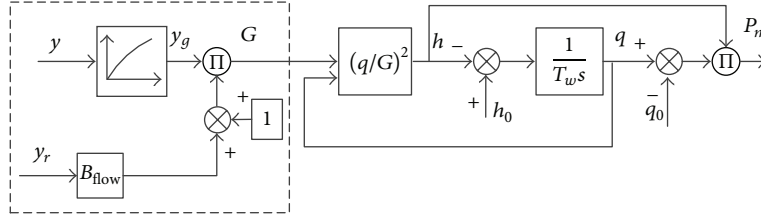
Figure 7: Kaplan turbine and water diversion system model.

servomotor time constant of the blade control system model. The default blade servomotor time constant given by value of one simplified the modeling and experimental work. The Kaplan turbine regulating system dynamic model and three different turbine and water diversion system models were compared for stability analysis in the grid connected power system.

4.1. On-Cam Relationship Validation of Wicket Gate Opening and Blade Opening. The fifth order polynomial fitting of the on-cam relationship between the wicket gate opening and the blade opening based on the measured data near the rated water head is expressed as

$$y_r = 24.36y^5 - 76.29y^4 + 83.92y^3 - 37.77y^2$$
$$+ 7.187y - 0.4525. \tag{11}$$

The blade opening is set to a certain value near zero when wicket gate opening is less than 40%. The best fitting function is obtained through measured data depending on different Kaplan turbines or different working heads. The relationship between the wicket gate opening and the blade opening needs to be revised with appropriate polynomial fitting at other working heads. Figure 8(a) shows the simulated blade opening of the blade control system model (see Figure 4) using the expression in (11) under given wicket gate opening disturbances of whole range.

As shown in Figure 8(a), when the wicket gate opening is less than 0.4 by per unit value, the blade opening is approximate to zero and the maximum blade opening is about 0.9 by per unit value. The piecewise linear fitting result fits with the measured data when the wicket gate opening is less than 0.6 and has a difference as the wicket gate opening over 0.6 by per unit value. The fifth order fitting shown in (11) is closest to the measured data which shows a better on-cam relationship than the other fitting methods. Given the wicket gate opening disturbances, the simulation result using fifth order fitting is consistent with the measured data in the whole operating range (see Figure 8(c)) and verifies the correctness of the on-cam relationship function.

4.2. Influence of Blade Servomotor Time Constant on the Blade Control System Model. The influence of blade servomotor time constant on the blade control system model was studied. Figure 9 shows the simulation results with different T_{yr} values.

The influence of different T_{yr} values on the simulation results shows that the bigger T_{yr} is, the slower adjusted rate

is (see Figure 9), and thus T_{yr} should be as small as possible. Due to the rate restrictions of VEL$_{\text{open1}}$ and VEL$_{\text{close1}}$, the result difference is not obvious when T_{yr} is less than 1 s. In conclusion, the blade servomotor time constant affects the adjustment of blade control system model and should be set in a reasonable range of value. T_{yr} in this case is 1 s considering that it is difficult to be tested. When T_{yr} is about 1 s (i.e., the blade servomotor response time link is ignored), it has little effect on the simulation result but simplifies the modeling and experimental work.

4.3. Turbine Regulating System Model Validation. The improved particle swarm optimization (IPSO) algorithm (i.e., the inertia weight is linearly changed and the chaos method is used to generate initial particle population) is applied to identify the model parameters [20]. The inertia weight is introduced to revise the speed update equation to improve search ability in the global scope. The chaos method is used to generate good initial particle population.

The error evaluation index to quantize the simulation agreement is defined as

$$\sigma = \frac{\sum_{i=1}^{N} \left(Y_i - \widehat{Y}_i \right)^2}{N}, \tag{12}$$

where Y_i is ith measured value, $\widehat{Y}_i$ is ith simulation value, and N is the total number of sampling points.

Figure 10 shows the graphical representation of measured behavior of power acquired at a certain head given in 2D and 3D space. y_r changes with y when the wicket gate opening is over a certain value. In addition, the relationship between the wicket gate opening and the turbine mechanical power is nonlinear.

The speed limits and amplitude limits of the vane control system and blade control system were calculated under large wicket gate opening disturbances while other parameters were identified by the improved particle swarm optimization (IPSO) algorithm under small wicket gate opening disturbances and the parameters were revised through grid connected simulation comparison. The maximum iteration step of IPSO algorithm is 100. For the error function with the extreme point of zero, the tolerance of error is set to 10^{-6}. The parameters generally range from slightly greater than 0 to 100 to get global solution as much as possible. Based on the turbine power disturbance test data and the unit frequency disturbance test data, Table 1 lists the cubic fitting relation between variables y_g and y (see Figure 7), the coefficient representing blade angle impact on the turbine water flow,

(a)

(b)

(c)
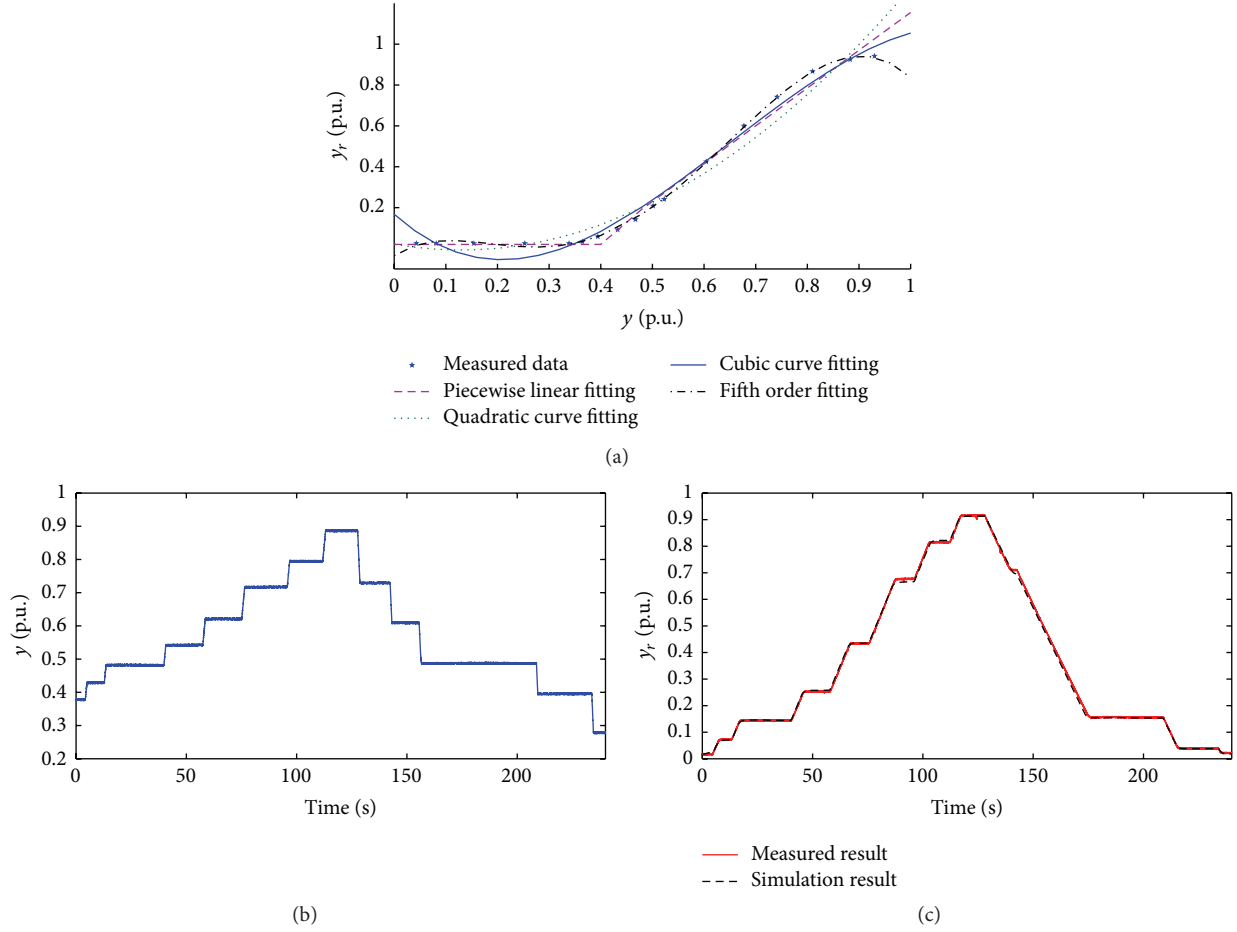
FIGURE 8: On-cam relationship contrast curves. (a) On-cam relationship fitting curves. (b) Measured gate opening. (c) Simulation and measured blade opening.
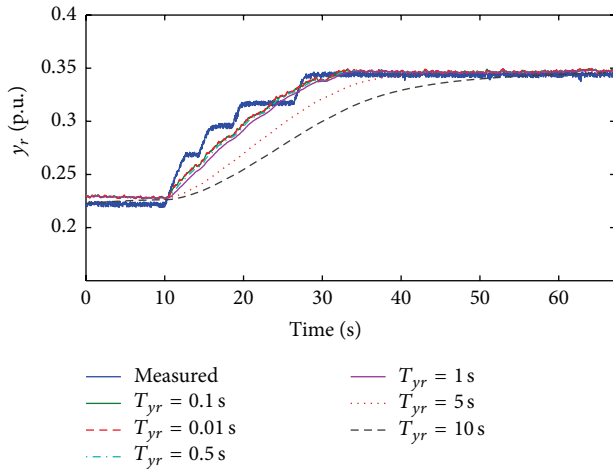
FIGURE 9: The influence of different T_{yr} on the blade opening.

TABLE 1: The parameter list.

Model	Parameter	Value
Vane control system model	K_{sp}	35
	T_y	28 s
	VEL_{open}	0.15
	VEL_{close}	−0.15
	y_{gMAX}	1
	y_{gMIN}	0
Blade control system model	T_{yr}	1 s
	VEL_{open1}	0.020
	VEL_{close1}	−0.017
	y_{rMAX}	0.9
	y_{rMIN}	0
Nonlinear Kaplan turbine model	B_{flow}	0.6
	$y_g = 0.267y^3 - 1.42y^2 + 2.383y - 0.5916$	

and other parameters of the double regulating system model identified by the IPSO algorithm. Figure 11 shows the Kaplan turbine regulating system model simulation result using nonlinear Kaplan turbine model for unit frequency disturbance test with disturbance values of ±0.2 Hz under a certain unit power value.

(a)

(b)
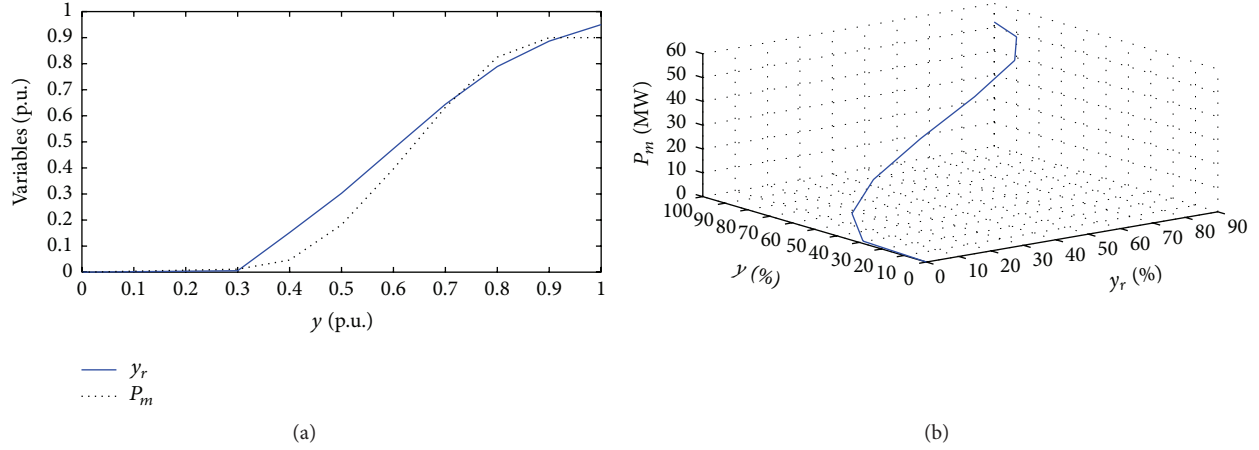
FIGURE 10: Graphical representation of measured behavior of power acquired at a certain head. (a) 2D space. (b) 3D space.

(a)

(b)

(c)
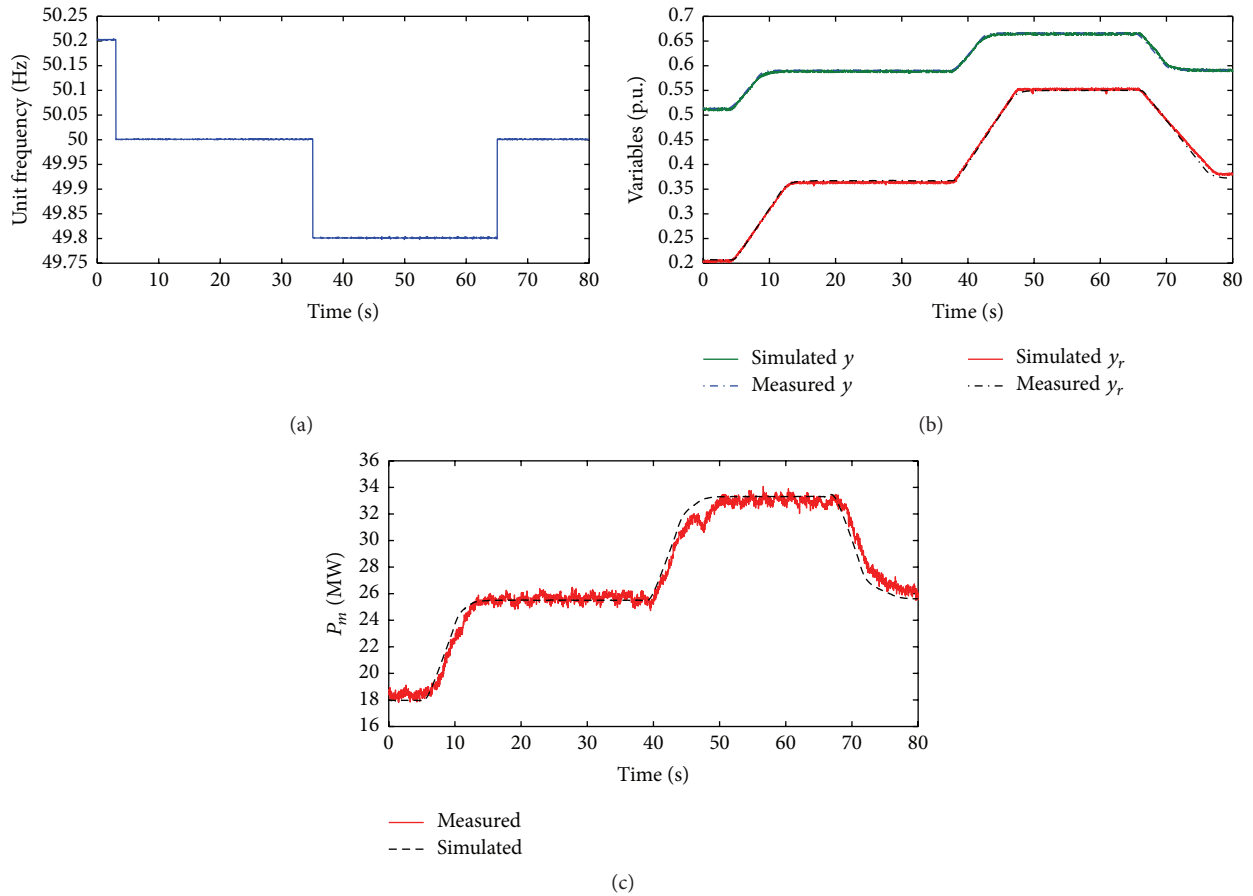
FIGURE 11: Simulation and measured results of unit frequency disturbance test. (a) Given unit frequency disturbance curve. (b) Wicket gate opening and blade opening curves. (c) Mechanical power curves.

From the simulation results shown in Figure 11, the initial recorded gate opening and blade opening are about 0.5 and 0.2 by per unit, respectively. The blade opening changes linearly under the speed limits effect in the model and the blade opening change is slower than the wicket gate opening change which can be seen from Figure 8 and speed limit differences listed in Table 1. There is little difference at the final mechanical power value due to measured blade opening difference when the unit frequency is recovered to 50 Hz. The wicket gate opening response and the blade opening response
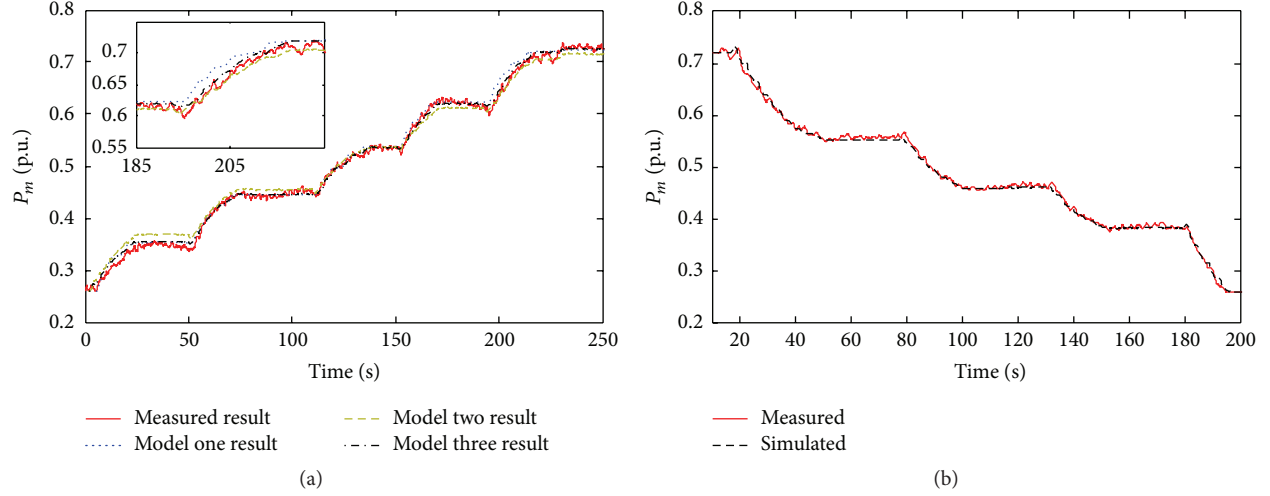
FIGURE 12: Turbine power output comparison. (a) Turbine power rising disturbance test. (b) Turbine power decreasing disturbance test.

TABLE 2: The parameter list of Kaplan turbine model.

Model	Parameter
	$a = 2.3238$
	$b = 1.7365$
Model one	$c = 2.6712$
	$d = 0.5552$
	$f = 2.3238$
	$y_{NL} = 0.3$
Model two	$A_t = 1.6$
	$T_w = 1.8 \text{ s}$

TABLE 3: Simulation error comparison of power rising disturbance.

Turbine model	σ
Model one	5.5229×10^{-5}
Model two	1.7459×10^{-4}
Model three	4.4294×10^{-5}

results agree with the measured curve indicating that the established vane and blade control model can effectively reflect the dynamic characteristics of Kaplan turbine.

Table 2 shows the identified parameters of the Kaplan turbine and water diversion system model using model one and model two.

Figure 12 shows the comparison results of the turbine power rising test using these three turbine models (i.e., model one, model two, and model three) and decreasing disturbance test using model three with the same Kaplan turbine regulating system mathematical model, respectively. Also, Table 3 lists the power rising disturbance simulation errors of three models.

The Kaplan turbine and water diversion system model using model one (i.e., the linearized Kaplan turbine model) to a certain extent does not accurately reflect the hydroturbine dynamic process when the power output is between 0.62 and 0.72 by per unit value (see Figure 12). As shown in Table 3,

model two failed to completely reflect dynamic change of the blade opening and its simulation result has a significant difference with the measured curve. The output of model three (i.e., the nonlinear Kaplan turbine model) is basically the same with the measured curve which is more accurate than model two. In summary, model one and model three apply dual input of the wicket gate opening and the blade opening. These two models regard the effect of blade opening on Kaplan turbine as increasing flow amount and, as a result, influence the unit power output. Through verification of the frequency disturbance test and turbine power disturbance test, the nonlinear Kaplan turbine model can meet the needs of the Kaplan turbine dynamic simulation better.

The generator rotor inertia time constant is obtained by load rejection test. Figure 13 shows the measured curves and the speed fitting result under load rejection of 25 MW.

The rotor motion equation is given by

$$\Delta T_M^* = T_j \cdot \alpha^*, \tag{13}$$

where T_j is the rotor inertia time constant. α^* and ΔT_M^* are the rotor acceleration and rotor mechanical torque change by per unit value, respectively. The calculated rotor inertia time constant from linear segment of the speed change between $t1$ and $t2$ (see Figure 13(b)) is 5.3 s.

4.4. Grid Connected Simulation Analysis of Kaplan Turbine. The Kaplan turbine regulating system model was established and the simulation was performed on the IEEE 39-bus system [21] with three different kinds of Kaplan turbine and water diversion system models using model one, model two, and model three in Section 3 and the same Kaplan turbine regulating system mathematical model in Section 2, respectively. The Kaplan turbine is connected to bus 18. Figure 14 shows the revised 39-bus system diagram.

The generator parameters of bus 40 herein are as follows: $X_d = 1.217$, $X_d = 0.600$, $X_d' = 0.349$, $X_q' = 0.600$, $X_d'' = 0.250$, and $X_q'' = 0.250$, and the rotor inertia time constant of
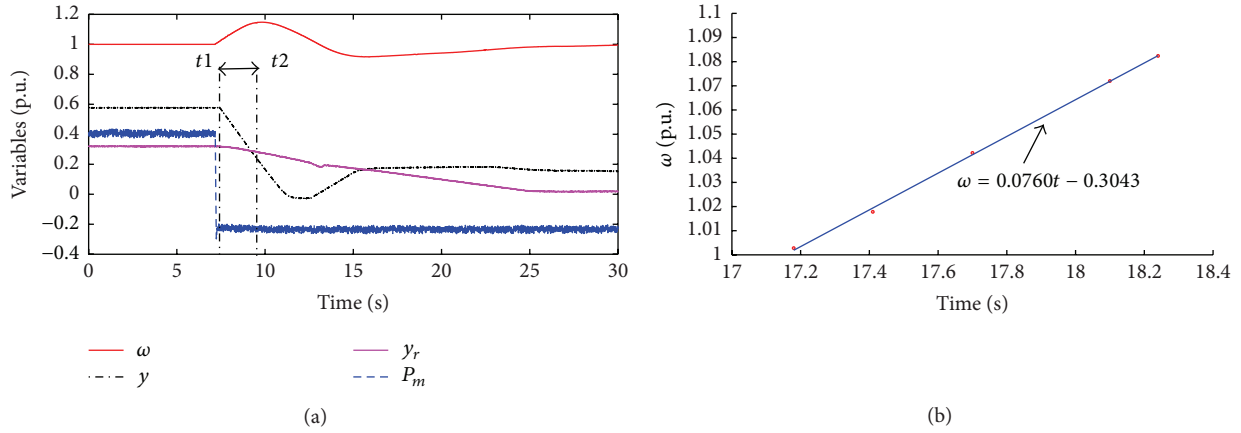
FIGURE 13: Load rejection test. (a) Measured variables curves. (b) Linear segment of the speed change and the fitting curve.
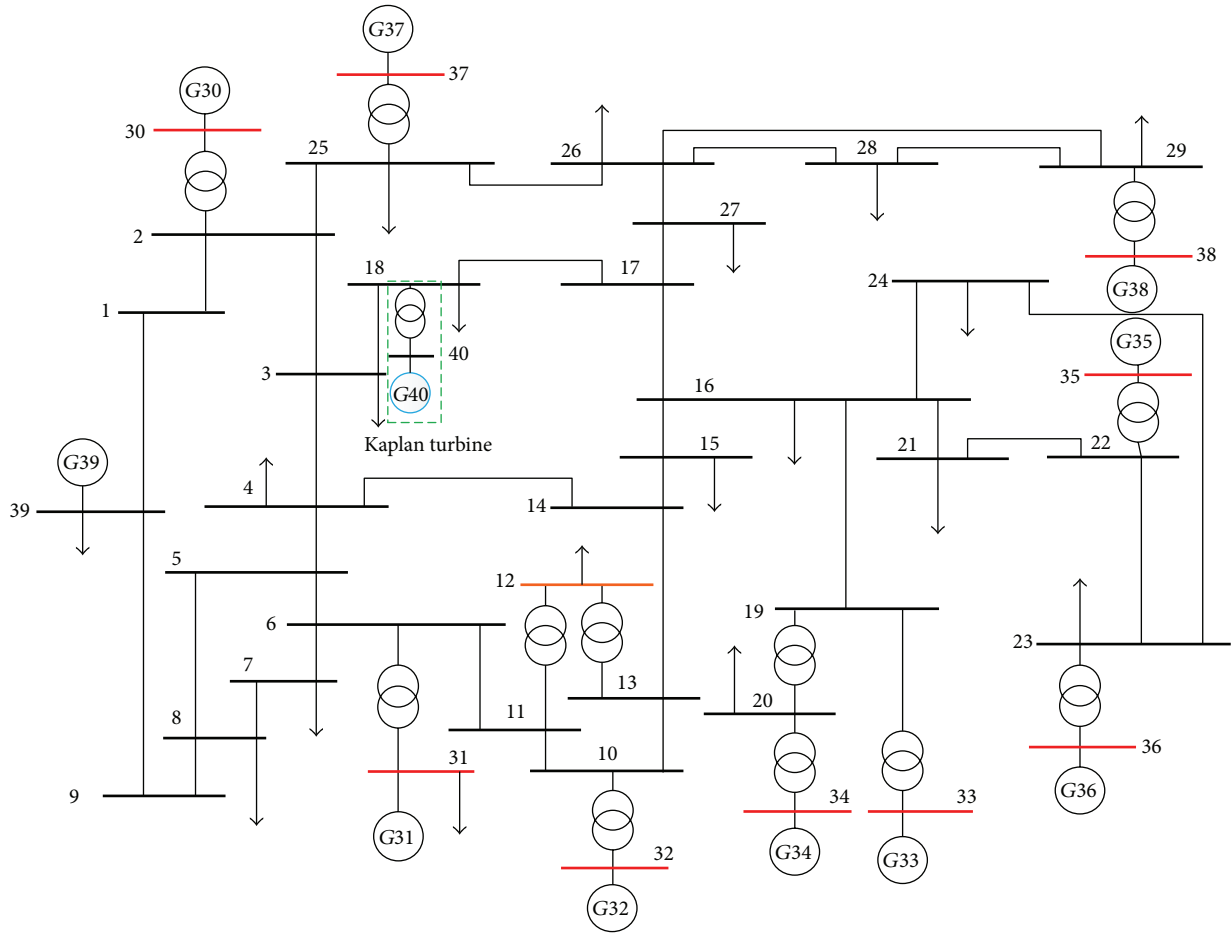
FIGURE 14: Revised 39-bus system diagram.

the generator is 6 s while the other parameters in the system are not modified.

Assume that there is a three-phase short-circuit fault on lines 17-18 near bus 18 at 1 s, and lines 17-18 are disconnected at 1.1 s. Figure 15 shows the simulation results.

As shown in Figure 15, the terminal voltage of Kaplan turbine is rapidly reduced to around zero when a three-phase short-circuit fault occurs. Each observation returns to the steady-state value after removing the fault line besides the reactive power with value of a little increase. The terminal voltage changes with the three kinds of the turbine regulating system dynamic model (i.e., the Kaplan turbine and water diversion system model using model one, model two, and model three) are basically the same. Under the initial turbine

(a)

(b)

(c)

(d)

(e)

(f)
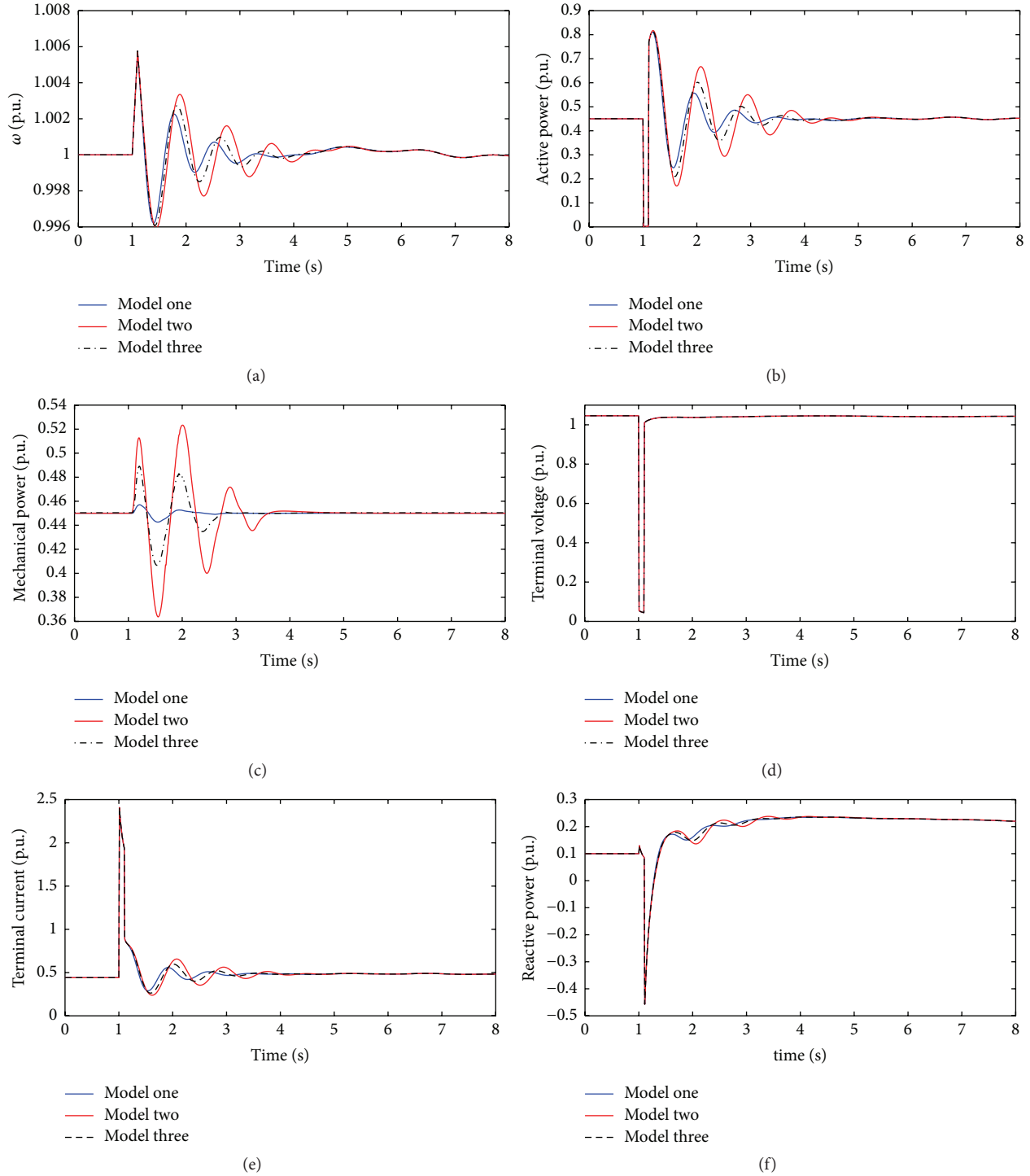
FIGURE 15: Power grid fault simulation results. (a) Unit speed curves. (b) Active power curves. (c) Mechanical power curves. (d) Terminal voltage curves. (e) Terminal current curves. (f) Reactive power curves.

mechanical power output of 0.45 (i.e., the mechanical power is 27 MW), the angle speed and the mechanical power fluctuations of model two are larger than those of model one and model three with the maximum deviation being about 0.3 Hz. The maximum mechanical power fluctuation value of model two is about 0.08 by per unit value (the reference capacity of

100 MVA). In addition to different turbine and water diversion system models, the governor parameters (the governor converts the frequency deviation to wicket gate opening) also influence the amount of mechanical power adjustment.

The adjustment of nonlinear turbine model was finer than that of the improved ideal turbine model since the latter was

derived based on the linearized model near the rated condition and its transfer function was simple. From the turbine operating characteristics, the coefficients in (3) may change significantly with turbine running in other conditions; thus, the original parameters may cause a large amount of power fluctuation.

5. Conclusion

Kaplan turbine model in this paper is divided into the governor system model, the blade control system model, and the turbine and water diversion system model. The detailed dual-regulated vane control system model and blade control system model were developed. The comparison of simulation and measured results showed the reasonableness and effectiveness of the established model and parameters. The on-cam relationship between the wicket gate opening and the blade opening and the influence of the blade servomotor time constant on the blade control system model were studied. The blade opening was transferred from the blade angle and the fifth order polynomial fitting function could characterize the on-cam relationship with amplitude limits set of the blade opening in this study. The blade servomotor time constant should be set in a reasonable range of value, but it was difficult to be tested. In this study, T_{yr} was 1 s, and thus the blade servomotor response time link could be ignored. However, it had a little effect on the simulation result and simplified the modeling and the experimental work.

Progressively the linearized Kaplan turbine model, improved ideal turbine model, and nonlinear Kaplan turbine model are developed. The nonlinear Kaplan turbine model is proposed which reflects the effects of the wicket gate opening and the blade opening changes on the turbine mechanical power (i.e., it regards the effect of blade opening on Kaplan turbine as increasing flow amount) and has a high accuracy through the comparison of the simulation and the measured results. The power disturbance test and the frequency disturbance test confirmed that the established governor model, the dual-regulated vane and blade control system model, and the nonlinear Kaplan turbine model reflected the dynamic response of the Kaplan turbine adequately, which could be applied in the power system analysis.

Conflict of Interests

The authors have declared that no conflict of interests exists.

Acknowledgment

References

[1] S. Patterson, "Importance of hydro generation response resulting from the new thermal modeling and required hydro modeling improvements," in *Proceedings of the IEEE Power Engineering Society General Meeting*, vol. 2, pp. 1779–1783, Denver, Colo, USA, June 2004.

[2] D. M. Dobrijevic and M. V. Jankovic, "An improved method of damping of generator oscillations," *IEEE Transactions on Energy Conversion*, vol. 14, no. 4, pp. 1624–1629, 1999.

[3] I. Vojtko, V. Fecova, M. Kocisko, and J. Novak-Marcincin, "Proposal of construction and analysis of turbine blades," in *Proceedings of the 4th IEEE International Symposium on Logistics and Industrial Informatics (LINDI '12)*, pp. 75–80, IEEE, Smolenice, Slovakia, September 2012.

[4] D. Jošt, A. Škerlavaj, and A. Lipej, "Improvement of efficiency prediction for a Kaplan turbine with advanced turbulence models," *Journal of Mechanical Engineering*, vol. 60, no. 2, pp. 124–134, 2014.

[5] E. Malenovsky and F. Pochyly, "Dynamic analysis of kaplan turbine caused by short circuit," in *Proceedings of the 14th International Symposium on Transport Phenomena and Dynamics of Rotating Machinery*, Honolulu, Hawaii, USA, March 2012.

[6] Y. C. Choo, K. M. Muttaqi, and M. Negnevitsky, "Modelling of hydraulic turbine for dynamic studies and performance analysis," in *Proceedings of the Australasian Universities Power Engineering Conference*, pp. 1–6, December 2007.

[7] L. N. Hannett, J. W. Feltes, and B. Fardanesh, "Field tests to validate hydro turbine-governor model structure and parameters," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1744–1751, 1994.

[8] Working Group, "Hydraulic turbine and turbine control models for system dynamic studies," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 167–179, 1992.

[9] D. H. Thorne and E. F. Hill, "Field testing and simulation of hydraulic turbine governor performance," *IEEE Transactions on Power Apparatus and Systems*, vol. 93, no. 4, pp. 1183–1191, 1974.

[10] Q. Lu, W. Hu, J. Li, and J. Zhou, "Research on frequency simulation model and control strategy of hydropower islands," in *Proceedings of the International Conference on Advanced Power System Automation and Protection (APAP '11)*, vol. 3, pp. 1896–1901, IEEE, Beijing, China, October 2011.

[11] J. L. Agüero, M. C. Beroqui, and F. Issouribehere, "Grid frequency control. secondary frequency control tuning taking into account distributed primary frequency control," in *Proceedings of the IEEE Power and Energy Society General Meeting*, pp. 1–8, IEEE, Minneapolis, Minn, USA, July 2010.

[12] P. Schniter and L. Wozniak, "Efficiency based optimal control of Kaplan hydrogenerators," *IEEE Transactions on Energy Conversion*, vol. 10, no. 2, pp. 348–353, 1995.

[13] P. Huvet, "Influence of runner blade angle variations on Kaplan turbine transients," *Houille Blanche*, vol. 41, no. 1, pp. 137–147, 1986.

[14] D. N. Kosterev, "Hydro turbine-governor model validation in pacific northwest," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1144–1149, 2004.

[15] M. Brezovec, I. Kuzle, and T. Tomisa, "Nonlinear digital simulation model of hydroelectric power unit with Kaplan turbine," *IEEE Transactions on Energy Conversion*, vol. 21, no. 1, pp. 235–241, 2006.

[16] W. Zhao and D. Wei, "Relevance vector machine combined with glowworm swarm optimization for cam relationship of Kaplan turbine," *Advanced Science Letters*, vol. 11, no. 1, pp. 244–247, 2012.

[17] D. Kranjcic and G. Stumberger, "Differential evolution-based identification of the nonlinear kaplan turbine model," *IEEE Transactions on Energy Conversion*, vol. 29, no. 1, pp. 178–187, 2014.

[18] M. Djukanovic, M. Novicevic, D. Dobrijevic, B. Babic, D. J. Sobajic, and Y.-H. Pao, "Neural-net based coordinated stabilizing control for the exciter and governor loops of low head hydropower plants," *IEEE Transactions on Energy Conversion*, vol. 10, no. 4, pp. 760–767, 1995.

[19] J. Zhao, L. Wang, D. Liu, and J. Wang, "Piecewise model and parameter obtainment of governor actuator in turbine," *Journal of Applied Mathematics*, vol. 2015, Article ID 709272, 9 pages, 2015.

[20] E. De Jaeger, N. Janssens, B. Malfliet, and F. Van De Meulebroeke, "Hydro turbine model for system dynamic studies," *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1709–1715, 1994.

[21] M. D. Amadou, H. Mehrjerdi, S. Lefebvre, M. Saad, and D. Asber, "Area voltage control analysis in transmission systems based on clustering technique," *IET Generation, Transmission & Distribution*, vol. 8, no. 12, pp. 2134–2143, 2014.

Research Article

Bus Travel Time Deviation Analysis Using Automatic Vehicle Location Data and Structural Equation Modeling

Xiaolin Gong, Xiucheng Guo, Xueping Dou, and Lili Lu

School of Transportation, Southeast University, Sipailou 2, Nanjing, Jiangsu 210096, China

Correspondence should be addressed to Xiucheng Guo; seuguo@163.com

To investigate the influences of causes of unreliability and bus schedule recovery phenomenon on microscopic segment-level travel time variance, this study adopts Structural Equation Modeling (SEM) to specify, estimate, and measure the theoretical proposed models. The SEM model establishes and verifies hypotheses for interrelationships among travel time deviations, departure delays, segment lengths, dwell times, and number of traffic signals and access connections. The finally accepted model demonstrates excellent fitness. Most of the hypotheses are supported by the sample dataset from bus Automatic Vehicle Location system. The SEM model confirms the bus schedule recovery phenomenon. The departure delays at bus terminals and upstream travel time deviations indeed have negative impacts on travel time fluctuation of buses en route. Meanwhile, the segment length directly and negatively impacts travel time variability and inversely positively contributes to the schedule recovery process; this exogenous variable also indirectly and positively influences travel times through the existence of signalized intersections and access connections. This study offers a rational approach to analyzing travel time deviation feature. The SEM model structure and estimation results facilitate the understanding of bus service performance characteristics and provide several implications for bus service planning, management, and operation.

1. Introduction

Bus service reliability can have significant impacts on the service providers and the existing and potential users [1]. From the passenger's perspective, reliable bus services present predictable travel times and wait times; from bus agencies point of view, they benefit from stable ridership of passengers who are satisfied with reliable services. As a result, the public transit administrative authorities take service reliability as one of the vital performance measures [2]; and transportation researchers take into account the bus traveling randomness in bus assignment modeling and network and operation design [3–5]. For fixed-route bus services with fixed timetables and trajectories, on-time performance, and headway regularity are the most commonly used reliability measures [2], while travel time variability can be an important agencies-concerned issue relating to these two service reliability measures. The focus of bus service operation and management is on travel time reliability; travel times are core components of travelers' travel cost in transit assignment modeling. Thus, it

is of great importance to investigate bus travel time reliability.

Many researchers have made efforts to explore the indices definition, overall features, and descriptive cause analysis of travel time reliability [1, 6]. However, this study adopts Structural Equation Modeling (SEM) method to specify, estimate, and measure the proposed theoretical model for analyzing travel time deviation from schedules on the microscopic bus route segment level. Compared with the previous studies on transit reliability analysis employing regression methods [7, 8], the SEM model establishes and verifies hypotheses representing interrelationships among observed variables based on existing theories and empirical results. The relevant variables, denoting departure delays and upstream travel time deviations, are embedded into the SEM models so as to reveal bus schedule recovery phenomenon first investigated by Kalaputapu and Demetsky [9]. Meanwhile, the availability of bus Automatic Vehicle Location (AVL) systems makes it feasible to conduct the microscopic modeling and analysis on the bus segment level.

This study begins with establishing hypotheses based on literature review about bus service reliability analysis, followed by Structural Equation Models specification. Then, it conducts the SEM model testing and modification by examining model estimation results in terms of estimates statistics and multiple fitness measures. With the respecified SEM model, the fitness of the entire model and estimates of path coefficients are discussed. Finally, research conclusions and relevant implications for bus service planning and operation are present.

2. Literature Review and Research Hypotheses

In Structural Equation Modeling, five basic steps should be followed, namely, model specification, model identification, model estimation, model testing, and model modification. The meaningfulness of correlation relationships in specified models depends on the employed variables and reasonable hypotheses. Hence, the theoretical hypotheses are very important and should be based on previous research. This section reviews related literature and proposes theoretical assumptions for model specification.

As aforementioned, on-time performance and headway regularity are key measures of bus service reliability, while travel time variability performs as an important and essential issue relating to these two reliability measures. This research gets insight into the internal and external factors influencing bus service reliability especially travel time deviations on the bus route segment level.

2.1. Effects of Internal and External Factors on Bus Service Reliability. Causes of unreliability analysis for bus service have been well documented by Cham [1], TCRP-88 [2], TCQSM [10], Abkowitz and Engelstein [11], and Abkowitz et al. [12]. Deriving from the internal bus systems or external traffic conditions, a number of factors affect bus travel times resulting in travel time variability and service unreliability. According to the previous research [1, 2], travel time delays are impacted by major factors involving departure delays, number of stops made, dwell times, number of traffic signals, and so forth. Intuitively, the existence of signalized intersections leads to the variability of travel times due to bus random arrivals at traffic signals; access connections on the road represent conflict points where buses interact with the merging and diverging vehicles. Consequently, the following hypotheses are inferred and present:

(H1) The dwell time has a direct and positive impact on travel time deviation.

(H2) Number of signalized intersections has a direct and positive impact on travel time deviation.

(H3) Number of access connections has a direct and positive impact on travel time deviation.

Apart from the above major interrelationships, it is likely that departure deviations at bus terminals cause an increase in passenger boarding (namely, dwell times) at bus stops further downstream. Increased boarding at bus stops results in longer dwell times, which increase total travel times [1]. Meanwhile, longer bus stop spacing makes it more likely for buses to traverse more traffic signals and access connections. Thus, the following assumptions are proposed:

(H4) Departure delays directly and positively impact dwell times.

(H5) Segment length directly and positively impacts number of traffic signals.

(H6) Segment length directly and positively impacts number of access connections.

2.2. Bus Drivers Schedule Recovery Behaviors. Provided that bus travel time deviations from schedules exist, bus drivers could be motivated to adjust travel speeds to ensure the schedule adherence. This schedule recovery behavior of bus drivers was first investigated by Kalaputapu and Demetsky [9]. Other researchers considered the schedule recovery effort as a control factor in modeling bus arrival time prediction and schedule optimization problems [13–15]. Chen et al. [14] correlated the travel time delays on upstream segments with the travel time deviation on the segment under consideration. Similarly, Lin and Bertini [15] deem it reasonable that arrival time delays at two adjacent stops are strongly correlated, but delays for two stops far apart are usually weakly correlated. Besides the upstream delays having been considered above, departure punctuality at terminals is an important measure of bus service performance, impacts dwell times and travel times on downstream segments, and contributes to bus drivers schedule recovery efforts. Therefore, this study raises another two variables in SEM, departure delays at bus terminals and accumulated delays (of travel time on upstream segments), to explore schedule recovery phenomenon.

(H7) The accumulated delay has a direct and negative influence on travel time deviation.

(H8) The departure delay has a direct and negative influence on travel time deviation.

(H9) The departure delay has a direct and negative influence on the accumulated delay.

According to Lin and Bertini [15], how fast bus drivers can bring the bus back on schedule depends on the magnitude of deviation and the length of the remaining trip. Based on this inference, the following hypotheses are established:

(H10) The segment length has a direct and negative influence on travel time deviation.

(H11) Percentage of completed trip has a direct and negative influence on travel time deviation.

Based on the above eleven research hypotheses, there are three exogenous variables which are assumed to be not affected by other variables, and five endogenous variables supposed to have unidirectional causal relationships with exogenous variables or other endogenous variables. Assume that the ith bus trip on the kth segment is under consideration. The bus route segment k originates from the bus stop k and terminates at bus stop $k + 1$, where $k = 1, 2, \ldots, n$.
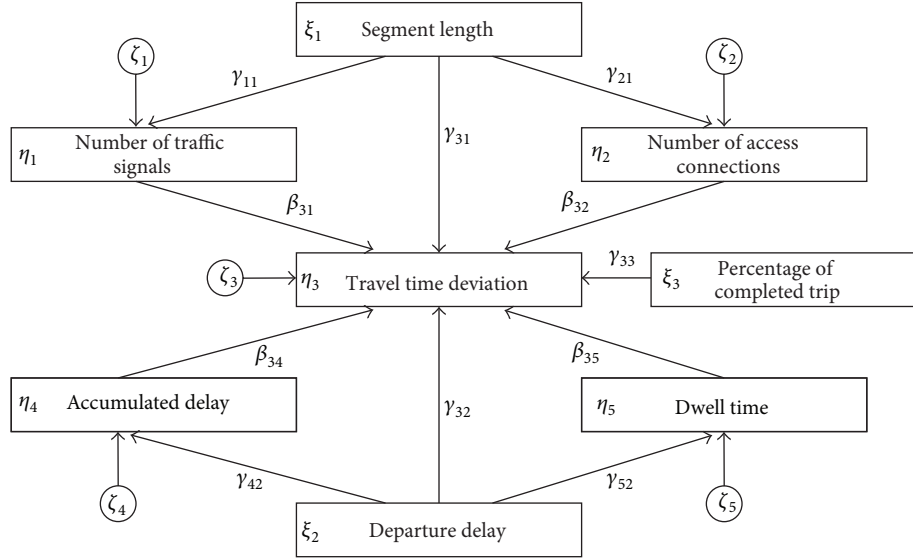
FIGURE 1: Path diagram of the original SEM model for bus travel time deviation analysis.

3. SEM Model Specification

Structural Equation Modeling uses various types of models to depict relationships among observed variables, with the same basic goal of providing a quantitative test for the hypothesized theoretical models [16]. In detail, this approach refers to a series of statistical methodologies, including path analysis, confirmatory factor analysis, and structural regression models. In the 1980s, researchers introduced this approach to travel behavior studies [17, 18]. Until now, SEM methods have been applied to transportation market segmentation [19, 20], travel behavior analysis [21–23], and service quality and satisfaction study [24–26].

Based on the inferred hypotheses in the previous section and the corresponding variables depicted in Table 1, this research develops a SEM model as follows.

$$\eta = B\eta + \Gamma\xi + \zeta, \tag{1}$$

where η is the column vector of the five endogenous variables, ξ is the column vector of the three exogenous variables, B is the matrix of path coefficients denoting the direct effects of endogenous variables on other endogenous variables, Γ is the matrix of path coefficients indicating the direct effects of exogenous variables on endogenous variables, and ζ is the column vector of estimation errors for five endogenous variables.

Equation (1) can be expressed in the vector and matrix form as (2). The relevant path diagram of the proposed theoretical SEM model is shown in Figure 1. Consider

The regarding variables for the SEM models are notated and described in Table 1.

$$
\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix}_{5\times1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 & \beta_{34} & \beta_{35} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}_{5\times5} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix}_{5\times1}
$$
$$
+ \begin{bmatrix} \gamma_{11} & 0 & 0 \\ \gamma_{21} & 0 & 0 \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \\ 0 & \gamma_{42} & 0 \\ 0 & \gamma_{52} & 0 \end{bmatrix}_{5\times3} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}_{3\times1} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \end{bmatrix}_{5\times1}. \tag{2}
$$

Suppose that S is the sample covariance matrix of the exogenous and endogenous variables, and Σ is the theoretical SEM (see (1) and (2)) implied covariance matrix. The SEM models estimation process adopts particular fitting functions to minimize the discrepancy between Σ and S and to obtain estimates for each of the parameters specified by SEM models. In this study, the generalized least squares (GLS) method is employed in SEM models estimation, given the specific sample size [27].

According to previous research, a sample size needs to be sufficient to achieve the desired precision level of path coefficients estimates and model fit. On one hand, the sample size should be greater than 200 for an acceptable model [28]; on the other hand, it should be ten times or fifteen times the number of the observed variables [29]. This study focuses on peak hour periods when bus travel time deviations occur frequently. The first sample used to test the proposed original model includes 209 observations, deriving from the eleven

TABLE 1: Exogenous and endogenous variables for SEM models.

	Name	Notation	Description
Endogenous variables	Dwell time	DT_{ik}	Difference between actual departure time (ADT_{ik}) and actual arrival time (AAT_{ik}) for bus trip i at the bus stop k, that is, $DT_{ik} = ADT_{ik} - AAT_{ik}$
	Travel time deviation	TTD_{ik}	Deviation of actual travel time (ATT_{ik}) from scheduled travel time (STT_{ik}) for bus trip i on the bus route segment k, that is, $TTD_{ik} = ATT_{ik} - STT_{ik}$, where $ATT_{ik} = AAT_{i,k+1} - AAT_{ik}$
	Accumulated delay	AD_{ik}	Sum of travel time deviation on the upstream segments from 1 to $k-1$, that is, $AD_{ik} = \sum_{j=1}^{k-1} TTD_{ij}$
	Number of traffic signals	TS_k	Number of traffic signals on the bus route segment k
	Number of access connections	AC_k	Number of access connections on the bus route segment k
Exogenous variables	Segment length	L_k	Length of the bus route segment k between adjacent bus stops
	Departure delay	DD_i	Deviation of actual departure time (ADT_i) from scheduled departure time (SDT_i) at the bus terminal for the bus trip i, that is, $DD_i = ADT_i - SDT_i$
	Percentage of completed trip	PCR_{ik}	Percentage of travel time that buses spent on upstream segments, that is, $PCR_{ik} = \sum_{j=1}^{k-1} ATT_{ij} / \sum_{j=1}^{n} ATT_{ij}$

TABLE 2: Model fit measures for the original and modified SEM models.

Measure name	Acceptable level	Original model	Modified Model A	Modified Model B
Chi-square χ^2	The smaller the nonsignificant ($p > 0.050$) value the better	96.500 ($p = 0.000$)	78.066 ($p = 0.000$)	21.399 ($p = 0.045$)
Ratio of Chi-square to the degrees of freedom (df)	$1 < \chi^2/\text{df} < 3$ reflects a good model fit	5.676	4.337	1.783
Goodness-of-fit (GFI)	GFI > 0.95 reflects a good model fit	0.884	0.906	0.971
Adjusted GFI (AGFI)	AGFI > 0.95 reflects a good model fit	0.754	0.812	0.931
Root-mean-square error of approximation (RMSEA)	RMSEA < 0.05 indicates a good model fit 0.05~0.08 indicates reasonable fit	0.150	0.127	0.061

bus trips in peak hour period (7:00-8:00) on May 14, 2012. Another sample with 209 observations from eleven bus trips during the same service period on May 15 is collected to validate the modified models.

The data in the above samples derives from the AVL archived records, the schemed timetable, and the field survey data of a bus route numbered 102 in Suzhou City, China. Specifically, the bus trip information (namely, actual departure times ADT and actual arrival times AAT) is directly extracted from AVL archived records; the bus route timetable presents scheduled departure times SDT and scheduled travel times STT; the data concerning bus route segments (lengths L, number of traffic signals TS, and number of access connections AC) is collected by field surveys. Data entry and editing are conducted in the statistical software package SPSS. With the sample data and path diagram of theoretical SEM model as inputs, the step of model estimation is performed by using the SEM software, Amos of version 17.0.

4. SEM Model Testing and Modification

In order to inspect how well the sample data supports the proposed theoretical SEM model, the model testing procedure needs to be carried out by examining the goodness of fit for the entire model and the statistical significance for the individual parameters. The original model is estimated by inputting the first sample which includes 209 observations, deriving from the eleven bus trips in peak hour period (7:00-8:00) on May 14, 2012. Another sample with 209 observations from eleven bus trips during the same service period on May 15 is collected to validate the modified models.

As to the entire model fit test, SEM has a large number of model fit measures. Most of these measures are established based on the discrepancy between Σ and S, which is referred to as Chi-square χ^2 [16, 21, 25]. The model fit measures typically used are listed in Table 2. The third column in Table 2 shows that the five fitness measures for the original model defined by Figure 1 cannot reach the acceptable levels, illustrating the poor fitness of the proposed original model. To improve the goodness of model fit, researchers are inclined to add or remove paths in the originally proposed model based on the statistical significance of path coefficients. As a result, the following model modification and testing are conducted.

4.1. Model Modification A. According to the critical ratios (CR) and p values for path coefficients in columns 2 and 3 of Table 3, most estimates of the path coefficients have values significantly different from zero. But the p value (0.688) for the path, departure delay $\rightarrow$ dwell time, is extremely great compared with 0.000. Correspondingly, the correlation between departure delay and dwell time in the sample correlation matrix of Table 4 is −0.036 indicating low correlativity. Accordingly, the path (denoting hypothesis (H4)) is removed from the path diagram of the original model and Modified Model A is raised.

4.2. Model Testing A. Model estimation process is performed for this modified SEM model with the new sample as inputs. For Modified Model A, all of the model fit measures in the fourth column in Table 2 cannot reach the good model fit thresholds, reflecting that Modified Model A need to be respecified further.

4.3. Model Modification B. By analyzing the CR and p values for path coefficients in columns 4 and 5 of Table 3, it is found that the path, completed trip $\rightarrow$ travel time deviation, shows low significance with the p value of 0.401. Therefore, the corresponding path (denoting hypothesis (H11)) is removed to specify Modified Model B.

4.4. Model Testing B. It is shown in the fifth column of Table 2 that the vital fit measures for Modified Model B reflect good or reasonable model fit. Meanwhile, the path coefficients are statistically significant with all of the p values less than 0.100 shown in columns 6 and 7 of Table 3. As a result, the specification of Modified Model B with better parsimony and model fit can be accepted finally.

The finally modified and accepted model denoted as Modified Model B in Tables 2 and 3 can be represented by the path diagram in Figure 2.

5. Empirical Results

5.1. Theoretical SEM Model Fitness. For the finally modified and accepted model shown in Figure 2, the Chi-square value is 21.399, the degrees of freedom are 12, and thus the ratio of Chi-square to df, χ^2/df, is 1.783. Many researchers have

TABLE 3: Path coefficients for original and modified SEM models.

Hypothesized paths	Original model		Modified Model A		Modified Model B	
	CR	p	CR	p	CR	p
H1: dwell time $\rightarrow$ travel time deviation	6.693	$***$	3.368	$***$	3.419	$***$
H2: number of traffic signals $\rightarrow$ travel time deviation	3.397	$***$	3.028	0.002	3.053	0.002
H3: number of access connections $\rightarrow$ travel time deviation	−1.183	0.237	−2.255	0.024	−2.352	0.019
H4: departure delay $\rightarrow$ dwell time	*−0.402*	*0.688*	na	na	na	na
H5: segment length $\rightarrow$ number of traffic signals	11.132	$***$	12.054	$***$	16.534	$***$
H6: segment length $\rightarrow$ number of access connections	4.839	$***$	4.370	$***$	4.472	$***$
H7: accumulated delay $\rightarrow$ travel time deviation	−1.737	0.082	−1.813	0.07	−1.885	0.059
H8: departure delay $\rightarrow$ travel time deviation	−2.563	0.01	−3.410	$***$	−3.483	$***$
H9: departure delay $\rightarrow$ accumulated delay	−10.471	$***$	−12.560	$***$	−12.432	$***$
H10: segment length $\rightarrow$ travel time deviation	−3.882	$***$	−3.577	$***$	−3.694	$***$
H11: completed trip $\rightarrow$ travel time deviation	1.062	0.288	*0.839*	*0.401*	na	na

Note: $***$ denotes that p value < 0.001. "na" denotes "not applicable."

TABLE 4: Sample correlation matrix.

	Percentage of completed trip	Departure delay	Segment length	Dwell time	Accumulated delay	Number of access connections	Number of traffic signals	Travel time deviation
Percentage of completed trip	1.000	−0.035	0.426	0.128	0.340	0.145	0.282	*0.035*
Departure delay	na	1.000	0.000	*−0.036*	*−0.566*	0.000	0.000	**−0.121**
Segment length	na	na	1.000	0.008	0.027	**0.326**	**0.737**	−0.134
Dwell time	na	na	na	1.000	−0.051	−0.036	−0.020	**0.450**
Accumulated delay	na	na	na	na	1.000	0.166	−0.122	**−0.133**
Number of access connections	na	na	na	na	na	1.000	0.200	**−0.176**
Number of traffic signals	na	na	na	na	na	na	1.000	**0.070**
Travel time deviation	na	na	na	na	na	na	na	1.000

Note: "na" denotes "not applicable."
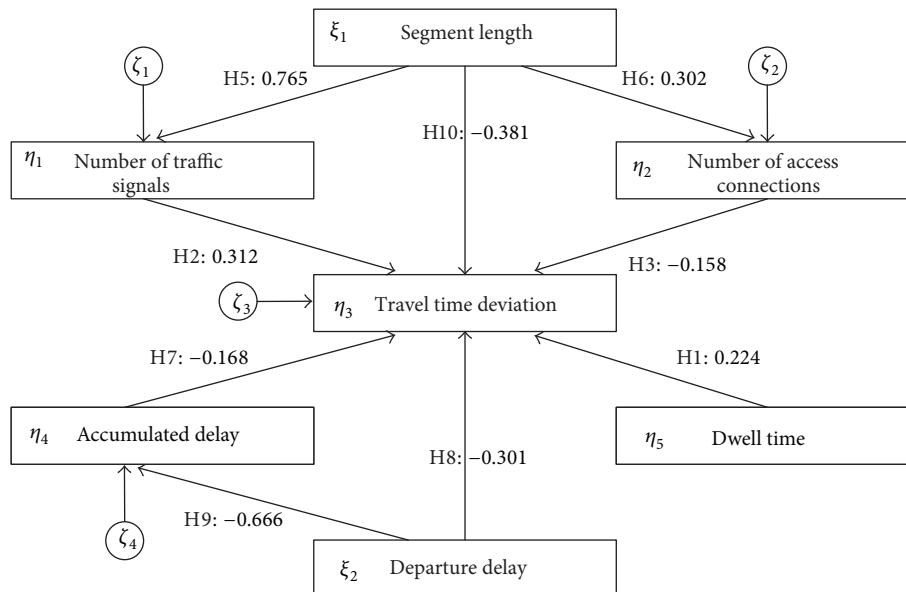
FIGURE 2: Path diagram of the modified SEM model for bus travel time deviation analysis.

TABLE 5: Standardized direct, indirect, and total effects.

	Departure delay	Dwell time	Segment length	Accumulated delay	Number of access connections	Number of traffic signals
Accumulated delay						
Direct	−0.666	na	na	na	na	na
Indirect	0.000	na	na	na	na	na
Total	−0.666	na	na	na	na	na
Number of access connections						
Direct	na	na	0.302	na	na	na
Indirect	na	na	0.000	na	na	na
Total	na	na	0.302	na	na	na
Number of traffic signals						
Direct	na	na	0.765	na	na	na
Indirect	na	na	0.000	na	na	na
Total	na	na	0.765	na	na	na
Travel time deviation						
Direct	*−0.301*	0.224	*−0.381*	−0.168	−0.158	0.312
Indirect	*0.112*	0.000	*0.191*	0.000	0.000	0.000
Total	*−0.189*	0.224	*−0.190*	−0.168	−0.158	0.312

Note: "na" denotes "not applicable."

suggested the use of the ratio as a measure of fit but different values for acceptable levels to indicate a reasonable fit [25, 30]. However, the ratio less than 2.0 generally indicates a reasonable fit between the hypothetical SEM model and the sample data. The goodness of fit (GIF) and adjusted goodness of fit (AGIF) for the modified SEM model are 0.971 and 0.931, respectively, very close to 1.0. The value of RMSEA is 0.061. These measures in the fifth column of Table 2 yield supportive indices for the reasonable SEM model structure and also suggest that the sample data fits the final model well.

5.2. Hypothesis Testing Results. Figure 2 presents the path diagram with the path coefficients representing standardized estimates of regression weights. The standardized path coefficients are useful in determining the relative importance of each variable to other variables for a given sample. In addition, standardized path coefficients make it feasible to interpret interrelationships on the same scale of measurement [16]. Causal relationships between the physical features of the bus route segment and travel time deviation, between the departure and arrival delay and travel time deviation, can be illustrated by the magnitude and sign of standardized coefficients.

The path coefficients, from departure delay to accumulated delay, from departure delay to travel time deviation, and from accumulated delay to travel time deviation, are −0.666, −0.301, and −0.168, respectively. The negative signs verify hypotheses (H7), (H8), and (H9). They also imply that bus drivers attempt to reduce travel time deviation and pursue schedule adherence by schedule recovery behavior, in cases where departure delay and travel time deviations from upstream segments occur.

Dwell time does have a vital impact on travel time variability as the path coefficient from dwell time to travel time deviation takes a medium value of 0.224. It suggests that passenger boarding and alighting at bus stops should be paid attention to and treated as independent variable in service reliability analysis and service planning modeling.

Causal relationships between physical feature of roadway segment and travel time deviation are implied by path coefficients for hypotheses (H2), (H3), (H5), (H6), and (H10). The hypotheses, (H5) and (H6), take positive values of 0.765 and 0.302, respectively, consistent with the common sense that buses traverse more intersections and access connections with longer stop spacing. Compared with number of access, traffic signals lead to the fluctuation of travel time more intensively. We can infer that the stop delays for buses at traffic signals make great contribution to the total bus delay on segment. As supposed, the direct effect of segment length on travel time deviation is negative. The greatest absolute value of 0.381, among the coefficients for paths from other variables to travel time deviation, proves that travel distance plays an essential role in bus drivers' schedule recovery behavior. It is more likely for bus drivers to bring the buses back to schedule with long travel distance.

5.3. Direct and Indirect Effects. It is shown that, in Table 5, the exogenous variables, departure delay and segment length, both have direct and indirect effects on the endogenous variable travel time deviation.

The correlation (−0.189) between departure delay and travel time deviation is the sum of (i) the direct effect (−0.301) of departure delay on travel time deviation and (ii) the indirect effect ($0.112 = −0.666 * (−0.168)$) of departure

delay on travel time deviation through accumulated delay. The direct and indirect effects take reverse signs, representing negative and positive influences, respectively.

Rather than comparing path coefficients −0.301 and −0.168 directly, we suggest that departure delay and travel time delay of upstream segments have a similar influence on travel time variability as they have the correlations or total effects of −0.189 and −0.168, respectively.

The correlation (−0.190) between segment length and travel time deviation also consists of direct and indirect effects. On one hand, segment length directly and negatively impacts travel time variability and inversely positively contributes to the schedule recovery process; on the other hand, this exogenous variable indirectly and positively influences the variance of segment travel times through the existence of signalized intersections and access connections on the road.

6. Conclusions and Implications

This study investigates the influences of causes of unreliability and bus drivers' schedule recovery efforts on travel time variance. The theoretical hypothesized SEM modeling these interrelationships demonstrates excellent fitness with multiple measures. In other words, most of the preestablished hypotheses are adequately supported by the research sample dataset from bus AVL system. The SEM model structure and estimation results facilitate the understanding of bus service performance characteristics and provide several implications for bus service planning, management, and operation.

The final SEM model confirms the schedule recovery phenomenon, namely, bus drivers' active schedule adherence behavior. The departure delays at the bus terminals and upstream travel time deviations indeed have a negative impact on travel time fluctuation of buses en route. It also shows that these two portions of delays have similar magnitudes of total effects on travel time deviations. Given that upstream travel time delays have been taken into consideration in bus service planning and arrival time prediction of the existing research, particular emphasis should be placed on departure delays at bus terminals. Thus, there is a need to embed departure punctuality into bus operation modeling.

It is known that traffic signals on the road cause additional travel time delays for the passing buses. In this study, the number of signalized intersections on bus segments is taken as an observed variable in SEM model and found to positively affect travel time deviation. It comes to a conclusion that treatments reducing the stop delays of buses at traffic signals will make travel time deviations decrease. This kind of treatments often refers to active and passive transit signal priority controls.

As discussed in the last section, the segment length or bus stop spacing directly and negatively impacts travel time variability and inversely positively contributes to the schedule recovery process; on the other hand, it indirectly and positively influences travel time variance through the existence of signalized intersections and access connections. To optimize bus stop spacing, bus service researchers and planners can refine stop spacing model by taking account of its effects on travel time reliability and bus schedule recovery.

The parsimony for the proposed SEM model in this study is promised, and all of the hypothesized paths are based on well-known empirical research and supported by real-world data. But in the future work, it is advised to explore the correlation between bus service reliability and additional observed variables, such as those regarding transit preferential treatments.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] L. C. Cham, *Understanding Bus Service Reliability: A Practical Framework Using AVL/APC Data*, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2006.

[2] Transportation Research Board of the National Research Council, "A guidebook for developing a transit performance-measurement system," TCRP Report 88, Transportation Research Board of the National Research Council, Washington, DC, USA, 2003.

[3] W. Y. Szeto, Y. Jiang, K. I. Wong, and M. Solayappan, "Reliability-based stochastic transit assignment with capacity constraints: formulation and solution method," *Transportation Research C: Emerging Technologies*, vol. 35, pp. 286–304, 2013.

[4] Z. Liu, Y. Yan, X. Qu, and Y. Zhang, "Bus stop-skipping scheme with random travel time," *Transportation Research Part C: Emerging Technologies*, vol. 35, pp. 46–56, 2013.

[5] Y. Yan, Z. Liu, Q. Meng, and Y. Jiang, "Robust optimization model of bus transit network design with stochastic travel time," *Journal of Transportation Engineering*, vol. 139, no. 6, pp. 625–634, 2013.

[6] X. Chen, L. Yu, Y. Zhang, and J. Guo, "Analyzing urban bus service reliability at the stop, route, and network levels," *Transportation Research A: Policy and Practice*, vol. 43, no. 8, pp. 722–734, 2009.

[7] G. S. Rutherford and K. E. Watkins, "Measurement and evaluation of transit travel time reliability," in *Transportation Northwest Final Report*, Transportation Northwest, Seattle, Wash, USA, 2011.

[8] A. M. El-Geneidy, J. Horning, and K. J. Krizek, "Analyzing transit service reliability using detailed data from automatic vehicular locator systems," *Journal of Advanced Transportation*, vol. 45, no. 1, pp. 66–79, 2011.

[9] R. Kalaputapu and M. J. Demetsky, "Modeling schedule deviations of buses using automatic vehicle-location data and

artificial neural networks," *Transportation Research Record*, no. 1497, pp. 44–52, 1995.

[10] Transportation Research Board of the National Research Council, "TCQSM, transit capacity and quality of service manual," TCRP Report, Transportation Research Board of the National Research Council, Washington, DC, USA, 2013.

[11] M. Abkowitz and I. Engelstein, "Empirical methods for improving transit scheduling," in *Proceedings of the World Conference on Transport Research: Research for Transport Policies in a Changing World*, pp. 844–856, Hamburg, Germany, April 1983.

[12] M. Abkowitz, H. Slavin, and R. Waksman, *Transit Service Reliability*, U.S. Department of Transportation, Washington, DC, USA, 1978.

[13] Y. Yan, Q. Meng, S. Wang, and X. Guo, "Robust optimization model of schedule design for a fixed bus route," *Transportation Research C: Emerging Technologies*, vol. 25, pp. 113–121, 2012.

[14] M. Chen, X. Liu, and J. Xia, "Dynamic prediction method with schedule recovery impact for bus arrival time," *Transportation Research Record*, no. 1923, pp. 208–217, 2005.

[15] W.-H. Lin and R. L. Bertini, "Modeling schedule recovery processes in transit operations for bus arrival time prediction," in *Proceedings of the 5th IEEE International Conference on Intelligent Transportation Systems*, pp. 857–862, IEEE, Singapore, September 2002.

[16] R. E. Schumacker and R. G. Lomax, *A Beginner's Guide to Structural Equation Modeling*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, USA, 2004.

[17] P. M. Allaman, T. J. Tardiff, and F. C. Dunbar, "New approaches to understanding travel behavior," NCHRP Report, Transportation Research Board, Washington, DC, USA, 1982.

[18] P. K. Lyon, *Dynamic Analysis of Attitude-Behavior Response to Transportation Service Innovation*, Transportation Center, Northwestern University, Evanston, Ill, USA, 1981.

[19] Z. Li, W. Wang, C. Yang, and D. R. Ragland, "Bicycle commuting market analysis using attitudinal market segmentation approach," *Transportation Research Part A: Policy and Practice*, vol. 47, pp. 56–68, 2013.

[20] Y. Shiftan, M. L. Outwater, and Y. Zhou, "Transit market research using structural equation modeling and attitudinal market segmentation," *Transport Policy*, vol. 15, no. 3, pp. 186–195, 2008.

[21] V. Van Acker and F. Witlox, "Car ownership as a mediating variable in car travel behaviour research using a structural equation modelling approach to identify its dual relationship," *Journal of Transport Geography*, vol. 18, no. 1, pp. 65–74, 2010.

[22] J. de Abreu e Silva, C. Morency, and K. G. Goulias, "Using structural equations modeling to unravel the influence of land use patterns on travel behavior of workers in Montreal," *Transportation Research A: Policy and Practice*, vol. 46, no. 8, pp. 1252–1264, 2012.

[23] K. Deutsch, S. Y. Yoon, and K. G. Goulias, "Modeling travel behavior and sense of place using a structural equation model," *Journal of Transport Geography*, vol. 28, pp. 155–163, 2013.

[24] J. Liu and P. Zhao, "Constructing quality-satisfaction-loyalty chain system: an empirical study of insurance industry in China," in *Proceedings of the International Conference on Services Systems and Services Management (ICSSSM '05)*, vol. 2, pp. 1237–1241, June 2005.

[25] J.-S. Chou and C. Kim, "A structural equation analysis of the QSL relationship with passenger riding experience on high speed rail: an empirical study of Taiwan and Korea," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6945–6955, 2009.

[26] J. de Oña, R. de Oña, L. Eboli, and G. Mazzulla, "Perceived service quality in bus transit service: a structural equation approach," *Transport Policy*, vol. 29, pp. 219–226, 2013.

[27] L.-T. Hu, P. M. Bentler, and Y. Kano, "Can test statistics in covariance structure analysis be trusted?" *Psychological Bulletin*, vol. 112, no. 2, pp. 351–362, 1992.

[28] T. F. Golob, "Structural equation modeling for travel behavior research," *Transportation Research Part B: Methodological*, vol. 37, no. 1, pp. 1–25, 2003.

[29] B. Thompson, "Ten commandments of structural equation modeling," in *Reading and Understanding MORE Multivariate Statistics*, L. G. Grimm and P. R. Yarnold, Eds., American Psychological Association, Washington, DC, USA, 2000.

[30] X. Cao, P. L. Mokhtarian, and S. L. Handy, "Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach," *Transportation*, vol. 34, no. 5, pp. 535–556, 2007.

Research Article

Geometric-Process-Based Battery Management Optimizing Policy for the Electric Bus

Yan Li, Jin-kuan Wang, Peng Han, and Ying-hua Han

School of Information Science and Engineering, Northeastern University, Shenyang 110819, China

Correspondence should be addressed to Yan Li; liyan@mail.neuq.edu.cn

With the rapid development of the electric vehicle industry and promotive policies worldwide, the electric bus (E-bus) has been adopted in many major cities around the world. One of the most important factors that restrain the widespread application of the E-bus is the high operating cost due to the deficient battery management. This paper proposes a geometric-process-based (GP-based) battery management optimizing policy which aims to minimize the average cost of the operation on the premise of meeting the required sufficient battery availability. Considering the deterioration of the battery after repeated charging and discharging, this paper constructs the model of the operation of the E-bus battery as a geometric process, and the premaintenance time has been considered with the failure repairment time to enhance the GP-based battery operation model considering the battery cannot be as good as new after the two processes. The computer simulation is carried out by adopting the proposed optimizing policy, and the result verifies the effectiveness of the policy, denoting its significant performance on the application of the E-bus battery management.

1. Introduction

With the intensification of the environmental pollution as well as the fast development of the smart grid and the vehicle industry, the electric vehicle (EV) will be a significant trend in the vehicle industry for its environment-friendly, energy saving, and high energy utilization rate characteristics [1, 2]. With the massive access of the EVs in the near future, the electric bus, that is, E-bus, will constitute a significant part in the public transportation [3]. And, with the development of the battery-switching technology, the battery-switching station (BSS) is the necessary refueling infrastructure for the adoption of the E-bus [4, 5].

Due to the aging of the E-bus batteries through repeated charging and discharging as well as stochastic impacts of the daily usage, the state of health (SOH) of the batteries degenerates, causing the rising of batteries failing probability. Those disadvantages above will result in the reduction of the battery life-cycle, the decline of the battery's availability, and the increasement of the E-bus daily cost-in-use. Therefore, the management of the battery should be optimized to further enhance the operating availability and reduce the long-run average cost per unit time, and many related studies have been carried out focusing on the problems of battery operation and repairment.

Gould et al. [6] propose an adaptive battery model to observe the battery voltages and monitor the degradation of the estimated dynamic model parameters, and Agarwal et al. [7] build the battery model that incorporates with the recovery effect for accurate life-cycle estimation and make a suggestion about the maximum available energy approximated at charge/discharge nominal power level.

In the state of charge (SOC) estimation, Wang and Liu [8] gather the parameters of battery voltage, current, and temperature in real time and estimate the battery state of charge as well as implementing the fault diagnosis or alarming according to the battery status to achieve the full and efficient use of battery power.

In the SOH monitoring, Li et al. [9] study the EV battery monitoring and management system based on monitoring and analyzing the performance of Li-ion battery and the battery fault in the operation of electric vehicle; Han et al. [10] propose a novel health prediction model of Li-ion battery based on sample entropy and establish the prediction
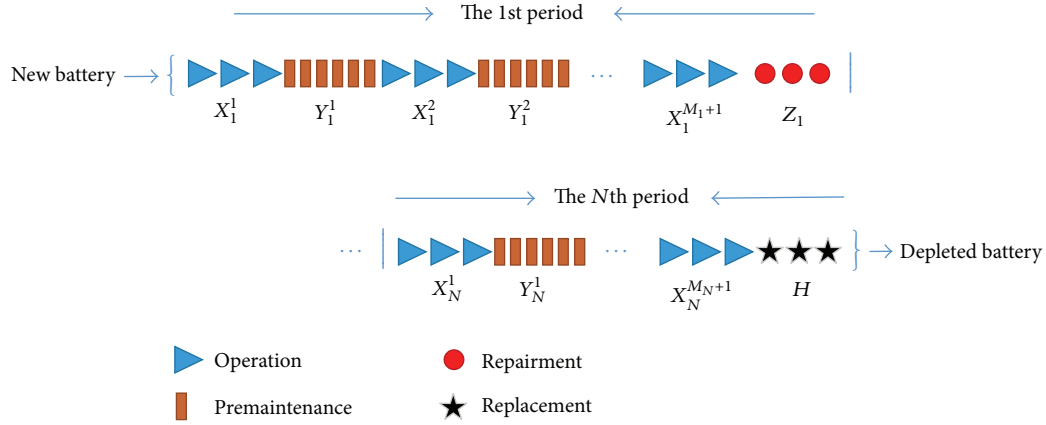
FIGURE 1: Life-cycle of the E-bus battery.

model by calculating the sample entropy, using Arrhenius formula and optimizing and fitting polynomial; Paul-Henri and Vincent [11] propose a scheme of state of charge and state of health estimation using a hybridization of Kalman filtering, Recursive Least Squares approach, and Support Vector Machines learning; and Dung et al. [12] propose a state of health estimation system based on time-constant-ratio measurement, which achieves the purpose of an environmental-impedance-free and fast SOH estimation.

However, the previous modeling, estimation, and monitoring studies related to the EV batteries mainly concentrate on the prediction of the SOC or SOH and seldom consider the effects of the battery premaintenance, which is critical in prolonging the life-cycle of the battery. And the geometric-process-based (GP-based) method, with its practical advantages in characterizing the whole operation process of the components in a system [13], gains worldwide scholar's attention in recent years, and many researches have been carried out based on GP in different application domains. Tan et al. [14] and Jia et al. [15] study an optimal maintenance strategy for one component and present the optimal tradeoff model of cost and availability. Lam [16] introduces a geometric process δ-shock maintenance model for a repairable system and adopts a replacement policy for minimizing the long-run average cost per unit time. Wang and Zhang [17–20] as well as Zhang [21] propose a GP-based preventive repairment policy to solve the efficiency for a deteriorating and valuable system to minimize the average cost rate, and Han et al. [22] present a generalized formulation for determining the optimal operating strategy and cost optimization for battery and formulate the operating strategy as a nonlinear optimization problem. Most of the studies based on GP denote a good applicability in the related fields, but their assumption that the property of the component, for example, the EV battery, after premaintenance is as good as new is not fully compatible with the E-bus battery considering its chemical reactions within. Aiming at improving the insufficient aspects mentioned above, the main research work of this paper is carried out as follows:

(1) A monotone process model of the E-bus battery is proposed by applying GP theory.

(2) The process of the premaintenance is modelled on condition that the battery, after premaintenance, cannot be as good as new.

(3) The long-run average cost per unit time is optimized on the premise of meeting the required availability under the proposed E-bus battery management optimizing policy.

The content of the paper is arranged as follows: in Section 2, the mathematical model is proposed for the E-bus battery life-cycle based on GP, and the necessary definitions and the assumptions are given; in Section 3, E-bus battery management optimizing policy is applied to estimate the long-run average cost per unit time under the constraint of the availability; in Section 4, the simulation results are evaluated for the 220 Ah Li-ion battery and the effectiveness of the proposed E-bus battery management optimizing policy is verified; and the paper concludes in Section 5.

2. Mathematical Model

The mathematical model of the E-bus battery life-cycle based on GP is firstly proposed. Assume that the E-bus battery is new at the beginning, and the initial state of health (SOH) is R_0. When a single operation is completed, the E-bus battery will be premaintained. Then the maintenance policy (M, N) is applied, where the amount of the premaintenance within a single failure repairment is M while the failure repairment amount during the battery life-cycle is N.

By the definition of GP, a complete life-cycle of an E-bus battery is actually a time interval between the beginning of the battery's initiated utilization and the first replacement or a time interval between two consecutive replacements, and the life-cycle of the E-bus battery is illustrated in Figure 1. For $n = 1, 2, \ldots, N - 1$, the time interval between the $(n - 1)$th and the nth failure repairment in a cycle can be defined as the nth period of the E-bus battery life-cycle. Given the battery operating time after the ith premaintenance in the nth period $\{X_n^i, n = 1, 2, \ldots, i = 1, 2, \ldots\}$, the ith premaintenance time in the nth period $\{Y_n^i, n = 1, 2, \ldots, i = 1, 2, \ldots\}$, and the failure repairment time in the nth period $\{Z_n, n = 1, 2, \ldots\}$,

let $\{M_n, \; n = 1, 2, \ldots\}$ be the amount of the premaintenance in the nth period and H the replacement time. Finally, the replacement policy (M, N) is applied by which the battery is replaced by a new one at the time when the SOH is less than or equal to the threshold R after the Nth period. And the depleted battery that has been replaced will be sold or recycled for the secondary usage.

Considering the chemical property of the battery, the battery's SOH will deteriorate in the process of repeated charging and discharging. The following definitions and assumptions are given.

Definition 1. The depreciation rate of the battery after a single operation is w_1, the recovery rate after a single premaintenance is w_2, and the descent rate after the failure repairment is w_3, where $(1 + w_2)(1 - w_1) < 1$ denotes that the premaintenance process will improve the performance of the battery by reducing but not eliminating the loss of the SOH during the operating process.

Definition 2. The successive battery operating time $\{X_n^i, \; i = 1, 2, \ldots\}$ after premaintenance in the nth period forms a GP with ratio a and $E(X_1^1) = \lambda$, and the premaintenance time $\{Y_n^i, \; i = 1, 2, \ldots\}$ in the nth period forms a GP with ratio b and $E(Y_1^1) = \mu$, while the consecutive failure repairment time $\{Z_n, \; n = 1, 2, \ldots\}$ constitutes a GP with ratio c and $E(Z_1) = v$. Besides, the replacement time H is a random variable with $E(H) = \tau$.

Definition 3. The premaintenance cost rate is c_m, the failure repairment cost rate is c_f, the benefit rate of the battery operating is c_b, the replacement cost rate is c_r corresponding to the replacement time H, the cost of the new battery is c_N, and the depleted battery recycling benefit is c_c.

Assumption 4. The processes of $\{X_n^i, \; i = 1, 2, \ldots\}$, $\{Y_n^i, \; i = 1, 2, \ldots\}$, and $\{Z_n, \; n = 1, 2, \ldots\}$ are independent. Besides, X_n^i, $i = 1, 2, \ldots, M + 1$, $n = 1, 2, \ldots, N$ and Y_n^i, $i = 1, 2, \ldots, M$, $n = 1, 2, \ldots, N$ are independent and identically distributed (i.i.d.) random variables.

Assumption 5. Neither the premaintenance nor the repairment can repair the battery to be as good as new.

Based on Definitions 1–3 and Assumptions 4-5, the studied policy is based on the GP model with premaintenance. The battery after premaintenance is not as good as new and the successive operating time in one period will form a GP. Furthermore, the GP model is a deteriorating system for $a > 1$, and it is an improving system for $b < 1$ and $c < 1$.

3. E-Bus Battery Management Optimizing Policy

The optimization of the E-bus battery management policy is to find the proper M and N that minimize the long-run average cost per unit time on the premise of meeting the required availability. In the defined mathematical model, the related functions are deduced.

The successive operating time with the ratio a can be derived through the following function:

$$
\begin{aligned}
E\left[\sum_{i=1}^{M+1} X_1^i\right] &= \sum_{i=1}^{M+1} \frac{\lambda}{a^i}, \\
E\left[\sum_{i=1}^{M+1} X_2^i\right] &= A \sum_{i=1}^{M+1} \frac{\lambda}{a^i}, \\
&\;\;\vdots \\
E\left[\sum_{i=1}^{M+1} X_N^i\right] &= A^{N-1} \sum_{i=1}^{M+1} \frac{\lambda}{a^i},
\end{aligned}
\tag{1}
$$

where $A = a^{-M}$ and M is the amount of the premaintenance; then

$$
E \sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i = \sum_{n=1}^{N} \left(E \sum_{i=1}^{M+1} X_n^i \right) = \frac{1 - A^N}{1 - A} \sum_{i=1}^{M+1} \frac{\lambda}{a^i}. \tag{2}
$$

And the successive premaintenance time with the ratio b can be derived through the following function:

$$
\begin{aligned}
E\left[\sum_{i=1}^{M} Y_1^i\right] &= \sum_{i=1}^{M} \frac{\mu}{b^{i-1}}, \\
E\left[\sum_{i=1}^{M} Y_2^i\right] &= B \sum_{i=1}^{M} \frac{\mu}{b^{i-1}}, \\
&\;\;\vdots \\
E\left[\sum_{i=1}^{M} Y_N^i\right] &= B^{N-1} \sum_{i=1}^{M} \frac{\mu}{b^{i-1}}, \\
E \sum_{n=1}^{N} \sum_{i=1}^{M} Y_n^i &= \sum_{n=1}^{N} \left(E \sum_{i=1}^{M} Y_n^i \right) = \frac{1 - B^N}{1 - B} \sum_{i=1}^{M} \frac{\mu}{b^{i-1}},
\end{aligned}
\tag{3}
$$

where $B = b^{1-M}$.

The long-run average cost per unit time on the premise of meeting the required availability is the major target for optimizing the E-bus battery management. For the commercial operation of the BCS, the availability of the E-bus is the ratio of the battery operation time to its total life-cycle. In the proposed policy, firstly, the long-run average cost per unit time $C(M, N)$ is defined in function (4). Consider

$$
C(M, N) = \frac{C_{\text{Sum}}}{T}, \tag{4}
$$

where C_{Sum} and T are given by function (5). Consider

$$
\begin{aligned}
C_{\text{Sum}} = E\Bigg(& c_m \sum_{n=1}^{N} \sum_{i=1}^{M} Y_n^i - c_b \sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i + c_f \sum_{n=1}^{N-1} Z_n + c_r H \\
& + c_N - c_c \Bigg),
\end{aligned}
\tag{5}
$$

$$
T = E\left(\sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i + \sum_{n=1}^{N} \sum_{i=1}^{M} Y_n^i + \sum_{n=1}^{N-1} Z_n + H \right).
$$

Then,

$$C(M, N) = \frac{E\left(c_m \sum_{n=1}^{N} \sum_{i=1}^{M} Y_n^i - c_b \sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i + c_f \sum_{n=1}^{N-1} Z_n + c_r H + c_N - c_c\right)}{E\left(\sum_{n=1}^{N} \sum_{i=1}^{M} Y_n^i + \sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i + \sum_{n=1}^{N-1} Z_n + H\right)}$$

$$= \frac{c_m\left(\left(1 - B^N\right)/(1 - B)\right) \sum_{i=1}^{M} \left(\mu/b^{i-1}\right) - c_b\left(\left(1 - A^N\right)/(1 - A)\right) \sum_{i=1}^{M+1} \left(\lambda/a^i\right) + c_f \sum_{n=1}^{N-1} \left(v/c^n\right) + c_r \tau + c_N - c_c}{\left(\left(1 - B^N\right)/(1 - B)\right) \sum_{i=1}^{M} \left(\mu/b^{i-1}\right) + \left(\left(1 - A^N\right)/(1 - A)\right) \sum_{i=1}^{M+1} \left(\lambda/a^i\right) + \sum_{n=1}^{N-1} \left(v/c^n\right) + \tau}. \quad (6)$$

According to the update process with the E-bus battery management optimizing policy, the availability $A(M, N)$, as another key constraint in the battery life-cycle, is calculated through function (7). Consider

$$A(M, N) = \frac{T_{\text{Operation}}}{T} = \frac{E\left(\sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i\right)}{E\left(\sum_{n=1}^{N} \sum_{i=1}^{M} Y_n^i + \sum_{n=1}^{N} \sum_{i=1}^{M+1} X_n^i + \sum_{n=1}^{N-1} Z_n + H\right)}$$

$$= \frac{\left(\left(1 - A^N\right)/(1 - A)\right) \sum_{i=1}^{M+1} \left(\lambda/a^i\right)}{\left(\left(1 - B^N\right)/(1 - B)\right) \sum_{i=1}^{M} \left(\mu/b^{i-1}\right) + \left(\left(1 - A^N\right)/(1 - A)\right) \sum_{i=1}^{M+1} \left(\lambda/a^i\right) + \sum_{n=1}^{N-1} \left(v/c^i\right) + \tau}. \quad (7)$$

Considering the deteriorating performance of the battery in the case of repeated charging and discharging, it will be replaced by a new one when its SOH drops to R, while $R(M, N)$ is correlated with the depreciation rate of the battery after single operating w_1 and the recovery rate after the premaintenance w_2. Firstly, at the end of the 1st period, $R(M, 1)$ is reduced to

$$R(M, 1) = \left(1 - w_1\right)^{M+1} \left(1 + w_2\right)^M \left(1 - w_3\right) R_0. \quad (8)$$

Then, before the replacement, the SOH after Nth repairment $R(M, N)$ is shown in function (9) and is subject to $R(M, N) \geq R$. Consider

$$R(M, N) = R(M, 1)^{N-1} \left(1 - w_1\right)^{M+1} \left(1 + w_2\right)^M$$

$$= \left(\left(1 - w_1\right)^{M+1} \left(1 + w_2\right)^M\right)^N \left(1 - w_3\right)^{N-1} \quad (9)$$

$$= \left(1 - w_1\right)^{N(M+1)} \left(1 + w_2\right)^{NM} \left(1 - w_3\right)^{N-1}.$$

To acquire the relationship between variable M and N, take the logarithm for both sides of the inequation below:

$$N(M + 1) \ln\left(1 - w_1\right) + NM \ln\left(1 + w_2\right)$$

$$+ (N - 1) \ln\left(1 - w_3\right) \geq \ln R,$$

$$MN\left(\ln\left(1 - w_1\right)\left(1 + w_2\right)\right) + N \ln\left(1 - w_1\right)\left(1 - w_3\right)$$

$$- \ln\left(1 - w_3\right) \geq \ln R, \quad (10)$$

$$MN\left(\ln\left(1 - w_1\right)\left(1 + w_2\right)\right)$$

$$+ N \ln\left(1 - w_1\right)\left(1 - w_3\right) \geq \ln R\left(1 - w_3\right);$$

then the relationship between variable M and N will be

$$N = \left\lfloor \frac{\ln R\left(1 - w_3\right)}{M\left(\ln\left(1 - w_1\right)\left(1 + w_2\right)\right) + \ln\left(1 - w_1\right)\left(1 - w_3\right)} \right\rfloor. \quad (11)$$

With the certain parameter R being the SOH of the battery at the time replacement occurred, functions (6) and (7) are the bivariate functions about M and N. When M is fixed, the bivariate functions turn into univariate functions, and the management optimizing policy N^* is determined analytically or numerically. However, if M is a variable to be determined, the optimizing about $C(M^*, N^*)$ and $A(M^*, N^*)$ needs to be set up based on the rationalized configuration of M^* and N^*. Therefore, the objective function is to determine an optimal management policy (M^*, N^*) for minimizing the long-run average cost per unit time $C(M, N)$ and limiting the availability $A(M, N)$ in a certain range, which can be expressed as

$$C\left(M^*, N^*\right): \min \quad C(M, N) = \min\left(\frac{C_{\text{Sum}}}{T}\right)$$

$$\text{s.t.} \quad A\left(M^*, N^*\right) = \frac{T_{\text{Operation}}}{T} \geq A_0, \quad (12)$$

where A_0 is the threshold of the availability that meets the minimum demand of the E-bus.

4. Simulation Results

The computer simulation is adopted to verify the proposed GP-based E-bus battery management optimizing policy. Originally from the 220 Ah Li-ion battery, the main parameters of the GP [22] are listed in Table 1, and the key parameters of the cost in the simulation are listed in Table 2.

TABLE 1: Parameters setting of the GP.

Parameter	a	b	c	λ	μ	υ	τ	w_1	w_2	w_3	R_0
Value	1.02	0.995	0.8	10	3.5	9	5	0.8%	0.7%	1.5%	1

TABLE 2: Parameters setting of the cost.

Parameter	c_b	c_m	c_f	c_r	c_c	c_N
Value	10	12	25	40	500	2000
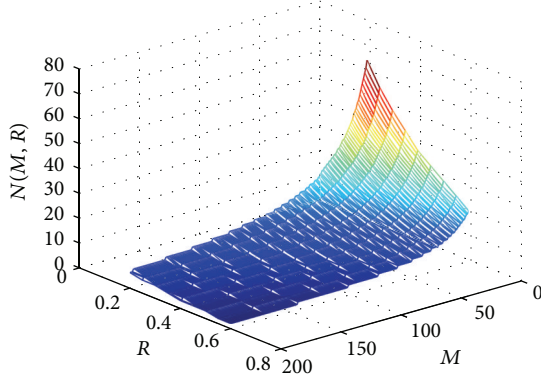
FIGURE 2: The amount of the failure repairment N with different M and R.

FIGURE 3: The average costs per unit time corresponding to the different M and N with $R = 0.4$.

FIGURE 4: The availabilities with different M and N.

In order to discover the relationship between M and N with different SOH thresholds, R, which empirically ranges from 0.2 to 0.6, and its corresponding simulation are firstly carried out and the results are illustrated in Figure 2.

Considering the long-run average cost per unit time $C(M, N)$ according to the battery management optimizing policy (M, N), the amounts of premaintenance and failure repairment as well as their correspondence relations are gained from the above simulation. Assume the SOH replacement threshold of the battery is 0.4; that is, $R = 0.4$; then the average costs per unit time corresponding to the different amounts of premaintenance and failure repairment are shown in Figure 3.

Due to the fact that the failure repairment cost is much more than the cost of the premaintenance, the higher average cost per unit time appears at the point of the lower value of M but higher value of N. And the lower average cost per unit time appears mostly at the point of higher value of M but lower value of N. Meanwhile, for the decline of the battery's SOH is presented in the process of repeated charging and discharging, the operating time of the model is a deteriorating system while the premaintenance time and failure repairment time are improving systems. The average cost per unit time declines at first and then increases gradually. So the lowest average cost per unit time exists and the lowest value is at the point where $(M^*, N^*) = (14, 4)$, as is shown in Table 3.

The optimal value of the average cost per unit time is further calculated as in Table 3. The following simulation is to verify if the availability corresponding to the selected point

$(M^*, N^*) = (14, 4)$ meets the proposed threshold A_0, which is set to be 0.5 according to the empirical statistics. Under the same conditions of the previous simulations, the availabilities in the different values of M and N are shown in Figure 4.

As is indicated in Figure 4, in the deteriorating system of the operating time, the availability declines with the increasing of the failure repairment times in a certain amount of premaintenance, which fits the entire trend of the availability and displays a strong consistency as is shown in Figure 4. The exact values corresponding to Table 3 are shown in Table 4, where the optimized value that meets the 0.5 availability threshold is $A(14, 4) = 0.567$.

5. Conclusions

Oriented towards the batteries' deteriorating characteristics after times of repairment and the demand of the BSS profit maximization while ensuring its battery availability, this paper proposes a GP-based E-bus battery management

TABLE 3: Average cost per unit time when $R = 0.4$.

M	N									
	36	23	19	16	12	9	6	5	4	3
2	24.806	22.592	20.357	17.928	14.061	11.418	10.098	10.323	11.243	14.532
6	24.697	21.454	18.379	15.241	10.572	7.403	5.155	4.767	4.709	6.250
10	24.490	20.560	17.420	14.430	10.075	6.961	4.374	3.734	3.303	4.253
14	24.079	19.474	16.556	13.976	10.230	7.335	4.561	3.744	3.058[*]	3.635
16	23.752	18.860	16.126	13.786	10.372	7.618	4.797	3.908	3.118	3.545
24	21.125	16.270	14.540	13.135	10.930	8.799	6.025	4.956	3.857	3.818
28	18.976	15.124	13.903	12.879	11.150	9.309	6.655	5.548	4.352	4.141
32	16.704	14.192	13.392	12.671	11.327	9.746	7.246	6.129	4.866	4.516
36	14.819	13.489	13.000	12.506	11.468	10.116	7.786	6.678	5.377	4.918
44	12.802	12.644	12.497	12.280	11.669	10.687	8.704	7.657	6.342	5.742

[*]The lowest average cost per unit time.

TABLE 4: Battery availability when $R = 0.4$.

M	N									
	36	23	19	16	12	9	6	5	4	3
2	0.005	0.067	0.131	0.205	0.330	0.432	0.529	0.559	0.588	0.518
6	0.005	0.074	0.148	0.229	0.361	0.463	0.558	0.587	0.616	0.544
10	0.005	0.067	0.132	0.202	0.322	0.424	0.528	0.563	0.597	0.531
14	0.004	0.058	0.109	0.167	0.271	0.370	0.485	0.525	0.567[*]	0.509
16	0.004	0.052	0.098	0.150	0.246	0.343	0.462	0.505	0.550	0.496
24	0.003	0.032	0.059	0.092	0.162	0.246	0.371	0.423	0.480	0.442
28	0.003	0.023	0.044	0.070	0.130	0.207	0.330	0.384	0.446	0.414
32	0.002	0.016	0.032	0.053	0.104	0.174	0.293	0.348	0.413	0.387
36	0.001	0.011	0.023	0.039	0.083	0.145	0.259	0.315	0.381	0.360
44	0.000	0.005	0.011	0.022	0.052	0.102	0.202	0.256	0.324	0.309

[*]The availability corresponding to the lowest average cost per unit time.

optimizing policy on condition that the battery, after premaintenance and failure repairment, cannot be as good as new. In addition, a deteriorating system is modeled for the operating time and an improving system is modeled for the premaintenance time as well as failure repairment time. The application of the GP to the battery management optimization reveals a good consistency and well reflects the operation demand of BSS. Moreover, the simulation is carried out for the analysis of the relationship between the amounts of premaintenance and failure repairment with different SOH thresholds when replacing the battery. Then the optimization of the long-run average cost is taken by importing those values. Finally, the simulation with the required availability threshold is carried out to verify its effectiveness. The results denote that the proposed management optimizing policy will prolong the life-cycle of the batteries and reduce the long-run average cost on the premise of the high availability, which is of much applicability on the batteries' optimizing management. And the future work is to apply the proposed GP-based E-bus battery management optimizing policy to the BSS and the microgrid to study the total cost aiming at the massive access of the E-buses and the EVs.

Nomenclature

E-bus:	Electric bus
GP-based:	Geometric-process-based
SOC:	State of charge
SOH:	State of health
c_m:	Premaintenance cost rate
c_f:	Failure repairment cost rate
c_r:	Replacement cost rate
c_b:	Benefit rate
c_N:	New battery cost
c_c:	Depleted battery recycling benefit
w_1:	Depreciation rate of the battery after a single operation
w_2:	Recovery rate after a single premaintenance
w_3:	Descent rate after the failure repairment
$A(M, N)$:	Battery availability
A_0:	Threshold of the battery availability
$C(M, N)$:	Long-run average cost per unit time

H:	Replacement time
M:	Amount of the premaintenance before one failure repairment
N:	Failure repairment amount during the battery life-cycle
R_0:	Initial value of SOH
R:	Threshold of SOH
T:	Life-cycle of the E-bus battery
$\{X_n^i,\ n = 1, 2, \ldots,\ i = 1, 2, \ldots\}$:	Battery operating time after the ith premaintenance in the nth period
$\{Y_n^i,\ n = 1, 2, \ldots,\ i = 1, 2, \ldots\}$:	The ith premaintenance time in the nth period
$\{Z_n,\ n = 1, 2, \ldots\}$:	Failure repairment time in the nth period.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] P. Han, J. Wang, Y. Han, and Y. Li, "Resident Plug-In Electric Vehicle charging modeling and scheduling mechanism in the smart grid," *Mathematical Problems in Engineering*, vol. 2014, Article ID 540624, 8 pages, 2014.

[2] K. Rajashekara, "Present status and future trends in electric vehicle propulsion technologies," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 1, no. 1, pp. 3–10, 2013.

[3] Y. Li, J.-K. Wang, P. Han, and Y.-H. Han, "Modeling and analysis on coordinated scheduling of E-bus recharging station participated wind-power generation," in *Proceedings of the 32nd Chinese Control Conference (CCC '13)*, pp. 8592–8596, Xi'an, China, July 2013.

[4] Q. Dai, T. Cai, S. Duan, and F. Zhao, "Stochastic modeling and forecasting of load demand for electric bus battery-swap station," *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1909–1917, 2014.

[5] P. Elbert, T. Nuesch, A. Ritter, N. Murgovski, and L. Guzzella, "Engine on/off control for the energy management of a serial hybrid electric bus via convex optimization," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3549–3559, 2014.

[6] C. R. Gould, C. M. Bingham, D. A. Stone, and P. Bentley, "New battery model and state-of-health determination through subspace parameter estimation and state-observer techniques," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 3905–3916, 2009.

[7] V. Agarwal, K. Uthaichana, R. A. Decarlo, and L. H. Tsoukalas, "Development and validation of a battery model useful for discharging and charging power control and lifetime estimation,"

IEEE Transactions on Energy Conversion, vol. 25, no. 3, pp. 821–835, 2010.

[8] Y. Wang and Y. Liu, "Electronic control system design and test of pure electric vehicle battery management system," in *Proceedings of the 2nd International Conference on Mechanic Automation and Control Engineering (MACE '11)*, pp. 1289–1292, IEEE, Hohhot, China, July 2011.

[9] H. Y. Li, Y. L. Jia, D. Zhang et al., "Application of electric vehicle battery intelligent monitoring and management system," in *Proceedings of the IEEE Conference and Expo Transportation Electrification Asia-Pacific*, pp. 1–5, Beijing, China, 2014.

[10] H. C. Han, H. X. Xu, Z. Q. Yuan, and Y. Shen, "A new SOH prediction model for lithium-ion battery for electric vehicles," in *Proceedings of the 17th International Conference on Electrical Machines and Systems (ICEMS '14)*, pp. 997–1002, Hangzhou, China, October 2014.

[11] M. Paul-Henri and H. Vincent, "An adaptive sigma point kalman filter hybridized by support vector machine algorithm for battery SoC and SoH estimation," in *Proceedings of the IEEE 81st Vehicular Technology Conference*, pp. 1–7, Glasgow, UK, May 2015.

[12] L.-R. Dung, S.-H. Wu, and H.-F. Yuan, "An SOH estimation system based on time-constant-ratio measurement," in *Proceedings of the IEEE 23rd International Symposium on Industrial Electronics (ISIE '14)*, pp. 1784–1787, Istanbul, Turkey, June 2014.

[13] Y. Lam and Y. L. Zhang, "A geometric-process maintenance model for a deteriorating system under a random environment," *IEEE Transactions on Reliability*, vol. 52, no. 1, pp. 83–89, 2003.

[14] L. Tan, T. Cheng, and B. Guo, "Optimal maintenance strategy for the repairable system consisting of one component based on geometric process," *Systems Engineering*, vol. 26, no. 6, pp. 88–92, 2008.

[15] J. Jia, D. Wang, and Z. Duan, "Study on the mathematical model under optimal replacement for a deteriorative repairable system," *Mathematics in Practice and Theory*, vol. 36, no. 4, pp. 1–4, 2006.

[16] Y. Lam, "A geometric process δ-shock maintenance model," *IEEE Transactions on Reliability*, vol. 58, no. 2, pp. 389–396, 2009.

[17] G. J. Wang and Y. L. Zhang, "A bivariate optimal replacement policy for a cold standby repairable system with preventive repair," *Applied Mathematics and Computation*, vol. 218, no. 7, pp. 3158–3165, 2011.

[18] Y. L. Zhang and G. J. Wang, "A geometric process repair model for a series repairable system with k dissimilar components," *Applied Mathematical Modelling*, vol. 31, no. 9, pp. 1997–2007, 2007.

[19] Y. L. Zhang and G. J. Wang, "An extended replacement policy for a deteriorating system with multi-failure modes," *Applied Mathematics and Computation*, vol. 218, no. 5, pp. 1820–1830, 2011.

[20] G. J. Wang and Y. L. Zhang, "Optimal periodic preventive repair and replacement policy assuming geometric process repair," *IEEE Transactions on Reliability*, vol. 55, no. 1, pp. 118–122, 2006.

[21] Y. L. Zhang, "A geometric-process repair-model with good-as-new preventive repair," *IEEE Transactions on Reliability*, vol. 51, no. 2, pp. 223–228, 2002.

[22] Y. Han, J. Wang, Q. Zhao, and P. Han, "An optimal operating strategy for battery life cycle costs in electric vehicles," *Journal of Applied Mathematics*, vol. 2014, Article ID 305905, 6 pages, 2014.

Research Article

Model Predictive Control for Connected Hybrid Electric Vehicles

Kaijiang Yu,[1] Xiaozhuo Xu,[1] Qing Liang,[2] Zhiguo Hu,[1] Junqi Yang,[1] Yanan Guo,[3] and Hongwei Zhang[1]

[1]*School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China*
[2]*College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China*
[3]*Graduate School of Integrated Frontier Sciences, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 8190395, Japan*

Correspondence should be addressed to Kaijiang Yu; yu.kaijiang@163.com

This paper presents a new model predictive control system for connected hybrid electric vehicles to improve fuel economy. The new features of this study are as follows. First, the battery charge and discharge profile and the driving velocity profile are simultaneously optimized. One is energy management for HEV for P_{batt}; the other is for the energy consumption minimizing problem of acc control of two vehicles. Second, a system for connected hybrid electric vehicles has been developed considering varying drag coefficients and the road gradients. Third, the fuel model of a typical hybrid electric vehicle is developed using the maps of the engine efficiency characteristics. Fourth, simulations and analysis (under different parameters, i.e., road conditions, vehicle state of charge, etc.) are conducted to verify the effectiveness of the method to achieve higher fuel efficiency. The model predictive control problem is solved using numerical computation method: continuation and generalized minimum residual method. Computer simulation results reveal improvements in fuel economy using the proposed control method.

1. Introduction

In recent years, the energy and environmental problems are emphasized. In particular, energy consumption of vehicles accounts for a substantial amount in the transportation sector. There are various approaches to reduce the fuel consumption of vehicles [1–5]. High efficient vehicles are being developed to increase fuel economy using lightweight automobiles, efficient power train systems, electric vehicles, and hybrid vehicles [1]. On the other hand, the so-called ecodriving can also reduce the fuel consumption [5–9]. Ecodriving can be characterized as avoiding aggressive acceleration or braking at any road-traffic situations, cruising at steady speed, decelerating smoothly at stops with little or no braking, and maintaining an optimal distance from the preceding vehicle. An ecological control of a single vehicle on a road with up-down shapes [2] and efficient spacing control of multiple vehicles [10] were presented.

A lot of works have been published on the energy management problem of hybrid electric vehicle (HEV) and plug-in hybrid electric vehicle (PHEV) systems. These approaches

are typical in a family of optimal control techniques. They can be subdivided into four categories: numerical optimization, analytical optimal control theories, instantaneous optimization, and heuristic control techniques [11]. The most representative of numerical optimization is dynamic programming (DP) [11, 12]. However DP is based on fixed speed patterns which are impossible to get in reality. A kind of analytical optimal control techniques is Pontryagin's minimum principle [13]. It gives necessary conditions that the optimal solution must satisfy. It also needs to know the entire driving cycle in advance. The convex optimization method [14] is also a kind of analytical optimal control techniques. The global optimality is guaranteed and the optimal solution can be rapidly and efficiently attained by solvers available. The instantaneous optimization includes the equivalent consumption minimization strategy (ECMS) [2, 15]. It is based on instantaneous optimization and is easy to implement in real-time. However it cannot guarantee the optimality over the whole driving cycle. Heuristic control techniques like rule-based control strategies [2] are robust, but they are impossible to guarantee the optimality.

To obtain even more fuel economy improvements, connected hybrid electric vehicles can be considered to reduce the air resistance. The air resistance of a vehicle is proportional to the square of the vehicle speed. When a vehicle runs at 100 km/h, its aerodynamic drag reaches more than sixty percent of total motion resistance forces [16]. It is obvious that its air resistance causes high fuel consumption. However, the air resistance can be reduced by maintaining a short spacing between two vehicles. Connected vehicles in an automated highway system can lead to increased driver safety, decreased road congestion, and improved fuel economy [17]. Connected vehicles can improve fuel economy through reduced wind resistance [18–20].

A low pressure area distributes in the rear of the lead vehicle. The size of the area can be different by changing the spacing between the vehicles. If the following vehicle runs at the back of the lead vehicle with a short spacing, its air resistance is decreased owing to improved airflow profile between the vehicles. Furthermore, the air resistance of the lead vehicle is also decreased by the smooth airflow [20]. Therefore, fuel consumption of both vehicles can be reduced. However, it is difficult to follow the lead vehicle with a short spacing at high speed by a human driver. Automated cruise control of the vehicle should be introduced to achieve this. Various conventional longitudinal control systems have been proposed such as vehicle following method using information of other vehicles [10] and point following method using a certain decided phase point [21]. A control law for internal combustion engine vehicles is proposed which uses relative speed and spacing information from the preceding and following vehicles in order to choose the proper control action for smooth vehicle following and for maintaining a desired intervehicle spacing specified by the driver [22]. Connected automatic guided electric vehicles to solve problems of traffic saturation, relying on GPS sensors and intervehicle communication, are addressed in [23]. However, these conventional methods consider string stability only. The quantitative effect of road shape and air resistance on fuel consumption for hybrid electric vehicles (HEVs) has not been researched.

For connected hybrid electric vehicles, it is necessary to compute the optimum control inputs of the vehicles by anticipating the future situations including road shape, vehicles' states, and road loads. Therefore, model predictive control (MPC) method can be used.

This paper extends HEV energy management research by adding two novel contributions. First, the battery charge and discharge profile and the driving velocity profile are simultaneously optimized. We make the two connected problems together: one is energy management for HEV for the battery; the other is for the energy consumption minimizing problem of speed control of two vehicles. In reality, the two connected problems are coupled together and affect each other always. The speed of the vehicle affects the charge and discharge profile of the battery. The charge and discharge profile of the battery affects the speed of the vehicle. Second, a new policy between the global optimization method and the instantaneous optimization method is developed. The global optimization method like dynamic programming needs all the information in the future to compute the global optimal

control input. The instantaneous optimization method needs no information in the future to compute the control input. The easiest way to deal with the complicated control system is to divide the longitudinal vehicle control system into an upper and lower level controller. The upper level controller determines the desired acceleration of the vehicle on the basis of the position and velocity relative to the other vehicles in the string. The lower level controller determines the input commands to the engine and the braking system, to accomplish the desired acceleration. Also, there is possible to consider road slope, wind, and so forth as a disturbance for the problem. However, in this work we intended to optimize the fuel economy and the speed profile for high fuel efficiency and safety simultaneously. In the HEV operation it is desirable to charge or discharge the battery properly according to the road loads. There is a problem between the fast dynamics components like the engine and the slow dynamics components like the battery. The prediction horizon of the battery state is limited. We developed a new policy to predict the battery state in a longer future for better performance. The desired battery state of charge is designed according to the road slopes for better recuperation of free braking energy. The battery state of charge profile is scheduled systematically to improve fuel economy inside the HEV considering the effect of different parameters, that is, road conditions, battery state of charge, and real-time implementation ability. The quantitative analysis of the vehicle spacing influence and the battery state of charge profile influence for the fuel economy is presented. Performance of the proposed system has been evaluated by computer simulation. The proposed system is found to be more fuel efficient and safer for running over several typical roads with up-down slopes.

The rest of this paper is organized as follows. In Section 2, the nonlinear model of two connected power-split HEVs is derived. Section 3 formulates the nonlinear model predictive control algorithm. Section 4 presents comparative simulation results. Section 5 provides conclusions.

2. Modeling of Two Connected HEVs

The configuration of the HEV system is shown in Figure 1. FD represents the final drive. The power-split device (PSD) is the key component of the power-split HEV system and has both functionality of speed coupler and continuously variable transmission (CVT). There are five dynamic components: the engine, the battery, two motor/generators (M/G), and the wheels in this power-split HEV system. The only dynamic state to be considered in the optimal control problem based on known driving cycle is the battery state of charge (SOC) which can simplify the MPC algorithm for implementation. This simplification is possible because this paper introduces four constraints: the road load, the torque and speed relationship of the speed coupler, the power flow relationship among the five components, and the engine optimal operating line (OOL) using CVT. In this work, we assume that the engine works along its OOL using CVT. For simplicity, we assume the two vehicle configurations are the same. It is assumed that the central controller set in the lead vehicle controls the two vehicles. The central controller computes the control inputs
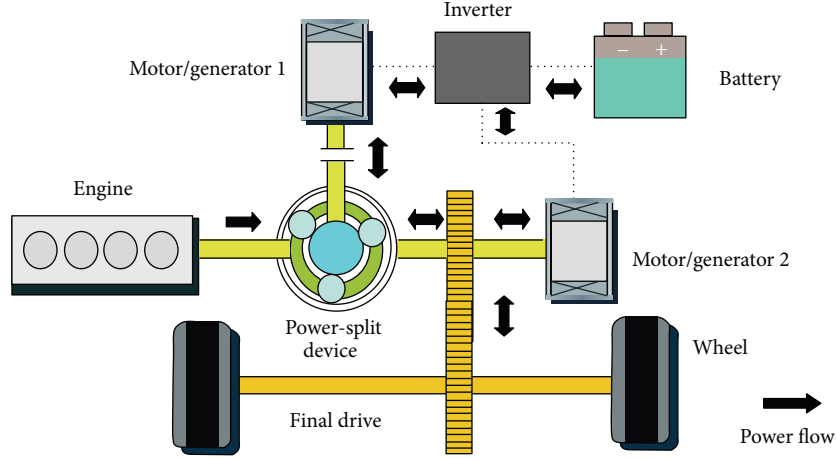
FIGURE 1: Configuration of the power-split HEV system. Diagram adapted from [12].

of the two vehicles. The control inputs of the two vehicles are fed into the two vehicles, respectively. The states of the two vehicles are measured and sent to the central controller. In this way a closed control loop is formed. Here, we call it central control system of connected vehicles. In a distributed control system of connected vehicles [18–20], the individual vehicles are controlled separately by its own controller. It cannot predict other vehicles precisely. In a central control system of connected vehicles, all the information of the vehicles is shared, and the global optimality of all the vehicles can be obtained. The distance between the two vehicles changes (which affect the air drag coefficient) and the slope changes; therefore the stability of this controller is very important. However, the control scheme proposed in this work is brand new; the stability of this controller is completely different from that of the distributed control system of connected vehicles. Hence, we would like to add the stability problem as our future directions because of its complexity. The control signals are transmitted to the vehicles through intervehicle communication. It is assumed that there is no delay of the communication. The proposed methodology will work independently of other kinds of vehicles on the roadway in the network if the vehicle has the functionality of CVT. This paper divided the optimal control problem into two levels. The high-level controller determines the optimal battery power and the low-level controller determines the optimal torque and speed of the engine and the motor/generators. This paper focuses on the high-level controller.

The torque and speed relationship of the speed coupler can be expressed as [24]

$$
\tau_{\text{eng}}(t) = -\left(1 + \frac{R}{S}\right)\tau_{M/G1}(t),
$$

$$
\tau_{\text{eng}}(t) = -\left(1 + \frac{S}{R}\right)\left(\tau_{M/G2}(t) - \frac{\tau_{\text{req}}(t)}{g_f}\right), \tag{1}
$$

$$
S\omega_{M/G1}(t) + R\omega_{M/G2}(t) - (S+R)\omega_{\text{eng}}(t) = 0,
$$

where S and R are the number of sun gear and ring gear teeth, respectively, $\tau_{M/G1}$, $\tau_{M/G2}$, τ_{req}, and τ_{eng} are the torques

of $M/G1$, $M/G2$, the road load, and the engine, respectively, and $\omega_{M/G1}$, $\omega_{M/G2}$, and ω_{eng} are the angular speeds of $M/G1$, $M/G2$, and the engine, respectively.

The power flow relationships among the five components at the inverter and the power-split device in Figure 1 are given as

$$
P_{\text{batt}}(t) = P_{M/G1}(t) + P_{M/G2}(t),
$$

$$
P_{\text{req}}(t) = P_{M/G1}(t) + P_{M/G2}(t) + P_{\text{eng}}(t), \tag{2}
$$

where P_{batt}, $P_{M/G1}$, $P_{M/G2}$, P_{eng}, and P_{req} are the power of the battery, $M/G1$, $M/G2$, the engine, and the road load.

This paper assumes that the engine always works along its OOL using CVT which can also be considered as a constraint. When the engine power is known, by looking up the table of OOL, the engine speed and torque can be obtained.

This paper evaluates the fuel consumption using Willans line method to reduce the complexity of the engine fuel consumption model. It was found that good approximations are obtained using the Willans line method [25]. The fuel consumption can be expressed as

$$
\dot{m}_f(t) = \dot{m}_f\left(P_{\text{req}}(t) - P_{\text{batt}}(t)\right)
$$

$$
\approx c_f\left(P_{\text{req}}(t) - P_{\text{batt}}(t)\right), \tag{3}
$$

where c_f is a constant. The detailed explanation of this fuel consumption model is included in Appendix A.

The road loads which are the vehicle speed and the required power at the wheels are known when the driving cycle is known. From the configuration of the power-split HEV system, $M/G2$ speed is also known as

$$
\omega_{M/G2}(t) = \frac{g_f}{r_w}v_{\text{req}}(t), \tag{4}
$$

where $\omega_{M/G2}$ is the speed of $M/G2$, g_f is the final drive gear ratio, r_w is the wheel radius, and v_{req} is the required vehicle speed by the driving cycle.

For simplicity, it is assumed that there are two cars in this central control system. When the driving cycle is unknown,

the system dynamics includes the battery and the vehicle dynamics. Both the fuel economy and the driving profile are optimized. The system model is then represented by

$$\dot{x}$$

$$= \begin{bmatrix} v_p \\ z_p - \dfrac{(1/2)\,\rho C_{Dp} A_p v_p^2}{m_p} - g\mu - g\sin\left(\theta_p\left(p_p\right)\right) \\ k_{pp}\left(u_p - z_p\right) \\ -\dfrac{V_{OCp} - \sqrt{V_{OCp}^2 - 4P_{battp}R_{battp}}}{2R_{battp}Q_{battp}} \\ v_h \\ z_h - \dfrac{(1/2)\,\rho C_{Dh} A_h v_h^2}{m_h} - g\mu - g\sin\left(\theta_h\left(p_h\right)\right) \\ k_{ph}\left(u_h - z_h\right) \\ -\dfrac{V_{OCh} - \sqrt{V_{OCh}^2 - 4P_{batth}R_{batth}}}{2R_{batth}Q_{batth}} \end{bmatrix},\tag{5}$$

$$x = \begin{bmatrix} p_p & v_p & z_p & x_{SOCp} & z_h & v_h & w_h & x_{SOCh} \end{bmatrix}^T,$$

$$u = \begin{bmatrix} u_p & P_{battp} & u_h & P_{batth} \end{bmatrix}^T,\tag{6}$$

where p, v, and z are the vehicle position, speed, and acceleration or deceleration converted from the traction force or brake force. The parameters ρ, C_D, A, m, g, μ, and $\theta(p)$ are the air density, the air drag coefficient, the frontal area of the vehicle, the vehicle mass, the gravity acceleration, the rolling resistance coefficient, and the road grade. u_p, u_h, and k_p are the vehicle acceleration or deceleration control inputs and the delay constant. V_{OC}, R_{batt}, and Q_{batt} are the open-circuit voltage, the internal resistance, and the capacity of the battery. The suffixes p and h denote the parameters of the preceding vehicle and the host vehicle.

The slope information from GPS or the digital map is approximated by the sigmoid functions. This modeling method of road slope is firstly proposed in this work. A description of the modeling method is provided in Appendix B.

3. Model Predictive Control

The driving control inputs are derived using model predictive control algorithm. The optimal control problem is defined as

$$\text{Min.} \quad J = \int_t^{t+T} L\left(x\left(\tau \mid t\right), u\left(\tau \mid t\right)\right) d\tau \tag{7}$$

$$\text{subject to} \quad SOC_{pmin} \le x_{SOCp}\left(\tau \mid t\right) \le SOC_{pmax}$$

$$P_{battpmin} \le P_{battp}\left(\tau \mid t\right) \le P_{battpmax}$$

$$u_{pmin} \le u_p\left(\tau \mid t\right) \le u_{pmax}$$

$$SOC_{hmin} \le x_{SOCh}\left(\tau \mid t\right) \le SOC_{hmax} \tag{8}$$

$$P_{batthmin} \le P_{batth}\left(\tau \mid t\right) \le P_{batthmax}$$

$$u_{hmin} \le u_h\left(\tau \mid t\right) \le u_{hmax},$$

where T is the prediction horizon and $_{min}$ and $_{max}$ denote the minimum and maximum bounds of the parameters.

The following objectives are considered in this optimal control problem.

The term L_x: acceleration or deceleration of vehicles is moderated.

The term L_y: the vehicle speed is kept near to its desired value.

The term L_z: the fuel consumption is minimized.

The term L_d: the battery SOC is kept near to its desired value. This is one of the cores of the proposed approach. This paper adapts the battery energy to the vehicle future energy requirements by setting the desired battery SOC as a function of road slopes which represent the main part of the future road load.

The term L_e: the battery energy is made best use of. This is one of the cores of the proposed approach. The battery energy is firstly used to satisfy the required road load. If it is not enough, the engine energy should be used, and the engine can work along its OOL.

The term L_f: the battery SOC constraint is kept satisfied.

The term L_g: the desired vehicle spacing is kept. This is one of the cores of the proposed approach. The following distance constraint is kept in a predictive controller structure. The following distance is varied above the minimum following distance, which improves the freedom of ecodriving car following control to optimize the driving profile for better fuel economy.

The cost function L is defined as follows:

$$L = w_x L_x + w_y L_y + w_z L_z + w_d L_d + w_e L_e + w_f L_f$$
$$\quad + w_g L_g,$$

$$L_x = \left(z_p - \dfrac{(1/2)\,\rho C_{Dp} A_p v_p^2}{m_p} - g\mu\right)^2$$
$$\quad + \left(z_h - \dfrac{(1/2)\,\rho C_{Dh} A_h v_h^2}{m_h} - g\mu\right)^2,$$

$$L_y = \left(v_p - v_d\right)^2 + \left(v_h - v_d\right)^2,$$

$$L_z = \dfrac{c_{fp}\left(m_p z_p v_p - P_{battp}\right)}{\left(1 + e^{\left(-\beta\left(m_p z_p v_p - P_{battp}\right)\right)}\right)}$$
$$\quad + \dfrac{c_{fh}\left(m_h z_h v_h - P_{batth}\right)}{\left(1 + e^{\left(-\beta\left(m_h z_h v_h - P_{batth}\right)\right)}\right)},$$

$$L_d = \left(x_{SOCp} - SOC_d\left(p_p\right)\right)^2$$
$$\quad + \left(x_{SOCh} - SOC_d\left(p_h\right)\right)^2,$$

$$L_e = \left(m_p w_p v_p - P_{battp}\right)^2 + \left(m_h w_h v_h - P_{batth}\right)^2,$$

$$L_f = -\ln\left(x_{\text{SOC}p} - \text{SOC}_{p\min}\right)$$
$$-\ln\left(\text{SOC}_{p\max} - x_{\text{SOC}p}\right)$$
$$-\ln\left(x_{\text{SOC}h} - \text{SOC}_{h\min}\right)$$
$$-\ln\left(\text{SOC}_{h\max} - x_{\text{SOC}h}\right),$$
$$L_g = \frac{1}{2}\left(d - d_d\right)^2,$$
$$d = p_p - p_h - l_p,$$

(9)

where $w_x, w_y, w_z, w_d, w_e, w_f$, and w_g are the weights and v_d is the desired vehicle speed. The parameters d, d_d, and l_p are the vehicle spacing, the desired vehicle spacing, and the length of the preceding vehicle. The parameter SOC_d is the desired SOC value. The sigmoid function is chosen to evaluate the vehicle brake fuel consumption. The log barrier function is used as a penalizing term for violations of state constraints.

The structure of the nonlinear model predictive control system is shown in Figure 2. The system inputs contain the control inputs. The system outputs consist of the vehicle states. The predictive controller uses terrain information from a digital map to calculate $\text{SOC}_d(p)$ and $\theta(p)$.

At each time t, the optimal control input is computed by solving the above optimal control problems during the prediction horizon T. Only the first element of the optimal control sequence is applied. At the next time step, the prediction horizon moves forward, and the process is repeated.

4. Computer Simulations

4.1. Comparison Controllers. There are two simulations in this work. They are the MPC approaches with fixed desired battery SOC and variable desired battery SOC. The aim is to demonstrate how the desired battery SOC affects the fuel economy, the power-split profile, and the drag coefficients.

The desired battery SOC value is set according to the road elevation. The authors think it is reasonable to utilize the road elevation information since this future road load information is known already. The desired battery SOC is assumed to use the function as

$$\text{SOC}_d(p)$$
$$= k_{\text{SOC}}\left(\frac{s_1}{1 + e^{(s_3(p - s_2))}} + \frac{s_4}{1 + e^{(s_6(p - s_5))}} + \cdots\right)$$

(10)

$$+ \text{SOC}_k,$$

where k_{SOC} and SOC_k are constant parameters set as $k_{\text{SOC}} = -2$ and $\text{SOC}_k = 0.7$, respectively.

4.2. Simulation Conditions. In these simulations, the parameters of both HEVs are used from ADVISOR 2002 Toyota Prius data (see Table 1). Seven tuning weights are used in this cost function; this makes the performance very subjective to choice of these weights. The goal is to minimize total fuel used, so the real cost function should be integral of
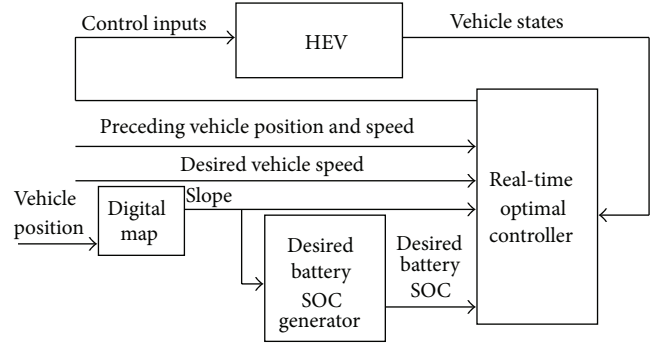
FIGURE 2: Structure of the model predictive control system.

TABLE 1: Simulation parameters.

Parameters	Values
m	1504
c_f	0.0874
g	9.8 [m/s^2]
V_{OC}	307.9 [V]
Q_{batt}	6 [Ah]
g_f	3.93
l_p	4.31 [m]
h_t	0.1 [s]
$\text{SOC}_{\min}$	0.6
w_x	100000
w_z	20
w_e	100
w_g	3000
ρ	1.23 [kg/m^3]
A	1.746 [m^2]
μ	0.015
R_{batt}	1.0 [Ω]
r_w	0.287 [m]
k_p	10
d_d	1 [m]
SOC_d	0.7
$\text{SOC}_{\max}$	0.8
w_y	2000
w_d	67000000
w_f	200000

fuel rate plus an equivalent fuel cost at the end of the MPC horizon. Any other choice makes the cost function very subjective and the optimal controller will not minimize fuel use. Tuning of the weight parameters is an important issue for attaining fuel efficient and safe behavior in the complex system. Weight parameters are tuned manually by observing the fuel economy and driving performance. w_x is tuned for minimal fuel consumption by dynamic acceleration of the vehicle. w_z is tuned for minimal fuel consumption by static maps of the engine which cannot evaluate the dynamic fuel consumption of the engine. w_y and w_g are tuned for safety of driving to avoid real-ends collision. w_d and w_e are tuned
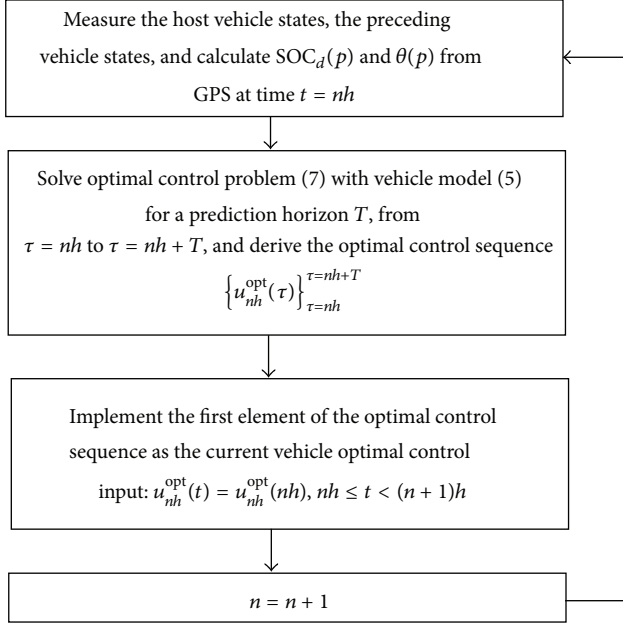
FIGURE 3: Flowchart of the nonlinear real-time optimal control algorithm.

TABLE 2: Comparison of the energy used by motion resistance forces.

c_{11}	4.498×10^{-8}
c_{12}	-1.475×10^{-6}
c_{13}	1.139×10^{-5}
c_{14}	9.373×10^{-5}
c_{15}	-0.002
c_{16}	0.003
c_{17}	0.035
c_{18}	0.205
c_{21}	1.326×10^{-7}
c_{22}	-6.593×10^{-6}
c_{23}	0.0001
c_{24}	-0.0014
c_{25}	0.0080
c_{26}	-0.026
c_{27}	0.046
c_{28}	0.244

for minimizing the equivalent fuel cost. w_f does not need to be tuned ordinarily. The reason is that it is for the state constraint. In reality, it is a physical constraint which cannot be violated.

The model predictive control problem is solved using the numerical computation method: the continuation and generalized minimum residual (C/GMRES) method [26]. The C/GMRES method uses forward difference approach and discretizes the HEV plant with a sampling interval h_t to implement the nonlinear real-time optimal control algorithm. A brief description of the solution of the model predictive control problem using the C/GMRES method is included in Appendix C. The flowchart of the nonlinear model predictive control algorithm implementation is shown in Figure 3. Since the optimization problem is nonlinear and nonconvex, its solution can be local optimal; it is assumed that it can only be solved numerically. A detailed mathematical analysis of robust stability and performance for the proposed method can be found in [26]. The fuel economy is calculated using the engine fuel consumption map which is obtained from ADVISOR 2002.

The MPC algorithm is realized by utilizing the C MEX S-function builder in MATLAB/Simulink. First, the optimal battery power is calculated by the high-level controller. Next, this optimal value is fed into the low-level controller where the optimal torque and speed of the engine and M/Gs are determined. Finally, these actual control input signals are applied to the vehicle. The fuel economy is calculated using the quasi-static map of ADVISOR. The backward simulation approach has been employed in this work, as ADVISOR software based on quasi-static maps of power train elements is utilized. However, (3) presents a different formula for fuel consumption estimation which is not consistent with the way

ADVISOR calculates the fuel consumption. The reason is that (3) is for the control input calculation of the model predictive control which needs to be continuous and have derivative, and quasi-static maps of power train elements are used for the output evaluation.

The parameters predicted are the road slope based on the GPS data and the traffic conditions. They are not embedded in ADVISOR, as in this software the slope is assumed to be zero. Quasi-static maps of power train elements in ADVISOR are used only for the output evaluation. The traffic conditions and their impact on the fuel consumption are predicted using the connected vehicle model to achieve optimal vehicle spacing for reducing air drag.

A set of data representing the relationship of the aerodynamic drag coefficient and the spacing obtained from a wind tunnel experiment [20] is adopted. By using the seventh-degree polynomial representation, the aerodynamic drag coefficient C_i can be represented as

$$
\begin{aligned}
C_i\left(d\left(t\right)\right) = & c_{i1}d^7\left(t\right) + c_{i2}d^6\left(t\right) + c_{i3}d^5\left(t\right) + c_{i4}d^4\left(t\right) \\
& + c_{i5}d^3\left(t\right) + c_{i6}d^2\left(t\right) + c_{i7}d\left(t\right) + c_{i8},
\end{aligned}
\tag{11}
$$

where parameters c_{i1}–c_{i8} are shown in Table 2. The approximation results of C_i are shown in Figure 4.

4.3. Simulation Results. The driving profile of the HEV using the MPC algorithm with fixed desired battery SOC and unfixed desired battery SOC (see Figures 5 and 6) shows that the MPC algorithm can use the road slope information well to reduce the fuel consumption. The rows of Figures 5 and 6 from the top are the slope of the road, the speed of the preceding vehicle and the host vehicle, the battery SOC of the preceding vehicle and the host vehicle, and the vehicle distance between the preceding vehicle and the host vehicle. The MPC algorithm simultaneously controls both vehicles by predicting their states, and fast convergence of their spacing is achieved. The vehicle accelerates before the up slope to make use of the kinetic energy. The battery

TABLE 3: Fuel economy comparison results.

Method	Preceding vehicle mileage (km/L)	Host vehicle mileage (km/L)	Total mileage (km/L)
MPC with unfixed SOC	31.9 (+2.9%)	20.4 (+3.0%)	12.4 (+2.5%)
MPC with fixed SOC	31.0	19.8	12.1
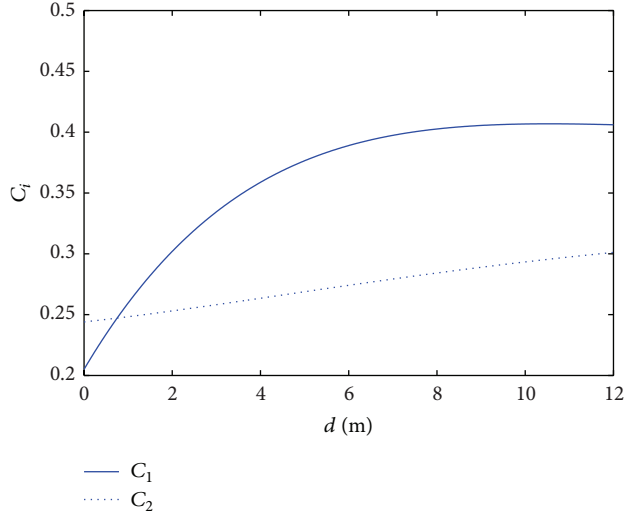
FIGURE 4: Approximation of C_1 and C_2 with respect to the spacing d.

FIGURE 5: Driving profile of the HEV using the MPC algorithm with fixed desired battery SOC.

recuperates vehicle braking power during the vehicle down slope driving. The preceding vehicle SOC variation range is smaller than that of the host vehicle. The host vehicle SOC variation range with unfixed desired battery SOC is smaller than that with fixed desired battery SOC. Since the engine needs to charge the battery too often, the overuse of the battery leads to worse fuel economy.

The energy profile of the HEV using the MPC algorithm with fixed desired battery SOC and unfixed desired battery SOC (see Figures 7 and 8) shows that the MPC algorithm can make the vehicle drag coefficients converge to minimum. The rows of Figures 7 and 8 from the top are the air drag power of the preceding vehicle and the host vehicle, the drag coefficients of the preceding vehicle and the host vehicle, the fuel consumption rate of the preceding vehicle and the host vehicle, and the total cumulative fuel consumption. The drag coefficients of both vehicles are reduced significantly as a result of the fast convergence of the vehicle distance. Therefore, the drag force is reduced, and fuel savings are achieved. The air drag power of the host vehicle is smaller than that of the preceding vehicle because of the vehicle platooning. The fuel consumption of the HEV with unfixed desired battery SOC is smaller than that with fixed desired battery SOC. The reason is that, during the down slope driving, the MPC algorithm with unfixed desired battery SOC makes better use of the motor regenerative braking than that with fixed desired battery SOC.

All the constraints are satisfied in the simulation. The overall fuel economy results are presented in Table 3. It is

shown that the MPC approach with unfixed desired battery SOC can improve fuel economy and keep the final SOC near the initial SOC compared to that with fixed desired battery SOC. There are two reasons. The first reason is that the MPC approach with unfixed desired battery SOC makes the battery use less often than that with fixed desired battery SOC. The second reason is that the MPC approach with unfixed desired battery SOC reduces the duration and magnitude of the braking during the down slope driving period compared with that with fixed desired battery SOC. To sum up, the reason why MPC approach improves the fuel economy while keeping the SOC near initial status is the variation of the battery SOC. This variation can facilitate charging when the vehicles decelerate. The proposed method can improve the fuel economy of both the preceding vehicle and the host vehicle.
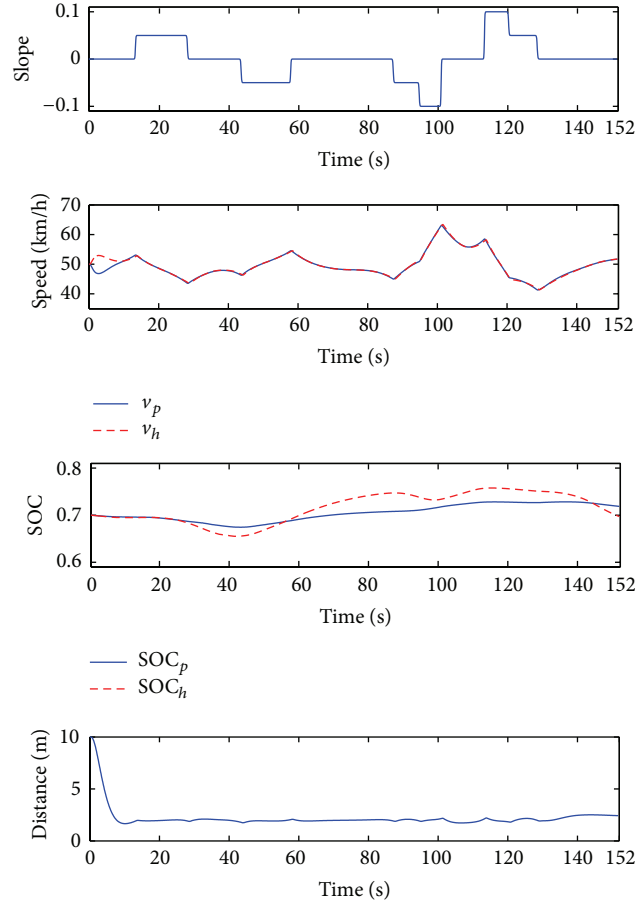
FIGURE 6: Driving profile of the HEV using the MPC algorithm with unfixed desired battery SOC.

The fuel economies with different desired vehicle speed, prediction horizon, and control horizon using the proposed method were shown in Figures 10, 11, and 12. It is shown that the best fuel economy occurs when the desired vehicle speed is equal to 50 km/h, the prediction horizon is equal to 2 s, and the control horizon is equal to 0.2 s.

MPC is sensitive to process-model mismatch. Process-model mismatch is always present: wheels, weather and road conditions, sensor accuracy, and so on. Therefore performance and stability under nominal conditions do not guarantee the robust performance and stability of the real car. In order to analyze the robust stability and performance of the MPC method, errors in slope sensor are assumed to exist while traveling. The proposed MPC method is found to be very robust against slope-sensing error. Figure 9 shows the deviation of the driving and power-split profiles due to errors in slope sensor for the lead vehicle. An error of 20% means that the sensor provides 1.2 times the actual value, whereas −20% error means that the sensor provides 0.8 times the actual value of the slope. Due to a sensing error of 20%, the vehicle speed is a bit lower; the battery SOC is a bit higher; and compared with the vehicle with no slope-sensing error, fuel savings dropped 3%. Similarly, due to a sensing error of −20%, the vehicle speed is a bit higher; the battery SOC is a bit lower; and compared with the vehicle with no

slope-sensing error, fuel savings increased 0.8%. Therefore, it can be concluded that, within a reasonable sensing error, the system is robust to maintain its ecological performance without significant deviation. Since the optimization problem is nonlinear and nonconvex, its solution can be local optimal; it is assumed that it can only be solved numerically. A detailed mathematical analysis of robust stability and performance for the proposed method can be found in [26].

The process-model mismatch can also give rise to important unfeasibility issues when computing the MPC control law. A policy to recover from unfeasibility issues is needed to be developed. The numerical computation method cannot manage unfeasibility by itself. Since we are dealing here with vehicles (i.e., people), a policy to recover from unfeasibility issues cannot be developed by reformulating the MPC problem with soft constraints. A switch policy is proposed to cope with the unfeasibility issues. It is summarized as follows.

Step 1. Switch the MPC algorithm to the rule-based algorithm [1] when unfeasibility issues occur.

Step 2. Use the rule-based algorithm for 5 s, and switch back to the MPC algorithm.

Step 3. Go back to Step 1 when unfeasibility issues occur.
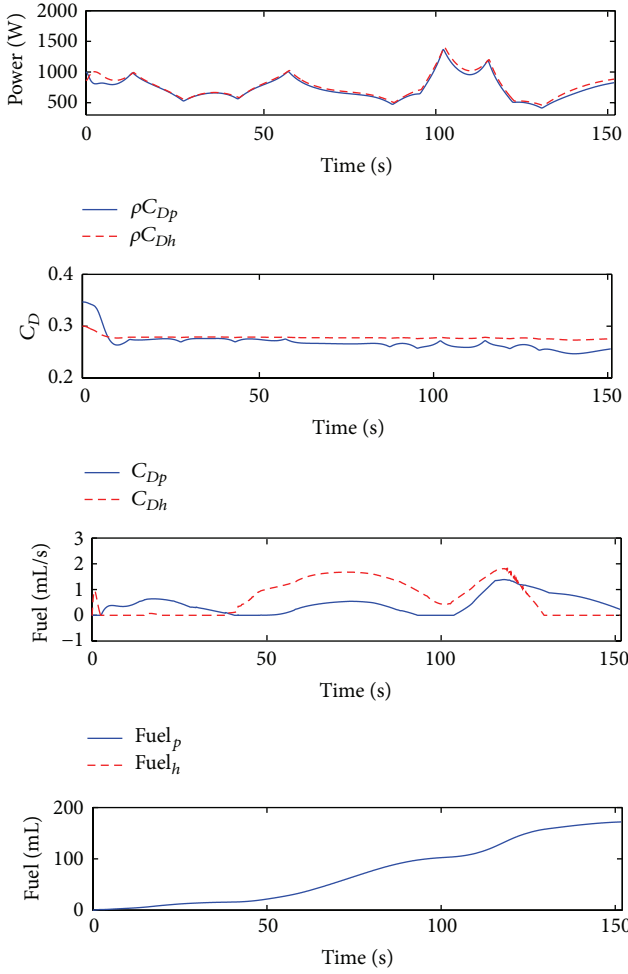
FIGURE 7: Energy profile of the HEV using the MPC algorithm with fixed desired battery SOC.

FIGURE 8: Energy profile of the HEV using the MPC algorithm with unfixed desired battery SOC.

The proposed MPC algorithm is fast for computation. The computer simulation time is 153 [s]. The computation time of the proposed MPC algorithm is 14 [s]. The simulation is run in a MATLAB/Simulink environment using a laptop with an Intel processor at 2.27 [GHz] processing speed and 2 [GB] of RAM. The sampling interval is 100 [ms]. The computation time per sampling interval of the proposed MPC algorithm is 9 [ms]. So it is concluded that the MPC algorithm has the potential for real-time vehicle control.

5. Conclusions

A model predictive control system for two connected power-split HEVs considering the fuel economy, the aerodynamic drag varied by vehicle spacing, and the road shape information has been presented. The performance of the proposed control system was confirmed by the computer simulations. The proposed control method has produced the fast convergence of the vehicle spacing. The excessive acceleration and deceleration have been avoided by predicting the road shapes. The results revealed improvements of the fuel economy considering the effect of different parameters, that is, road conditions, battery state of charge, and real-time implementation

ability. Since experiments of vehicles are expensive, we would like to conduct experiments in the future. For the sake of simplicity, we have considered only two cars. In the future, we will add more vehicles to make the model more realistic.

Appendices

A. Engine Fuel Consumption Model

The proposed engine fuel consumption modeling method is a special method using both Willans line method and the assumption of operating the engine along the engine optimal operating line and is introduced as follows. The HEV parameters are used from the ADVISOR 2002 Toyota Prius HEV data [27].

The Willans line model consists of an affine representation relating the available energy, that is, the energy that is theoretically available for conversion, to the useful energy that is actually present at the output of the energy converter [25]. Formally

$$W_{\text{out}} = eW_{\text{in}} - W_{\text{loss}}, \tag{A.1}$$
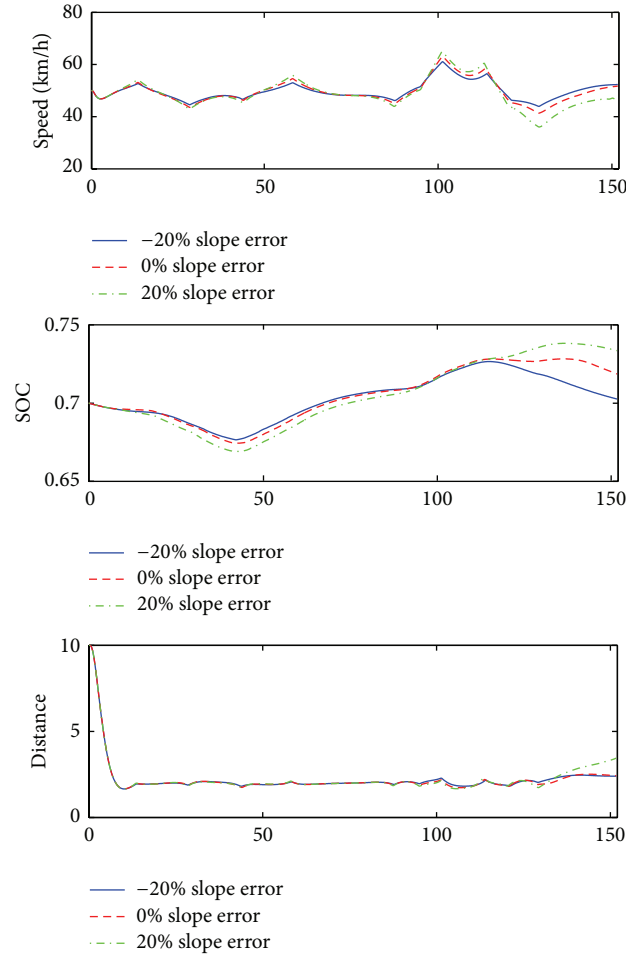
FIGURE 9: Driving and power-split profile deviation due to the slope-sensing error.
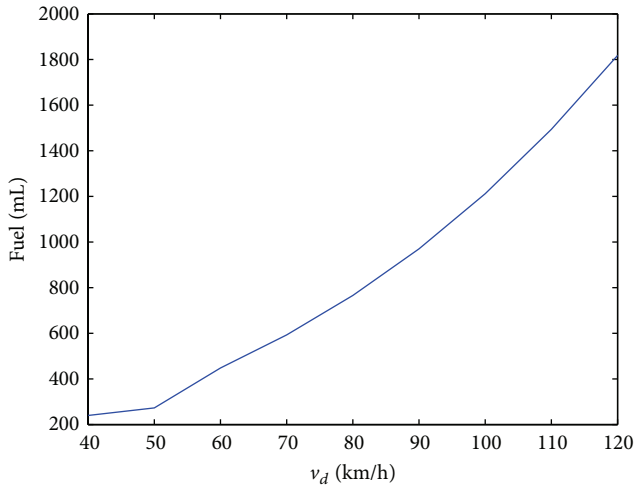
FIGURE 10: Effects of the desired vehicle speed on fuel economy.
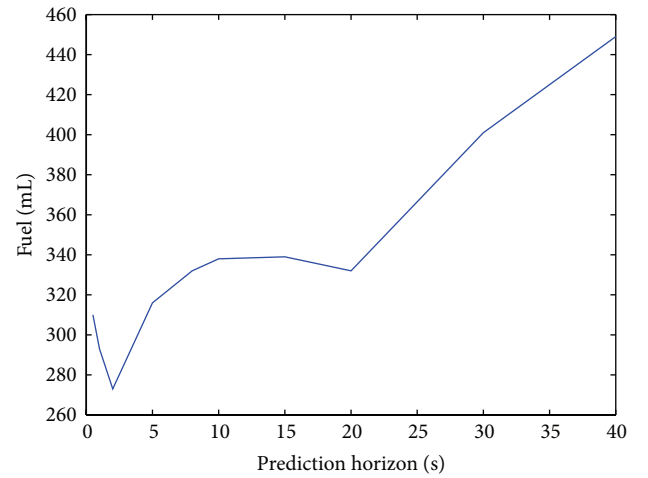
FIGURE 11: Effects of the prediction horizon on fuel economy.

where the parameter e represents the peak intrinsic energy conversion efficiency of the converter and W_{loss} represents external (parasitic) losses. In fact, this model of energy conversion efficiency is nonlinear, in that the parameters e and W_{loss} are represented as explicit functions of the output flow variable (e.g., engine speed) and are also implicit functions of the effort variable.
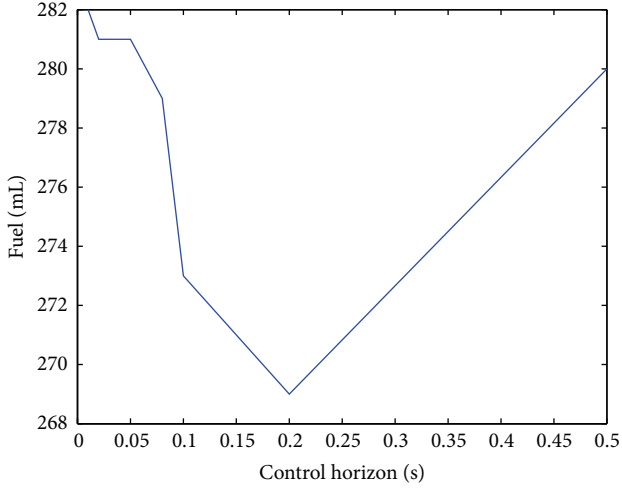
FIGURE 12: Effects of the control horizon on fuel economy.
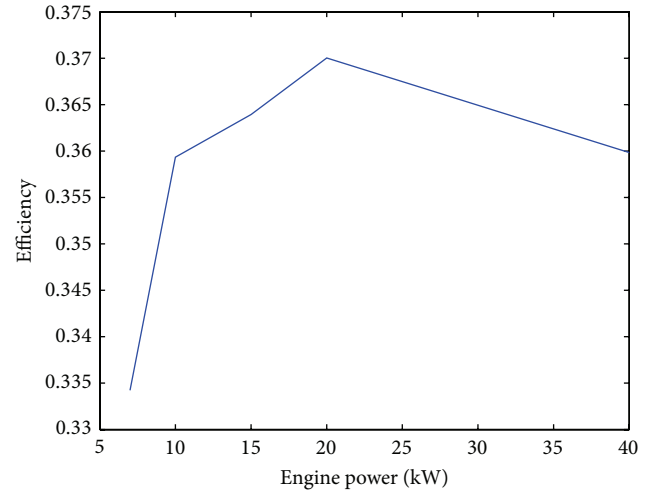
FIGURE 14: The engine efficiency curve to the best engine operating points.
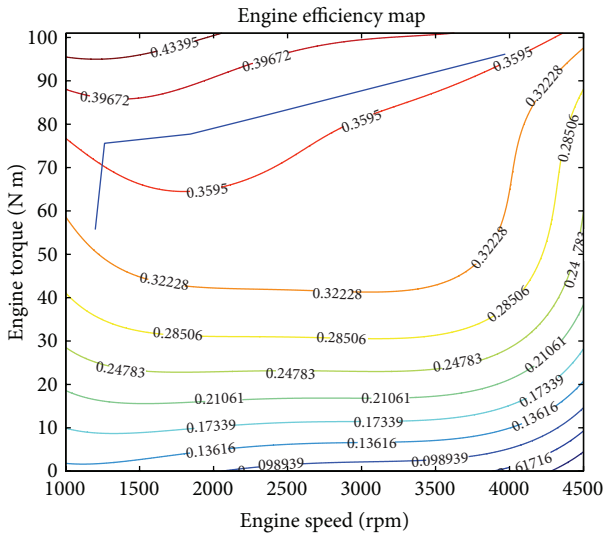
FIGURE 13: The engine efficiency map to the best engine operating points.

Original data
- - - Linear fit

FIGURE 15: The engine fuel consumption rate to the best engine operating points.

The modeling method given above is for general engines. However, in this work, the electric CVT can realize idle stop, so W_{loss} becomes zero. When it is assumed that the engine operating points are maintained at the best efficiency, the parameters e can be approximated as a constant. In this case, the fuel consumption rate corresponding to the optimal operating line can be fitted using a linear function.

The engine optimal operating line can be plotted on the engine map as shown in Figure 13. The engine optimal operating points provide the highest efficiency for a given power level. The engine best efficiency related to the engine power according to the engine characteristics is shown in Figure 14.

The fuel consumption rate is estimated as (see Figure 15)

$$\dot{m}_f = \frac{P_{\text{eng}}}{C\eta} \approx c_f P_{\text{eng}}, \tag{A.2}$$

where C is the calorific value of the gasoline, which is equal to 34.5×10^6 [J/l], and η is the engine efficiency.

B. Road Slope Modeling Method

A brief description of the road slope modeling method is provided as follows. In this research, the sigmoid function is used to model the road slope. The general sigmoid function to model the road slope is expressed as follows:

$$\theta(p) = \frac{s_1}{1 + e^{(s_3(p - s_2))}} + \frac{s_4}{1 + e^{(s_6(p - s_5))}} + \cdots, \tag{B.1}$$

where s_1, s_2, s_3, s_4, s_5, and s_6 are slope shape parameters. The parameters $s_2, s_5, \ldots$ are position parameters where the
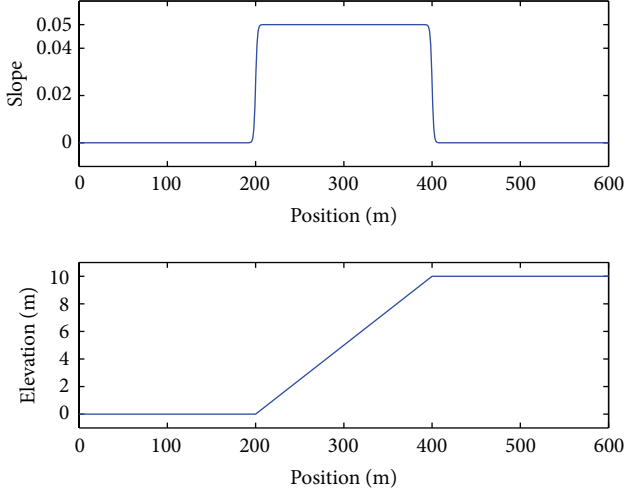
FIGURE 16: The sigmoid function of the up slope.

FIGURE 18: The sigmoid function of the up-down slope.

When it is the up-down slope case, the sigmoid function is expressed as follows:

$$\theta(p) = \frac{0.05}{1 + e^{(-(p-200))}} + \frac{-0.1}{1 + e^{(-(p-400))}}$$
$$+ \frac{0.05}{1 + e^{(-(p-600))}}. \tag{B.4}$$

The figure of the sigmoid function is showed in Figure 18.

C. Solution of the Model Predictive Control Problem

A brief description of the solution of the model predictive control problem is provided as follows.

To implement the model predictive control algorithm, the horizon T is divided into N steps, and the optimal control problem is discretized. The general discretized optimal control problem is formulated as

$$\min_u \quad J = \sum_{i=0}^{N-1} L\left(x_i\left(\tau \mid t\right), u_i\left(\tau \mid t\right)\right) \Delta \tau(t)$$

subject to $\quad x_{i+1}\left(\tau \mid t\right)$
$$= x_i\left(\tau \mid t\right) \tag{C.1}$$
$$+ f\left(x_i\left(\tau \mid t\right), u_i\left(\tau \mid t\right)\right) \Delta \tau(t)$$
$$G\left(x_i\left(\tau \mid t\right), u_i\left(\tau \mid t\right)\right) \le 0,$$

where u is the control input, x is the state, and L is the cost function. $f(x, u)$ is the state equation. $G(x, u)$ is the inequality constraint.

The inequality constraint in the optimal control problem is converted to an equality constraint by introducing a dummy input u_d for computation simplicity as follows:

$$C\left(x(t), u(t)\right) = u^2(t) + u_d^2(t) - u_{max}^2 = 0, \tag{C.2}$$

where u_{max} denotes the upper bound of the control input.

FIGURE 17: The sigmoid function of the down slope.

road slopes change. The parameters $s_3, s_6, \ldots$ are parameters to decide the abruptness of the road slope.

When it is the up slope case, the sigmoid function is expressed as follows:

$$\theta(p) = \frac{0.05}{1 + e^{(-(p-200))}} + \frac{-0.05}{1 + e^{(-(p-400))}}. \tag{B.2}$$

The figure of the sigmoid function is showed in Figure 16.

When it is the down slope case, the sigmoid function is expressed as follows:

$$\theta(p) = \frac{-0.05}{1 + e^{(-(p-200))}} + \frac{0.05}{1 + e^{(-(p-400))}}. \tag{B.3}$$

The figure of the sigmoid function is showed in Figure 17.

To solve this optimal control problem with the calculus of variation method, the Hamiltonian function is defined by

$$H(x, u, \lambda, \psi) = L(x, u) + \lambda^T f(x, u) + \psi^T C(x, u), \quad \text{(C.3)}$$

where λ denotes the costate and ψ denotes the Lagrange multiplier associated with the equality constraint.

The first-order necessary conditions for the optimal control input u, the multiplier ψ, and the costate λ are obtained using the calculus of variation as

$$x_{i+1}(t) = x_i(t) + f(x_i(t), u_i(t)) \Delta \tau(t)$$
$$x_0(t) = x(t),$$

$$\lambda_i(t)$$
$$= \lambda_{i+1}(t)$$
$$+ H_x(x_i(t), u_i(t), \lambda_{i+1}(t), \psi_i(t)) \Delta \tau(t) \quad \text{(C.4)}$$
$$\lambda_N(t) = 0,$$

$$H_u(x_i(t), u_i(t), \lambda_{i+1}(t), \psi_i(t)) = 0$$
$$C(x(t), u(t)) = 0,$$

where x_0 is the initial state.

To solve this optimal control problem, the continuation and GMRES (C/GMRES) method is employed for computation cost reduction. The necessary conditions of optimality for the constrained control input can be expressed as the following equation:

$$F(U(\tau \mid t), x(\tau \mid t), t)$$

$$:= \begin{bmatrix} H_u(u_0(\tau \mid t), x_0(\tau \mid t), \lambda_1(\tau \mid t), \psi_0(\tau \mid t)) \\ C(u_0(\tau \mid t), x_0(\tau \mid t)) \\ \vdots \\ H_u(u_{N-1}(\tau \mid t), x_{N-1}(\tau \mid t), \lambda_N(\tau \mid t), \psi_{N-1}(\tau \mid t)) \\ C(u_{N-1}(\tau \mid t), x_{N-1}(\tau \mid t)) \end{bmatrix} \quad \text{(C.5)}$$

$$= 0,$$

$$U(t)$$
$$:= \left[u_0^T(\tau \mid t), \psi_0^T(\tau \mid t), \ldots, u_{N-1}^T(\tau \mid t), \psi_{N-1}^T(\tau \mid t) \right]^T.$$

$F(U(t), x(t), t) = 0$ is identical to

$$F(U(0), x(0), 0) := 0,$$
$$\dot{F}(U, x, t) = -A_s F(U(t), x(t), t), \quad \text{(C.6)}$$

where A_s is a stable matrix introduced to stabilize $F = 0$. If F_U is nonsingular, a differential equation for $U(t)$ can be obtained as

$$\dot{U} = -F_U^{-1}(A_s F - F_x \dot{x} - F_t). \quad \text{(C.7)}$$

The above differential equation can be solved by the GMRES method. The presented approach is also a kind of continuation method. The solution curve $U(t)$ is traced by integrating the above differential equation. Because there is no need to calculate the Jacobians and the linear equation iteratively, C/GMRES method assures the real-time optimal control ability because of small computational cost. The detailed description of the solution for the model predictive control algorithm can be found [26].

Notations and Abbreviations

HEVs:	Hybrid electric vehicles
MPC:	Model predictive control
CVT:	Continuously variable transmission
OOL:	Optimal operating line
SOC:	State of charge
PSD:	Power-split device
M/G:	Motor/generator
FD:	Final drive
A:	Frontal area
C_D:	Aerodynamic drag coefficient
g:	Gravitational acceleration
g_f:	Final drive gear ratio
m:	Vehicle mass
P_{batt}:	Battery power
P_{eng}:	Engine power
Q_{batt}:	Battery nominal capacity
r_w:	Wheel radius
R_{batt}:	Battery internal resistance
V_{OC}:	Battery open-circuit voltage
θ:	Road inclination
μ:	Rolling friction coefficient
ρ:	Density of air
S:	The number of sun gear teeth
R:	The number of ring gear teeth
$\tau_{M/G1}$:	The torque of motor/generator 1
$\tau_{M/G2}$:	The torque of motor/generator 2
τ_{req}:	The torque of the road load
τ_{eng}:	The torque of the engine
$\omega_{M/G1}$:	The angular speed of motor/generator 1
$\omega_{M/G2}$:	The angular speed of motor/generator 2
ω_{eng}:	The angular speed of the engine
P_{req}:	The power of the road load
c_f:	A constant for the fuel consumption
p:	The vehicle position
v:	The vehicle speed
w:	The acceleration or deceleration converted from the traction force or brake force
u_p:	The acceleration or deceleration control input of the preceding vehicle
u_h:	The acceleration or deceleration control input of the host vehicle
k_p:	The delay constant
$s_1, s_2, s_3, s_4, s_5,$ and s_6:	Slope shape parameters
T:	The prediction horizon
min and max:	The minimum and maximum bounds of the parameters
$w_x, w_y, w_z, w_d, w_e, w_f,$ and w_g:	The weights

v_d: The desired vehicle speed
d: The vehicle spacing
d_d: The desired vehicle spacing
l_p: The length of the preceding vehicle
SOC_d: The desired SOC value
k_{SOC} and SOC_k: Constant parameters
c_{i1}–c_{i8}:   Constant parameters for the aerodynamic drag coefficients
C: The calorific value of the gasoline
u: The control input
x: The state
L: The cost function
$f(x, u)$:   The state equation
$G(x, u)$:   The inequality constraint
u_d: The dummy input
$u_{\max}$:   The upper bound of the control input
λ: The costate
ψ: The Lagrange multiplier associated with the equality constraint
x_0:   The initial state
A_s: A stable matrix.

Acknowledgments

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] A. Sciarretta and L. Guzzella, "Control of hybrid electric vehicles," *IEEE Control Systems Magazine*, vol. 27, no. 2, pp. 60–70, 2007.

[2] C. Zhang, A. Vahidi, P. Pisu, X. Li, and K. Tennant, "Role of terrain preview in energy management of hybrid electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1139–1147, 2010.

[3] K. Yu and J. Yang, "Performance of a nonlinear real-time optimal control system for HEVs/PHEVs during car following," *Journal of Applied Mathematics*, vol. 2014, Article ID 879232, 14 pages, 2014.

[4] K. Yu, H. Yang, Q. Liang, J. Yang, and Y. Guo, "Model predictive control for hybrid electric vehicle platooning using route information," *Proceedings of the Institution of Mechanical Engineers Part D: Journal of Automobile Engineering*. In press.

[5] K. Yu, M. Mukai, and T. Kawabe, "Model predictive control of a power-split hybrid electric vehicle system," *Artificial Life and Robotics*, vol. 17, no. 2, pp. 221–226, 2012.

[6] K. Yu, M. Mukai, and T. Kawabe, "A battery management system using nonlinear model predictive control for a hybrid electric vehicle," in *Proceedings of the 7th IFAC Symposium on Advances in Automotive Control*, pp. 301–306, Tokyo, Japan, September 2013.

[7] K. Yu, J. Yang, and D. Yamaguchi, "Model predictive control for hybrid vehicle ecological driving using traffic signal and road slope information," *Control Theory and Technology*, vol. 13, no. 1, pp. 333–344, 2015.

[8] K. Yu, Q. Liang, Z. Hu, J. Yang, and H. Zhang, "Performance of an eco-driving model predictive control system for HEVs during car following," *Asian Journal of Control*, 2015.

[9] K. Yu, Q. Liang, J. Yang, and Y. Guo, "Model predictive control for hybrid electric vehicle platooning using slope information," *IEEE Transactions on Intelligent Transportation Systems*. In press.

[10] D. Swaroop, J. K. Hedrick, C. C. Chien, and P. Ioannou, "A comparison of spacing and headway control laws for automatically controlled vehicles," *Vehicle System Dynamics*, vol. 23, no. 8, pp. 597–625, 1994.

[11] L. Serrao, *A comparative analysis of energy management strategies for hybrid electric vehicles [Ph.D. thesis]*, The Ohio State University, Columbus, Ohio, USA, 2009.

[12] J. Liu and H. Peng, "Modeling and control of a power-split hybrid vehicle," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1242–1251, 2008.

[13] N. Kim, S. Cha, and H. Peng, "Optimal control of hybrid electric vehicles based on Pontryagin's minimum principle," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1279–1287, 2011.

[14] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[15] C. Musardo, G. Rizzoni, Y. Guezennec, and B. Staccia, "A-ECMS: an adaptive algorithm for hybrid electric vehicle energy management," *European Journal of Control*, vol. 11, no. 4-5, pp. 509–524, 2005.

[16] H. Yamazaki and K. Okamoto, "Fuel saving by trucks in single-file platooning on the expressway," *JARI Research Journal*, vol. 32, no. 3, pp. 139–143, 2010.

[17] R. Rajamani, H.-S. Tan, B. K. Law, and W.-B. Zhang, "Demonstration of integrated longitudinal and lateral control for the operation of automated vehicles in platoons," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 4, pp. 695–708, 2000.

[18] M. Barth, "An emissions and energy comparison between a simulated automated highway system and current traffic conditions," in *Proceedings of the IEEE Intelligent Transportation Systems Proceedings*, Dearborn, Mich, USA, October 2000.

[19] Y. Yamamura, Y. Seto, and M. Nagai, "Study on a string-stable ACC using vehicle-to-vehicle communication (2nd report, ACC design method using model predictive control)," *Transactions of the Japan Society of Mechanical Engineers Series C*, vol. 73, no. 7, pp. 1917–1922, 2007.

[20] M. Zabat, N. Stabile, S. Frascaroli, and F. Browand, *The Aerodynamic Performance of Platoons: A Final Report*, California Partners for Advanced Transit and Highways, 1995.

[21] R. E. Fenton and R. J. Mayhan, "Automated highway studies at the Ohio State University—an overview," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 100–113, 1991.

[22] Y. Zhang, E. B. Kosmatopoulos, P. A. Ioannou, and C. C. Chien, "Autonomous intelligent cruise control using front and back information for tight vehicle following maneuvers," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 1, pp. 319–328, 1999.

[23] J. Bom, B. Thuilot, F. Marmoiton, and P. Martinet, "Nonlinear control for urban vehicles platooning, relying upon a unique kinematic GPS," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '05)*, pp. 4138–4143, IEEE, April 2005.

[24] M. Ehsani, Y. Gao, and A. Emadi, *Modern Electric, Hybrid Electric, and Fuel Cell Vehicles: Fundamentals, Theory, and Design*, CRC, Boca Raton, Fla, USA, 2nd edition, 2009.

[25] G. Rizzoni, L. Guzzella, and B. M. Baumann, "Unified modeling of hybrid electric vehicle drivetrains," *IEEE/ASME Transactions on Mechatronics*, vol. 4, no. 3, pp. 246–257, 1999.

[26] T. Ohtsuka, "A continuation/GMRES method for fast computation of nonlinear receding horizon control," *Automatica*, vol. 40, no. 4, pp. 563–574, 2004.

[27] K. B. Wipke, M. R. Cuddy, and S. D. Burch, "ADVISOR 2.1: a user-friendly advanced powertrain simulation using a combined backward/forward approach," *IEEE Transactions on Vehicular Technology*, vol. 48, no. 6, pp. 1751–1761, 1999.

Research Article

Robust Online State of Charge Estimation of Lithium-Ion Battery Pack Based on Error Sensitivity Analysis

Ting Zhao,[1,2] Jiuchun Jiang,[1,2] Caiping Zhang,[1,2] Kai Bai,[3] and Na Li[3]

[1]*National Active Distribution Network Technology Research Center (NANTEC), Beijing Jiaotong University,*
 No. 3 Shang Yuan Cun, Haidian District, Beijing 100044, China
[2]*Collaborative Innovation Center of Electric Vehicles in Beijing, Beijing 100044, China*
[3]*State Grid Jibei Electric Power Co., Ltd., Research Institute, No. 1 Dizang'an Nanxiang Fuxingmenwai Street, Beijing 100045, China*

Correspondence should be addressed to Jiuchun Jiang; jcjiang@bjtu.edu.cn

Accurate and reliable state of charge (SOC) estimation is a key enabling technique for large format lithium-ion battery pack due to its vital role in battery safety and effective management. This paper tries to make three contributions to existing literatures through robust algorithms. (1) Observer based SOC estimation error model is established, where the crucial parameters on SOC estimation accuracy are determined by quantitative analysis, being a basis for parameters update. (2) The estimation method for a battery pack in which the inconsistency of cells is taken into consideration is proposed, ensuring all batteries' SOC ranging from 0 to 1, effectively avoiding the battery overcharged/overdischarged. Online estimation of the parameters is also presented in this paper. (3) The SOC estimation accuracy of the battery pack is verified using the hardware-in-loop simulation platform. The experimental results at various dynamic test conditions, temperatures, and initial SOC difference between two cells demonstrate the efficacy of the proposed method.

1. Introduction

As one of the most important performance parameters of traction batteries, real-time SOC estimation of battery becomes necessary in the field of application of battery-driven electric vehicles. Commonly used SOC estimation methods for single battery are as follows [1, 2]: ampere hour integration method; open circuit voltage method using the corresponding relation between the open circuit voltage and SOC; the algorithm based on electric circuit models or electrochemical models and the typical methods which are Kalman filters and methods based on some observer; estimation using fuzzy logic or methods of machine learning. Since the dynamic battery shows high nonlinearity, these mentioned methods have their own defects: ampere hour integration method has to know the exact SOC initial value and can bring accumulative error; the open circuit voltage scheme has strict requirements on measuring conditions and update of OCV curves can be a difficult task, and it is therefore not suitable for electric vehicle during driving and would better be used as auxiliary correction means; the accuracy of estimation method based on circuit model depends on the model parameters; however, adaptation of the model parameters over the battery lifetime to any given battery aging state or environment requires relatively complex algorithms and is practical for only quite simple models. Some researches based on Kalman filter have achieved online estimation [2–4], but online realization is at the expense of additional computing power and its robustness needs to be concerned; the electrochemical method cannot be applied into operation due to the complexity of the model itself [5]; fuzzy control or vector machine algorithm requires a large number of sample data to train the model and also is no longer applicable for aged battery [6].

A large number of batteries are connected in series and parallels when used in electric vehicles to reach the corresponding level of voltage, power, and energy [3]. Therefore, accurate estimation of state of charge for a battery pack is remaining challenging. Typical SOC calculation method for battery packs [7–10] takes the group as a large battery and
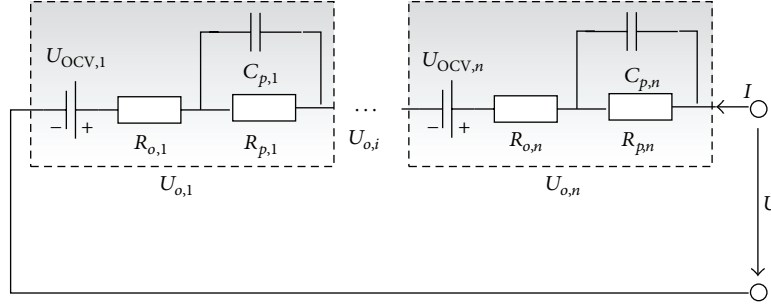
FIGURE 1: The model of series-connected battery pack.

then applies one of the methods described in previous sections, while inconsistent characteristics of cell capacity and resistance cannot be ignored or lead to low management efficiency. Some methods calculate the mean SOC of a battery pack based on mean model or difference model considering cells discordance [8, 11–13]; others estimate the SOC of every cell in battery pack using common calculation methods and then obtain the pack SOC [7, 14]. From the above references, methods considering multiple cells are more reliable for battery pack SOC estimation. However, in practical application of pack SOC estimation, the operation conditions, measurement accuracy, and computation of the algorithm need to be considered in integration.

A key contribution of this paper is solving problems in practical application of SOC estimation. In the paper, SOC estimation error model is established, where the crucial parameters on SOC estimation accuracy are determined by quantitative analysis. Battery pack SOC estimation method considered cell inconsistency is on the basis of cell SOC. And single cell SOC is estimated through PI observer combined ampere hour integration method. Also, online identification of the model parameters can make sure of the applicability of the algorithm under different operating conditions during the battery life cycle. In addition, a hardware-in-loop platform is built to fulfill real-time estimation. The verification results of the two series-connected battery packs under different circumstances indicate that this method has a high accuracy and achieves better convergence and stability.

The remainder of this paper is organized as follows: Section 2 analyzes the SOC estimation error theoretically. Section 3 builds series-connected battery pack model, gives the SOC definition, and describes online model parameters update. Our validation results and discussions for different cases of battery pack are presented in Section 4. Section 5 summarizes the conclusions of the paper.

2. Sensitivity Analysis of SOC Estimation

2.1. The Proposed SOC Definition for a Battery Pack. For series-connected battery packs, we adopt the model presented in Figure 1, which is a group model connected with many first-order Thevenin models in series. The definition for SOC of battery strings is given in formula (1), where $Q_{\max}^{B}$ denotes the maximum available capacity of the pack and Q_{rem}^{B} denotes residual capacity, namely, the maximum discharge capacity for the group. As long as we know SOC value and the maximum available capacity of each cell in packs, group SOC will be achieved. Through the proposed SOC pack model, battery pack SOC estimation with inconsistency among cells can be considered on the one hand; on the other hand, as long as the group SOC, namely, SOC^{B}, ranges from 0% to 100%, all single cell SOC will change in the range of 0% to 100%. Thus overcharging or overdischarging for the battery can be effectively avoided, and it also can provide data support and ensure secure use of the battery pack:

$$
\begin{aligned}
\mathrm{SOC}^{B} &= \frac{Q_{\mathrm{rem}}^{B}}{Q_{\max}^{B}} \times 100\% \\
&= \frac{\min\left(Q_{\max}\left[1\right]\mathrm{SOC}\left[1\right], \ldots, Q_{\max}\left[n\right]\mathrm{SOC}\left[n\right]\right)}{\min\left(Q_{\max}\left[1\right]\mathrm{SOC}\left[1\right], \ldots, Q_{\max}\left[n\right]\mathrm{SOC}\left[n\right]\right) + \min\left(Q_{\max}\left[1\right]\left(1 - \mathrm{SOC}\left[1\right]\right), \ldots, Q_{\max}\left[n\right]\left(1 - \mathrm{SOC}\left[n\right]\right)\right)} \\
&= \frac{Q_{\max}\left[i\right]\mathrm{SOC}\left[i\right]}{Q_{\max}\left[i\right]\mathrm{SOC}\left[i\right] + Q_{\max}\left[j\right]\left(1 - \mathrm{SOC}\left[j\right]\right)} \times 100\%.
\end{aligned}
\tag{1}
$$

2.2. SOC Estimation Error Modeling. Based on the definition of SOC for battery strings in Section 2.1, error of group SOC comes from single cell SOC error. Therefore, error analysis in this part aims for single cells. When the battery is in a different environment or state, the model parameters are also not the

same. Through analyzing impact factors of SOC estimation error and influencing degree of each impact factor, SOC estimation accuracy can be improved effectively. Consequently, SOC estimation error formula is deduced mathematically based on the first-order Thevenin battery model [15, 16]. In

the formula, parameters in the coefficient matrices $\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \widetilde{\mathbf{C}}, \widetilde{\mathbf{D}}$ are estimation value and have difference with the true value matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$, where $\mathbf{A} = \begin{bmatrix} -1/R_p C_p & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1/C_p \\ 1/(3600 \cdot C_n) \end{bmatrix}$, $\mathbf{C} = \begin{bmatrix} 1 & a_i \end{bmatrix}$, and $\mathbf{D} = R_o$. The actual and estimated state equations of battery are listed in (2), error is expressed in (3), and the final SOC error expression is represented in (9), where $\dot{\mathbf{x}} = \begin{bmatrix} \dot{U}_p \\ \text{SÒC} \end{bmatrix}$, $y = U_o - b_i$, $u = I$, and $\text{OCV} = a_i \cdot \text{SOC} + b_i$. Consider

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u,$$

$$y = \mathbf{C}\mathbf{x} + \mathbf{D}u,$$

$$\dot{\widetilde{\mathbf{x}}} = \widetilde{\mathbf{A}}\widetilde{\mathbf{x}} + \widetilde{\mathbf{B}}u + \mathbf{L}\left(U_o - \widetilde{U}_o\right),$$

$$\widetilde{y} = \widetilde{\mathbf{C}}\widetilde{\mathbf{x}} + \widetilde{\mathbf{D}}u,$$

$$\mathbf{e} = \widetilde{\mathbf{x}} - \mathbf{x},$$

$$\dot{\mathbf{e}} = \dot{\widetilde{\mathbf{x}}} - \dot{\mathbf{x}} = \widetilde{\mathbf{A}}\widetilde{\mathbf{x}} + \widetilde{\mathbf{B}}u + \mathbf{L}\left(U_o - \widetilde{U}_o\right) - [\mathbf{A}\mathbf{x} + \mathbf{B}u] \tag{2}$$

$$= \widetilde{\mathbf{A}}\widetilde{\mathbf{x}} - \mathbf{A}\mathbf{x} + \left(\widetilde{\mathbf{B}} - \mathbf{B}\right)u$$

$$+ \mathbf{L}\left[\mathbf{C}\mathbf{x} - \widetilde{\mathbf{C}}\widetilde{\mathbf{x}} + \left(\mathbf{D} - \widetilde{\mathbf{D}}\right)u + b_i - \widetilde{b}_i\right],$$

$$\dot{\mathbf{e}} = \widetilde{\mathbf{A}}\widetilde{\mathbf{x}} - \widetilde{\mathbf{A}}\mathbf{x} + \left(\widetilde{\mathbf{A}} - \mathbf{A}\right)\mathbf{x} + \left(\widetilde{\mathbf{B}} - \mathbf{B}\right)u$$

$$- \mathbf{L}\left[\widetilde{\mathbf{C}}\widetilde{\mathbf{x}} - \widetilde{\mathbf{C}}\mathbf{x} + \left(\widetilde{\mathbf{C}} - \mathbf{C}\right)\mathbf{x} + \left(\widetilde{\mathbf{D}} - \mathbf{D}\right)u + b_i - \widetilde{b}_i\right]$$

$$= \widetilde{\mathbf{A}}\left(\widetilde{\mathbf{x}} - \mathbf{x}\right) + \Delta\mathbf{A} \cdot \mathbf{x} + \Delta\mathbf{B} \cdot u$$

$$- \mathbf{L}\left[\widetilde{\mathbf{C}}\left(\widetilde{\mathbf{x}} - \mathbf{x}\right) + \Delta\mathbf{C} \cdot \mathbf{x} + \Delta\mathbf{D} \cdot u - \Delta b_i\right] \tag{4}$$

$$= \left(\widetilde{\mathbf{A}} - \mathbf{L}\widetilde{\mathbf{C}}\right)\mathbf{e} + (\Delta\mathbf{A} - \mathbf{L} \cdot \Delta\mathbf{C}) \cdot \mathbf{x}$$

$$+ (\Delta\mathbf{B} - \mathbf{L} \cdot \Delta\mathbf{D}) \cdot u - \mathbf{L} \cdot b_i,$$

$$\dot{\mathbf{e}} = \begin{bmatrix} \dot{e}_{U_p} \\ \dot{e}_{\text{SOC}} \end{bmatrix}$$

$$= \begin{bmatrix} -\dfrac{1}{\widetilde{R}_p \widetilde{C}_p} - L_1 & -L_1 \widetilde{a}_i \\ -L_2 & -L_2 \widetilde{a}_i \end{bmatrix} \begin{bmatrix} e_{U_p} \\ e_{\text{SOC}} \end{bmatrix}$$

$$+ \begin{bmatrix} -\Delta\dfrac{1}{R_p C_p} & -L_1 \cdot \Delta a_i \\ 0 & -L_2 \cdot \Delta a_i \end{bmatrix} \begin{bmatrix} U_p \\ \text{SOC} \end{bmatrix} \tag{5}$$

$$+ u \begin{bmatrix} \Delta\dfrac{1}{C_p} - L_1 \cdot \Delta R_o \\ \dfrac{1}{3600} \cdot \left(\Delta\dfrac{1}{C_n}\right) - L_2 \cdot \Delta R_o \end{bmatrix} - \begin{bmatrix} L_1 \cdot \Delta b_i \\ L_2 \cdot \Delta b_i \end{bmatrix}.$$

The SOC error matrix can be expanded as follows:

$$\dot{e}_{\text{SOC}} + L_2 \widetilde{a}_i \cdot e_{\text{SOC}} = -L_2 e_{U_p} - L_2 \cdot \Delta a_i \cdot \text{SOC}$$

$$+ u\left(\frac{1}{3600} \cdot \Delta\frac{1}{C_n} - L_2 \cdot \Delta R_o\right) \tag{6}$$

$$- L_2 \cdot \Delta b_i,$$

$$U_p = u \cdot R_p \left(1 - e^{-t/R_p C_p}\right) + U_p(0)$$

$$\cdot e^{-t/R_p C_p}, \quad U_p(0) = 0. \tag{7}$$

Equation (6) can be simplified to the following form: $\dot{e}_{\text{SOC}} + p(t) \cdot e_{\text{SOC}} = q(t)$.
The corresponding solution to the equation is as follows:

$$e_{\text{SOC}} = c \cdot e^{-\int L_2 \widetilde{a}_i dt} + e^{-\int L_2 \widetilde{a}_i dt} \int q \cdot e^{\int L_2 \widetilde{a}_i dt} dt$$

$$(c \text{ depends on } e_{\text{SOC}}(0)),$$

$$q = -L_2 \left(u \cdot \widetilde{R}_p \left(1 - e^{-t/\widetilde{R}_p \widetilde{C}_p}\right) - u \cdot R_p \left(1 - e^{-t/R_p C_p}\right)\right)$$

$$- L_2 \cdot \Delta a_i \cdot \text{SOC} + u\left(\frac{1}{3600} \cdot \Delta\frac{1}{C_n} - L_2 \cdot \Delta R_o\right)$$

$$- L_2 \cdot \Delta b_i, \tag{8}$$

$$e_{\text{SOC}} = c e^{-L_2 \widetilde{a}_i t}$$

$$+ \frac{1}{L_2 \widetilde{a}_i} \left[u\left(\frac{1}{3600} \cdot \Delta\frac{1}{C_n} - L_2 \cdot \left(\Delta R_o + \Delta R_p\right)\right)\right.$$

$$\left. - L_2 \cdot \Delta b_i - L_2 \cdot \Delta a_i \cdot \text{SOC}\right] + L_2 u \left[-\widetilde{R}_p^2 \widetilde{C}_p\right.$$

$$\left. \cdot e^{(-1/\widetilde{R}_p \widetilde{C}_p)t} + R_p^2 C_p \cdot e^{(-1/R_p C_p)t}\right].$$

When $t = 0$, assuming that current u is zero, the open circuit voltage estimation error is zero; then $c = e_{\text{SOC}}(0)$. In conclusion, SOC error equation may be obtained as follows:

$$e_{\text{SOC}} = \underbrace{e_{\text{SOC}}(0) e^{-L_2 \widetilde{a}_i t} + L_2 u \left[-\widetilde{R}_p^2 \widetilde{C}_p \cdot e^{(-1/\widetilde{R}_p \widetilde{C}_p)t} + R_p^2 C_p \cdot e^{(-1/R_p C_p)t}\right]}_{\text{transient error}}$$

$$+ \underbrace{\frac{1}{L_2 \widetilde{a}_i} \left[u\left(\frac{1}{3600} \cdot \Delta\frac{1}{C_n} - L_2 \cdot \left(\Delta R_o + \Delta R_p\right)\right) - L_2 \cdot \Delta b_i - L_2 \cdot \Delta a_i \cdot \text{SOC}\right]}_{\text{steady-state error}} = e_{\text{SOC}}1 + e_{\text{SOC}}2. \tag{9}$$

TABLE 1: SOC estimation error with different parameter error for a battery with the capacity of 8.8 Ah.

Factors	Factor error							
	0.2%	0.5%	1%	2%	5%	10%	20%	30%
	e_{SOC} (%)							
ΔQ	−0.0000037	−0.0000092	−0.0000183	−0.0000362	−0.0000879	−0.000168	−0.000308	−0.000426
ΔR_{total}	0.0296	0.074	0.148	0.296	0.74	1.48	2.96	4.44
Δa_i	−0.2	−0.5	−1	−2	−5	−10	−20	−30
Δb_i	1.45	3.62	7.24	14.5	36.2	—	—	—

For expression $e_{SOC}(0)e^{-L_2\tilde{a}_i t}$, because the maximum value of e_{SOC} is one and the value of $L_2\tilde{a}_i$ is not too small, thus the value of $e^{-L_2\tilde{a}_i t}$ will soon decline over time. Of course the expression $e_{SOC}(0)e^{-L_2\tilde{a}_i t}$ can be ignored. As for other part of the transient error $L_2 u[-\tilde{R}_p^2 \tilde{C}_p \cdot e^{(-1/\tilde{R}_p \tilde{C}_p)t} + R_p^2 C_p \cdot e^{(-1/R_p C_p)t}]$, it will get small as time goes by and tend to zero gradually after 360 s (6 min), while the value of SOC decreases only about 1.4% during this period of time. Transient error can be ignored considering the actual application and the analysis. And the example for transient error will be displayed in Figure 1.

On the other hand, from the expression of the steady-state error can we see that SOC estimation error is brought by the error of capacity, total internal resistance, and OCV. The variable $e_{SOC}2$ is expressed in (10) and according to the actual simulation data, the sensitivity of each factor for SOC estimation will be analyzed furtherly. Consider

$$
e_{SOC}2 = \frac{i}{L_2 \cdot \tilde{a}_i} \cdot \frac{1}{3600} \cdot \Delta\left(\frac{1}{Q}\right) - \frac{i}{\tilde{a}_i} \cdot \Delta R_{total}
$$
$$
- \frac{\Delta a_i \cdot SOC(t)}{\tilde{a}_i} - \frac{\Delta b_i}{\tilde{a}_i} = e_1 - e_2 - e_3 - e_4. \quad (10)
$$

2.3. Quantified Analysis.

The SOC estimation error is divided into transient and steady-state error from the analysis in Section 2.2 and both will be discussed, respectively. Equation (11) is used to calculate and analyze transient SOC error, where resistance R_p, capacitance C_p, observer coefficient L_2, and input current u are all based on the simulation value or test data and they are assigned the value of 0.6 milliohms, 80000 F, 10, and DST operating current condition, respectively, in this case. SOC transient error calculation results are presented in Figure 2, in which the black line represents the SOC transient error with 2% polarization resistance and capacitance error while the red line is on behalf of SOC transient error of 5% polarization resistance and capacitance error; that is, the values of R_p and C_p are assigned to $1.02R_p$, $1.02C_p/1.05R_p$, $1.05C_p$. Although the initial SOC transient error is large, it will hold the tendency of decline over time and decrease to almost zero within 360 s. And the value of SOC only reduced about 1.4% during this time; thus it can be ignored in practical application. Consider

$$
e_{SOC}1
$$
$$
= L_2 u \left[-\tilde{R}_p^2 \tilde{C}_p \cdot e^{(-1/\tilde{R}_p \tilde{C}_p)t} + R_p^2 C_p \cdot e^{(-1/R_p C_p)t} \right]. \quad (11)
$$

In terms of SOC steady-state error, conditions like different capacity, different charge, or discharge current rates are discussed, respectively.

(1) A Li(NiCoMn)O$_2$ battery with the capacity of 8.88 Ah is used for analysis. According to the parameter identification results for this cell, taking SOC as 55% as an example, total internal resistance equals 25 milliohms, the slope a_i of the OCV-SOC line is 0.5, and intercept is 3.62; the value of current is 1/3 C; observer coefficient is taken to be 10 according to the simulation. Considering the error caused by Δa_i alone, since the SOC ranges from 0 to 1, thus SOC maximum value is taken as 1 in this case. The results of e_1, e_2, e_3, e_4 with different variable errors are shown in Table 1:

$$
e_{SOC}(t) = \frac{2.96}{10 \cdot 0.5} \cdot \frac{1}{3600} \cdot \Delta\left(\frac{1}{Q}\right) - \frac{2.96}{0.5} \cdot \Delta R_{total}
$$
$$
- \frac{\Delta a_i \cdot SOC(t)}{0.5} - \frac{\Delta b_i}{0.5} \quad (12)
$$
$$
= 1.64 \times 10^{-4}\Delta\left(\frac{1}{Q}\right) - 5.92 \cdot \Delta R_{total} - 2\Delta a_i
$$
$$
\cdot SOC(t) - 2\Delta b_i.
$$

(2) A C/LiMn$_2$O$_4$ battery with the capacity of 90 Ah is considered as an example. Like in case one, total internal resistance is 1.5 milliohms when 55% SOC is taken as an example, the slope a_i of the OCV-SOC line is 0.4, and intercept b_i is 3.786; current value is 1/3 C and observer coefficient is taken to be 10 through simulation. The results of e_1, e_2, e_3, e_4 with different variable errors are shown in Table 2.

(3) In terms of the same battery as in case (2), considering the situations with different charge and discharge rates, since only e_1, e_2 are affected by the rate of current, we only discuss these two. Comparison results are listed in Table 3 and we can conclude e_1 is very small and tends to zero. On the basis of this, SOC steady-state error under DST condition is presented in Figure 3 which only considers the influence of internal resistance error e_2.

Both theoretical analysis and calculation results show that the influencing degree of four impact factors for SOC estimation error is closely related to the battery capacity and charging or discharging current rate. The degree of influence under the same error range ranks as follows: $\Delta a_i > \Delta b_i > \Delta R_{total} > \Delta Q$. However, in a practical situation, when OCV-SOC curve is linearized, error of a_i (the slope of the line) may reach several tens of percent and b_i (the intercept of the line)

TABLE 2: SOC estimation error with different parameter error for a battery with the capacity of 90 Ah.

Factors	Factor error							
	0.2%	0.5%	1%	2%	5%	10%	20%	30%
	e_{SOC} (%)							
ΔQ	−0.0000046	−0.000012	−0.000022	−0.000045	−0.00011	−0.00021	—	—
ΔR_{total}	0.0225	0.0563	0.113	0.225	0.563	1.125	2.25	3.375
Δa_i	−0.196	−0.49	−0.98	−1.96	−4.9	−9.8	−19.6	−29.4
Δb_i	1.893	4.733	9.465	18.93	43.25	—	—	—

TABLE 3: SOC estimation error with different parameter error and rate for a battery with the capacity of 90 Ah.

Factor error	Current rate			
	0.5 C	1 C	1.5 C	2 C
	e_{SOC} (%)			
ΔQ				
2%	−0.00007	−0.00014	−0.00020	−0.00027
5%	−0.00017	−0.00033	−0.00050	−0.00066
10%	−0.00032	−0.00063	−0.00095	−0.0013
20%	−0.00058	−0.00116	−0.00174	−0.00231
ΔR_{total}				
2%	0.3375	0.6750	1.013	1.350
5%	0.8438	1.688	2.531	3.375
10%	1.688	3.375	5.063	6.750
20%	3.375	6.750	10.13	13.50
30%	5.063	10.13	15.19	20.25

FIGURE 3: SOC estimation steady-state error with different resistance error under DST condition.

FIGURE 2: SOC estimation transient error with different polarization resistance error under DST condition with two cycles.

error reaches one-thousandth over aging. Therefore, the OCV-SOC curve has a great influence on SOC estimation error and the accurate measurement of OCV is very helpful to reduce the estimation error. Besides, the effect degree of total internal resistance error on estimation is following and the error of capacity is affected the least, while as charging and discharging current rates get large, SOC estimation error

caused by these two factors will become larger and the impact of total internal resistance error is especially marked. Based on the quantitative calculation and analysis above, we can see it is of profound significance for SOC estimation to improve the accuracy of model parameters.

3. Battery Pack SOC Estimation

3.1. The Proposed SOC Estimation Method for a Battery Pack. The group SOC estimation method adopted in this paper is listed in Figure 4. The battery model parameters of each cell are obtained by online identification. SOC of each cell is estimated with PI observer estimation combined ampere hour integration method; then group SOC can be calculated according to the SOC pack model.

3.2. Online Update of Battery Model Parameters. With the increase of actual use time of the battery and the change in temperature of the environment, battery model parameters like the internal resistance and OCV-SOC curve will also change. In practice, if battery model parameters are not updated, it will bring battery model error and lead to SOC estimation error furtherly. To deal with this problem, the recursive least square method will be employed for online battery model parameters identification. Achievement
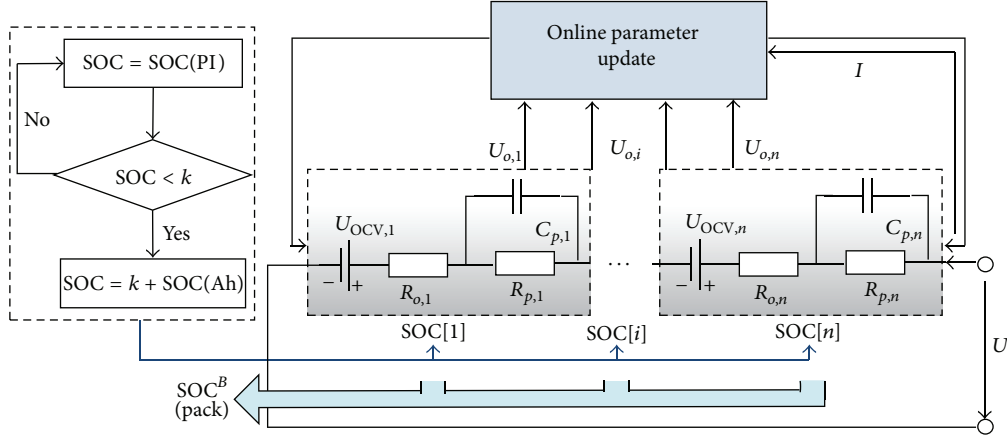
FIGURE 4: The proposed group SOC estimation method.

of real-time update reduces parameter error and thereby increases the SOC estimation accuracy. Principle of recursive least square method is explained in [17]. For a given system $y = \boldsymbol{\varphi} \cdot \boldsymbol{\theta}$, where y denote the output vector and $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ denote the data matrix and the parameter vector, separately. The implementation process is as follows:

$$U_o = U_{oc} + IR_o + U_p,$$

$$\dot{U}_o = \dot{U}_{oc} + \dot{I}R_o + \dot{U}_p,$$

$$\dot{U}_p = \frac{I}{C_p} + \frac{-U_p}{R_p C_p}, \tag{13}$$

$$U_{o,k} - U_{o,k-1} = (I_k - I_{k-1}) \cdot R_o + T$$
$$\cdot \left(\frac{I_k}{C_p} - \frac{U_{o,k} - U_{oc,k} - I_k R_o}{R_p C_p} \right), \tag{14}$$

$$\left(1 + \frac{T}{R_p C_p} \right) U_{o,k} = U_{o,k-1} + \left(R_o + \frac{T}{C_p} + \frac{T \cdot R_o}{R_p C_p} \right)$$
$$\cdot I_k - I_{k-1} \cdot R_o + \frac{V_{oc}}{R_p C_p} T, \tag{15}$$

$$U_{o,k} = \frac{R_p C_p}{R_p C_p + T} U_{o,k-1}$$
$$+ \frac{R_o R_p C_p + T(R_o + R_p)}{R_p C_p + T} I_k$$
$$- \frac{R_o R_p C_p}{R_p C_p + T} I_{k-1}$$
$$+ \frac{U_{oc,k}}{R_p C_p + T} T, \tag{16}$$

$$U_{o,k} = \theta_1 \cdot U_{o,k-1} + \theta_2 \cdot I_k + \theta_3 \cdot I_{k-1}$$
$$+ \theta_4, \tag{17}$$

$$\boldsymbol{\varphi}_k = \begin{bmatrix} U_{o,k-1} & I_k & I_{k-1} & 1 \end{bmatrix},$$
$$\boldsymbol{\theta}_k = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{bmatrix}^T. \tag{18}$$

Formula (13) describes the fundamental circuit relationship of first-order RC battery model. In formulas (14)–(16), T is sampling interval, as T is very small, so $U_{oc,k} \approx U_{oc,k-1}$, and resistance and capacity can be deduced by (18). Besides, traditional piecewise linear function relationship between SOC and OCV cannot be identified online over the whole range of SOC. Moreover, as open circuit voltage keeps strictly monotone increasing relationship with SOC, thus we refer to the new SOC-OCV functional model expressed in formula (19) which is proposed in [18]. Besides, this function model also applies to $Li(NiCoMn)O_2$ batteries; therefore, it is reasonable to adopt this model for online identification of open circuit curve. Take the SOC definition into formula (20) and assume $s(0) = 0$. Supposing that resistance and capacitance have been obtained through RLS identification and (20) may take formula (19) then coefficient of OCV function may be obtained through recursive least square method. Formula (21) also clarifies the fundamental circuit relationship of first-order RC battery model and formula (22) is the discretization of formula (21). The detailed solution procedure is as formula (22) to formula (25). Consider

$$U_{oc} = a + b \cdot (-\ln s)^{2.1} + c \cdot s + d \cdot e^s \quad (s = \text{SOC}), \tag{19}$$

$$s(t) = s(0) + \frac{\int i\,dt}{Q} = s(0) + \frac{q(t)}{Q}, \tag{20}$$

$$U_o - IR_o = U_1 = U_{oc} + U_p,$$
$$\dot{U}_p = -\frac{U_p}{R_p C_p} + \frac{I}{C_p}, \tag{21}$$
$$\dot{U}_1 = \dot{U}_p + \dot{U}_{oc} = \dot{U}_{oc} - \frac{U_1 - U_{oc}}{R_p C_p} + \frac{I}{C_p},$$

$$\frac{U_{1,k} - U_{1,k-1}}{T} = \frac{U_{oc,k} - U_{oc,k-1}}{T} + \frac{I_k}{C_p} - \frac{U_{1,k} - U_{oc,k}}{R_p C_p}, \tag{22}$$
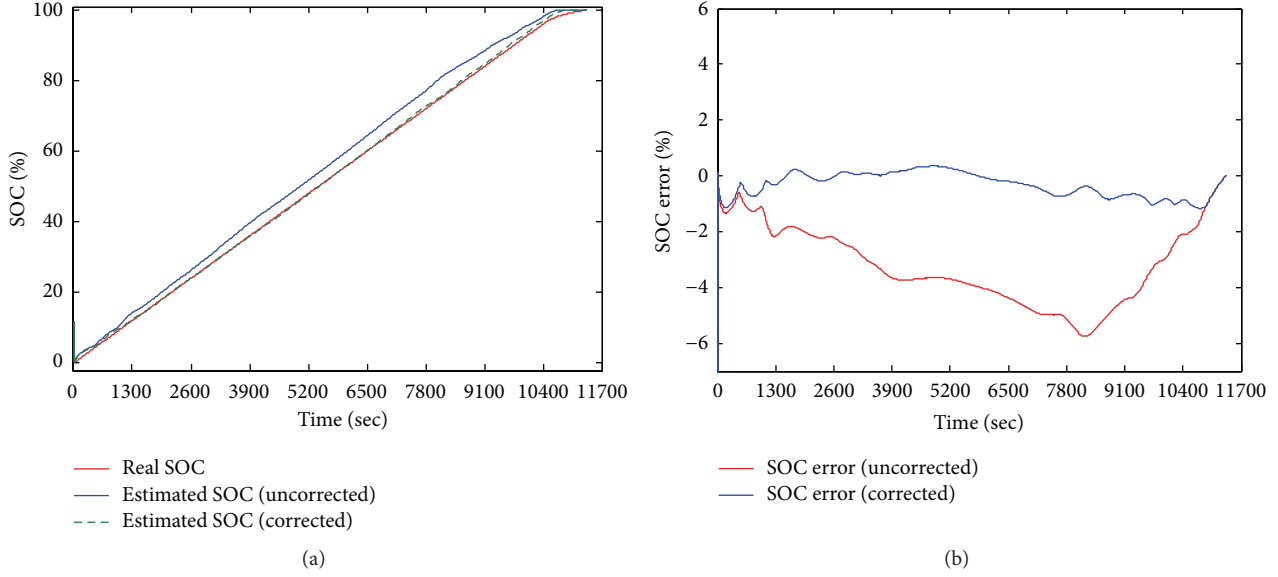
FIGURE 5: SOC estimation results before and after correction of the internal resistance.

$$U_{1,k} - U_{1,k-1} + \frac{T \cdot U_{1,k}}{R_p C_p}$$

$$= U_{oc,k} + \frac{T \cdot U_{oc,k}}{R_p C_p} - U_{oc,k-1} + T \cdot \frac{I_k}{C_p}, \tag{23}$$

$$U_{1,k}$$

$$= \frac{R_p C_p}{R_p C_p + T} \cdot U_{1,k-1} + \frac{T \cdot R_p}{R_p C_p + T} \cdot I_k + b$$

$$\cdot \left[\left(-\ln \frac{q_k}{Q} \right)^{2.1} - \left(-\ln \frac{q_{k-1}}{Q} \right)^{2.1} \right] + \frac{c}{Q} \tag{24}$$

$$\cdot (q_k - q_{k-1}) + d \cdot \left(e^{q_k/Q} - e^{q_{k-1}/Q} \right)$$

$$+ \frac{T \cdot a}{R_p C_p + T},$$

$$U_{1,k}$$

$$= \theta_1 \cdot U_{1,k-1} + \theta_2 \cdot I_k + \theta_3$$

$$\cdot \left[\left(-\ln \frac{q_k}{Q} \right)^{2.1} - \left(-\ln \frac{q_{k-1}}{Q} \right)^{2.1} \right] + \theta_4 \tag{25}$$

$$\cdot (q_k - q_{k-1}) + \theta_5 \cdot \left(e^{q_k/Q} - e^{q_{k-1}/Q} \right) + \theta_6.$$

Considering the real situation that the battery gets old but model parameters are not updated timely, simulation results before and after the parameter correction will be presented in following figures to validate the theoretical results. Comparison of two estimation results which only update the OCV-SOC curve but do not change the internal resistance of aged battery under CCCV charging mode at 25°C is shown in Figure 5. It can be observed that SOC error with uncorrected, namely, not updated, internal resistance is larger and

the maximum error reaches about 6%. However, SOC estimation error is greatly reduced and the maximum error is within 1.5% after parameter correction. On the contrary, Figure 6 presents comparison results for the same battery in which only OCV-SOC curves are amended. Maximum error reduces from 3.5% to 1% and the average error gets smaller. Comparing Figures 5 and 6, the resistance error has greater impact on SOC estimation precision in this case due to extent of change on battery resistance and OCV caused by aging is different. Simulation validation results at different temperatures can get similar conclusions and all results will not be enumerated here. All these verify the conclusion that accuracy of the model parameters can improve the precision of the SOC estimation.

4. Validation and Discussions

4.1. Hardware-in-Loop Experimental Platform. The real-time current/voltage acquisition experimental platform is illustrated in Figure 7. Arbin BTS2000 is the charger to charge/discharge the battery and the battery is laid in the thermostat to maintain constant temperature. BMS includes current/voltage acquisition board and CAN-TCP conversion card. The computer is a monitor to control the experiments, obtaining voltage/current data from sampling board, and carry out the simulation under MATLAB/Simulink. Sampling frequency of the voltage acquisition board is less than 30 ms, voltage sampling error is less than 10 mV, and current sampling precision is less than 1%. The interface between acquisition board and external communication is CAN buses and CAN-TCP conversion card is added to receive the CAN bus data frame. The S-function module is utilized to call the CAN-TCP conversion card driver program and return current/voltage data to Simulink space. Besides, real-time sync module is used for synchronizing the simulation clock with real-time clock. All modules in platform work together to

(a)

(b)
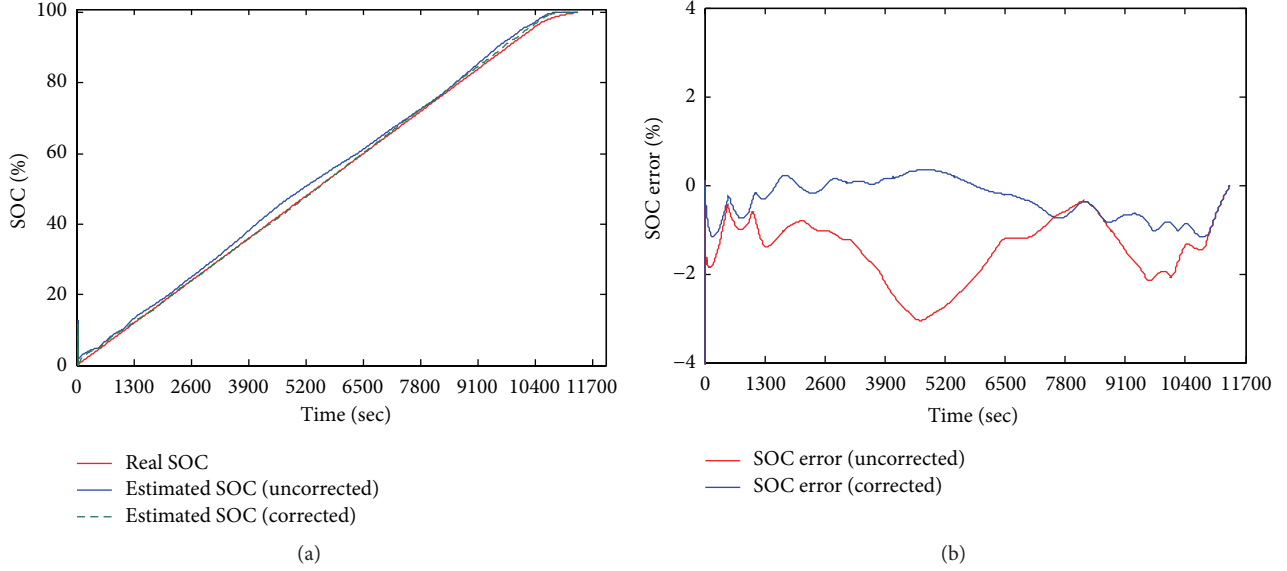
Figure 6: SOC estimation results before and after the correction of the OCV-SOC curve.
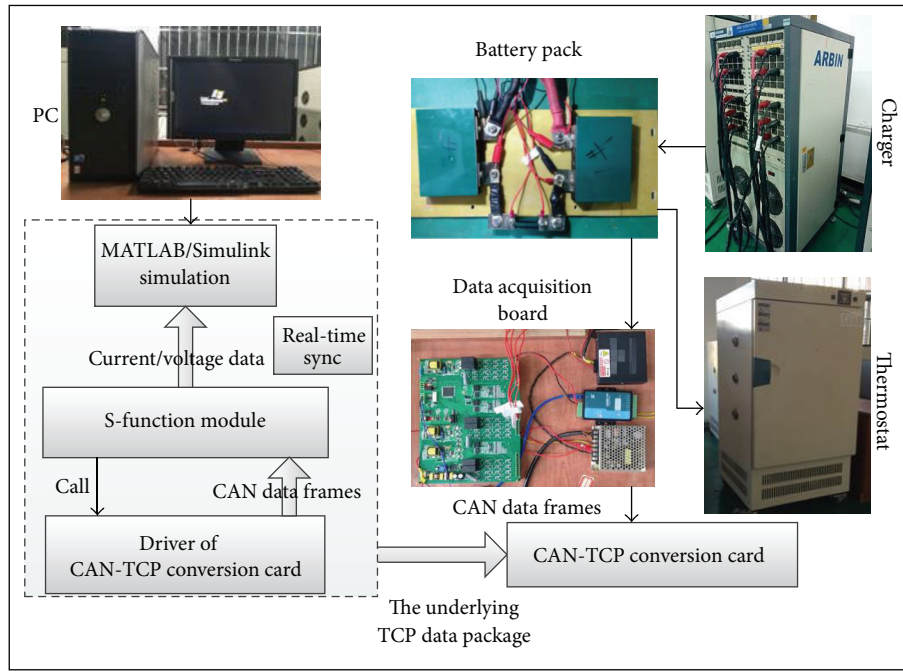
Figure 7: Experimental platform.

ensure that the current and voltage data flows into the simulation space and online real-time simulation is achieved eventually.

4.2. Experimental Results and Discussions. We take two series-connected Li(NiCoMn)O$_2$ batteries with the capacity of 28 Ah as test objects. Considering that online identification results of battery model parameters may lead to relatively larger model error when at lower SOC, which leads to SOC estimation error based on PI observer method increases at low SOC, therefore, PI observer and ampere hour integration

method are combined to obtain SOC for a battery pack. Besides, this combination also reduces computation cost to some extent. The transforming SOC value from PI observer to ampere hour integration method can be decided by the practical experience. For battery model, resistance and capacitor parameters can be obtained under a certain SOC. We can get some regular patterns that these battery parameters have fluctuation change under low SOC, and there would be some SOC value that can be defined as transforming SOC value. During discharging of the battery, when estimated SOC value based on PI observer is above 40%, we take the observer

(a)

(b)

(c)

(d)
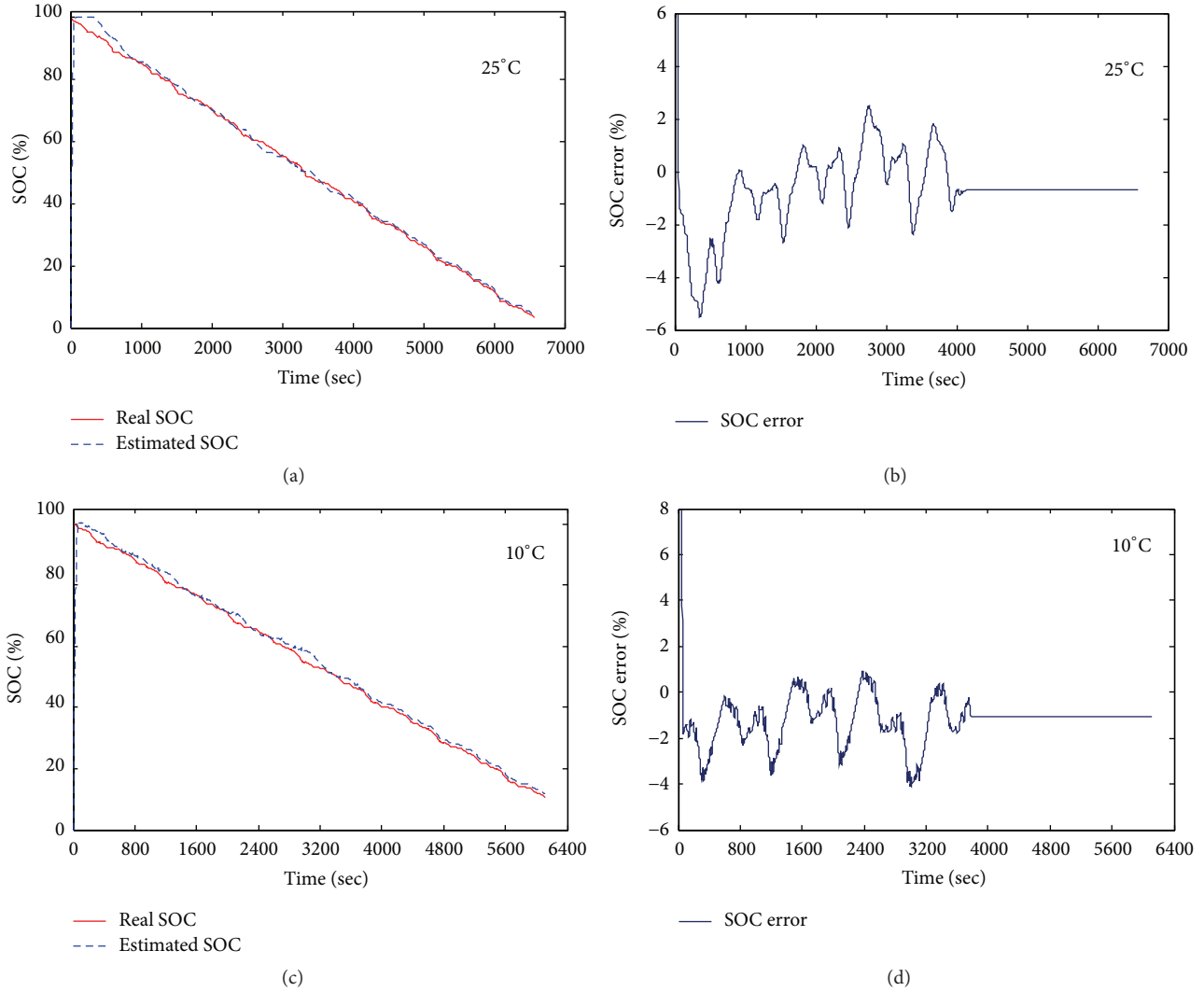
FIGURE 8: Real-time SOC estimation results under BJDST condition at different temperature.

estimation value as SOC output; however, as it is below 40%, ampere hour integration value with the initial value of 40% will be regarded as the SOC output. Figures 8 and 9 are SOC estimation results with this method under different temperature, dynamic working conditions, and initial SOC value, respectively.

Because the low temperature will have a greater impact on characteristics of batteries, we choose low temperature 10°C and room temperature 25°C to do comparison and verification experiments. Figure 8 shows SOC online estimation results under BJDST current conditions at 25°C and 10°C, respectively. Figures 8(b) and 8(d) are corresponding error curves. As shown in Figure 8, maximum estimation error is within 5% and average error is 2%. Besides, SOC error is relatively higher at 10°C because of low temperature influence. On the whole, this combination method notably increases estimation precision during the whole SOC interval. SOC calculation results with 10% and 30% divergences of SOC initial value between two cells are described in Figures 9(a) and 9(c) separately. Figures 9(b) and 9(d) are also corresponding error

curves. It is noted that this combination method proposed in the paper is also applicable to the case with inconsistencies between the cells in a battery pack. Additionally, the precision is within 3%; thus this method is reliable.

5. Conclusions

Recognizing that SOC estimation is affected by battery model parameters, temperature, operation conditions, and cell inconsistency, the estimation error sensitivity for model parameters is firstly qualified and analyzed, concluding that the SOC estimation error resulting from the internal resistance inaccuracy will be enlarged as charge/discharge current increase; the effect of polarization capacity on SOC estimation can be neglected since it is involved with transient error, and the SOC-OCV mapping is the most crucial function in SOC estimation. The SOC estimation method for a lithium-ion battery pack taking error sensitivity analysis and cell inconsistency into consideration is developed, and the recursive parameters update method is also presented. Using

(a)

(b)

(c)

(d)
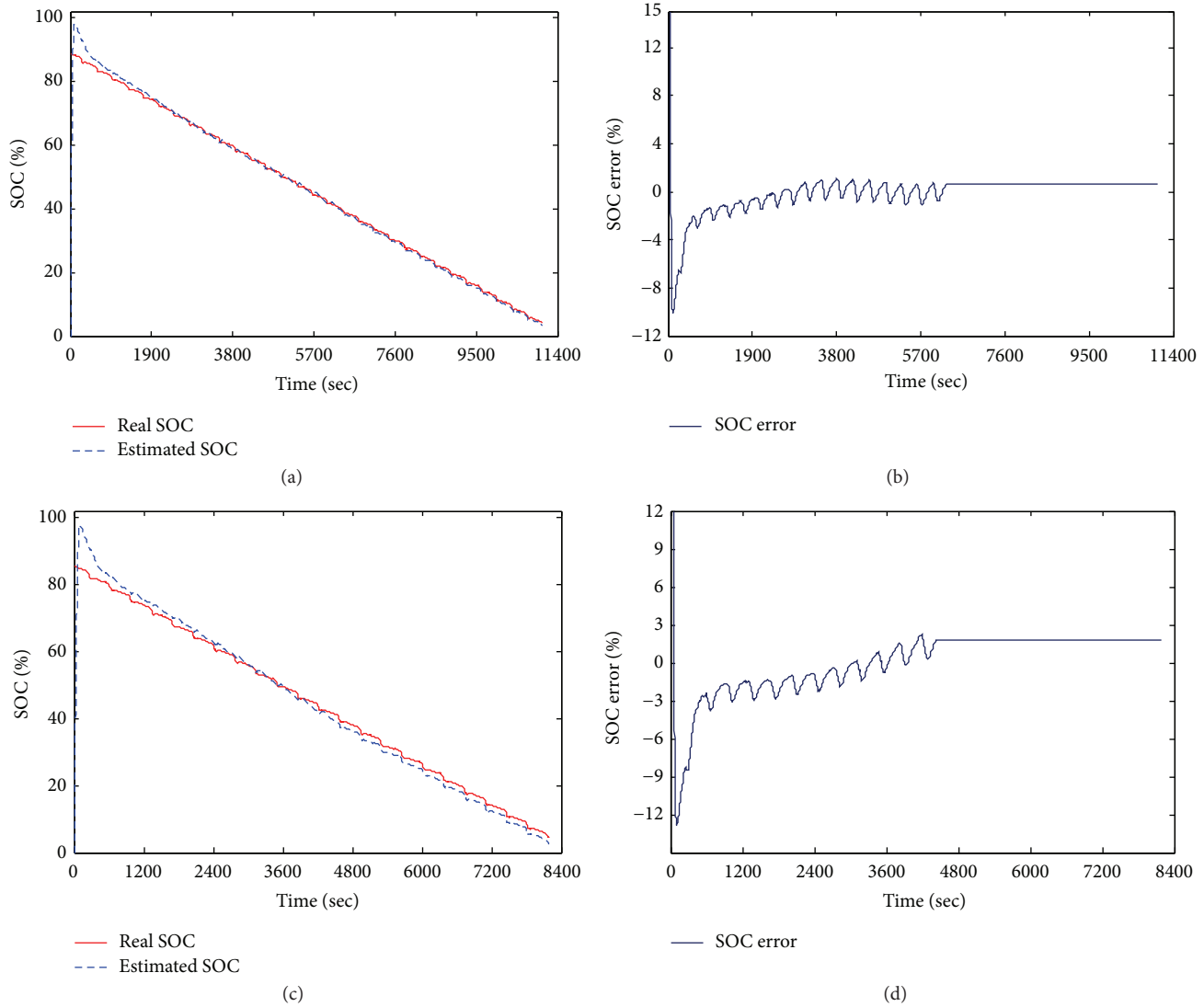
FIGURE 9: Real-time SOC estimation results under DST condition at 25°C with different SOC initial value.

the hardware-in-loop simulation platform, the experimental results at various dynamic test conditions, temperatures, and initial SOC difference between two cells demonstrate that the estimation error can be controlled within 4%, verifying the efficacy of the proposed method.

Conflict of Interests

The authors declare no conflict of interests.

Authors' Contribution

The theoretical analysis, experiments, and simulation verification were done by Ting Zhao. Besides, this work was also performed under the advisement from Jiuchun Jiang and Caiping Zhang, who also revised the paper critically. Kai Bai and Na Li gave some useful information and suggestions for this paper.

Acknowledgment

References

[1] W. Zhenpo and S. Fengchun, *Electric Vehicle Power Battery System and Its Applications*, Machinery Industry Press, Beijing, China, 2012.

[2] J. Li, J. Klee Barillas, C. Guenther, and M. A. Danzer, "A comparative study of state of charge estimation algorithms for LiFePO$_4$ batteries used in electric vehicles," *Journal of Power Sources*, vol. 230, pp. 244–250, 2013.

[3] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs—part 1. Background," *Journal of Power Sources*, vol. 134, no. 2, pp. 252–261, 2004.

[4] X. Hu, F. Sun, and Y. Zou, "Comparison between two model-based algorithms for Li-ion battery SOC estimation in electric

vehicles," *Simulation Modelling Practice and Theory*, vol. 34, pp. 1–11, 2013.

[5] W.-Y. Chang, "Estimation of the state of charge for a LFP battery using a hybrid method that combines a RBF neural network, an OLS algorithm and AGA," *International Journal of Electrical Power & Energy Systems*, vol. 53, no. 1, pp. 603–611, 2013.

[6] J. Gao and H. He, "Comparison of nonlinear filtering methods for estimating the state of charge of $Li_4Ti_5O_{12}$ lithium-ion battery," *Mathematical Problems in Engineering*, In press.

[7] H. Dai, X. Wei, Z. Sun, J. Wang, and W. Gu, "Online cell SOC estimation of Li-ion battery packs using a dual time-scale Kalman filtering for EV applications," *Applied Energy*, vol. 95, pp. 227–237, 2012.

[8] X. Liu, Y. He, and Z. Chen, "State-of-charge estimation for power Li-ion battery pack using Vmin-EKF," in *Proceedings of the 2nd International Conference on Software Engineering and Data Mining (SEDM '10)*, pp. 27–31, Chengdu, China, June 2010.

[9] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs—part 2. Modeling and identification," *Journal of Power Sources*, vol. 134, no. 2, pp. 262–276, 2004.

[10] Z. Chen, Y. Fu, and C. C. Mi, "State of charge estimation of lithium-ion batteries in electric drive vehicles using extended Kalman filtering," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 3, pp. 1020–1030, 2013.

[11] Y. Zheng, M. Ouyang, L. Lu et al., "Cell state-of-charge inconsistency estimation for $LiFePO_4$ battery pack in hybrid electric vehicles using mean-difference model," *Applied Energy*, vol. 111, pp. 571–580, 2013.

[12] Z. Cheng, J. Lv, Y. Liu, and Z. Yan, "Estimation of state of charge for lithium-ion battery based on finite difference extended Kalman filter," *Journal of Applied Mathematics*, vol. 2014, Article ID 348537, 10 pages, 2014.

[13] C. Truchot, M. Dubarry, and B. Y. Liaw, "State-of-charge estimation and uncertainty for lithium-ion battery strings," *Applied Energy*, vol. 119, pp. 218–227, 2014.

[14] M. A. Roscher, O. S. Bohlen, and D. U. Sauer, "Reliable state estimation of multicell lithium-ion battery systems," *IEEE Transactions on Energy Conversion*, vol. 26, no. 3, pp. 737–743, 2011.

[15] T. Zhao, J. Jiang, C. Zhang, L. Zheng, and F. Wen, "Error analysis of SOC estimation based on PI observer," in *Proceedings of the IEEE Transportation Electrification Conference and Expo (ITEC '14)*, pp. 1–5, Beijing, China, September 2014.

[16] O. Katsuhiko, *Modern Control Engineering*, Publishing House of Electronics Industry, Beijing, China, 2011.

[17] T. Xia, G. Xiong, and F. Li, *System Identification: Least Square Method*, Publishing House of Tsinghua University, Beijing, China, 1983.

[18] C. Zhang, L. Y. Wang, X. Li, W. Chen, G. G. Yin, and J. Jiang, "Robust and adaptive estimation of state of charge for lithium-ion batteries," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 4948–4957, 2015.

Research Article

Research on Improved Adaptive Control for Static Synchronous Compensator in Power System

Chao Zhang,[1] Aimin Zhang,[2] Hang Zhang,[1] Yingsan Geng,[1] and Yunfei Bai[1]

[1]*State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China*
[2]*School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China*

Correspondence should be addressed to Aimin Zhang; zhangam@mail.xjtu.edu.cn

This paper deals with the problems of "explosion of term," uncertain parameter in static synchronous compensator (STATCOM) system with nonlinear time-delay. An improved adaptive controller is proposed to enhance the transient stability of system states and reduce computational complexity of STATCOM control system. In contrast to backstepping control scheme in high order systems, the problem of "explosion of term" is avoided by designing dynamic surface controller. The low pass filter is included to allow a design where the model is not differentiated and thus has prevented the mathematical complexities effectively. In addition, unlike the traditional adaptive control schemes, the certainty equivalence principle is not required for estimating the uncertain parameter by system immersion and manifold invariant (I&I) adaptive control. A smooth function is added to ensure that the estimation error converges to zero in finite time. The effectiveness of the proposed controller is verified by the simulations. Compared with adaptive backstepping and proportion integration differentiation (PID), the oscillation amplitudes of transient response are reduced by nearly half, and the time of reaching steady state is shortened by at least 11%.

1. Introduction

As a member of flexible alternate current transmission system (FACTS), static synchronous compensator (STATCOM) is being widely used to compensate the reactive power with fast response time by generating or absorbing reactive power continuously in a power system [1–3]. Particularly in sustainable transportation systems, the STATCOM can also be used, such as energy storage [4], renewables integration [5], synergy between electric vehicles, and power systems [6]. It is an effective way of reducing voltage flicker emissions at the point of common coupling (PCC), removing the external fluctuations, and improving the transient stability of system states. However, as dealing with the real STATCOM control problem, the designers are unavoidably to face the difficulties involving uncertain parameter, nonlinear time delay, and complex mathematical models. The nonlinear time-delay is caused by smoothing the harmonic currents and harmonic voltage by using the digital filtering, calculating mean value of each DC voltage and transmitting to protection control unit.

The time-delay ranges from 0.01 s to 0.04 s. Furthermore, the damping coefficient is difficult to measure accurately in practice, which can be seen as uncertain parameter. Last but most important, since the mathematical modeling is more complicated, the computing complexity is improved significantly in designing STATCOM controller. Therefore, it is a hotspot to study an advanced, reliable, and low complex nonlinear control scheme for STATCOM.

Recently, adaptive backstepping control has been shown as an effective scheme to design STATCOM controllers while keeping better performance [7, 8]. This method is a vast of research, which is adapted by some intelligent methods, for example, least-squares estimation [9] and various kalman filtering approaches [10–12]. Some STATCOM controllers are designed by using backstepping technique involving the nonlinear robust control law and a new estimator to estimate the uncertain parameters [13–15]. However, in the real STATCOM system, nonlinear time-delay is an important factor that these adaptive backstepping schemes did not include. Moreover, the adaptive backstepping is

flawed with two shortcomings. Frist, the strong coupling between state variables and estimation errors existed in constructing the control Lyapunov function (CLF) for estimating the uncertain parameter. If the estimator is fixed, the estimation error will be accumulated and the error of coupling term will in turn be accumulated as running time increases [16]. Second, backstepping technique involves model differentiation in STATCOM control system and thus has suffered the "explosion of term." So the computational processes of nonlinear controller leads to high complexity.

The method named the system immersion and manifold invariant (I&I) adaptive control is developed in [17]. Since the knowledge of the CLF is not required by this method, the problem of the oscillation of states caused by the coupling between state variables and estimation error can be avoided. It has been proved that transient stability of the system can be guaranteed even if estimators reach the limit of its capacity [14]. Therefore, for nonlinear STATCOM system in adaptive law design, this I&I adaptive control can be adopted to estimate uncertain parameter by designing a smooth function to offset the estimation error. Moreover, the stability and the convergence of the adaptive law are achieved by requiring the estimation error to converge to zero in the finite time.

To overcome the "explosion of term" caused by adaptive backstepping control, the method called dynamic surface control was developed to simplify the controller design where model differentiation can be avoided in higher systems [18]. The low pass filters are included to allow a design where the model is not differentiated, at the same time, avoiding the high computing complexity caused by the "explosion of term." For a class of pure-feedback nonlinear systems, the problem explosion of complexity has been solved by using the dynamic surface technique in [19]. Moreover, semiglobal uniform ultimate boundedness of all signals is guaranteed in the closed-loop system. However, the dynamic surface control cannot be applied to nonlinear STATCOM systems with uncertain parameter; it also did not consider the problem of nonlinear time delay.

In this paper, an improved adaptive controller is proposed to improve stability of STATCOM system by simultaneously addressing the problems involving "explosion of term," uncertain parameter, and nonlinear time-delay. The low pass filters are included to allow a design where the model is not differentiated based on dynamic surface control, at the same time, avoiding the "explosion of term" caused by model differentiation. Furthermore, based on the I&I adaptive control, the existing uncertain parameter is estimated in adaptive law design. The designed adaptive law can ensure that the estimation error converges to zero in finite time. Moreover, the nonnegative time-delay function is introduced to overcome the effect of nonlinear time delay and achieve the best possible control performance. It is proved that all the state variables are globally bounded and converge to the equilibrium points by using the proposed controller. The simulation results show that transient stability and the convergence speed of the system state variables are improved effectively by the proposed controller.
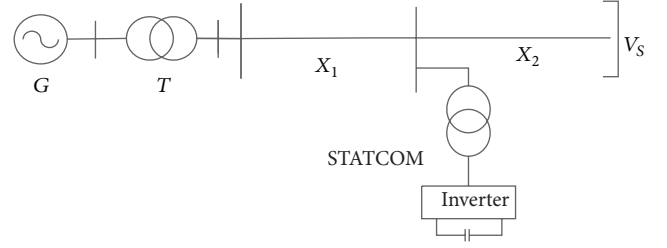
Figure 1: STATCOM single machine infinite system model.

2. System Model and Control Objective

Consider the single-machine infinite-bus system with STATCOM shown in Figure 1 [20], where G is the alternating current generator and T is the transformer.

It is clearly shown that the STATCOM is installed on the grid, which can instantly and continuously provide variable reactive power in response to voltage transients, supporting the stability of grid voltage.

Its mathematical equivalent system dynamic model can be expressed in (1) by the following nonlinear differential equations:

$$\dot{\delta} = \omega - \omega_0,$$

$$\dot{\omega} = \frac{\omega_0}{H} \left[P_m - \frac{E_q' V_s \sin \delta}{X_1 + X_2} \left(1 \right.\right.$$

$$\left. + \frac{X_1 X_2 I_q}{\sqrt{\left(X_2 E_q'\right)^2 + \left(X_1 V_s\right)^2 + 2X_1 X_2 E_q' V_S \cos \delta}} \right) \quad (1)$$

$$\left. - \frac{D}{H} \left(\omega - \omega_0\right) \right],$$

$$\dot{I}_q = \frac{1}{T_q} \left(-I_q \left(t - d\right) + I_{q0} + u_B \right),$$

where d is delay time and the three state variables are generator rotor angle δ, generator rotor angular speed ω, and reactive current I_q where transient responses will be tracked. It is noted that the vector $[\delta_0, \omega_0, I_{q0}]^T$ is the steady-state operating point. This implies that the steady operation point is the desired value or objective value [20, 21]. The parameters in (1) are expressed in Appendix A.

To simplify model (1), three state variables are redefined as $x_1 = \delta - \delta_0$, $x_2 = \omega - \omega_0$, and $x_3 = I_q - I_{q0}$. The model (1) can be rewritten as

$$\dot{x}_1 = x_2,$$

$$\dot{x}_2 = \theta x_2 + k_1 P_m$$

$$- k_2 \sin\left(\delta_0 + x_1\right)\left[1 + f\left(x_1\right)\left(x_3 + I_{q0}\right)\right],$$

$$\dot{x}_3 = k_3 \left(-x_3\left(t - d\right) + u_B\right),$$

$$(2)$$

where

$$f\left(x_1\right)$$

$$= \frac{X_1 X_2}{\sqrt{\left(X_2 E_q'\right)^2 + \left(X_1 V_S\right)^2 + 2 X_1 X_2 E_q' V_S \cos\left(x_1 + \delta_0\right)}},$$

$$k_1 = \frac{\omega_0}{H},$$

$$k_2 = \frac{\omega_0 E_q' V_S}{H\left(X_1 + X_2\right)},$$

$$(3)$$

$$k_3 = \frac{1}{T_q}.$$

The damping coefficient D cannot be measured accurately in STATCOM system and the inertia H is a constant. Therefore, the expression $\theta = -D/H$ is also the uncertain parameter.

The objective of designing STATCOM controller is to guarantee that all the state variables are globally bounded and converge to the desired points. This implies that generator rotor angle, generator rotor angular speed, and reactive current of the STATCOM can be adjusted to the equilibriums in the finite time.

3. Design of STATCOM Controller

Three are three sections to introduce our proposed controller in designing robust controller. In Section 3.1, I&I adaptive control is adopted for designing adaptive law. In Section 3.2, dynamic surface control is used for designing control law. In Section 3.3, the stability of STATCOM control system is verified.

3.1. Design of the Adaptive Law. The method I&I adaptive control can be adopted to the estimate uncertain parameter with the adaptive law design. By adopting this method, the uncertain parameter is estimated in the following steps.

Define a manifold as

$$e_\theta = \widehat{\theta} - \theta + \beta\left(x_1, x_2\right),$$

$$(4)$$

where θ is the uncertain parameter, $\widehat{\theta}$ is the estimation value of θ, and $\beta(x_1, x_2)$ is the smooth function to be designed. The derivative of (4) is

$$\dot{e}_\theta = \dot{\widehat{\theta}} + \sum_{k=1}^{2} \frac{\partial \beta}{\partial x_k} \times \dot{x}_k = \dot{\widehat{\theta}} + \frac{\partial \beta}{\partial x_1} x_2 + \frac{\partial \beta}{\partial x_2}\left(\theta x_2 + k_1 P_m\right)$$

$$- k_2 \sin\left(\delta_0 + x_1\right)\left(1 + f\left(x_1\right)\left(x_3 + I_{q0}\right)\right)\right).$$

$$(5)$$

In order to cancel the parameter-independent terms, $\dot{\widehat{\theta}}$ is designed as

$$\dot{\widehat{\theta}} = -\frac{\partial \beta}{\partial x_1} x_2 - \frac{\partial \beta}{\partial x_2}\left(\left(\widehat{\theta} + \beta\right) x_2 + k_1 P_m\right.$$

$$\left. - k_2 \sin\left(\delta_0 + x_1\right)\left(1 + f\left(x_1\right)\left(x_3 + I_{q0}\right)\right)\right).$$

$$(6)$$

Substituting (6) into (5), (5) can be rewritten as

$$\dot{e}_\theta = -\frac{\partial \beta}{\partial x_2} e_\theta x_2.$$

$$(7)$$

Lemma 1. *Define a candidate Lyapunov function (CLF)*

$$V\left(e_\theta\right) = \frac{1}{2} e_\theta^{\,2}.$$

$$(8)$$

By selecting the smooth function $\beta(x_1, x_2)$, we have $\lim_{t \to \infty} e_\theta(t) = 0$.

Proof. Theoretically, we have large flexibility in selecting $\beta(x_1, x_2)$. For simplicity, we let $\beta(x_1, x_2) = (1/2)\rho x_2^{\,2}$ with $\rho > 0$.

$$\dot{V}\left(e_\theta\right) = e_\theta \dot{e}_\theta = -\frac{\partial \beta}{\partial x_2} e_\theta^{\,2} x_2 = -e_\theta^{\,2} x_2^{\,2} \le 0.$$

$$(9)$$

Since the derivative of the CLF (8) is negative semidefinite, the manifold e_θ can converge to zero in finite time based on Lyapunov theorem. As a result, we have $\lim_{t \to \infty} e_\theta(t) = 0$ Lemma 1 holds. □

Remark 2. By using the designed smooth function, the manifold e_θ can converge to zero in finite time; that is $\lim_{t \to \infty} e_\theta(t) = 0$, based on Lyapunov stability theorem. Therefore, based on the theory of immersion and manifold invariant (I&I) adaptive control, the manifold $e_\theta(t) = \widehat{\theta} - \theta + \beta(x_1, x_2) = 0$ is invariant when $\lim_{t \to \infty} e_\theta(t) = 0$, and thus the parametric form manifold $I_e = \{(x, \widehat{\theta}) \in R^3 \times R^1 \mid \widehat{\theta} - \theta + \beta(x_1, x_2) = 0\}$ is invariant and attractive [9, 10].

3.2. Design of the Control Law. Based on dynamic surface control, we can design control law (u_B) in three steps.

Step 1. Error variables z_i $(i = 1, 2, 3)$ can be defined as the following:

$$z_1 = x_1,$$

$$z_2 = x_2 - x_2^{\,*},$$

$$z_3 = x_3 - x_3^{\,*},$$

$$(10)$$

where x_1, x_2, and x_3 are the virtual controls and x_2^* and x_3^* are the stabilizing functions. The derivative of error variables z_i $(i = 1, 2, 3)$ with (2) is

$$\dot{z}_1 = x_2,$$

$$\dot{z}_2 = \dot{x}_2 - \dot{x}_2^*$$
$$= \theta x_2 + k_1 P_m \tag{11}$$
$$- k_2 \sin(\delta_0 + x_1)(1 + f(x_1)(x_3 + I_{q0})) - \dot{x}_2^*,$$

$$\dot{z}_3 = \dot{x}_3 - \dot{x}_3^* = k_3(-x_3(t - d) + u_B) - \dot{x}_3^*.$$

Choose the first CLF as

$$V_1 = \frac{1}{2}z_1^2. \tag{12}$$

The derivative of V_1 along with (12) is

$$\dot{V}_1 = z_1 z_2 + z_1 x_2^*. \tag{13}$$

Take the stabilizing function x_2^* as

$$x_2^* = -c_1 x_1, \tag{14}$$

where c_1 is a nonnegative constant. It can be seen clearly that $\dot{V}_1 \leq 0$ if $z_2 = 0$.

Step 2. The second energy storage function with respect to Lyapunov function is

$$V_2 = \frac{1}{2}z_1^2 + \frac{1}{2}z_2^2. \tag{15}$$

The derivative of V_2 is

$$\dot{V}_2 = z_1 \dot{z}_1 + z_2 \dot{z}_2. \tag{16}$$

Substituting (11) and (14) into (16), the equation above is manipulated as

$$\dot{V}_2 = z_1 \dot{z}_1 + z_2 \dot{z}_2 = z_1 z_2 - c_1 z_1^2 + z_2(\dot{x}_2 - \dot{x}_2^*)$$
$$= z_1 z_2 - c_1 z_1^2 + z_2(\dot{x}_2 + c_1 z_2 - c_1^2 z_1) = (1 - c_1^2)$$
$$\cdot z_1 z_2 - c_1 z_1^2 + z_2[\theta x_2 + k_1 P_m \tag{17}$$
$$- k_2 \sin(\delta_0 + x_1)(1 + f(x_1)(x_3 + I_{q0})) + c_1 z_2].$$

To guarantee this second-order subsystem satisfying Lyapunov stability, the stabilizing function x_3^* must make (17) satisfy the inequality that $V_2 \leq 0$. And then, x_3^* is

$$x_3^*$$
$$= \left(\frac{(\hat{\theta} + \beta)x_2 + k_1 P_m + (1 - c_1^2)z_1 + c_1 z_2 + c_2 z_2}{k_2 f(x_1)\sin(\delta_0 + x_1)}\right) \tag{18}$$
$$- \frac{1}{f(x_1)} - I_{q0}.$$

Remark 3. It would be a tremendous expansion of terms if derivative of (18) is calculated. The problem of high computational complexity can be caused in the following control law design by using backstepping control. The dynamic surface control can be introduced to design the control law and solve the problem of "explosion of term."

The low pass filter $1/(\tau s + 1)$ is included to design control law without model differentiation, which can avoid the problem of "explosion of term" that has made other methods difficult to implement in practice.

The stabilizing function x_3^* is the output of low-pass filter, and the $\overline{x}_3$ is the input of low-pass filter. The relationship between x_3^* and $\overline{x}_3$ is

$$\tau(x_3^*)' + x_3^* = \overline{x}_3,$$
$$x_3^*(0) = \overline{x}_3(0). \tag{19}$$

From (19), we can obtain $(x_3^*)' = (\overline{x}_3 - x_3^*)/\tau$. The filtering error can be defined as

$$y = x_3^* - \overline{x}_3. \tag{20}$$

A CLF involving time-delay nonlinearity, error variables, and filtering error is designed as

$$V_3 = \frac{1}{2}z_1^2 + \frac{1}{2}z_2^2 + \frac{1}{2}z_3^2 + \frac{1}{2}y^2 + \int_{t-d}^{t} q(x(\alpha)) d\alpha, \tag{21}$$

where $q(x(t))$ is a nonnegative function. We have the derivative of (21) being

$$\dot{V}_3 = z_1 \dot{z}_1 + z_2 \dot{z}_2 + z_3 \dot{z}_3 + y \dot{y} + q(x(t))$$
$$- q(x(t - d)). \tag{22}$$

Substituting $\dot{z}_3 = k_3(-x_3(t) + u_B) - \dot{x}_3^*$, $\dot{y} = (\overline{x}_3 - x^*_3)/\tau - \dot{\overline{x}}_3 = -y/\tau + B_3$ into (22), we can obtain

$$\dot{V}_3 = z_1 \dot{z}_1 + z_2 \dot{z}_2 + z_3 \dot{z}_3 + y \dot{y} + q(x(t))$$
$$- q(x(t - d)) = (1 - c_1^2)z_1 z_2 - c_1 z_1^2 + z_2[\theta x_2$$
$$+ k_1 P_m - k_2 \sin(\delta_0 + x_1)(1 + f(x_1)(x_3 + I_{q0}))$$
$$+ c_1 z_2] + z_3(k_3(-x_3(t - d) + u_B) - \dot{x}_3^*) + y\left(\frac{-y}{\tau}\right) \tag{23}$$
$$+ B_3) + q(x(t)) - q(x(t - d)),$$

where $B_3 = -\dot{\overline{x}}_3$. Define $h(x_3(t - d)) = |k_3 x_3(t - d)|$, where $h(x_3(t-d))$ is a nonnegative time-delay function which can be compensated in the adaptive nonlinear controller design. The nonnegative function can be defined as $q(x(t)) = |k_3 z_3 x_3(t)|$, which is a reduced form of satisfying Lyapunov stability, but not the only form.

Based on Cauchy-Schwartz inequality theorem, a relational expression can be obtained

$$-z_3 h(x_3(t - d)) \leq |k_3 z_3 x_3(t - d)|$$
$$\leq |z_3| |k_3 x_3(t - d)|. \tag{24}$$

By $q(x(t)) = |k_3 z_3 x_3(t)|$, we can get $q(x(t-d)) = |k_3 z_3 x_3(t-d)|$. Substituting $q(x(t))$, $q(x(t-d))$ and (24) into (23), we can obtain

$$
\begin{aligned}
\dot{V}_3 = z_1 \dot{z}_1 &+ z_2 \dot{z}_2 + z_3 \dot{z}_3 + y\dot{y} + q(x(t)) \\
&- q(x(t-d)) \leq \left(1 - c_1^2\right) z_1 z_2 - c_1 z_1^2 + z_2 \left[\theta x_2\right. \\
&+ k_1 P_m - k_2 \sin(\delta_0 + x_1)\left(1 + f(x_1)(x_3 + I_{q0})\right) \\
&+ c_1 z_2 \left. \right] + z_3 \left(k_3 u_B - \dot{x}_3^*\right) + y\left(\frac{-y}{\tau} + B_3\right) \\
&+ \left|k_3 z_3 x_3(t)\right|.
\end{aligned}
\tag{25}
$$

The control law is designed as

$$
u_B = \frac{1}{k_3}\dot{x}_3^* - \frac{c_3}{k_3}z_3 + \lambda\left|x_3(t)\right|,
\tag{26}
$$

where $c_3 > 0$, and λ is a sign function, which is defined as $\lambda = -1$ when $z_3 > 0$ and $\lambda = 1$ when $z_3 < 0$.

3.3. Proof of System Stability

Lemma 4. *All of state variables of the closed-loop system are bounded and converge to the equilibrium point, if $V(0) \leq p$, $p > 0$.*

Proof. Let $V = (1/2)z_1^2 + (1/2)z_2^2 + (1/2)z_3^2 + (1/2)y^2 = p$, B_3 is bounded, which is denoted as M_3, and then we have $B_3^2/M_3^2 - 1 \leq 0$. Substituting (18) and (26) to (25), we can obtain

$$
\begin{aligned}
\dot{V} = z_1 \dot{z}_1 &+ z_2 \dot{z}_2 + z_3 \dot{z}_3 + y\dot{y} \\
&\leq -c_1 z_1^2 - c_2 z_2^2 - z_2 x_2 e_\theta - c_3 z_3^2 + y\left(\frac{-y}{\tau} + B_3\right).
\end{aligned}
\tag{27}
$$

Based on Cauchy-Schwartz inequality theorem, (27) can be rewritten as follows:

$$
\begin{aligned}
\dot{V} = z_1 \dot{z}_1 &+ z_2 \dot{z}_2 + z_3 \dot{z}_3 + y\dot{y} \\
&\leq -c_1 z_1^2 - c_2 z_2^2 - z_2 x_2 e_\theta - c_2 z_3^2 - \frac{y^2}{\tau} + |y||B_3| \\
&\leq -c_1 z_1^2 - c_2 z_2^2 - z_2 x_2 e_\theta - c_2 z_3^2 - \frac{y^2}{\tau} + \frac{1}{2}B_3^2 y^2 \\
&\quad + \frac{1}{2} \\
&= -c_1 z_1^2 - c_2 z_2^2 - c_3 z_3^2 - z_2 x_2 e_\theta \\
&\quad + \left(\frac{1}{2}B_3^2 - \frac{1}{\tau}\right)y^2 + \frac{1}{2}.
\end{aligned}
\tag{28}
$$

By designing $c_1 \geq r, c_2 \geq r, c_3 \geq r$, and $1/\tau \geq (1/2)M_3^2 + r$, $r \geq 0$, we can obtain

$$
\begin{aligned}
\dot{V} &\leq -rz_1^2 - rz_2^2 - rz_3^2 + \left(\frac{1}{2}B_3^2 - \frac{M_3^2}{2} - r\right)y^2 \\
&\quad + \frac{1}{2} = -2rV + \left(\frac{M_3}{2M_3}B_3^2 - \frac{M_3^2}{2}\right)y^2 + \frac{1}{2} \\
&= -2rV + \left(\frac{B_3^2}{M_3} - 1\right)\frac{M_3^2 y^2}{2} + \frac{1}{2}.
\end{aligned}
\tag{29}
$$

Substituting $r \geq 1/4p$ into (29), (29) can be rewritten as

$$
\dot{V} \leq -2\frac{1}{4p}p + \frac{1}{2} = 0.
\tag{30}
$$

From (30), we have $V(t) \leq p$, if $V(0) \leq p$, where $t \geq 0$. Lemma 4 holds. □

In addition, the convergence analysis is also given. From (29), we can get

$$
\dot{V} \leq -2rV + \frac{1}{2}.
\tag{31}
$$

Solve this differential equation as

$$
V \leq \frac{1}{4r} + \left(V(0) - \frac{1}{4r}\right)e^{-2rt}.
\tag{32}
$$

If $t \to \infty$, $V \to 1/4r$, and then we have $V \to 1/4r$, when $r \to \infty$. Furthermore, due to $1/\tau \geq (1/2)M_3^2 + 1/2 + r$, we can obtain $r \to \infty$, when $\tau \to 0$. It is an important basis for design of low pass filter $1/(\tau s + 1)$.

4. Simulation Results and Discussion

In this section, the simulation model of adaptive nonlinear controller has been established under the MATLAB/Simulink environment for nonlinear STATCOM with nonlinear time-delay. The parameters in (1) are given as follows.

Consider $H = 8$ s, $E_q' = 1.108$ pu, $P_m = 1.0$ kw, $V_s = 1$ pu, $X_1 = 0.84$ pu, $X_2 = 0.52$ pu, $T_q = 0.03$ s, $p = 2$, $c_1 = 1$, $c_2 = 1$, $c_3 = 1$, and $d = 0.02$ s, 0.04 s. The steady operation points are given as $\delta_0 = 57.1°$, $\omega_0 = 314.159$ rad/s, $I_0 = 0$, and $y(0) = 0$.

The transient responses of the nonlinear STATCOM system with time-delay are then discussed. A comparison analysis with the conventional nonlinear controller is also provided under the same conditions.

(1) Different Control Approaches. The comparison between the proposed dynamic surface method for nonlinear STAT-COM with time-delay based on system immersion and manifold invariant methodology (DSMII) approach and two approaches involving adaptive backstepping (AB) [21] and proportion integration differentiation (PID) [22] were investigated, when $d = 0.02$ s.

Figures 2(a)–2(c) show the comparison between the proposed controller and the two controllers when $d = 0.02$ s.

(a)

(b)

(c)
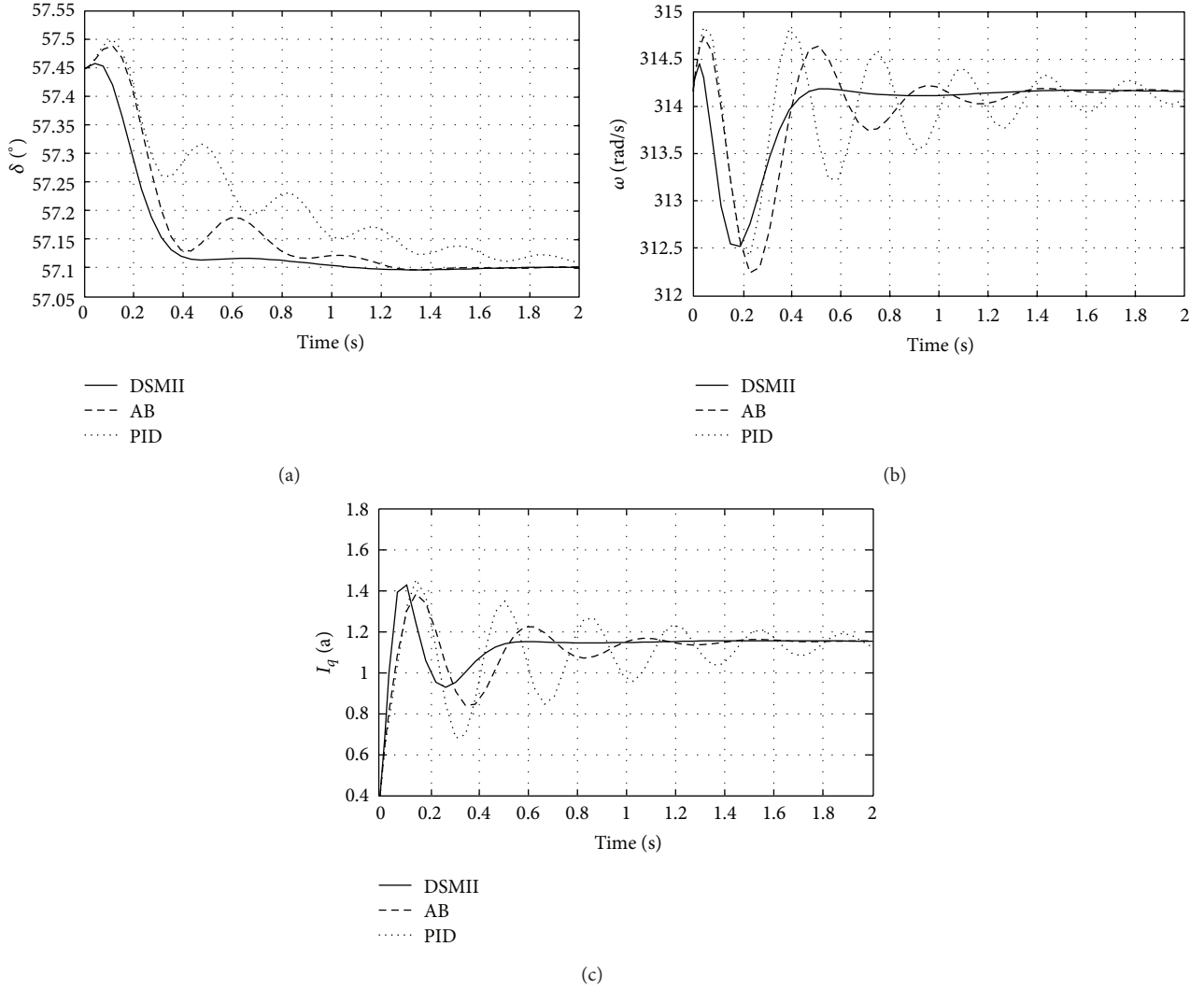
FIGURE 2: (a) Transient responses of rotor angle when $d = 0.02$ s. (b) Transient responses of rotor angular speed when $d = 0.02$ s. (c) Transient responses of reactive current when $d = 0.02$ s.

For the proposed controller, it can be clearly seen that the convergences of transient responses trajectories are achieved and the system tend to be stable state more rapidly after a very short time. Taking Figure 2(c) for example, the transient responses fluctuate fast and tend to be stable after 1.6 s or more under AB and PID. Instead, by using the proposed controller, transient responses fluctuate more smoothly and converge to stable state after 0.6 s, suggesting that the proposed controller results in better system performance.

(2) Different Time Delay. Simulations of our proposed controller are performed at $d_1 = 0.02$ and $d_1 = 0.04$, respectively.

In Figures 3(a)–3(c), we simulated the model in two different delay times to investigate its influences. The transient trajectories depart from the initial state and fluctuate strongly without an appropriate control. From the comparison between Figures 2(a)–2(c) and Figures 3(a)–3(c), all transient trajectories fluctuate faster, and system reaches the stable state more quickly when $d = 0.02$ s. Moreover, more

time is spent for the transient responses to converge to the stable state when $d = 0.04$ s. It is noted that the transient trajectories fluctuate powerfully and cannot reach steady state in finite time under AB. Consequently, the delay time d is a crucial nonlinear factor impacting the transient and steady performance of the STATCOM system. A larger d can result in a poorer robustness and worse convergence. This result is consistent with the theoretical analysis.

5. Conclusions

This paper presents an improved adaptive controller to address the problems of "explosion of term" and uncertain parameter in static synchronous compensator (STATCOM) with nonlinear time delay. Improvements are achieved in three aspects as follows.

  (1) The uncertain parameter is estimated by I&I adaptive control in designing adaptive law, which can ensure

(a)

(b)

(c)
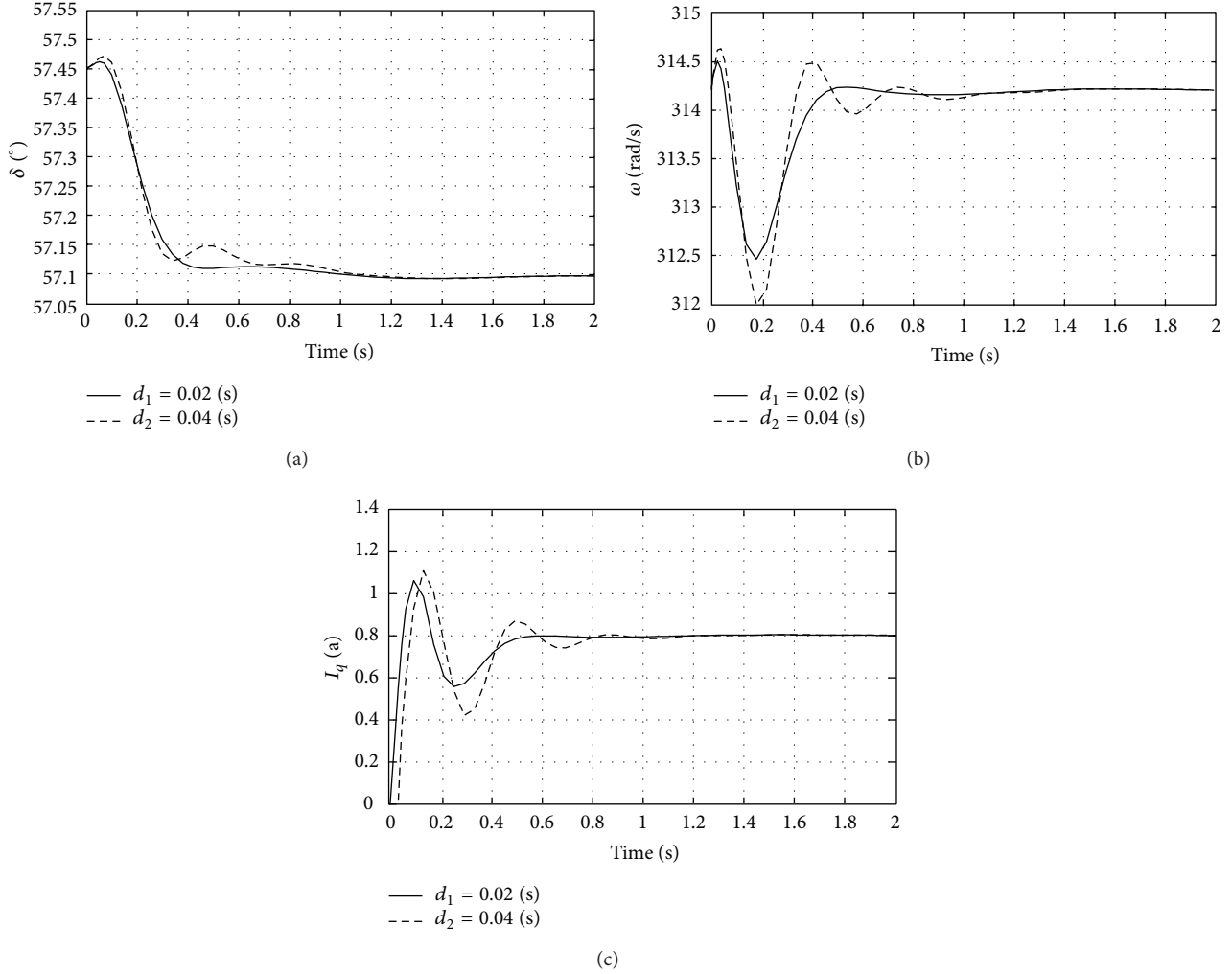
FIGURE 3: (a) Transient responses of rotor angle in different time delay. (b) Transient responses of rotor angular speed in different time delay. (c) Transient responses of reactive current in different time delay.

that the estimation error converges to zero in finite time.

(2) With regards to "explosion of term" caused by backstepping technology, a low pass filter is included to allow a design where the model is not differentiated by using dynamic surface control.

(3) Furthermore, the proposed method can add a nonnegative time-delay function to compensate the time-delay term, which can avoid the influence of time-delay term and achieve the best possible control performance.

By comparing with some conventional controller, the proposed controller has advantages in terms of enhancing transient stability and reducing computational complexity. Simulations results show that the proposed controller not only is insensitive to time-delay term but also reduces the convergence time and oscillation amplitude.

Appendices

A. Nomenclature

δ: Generator rotor angle

ω: Generator rotor angular speed

I_q: Reactive current

d: Time delay

c: Adjustable parameter

y: Output

H: Inertia constant

E_q': Transient electromotive force

P_m: Mechanical power

D: Damping coefficient

$X_{1,2}$: Equivalent impedance

u_B: Equivalence input

T_q: Time constant

V_s: Infinite bus voltage.

B. AB and PID Controller

The AB controller with the control law is

$$
\begin{aligned}
u_B = x_3 + T &\left\{ \left[\frac{1}{n_1} m_1 x_2 + \left(m_2 + \widehat{\theta} \right) \left(\theta \widehat{x}_2 + k_1 P_m \right. \right.\right. \\
&\left. - n_1 \left(1 + f\left(x_1\right)\left(x_3 + I_{q0}\right) \right) \right) + \dot{\widehat{\theta}} x_2 \Big] \\
&- \frac{1}{n_1{}^2} \left[x_2 n_2 \left(m_1 x_1 + m_2 x_2 + \widehat{\theta} x_2 + k_1 P_m \right) \right] \\
&\cdot \frac{1}{f\left(x_1\right)} + \frac{n_1 k_3 f\left(x_1\right) x_2}{k_1} \left[\frac{1}{n_1} \left(m_1 x_1 + m_2 x_2 \right.\right. \\
&\left.\left. + \widehat{\theta} x_2 + k_1 P_m \right) - 1 \right] - \left[\frac{\left(m_2 + \widehat{\theta} \right)}{n_1 f\left(x_1\right) \gamma} \right]^2 e^3 - \frac{e^3}{\gamma^2} \\
&\left. - \mu \left| k_3 x_3\left(t\right) \right| \right\}.
\end{aligned}
\tag{B.1}
$$

The parameters in simulations are given as follows.
$H = 8$ s, $E_q' = 1.108$ pu, $T_q = 0.03$ s, $c_1 = 2$, $\gamma = 0.2$, $\rho = 2$, $q_1 = 0.4$, $q_2 = 0.6$, $\sigma = 1$, $E_q' = 1.108$ pu, $V_s = 1$, $X_1 = 0.84$ pu, $X_2 = 0.52$ pu, $\delta_0 = 57.1°$, $\omega_0 = 314.159$ rad/s, and $I_0 = 0$.

The PID controller with the control law is

$$
\begin{aligned}
u_B = R\left(X\right) &+ \frac{H X_\Sigma{}^2 T_q}{E_q' V_s X_1 X_2 \sin \delta} \\
&\cdot \left(-k_l \int \Delta \omega' dt - k_p \Delta \omega' - \frac{k_d d\left(\Delta \omega'\right)}{dt} \right),
\end{aligned}
\tag{B.2}
$$

where

$$
\begin{aligned}
R\left(X\right) = &\frac{X_\Sigma{}^2 T_q}{E_q' V_s X_1 X_2 \sin \delta} \dot{P}_m - \frac{X_\Sigma{}^2 T_q D}{E_q' V_s X_1 X_2 \sin \delta} \Delta \dot{\omega} \\
&- \frac{X_\Sigma T_q \mathrm{con} \delta}{X_1 X_2 \sin \delta} \Delta \omega - I_q + I_{q0}.
\end{aligned}
\tag{B.3}
$$

The PID parameter is set as

$$
\left[k_d, k_p, k_l \right]^T = \left[1.3, 3.15, 4.2 \right]^T.
\tag{B.4}
$$

The parameters in simulations are given as follows: $H = 8$ s, $E_q' = 1.108$ pu, $T_q = 0.03$ s $P_m = 1.0$ kw, $V_s = 1$ pu, $X_1 = 0.84$ pu, $X_2 = 0.52$ pu, $\delta_0 = 57.1°$, $\omega_0 = 314.159$ rad/s, and $I_0 = 0$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. Rao, M. L. Crow, and Z. Yang, "STATCOM control for power system voltage control applications," *IEEE Transactions on Power Delivery*, vol. 15, no. 4, pp. 1311–1317, 2000.

[2] A. Astolfi, D. Karagiannis, and R. Ortega, *Nonlinear and Adaptive Control with Applications*, Springer Science & Business Media, 2007.

[3] W. Song and A. Q. Huang, "Fault-tolerant design and control strategy for cascaded H-bridge multilevel converter-based STATCOM," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 8, pp. 2700–2708, 2010.

[4] X. Hu, S. Li, H. Peng, and F. Sun, "Charging time and loss optimization for LiNMC and LiFePO$_4$ batteries based on equivalent circuit models," *Journal of Power Sources*, vol. 239, pp. 449–457, 2013.

[5] X. S. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 3, pp. 1193–1205, 2014.

[6] L. Zhang, Z. Wang, X. Hu, F. Sun, and D. G. Dorrell, "A comparative study of equivalent circuit models of ultracapacitors for electric vehicles," *Journal of Power Sources*, vol. 274, pp. 899–906, 2015.

[7] M. A. Mohd Basri, A. R. Husain, and K. A. Danapalasingam, "Intelligent adaptive backstepping control for MIMO uncertain non-linear quadrotor helicopter systems," *Transactions of the Institute of Measurement and Control*, vol. 37, no. 3, pp. 345–361, 2015.

[8] J. Zhou, C. Wen, and Y. Zhang, "Adaptive backstepping control of a class of uncertain nonlinear systems with unknown backlash-like hysteresis," *IEEE Transactions on Automatic Control*, vol. 49, no. 10, pp. 1751–1757, 2004.

[9] X.-S. Hu, F.-C. Sun, and Y. Zou, "Online model identification of lithium-ion battery for electric vehicles," *Journal of Central South University of Technology*, vol. 18, no. 5, pp. 1525–1531, 2011.

[10] X. S. Hu, F. Sun, and Y. Zou, "Comparison between two model-based algorithms for Li-ion battery SOC estimation in electric vehicles," *Simulation Modelling Practice and Theory*, vol. 34, pp. 1–11, 2013.

[11] F. Sun, X. Hu, Y. Zou, and S. Li, "Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles," *Energy*, vol. 36, no. 5, pp. 3531–3540, 2011.

[12] L. Zhang, Z. Wang, F. Sun, and D. G. Dorrell, "Online parameter identification of ultracapacitor models using the extended Kalman filter," *Energies*, vol. 7, no. 5, pp. 3204–3217, 2014.

[13] Y. Zhang, C. Wen, and Y. C. Soh, "Adaptive backstepping control design for systems with unknown high-frequency gain," *IEEE Transactions on Automatic Control*, vol. 45, no. 12, pp. 2350–2354, 2000.

[14] H. Farokhi Moghadam and N. Vasegh, "Robust PID stabilization of linear neutral time-delay systems," *International Journal*

of Computers Communications & Control, vol. 9, no. 2, pp. 201–208, 2014.

[15] L. Zhang, A. Zhang, Z. Ren, G. Li, C. Zhang, and J. Han, "Hybrid adaptive robust control of static var compensator in power systems," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 12, pp. 1707–1723, 2014.

[16] D. Karagiannis and A. Astolfi, "Nonlinear adaptive control of systems in feedback form: an alternative to adaptive backstepping," *Systems & Control Letters*, vol. 57, no. 9, pp. 733–739, 2008.

[17] A. Astolfi and R. Ortega, "Immersion and invariance: a new tool for stabilization and adaptive control of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 590–606, 2003.

[18] D. Swaroop, J. K. Hedrick, P. P. Yip, and J. C. Gerdes, "Dynamic surface control for a class of nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 45, no. 10, pp. 1893–1899, 2000.

[19] S. J. Yoo, J. B. Park, and Y. H. Choi, "Adaptive dynamic surface control for stabilization of parametric strict-feedback nonlinear systems with unknown time delays," *IEEE Transactions on Automatic Control*, vol. 52, no. 12, pp. 2360–2364, 2007.

[20] N. Jiang, S. Li, T. Liu, and X. Dong, "Nonlinear large disturbance attenuation controller design for the power systems with STATCOM," *Applied Mathematics and Computation*, vol. 219, no. 20, pp. 10378–10386, 2013.

[21] W. L. Li, Y. W. Jing, X. P. Liu, and B. Wang, "Nonlinear robust control based on adaptive backstepping design for STATCOM," *Journal of Northeastern University (Natural Science)*, vol. 24, pp. 221–224, 2003.

[22] C. Zhang, Z. Aimin, Z. Hang et al., "An advanced adaptive backstepping control method for STATCOM," in *Proceedings of the 26th Chinese Control And Decision Conference (CCDC '14)*, pp. 1822–1827, IEEE, Changsha, China, May 2014.

Research Article

Intelligent Ramp Control for Incident Response Using Dyna-Q Architecture

Chao Lu,[1,2] Yanan Zhao,[1] and Jianwei Gong[1]

[1]*School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China*
[2]*Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK*

Correspondence should be addressed to Chao Lu; tscl.lu@gmail.com

Reinforcement learning (RL) has shown great potential for motorway ramp control, especially under the congestion caused by incidents. However, existing applications limited to single-agent tasks and based on Q-learning have inherent drawbacks for dealing with coordinated ramp control problems. For solving these problems, a Dyna-Q based multiagent reinforcement learning (MARL) system named Dyna-MARL has been developed in this paper. Dyna-Q is an extension of Q-learning, which combines model-free and model-based methods to obtain benefits from both sides. The performance of Dyna-MARL is tested in a simulated motorway segment in the UK with the real traffic data collected from AM peak hours. The test results compared with Isolated RL and noncontrolled situations show that Dyna-MARL can achieve a superior performance on improving the traffic operation with respect to increasing total throughput, reducing total travel time and CO_2 emission. Moreover, with a suitable coordination strategy, Dyna-MARL can maintain a highly equitable motorway system by balancing the travel time of road users from different on-ramps.

1. Introduction

Traffic congestion occurs when the traffic demand for a road network approaches or exceeds its available road capacity. Even slight losses of the balance between demand and capacity on motorways can lead to long travel delays, high energy consumptions, and severe environmental problems. Therefore, how to alleviate traffic congestion and maintain the demand-capacity balance has become one of the main concerns of the transport community. To this end, a number of traffic control devices, such as variable speed limit (VSL), variable message sign (VMS), and ramp control systems, are developed under the umbrella of intelligent transportation systems (ITS). Among these advanced systems, ramp control (also known as ramp metering) has been widely used and proved to be an effective control method for different kinds of congestion on motorways [1].

Generally, traffic congestion can be classified into two categories: recurrent congestion and nonrecurrent congestion. Recurrent congestion is caused by the daily traffic operation with temporarily increased traffic demand in peak hours [2]. Considering the daily peak traffic on motorways, recurrent congestion is the main concern of many existing ramp control systems. For instance, fixed-time systems (also known as pretimed systems) use historical data collected from daily peak hours to generate control strategies offline and trigger these strategies at fixed times (e.g., morning or evening peak hours) of each day [1]. Local traffic-responsive systems such as demand-capacity method, ALINEA [3], and its variations [4] can respond to the real-time traffic and keep the outflow or road density of the motorway mainline close to some target value (e.g., road capacity or critical density). Usually, these target values should be defined in advance according to the so-called fundamental diagram which is derived from the daily traffic data. To deal with network-wide problems, traffic-responsive systems have been extended to coordinated ramp control systems, such as Flow [5], System Wide Adaptive Ramp Metering (SWARM) [6], and Zone algorithms [7]. Similar to local traffic-responsive systems, these coordinated systems also attempt to make

the outflow of motorway mainline approach a predetermined target value which is usually the road capacity. Another group of systems focuses on formulating different control scenarios as optimisation problems and using optimal control techniques (e.g., model predictive control) to solve them. The purpose of these systems is to maximise or minimise an objective function, not to achieve some predefined target value. Examples of these systems can be found in [8–12], where macroscopic traffic flow models were combined with control systems to formulate optimal control problems.

Although the aforementioned systems have shown their effectiveness in different scenarios, recurrent congestion is still the main focus of these systems and a component that can deal with nonrecurrent congestion is not included in these systems. Unlike recurrent congestion caused by the increased traffic demand in peak hours, nonrecurrent congestion is mainly induced by incidents, and thus, it is usually referred to as incident-induced congestion [2, 13]. Traffic incidents are nonrecurrent events such as road accidents, vehicle breakdown, and unexpected obstacles that may block one or more lanes of the motorway mainline. The temporary lane blockage will interrupt the normal operation of traffic flow and lead to a rapid reduction of road capacity [14]. In this case, fixed-time and simple traffic-responsive systems, which are dependent on the information collected from daily traffic operation or a predefined target value, are not applicable. Therefore, more sophisticated systems that can respond to incidents are required. During the last decades, a series of such kinds of ramp control systems have been designed, most of which are based on optimisation techniques. For example, an optimal control structure using a simple macroscopic traffic flow model was proposed in [15] to deal with incident-induced congestion. A more complex system with consideration of dynamic incident duration was developed in [16] which can be solved by the linear programming technique. In the research presented in [17, 18], both lane-changing and queuing behaviour during the incident were incorporated into a modelling structure and solved by a stochastic optimal control system. Although these systems are based on different technologies, they all need a model to predict traffic conditions and use these predictions to accomplish the control process.

Model-based methods usually have poor adaptability when the mismatch between simulation models and the real controlled environment emerges [19–21]. To overcome this limitation, another optimisation-based method, reinforcement learning (RL), was introduced to the ramp control area. This method is based on the Markov decision process (MDP) and dynamic programming (DP), which can approximately solve the optimisation problem through continuous learning without any models. The first ramp control system using RL to solve incident-induced problems was developed in [19, 22]. The basic RL algorithm named Q-learning was adopted by this system to alleviate traffic congestion caused by incidents. After this work, several Q-learning systems considering both local (e.g., [23, 24]) and coordinated (e.g., [25, 26]) control problems were proposed. However, Q-learning can only learn from real interactions with the traffic operation and cannot make full use of historical data (or models). Because

of this limitation, Q-learning usually has a low learning speed and needs a great number of trials to obtain the best control strategy in some complex scenarios, such as incident-induced congestion [27]. This problem is even worse in the coordinated ramp control problems with exponentially increased state and action spaces, which will lead to the so-called "curse of dimensionality" [28]. One solution to speed up the learning process and deal with incidents efficiently has been proposed in our previous work [27, 29]. This system used the Dyna-Q architecture to combine model-free Q-learning with a model-based method and can be used to accomplish single-agent tasks.

In this paper, the previous single-agent system is extended to a multiagent case that can deal with a network-wide problem with multiple ramp controllers. We refer this system to Dyna-MARL which adopts a multiagent RL (MARL) strategy based on Dyna-Q architecture. The rest of this paper is organised as follows. Section 2 briefly introduces the basic knowledge of RL including single-agent and multiagent cases. The architecture of Dyna-MARL is described in Section 3. After that, Sections 4 and 5 give the detailed description of the models, elements, and related algorithm of Dyna-MARL. The simulation experiments and relevant results are discussed in Section 6. Section 7 finally gives some conclusions and introduces the future work.

2. Reinforcement Learning

RL is a subclass of machine learning. In the following subsections, two kinds of RL problems, namely, single-agent and multiagent RL, will be briefly introduced.

2.1. Single-Agent RL. The problem of single-agent RL is usually defined as an MDP that can be represented by a tuple (S, P, R, C) [30]. S is the state space used to describe the external environment. C is the control action set containing executable actions of the agent. P is the state transition probability. For state pair $(s, s' \in S)$, $P^c(s, s')$ represents the probability of reaching state s' after executing action c at state s. $R : S \times C \rightarrow \mathbb{R}$ is the reward function. $R(s, c)$ denotes the immediate reward after taking action c at state s. Based on these definitions, Q value is defined for each state-action pair (s, c) and shown below:

$$
\begin{aligned}
&Q^\pi (s, c) \\
&= E \left\{ \sum_{n=0}^{\infty} \gamma^n R \left(s^{k+n+1}, c^{k+n+1} \right) \mid s^k = s, \ c^k = c \right\},
\end{aligned} \tag{1}
$$

where k is the time index and n is the number of time steps. $s^k \in S$ and $c^k \in C$ are the environment state and executed control action at time step k, respectively. $\gamma \in [0, 1]$ is the discount factor which indicates the importance of the following predicted rewards. For γ^n, n is the power. π is the policy corresponding to a sequence of actions. The optimal policy can be obtained by maximising the Q value.

The most widely used algorithm in literature for estimating the maximum Q value is Q-learning [31]. By using

the updating equation as given below, Q-learning can maximise Q value for each state-action pair:

$$Q^{k+1}\left(s^k, c^k\right) = Q^k\left(s^k, c^k\right) + \alpha\left[R^k\left(s^k, c^k\right)\right.$$
$$\left. + \gamma \max_{c^{k+1}} Q^k\left(s^{k+1}, c^{k+1}\right) - Q^k\left(s^k, c^k\right)\right], \tag{2}$$

where $Q^{k+1}(s^k, c^k)$ and $Q^k(s^k, c^k)$ are the Q value for state-action pair (s^k, c^k) at the $k+1$th step and kth step, respectively, and $Q^k(s^{k+1}, c^{k+1})$ is the Q value for the state-action pair (s^{k+1}, c^{k+1}) at the kth step. $\alpha \in [0, 1]$ is the learning rate. γ and α can be regulated according to different problems.

2.2. Multiagent Scenarios.

In multiagent scenarios, an MDP for single-agent case can be extended to a stochastic game (SG) or Markov game, in which a group of agents try to obtain some equilibrium solutions through coordination or competition [28].

In the absence of competition, all agents involved in a game have a common goal to maximise the global Q value, which forms a coordinated MARL problem. In this case, the policy optimisation is determined by actions executed by all agents.

For solving a coordinated MARL problem, the update equation (2) for Q-learning can be easily extended to represent the global Q value update [28]:

$$Q^{k+1}\left(s^k, c_1^k, \ldots, c_n^k\right) = Q^k\left(s^k, c_1^k, \ldots, c_n^k\right)$$
$$+ \alpha\left[R^k\left(s^k, c_1^k, \ldots, c_n^k\right)\right.$$
$$+ \gamma \max_{c_1^{k+1}, \ldots, c_n^{k+1}} Q^k\left(s^{k+1}, c_1^{k+1}, \ldots, c_n^{k+1}\right)$$
$$\left. - Q^k\left(s^k, c_1^k, \ldots, c_n^k\right)\right]. \tag{3}$$

The only difference with (2) is that Q and R in (3) relate to n actions $c_1, \ldots, c_n$ executed by n agents rather than to a single action c.

2.3. Solutions for Coordinated MARL.

It can be seen from (3) that as the number of agents grows, combinations of actions and the resultant computational complexity are increased exponentially, which may make the problem unsolvable within a required time limit [28]. Therefore, a commonly used method is to decompose the global Q value to several local Q values, each of which can be maximised by a few relevant agents rather than all agents [32]. Based on this distributed method, several strategies have been proposed. In [28], these strategies fall into three categories including coordination-based, coordination-free, and indirect coordination strategies.

Coordination-based strategies need local Q values to be updated according to actions executed by all relevant agents

(named joint actions) at each time step [28]. The decision making process of each agent is based on the information received from all other related agents with sufficient communication. This will complicate the problem. On the other hand, coordination-free (or independent) strategies, such as distributed Q-leaning algorithm, make each agent update the corresponding local Q values based on its own actions [33]. Therefore, each agent makes its decisions independently without increasing computational complexity. However, this computational efficiency is at the expense of nonguaranteed convergence [32]. Indirect coordination strategies try to find a balance between the above two methods. By applying indirect strategies, each agent can maintain models for its cooperative partners and update local Q values without knowing all the information of other agents at each step [28]. Based on high-quality models, this method can reduce the problem complexity and guarantee convergence with limited coordination.

3. Dyna-Q Based Indirect Coordination Strategy

Because of the benefits introduced in the above section, the indirect coordination strategy has been applied in [34] for solving urban traffic control problems. In their work, each agent maintains a model for estimating the action selection probability of its neighbours and uses this information to optimise control strategies. In this paper, we extend this method to motorway systems by applying Dyna-Q architecture.

Under the Dyna-Q architecture, a modified macroscopic flow model named asymmetric cell transmission model (ACTM) and Q-learning algorithm are combined together to deal with coordinated MARL problems. In this section, the application of Dyna-Q will be introduced.

3.1. Dyna-Q Architecture.

Dyna-Q architecture is an extension of standard Q-learning that integrates planning, acting, and learning together [30]. Unlike Q-learning which learns from the real experience without a model, Dyna-Q learns a model and uses this model to guide the agent [35]. After capturing the real experience, two loops run to learn optimal policies that can obtain the maximum Q value in Dyna-Q architecture (see Figure 1).

In loop I, direct RL is the standard Q-leaning process that can be used to interact with the real external environment. Loop II contains two main tasks: (1) model learning is used to improve the model accuracy through obtaining new knowledge from real experience; (2) planning is the same process of direct RL except that it is using the experience generated by a model. Acting is the action execution process.

Applying a model, the agent can predict reactions of its external environment and other agents before executing a specific action, which provides an opportunity for agent to update Q value before receiving the real feedback. Simultaneously, direct RL is running to update the Q value through the real interaction. Therefore, optimal policy is learned through both real experience and predictions. By using this strategy
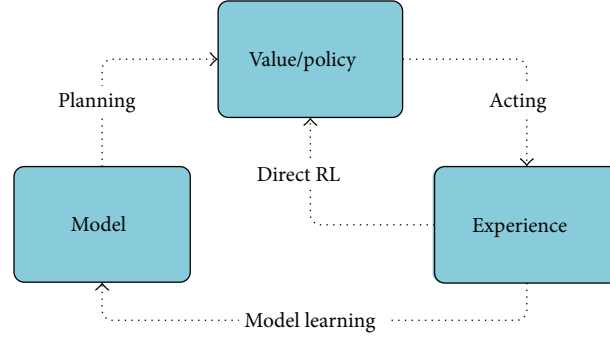
FIGURE 1: Dyna-Q architecture.
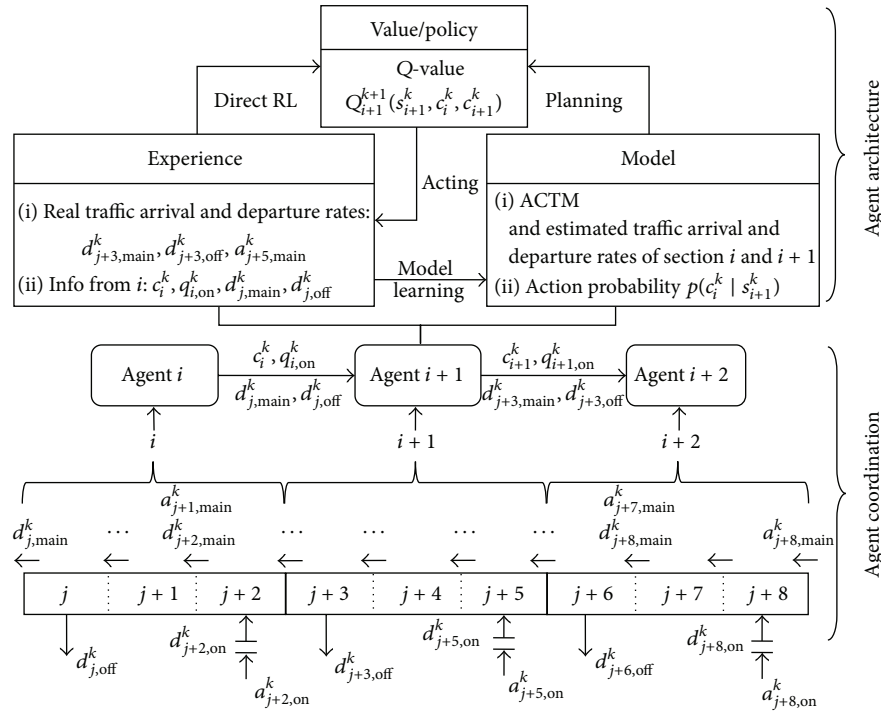
FIGURE 2: System architecture.

Dyna-Q can learn faster than Q-learning in many situations [30].

Although a model is maintained in the Dyna-Q architecture, the whole system is different from the model-based control method such as model predictive control (MPC). The model in Dyna-Q architecture is a complementary component, which is used to speed up the learning process and simplify the coordination of agents. The optimal control actions are learnt from both real and simulated experience. Without models, the Dyna-Q architecture is equivalent to the Q-learning technique and can still work as a model-free system. MPC, on the other hand, is dependent on the model, which means it cannot work without models. Therefore, Dyna-Q can be considered as a combination of model-free and model-based method [27].

3.2. System Architecture. Each agent in the motorway control system is designed on the basis of Dyna-Q architecture which controls one prespecified motorway section.

A simplified motorway segment is shown in Figure 2 for analysis. This segment contains three motorway sections ($i, i + 1, i + 2$) with detectors located at boundaries. Each motorway section is divided into a number of cells ($j, j + 1, \ldots, j + 8$) according to its layout and geometric features. Generally, three kinds of cells exist in the motorway, such as on-ramp cells that are linked with on-ramps ($j + 2, j + 5, j + 8$), off-ramp cells linked with off-ramps ($j, j + 1, j + 6$), and normal cells ($j + 1, j + 4, j + 7$). In this paper, we define that each motorway section can have at most one on-ramp cell.

The typical Dyna-Q architecture presented in Figure 2 is detailed for each agent here. Take agent $i + 1$, for example;
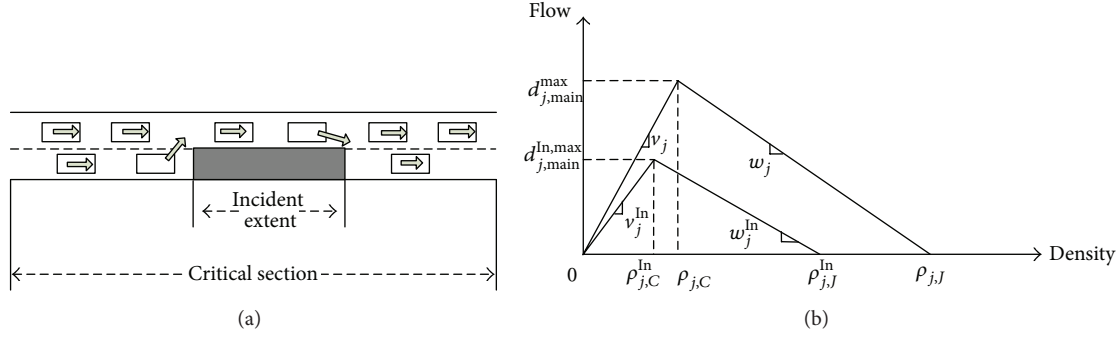
FIGURE 3: Fundamental diagram during the incident.

experience consists of traffic arrival and departure rates observed from the detectors of motorway section $i + 1$, as well as the information received from agent i, which is applied to improve models. In the model component, two models are maintained. An asymmetric cell transmission model (ACTM) with estimated traffic arrival and departure rates is used to simulate the traffic flow dynamics in relevant motorway sections. A probability model of action selection of agent i at the current state is updated for further planning process.

To reduce the complexity of MARL, like many real applications, some conventions are used to restrict the action selection of an agent [28]. Specifically, in our design, each agent only communicates with its spatial neighbours. For instance, agent $i + 1$ receives the control action and traffic information from agent i and sends its own information to agent $i + 2$. For the case shown in Figure 2, we assume motorway section i is the critical section where an incident occurs. In this situation, agent i plays a more important role than other agents for dealing with incidents. Agent i can be considered as the chief controller that makes decisions according to its own knowledge about the traffic and incident situations. Other agents should regulate their control policies based on the reaction of agent i.

Therefore, two Q values are defined for two kinds of agents. If motorway section i is the critical section, the Q value of agent i is only related to its own state and action space, which can be updated by the same equation denoted by (2).

If motorway section i is the normal section without incidents, the Q value of agent i can be calculated by

$$Q_i^{k+1}\left(s_i^k, c_{i-1}^k, c_i^k\right) = Q_i^k\left(s_i^k, c_{i-1}^k, c_i^k\right)$$

$$+ \alpha \left[R_i^k\left(s_i^k, c_{i-1}^k, c_i^k\right) + \gamma \max_{c_i^{k+1}} \sum_{c_{i-1}^{k+1}} p\left(c_{i-1}^{k+1} \mid s_i^{k+1}\right) \right.$$

$$\left. \cdot Q_i^k\left(s_i^{k+1}, c_{i-1}^{k+1}, c_i^{k+1}\right) - Q_i^k\left(s_i^k, c_{i-1}^k, c_i^k\right) \right], \tag{4}$$

$$p\left(c_{i-1}^{k+1} \mid s_i^{k+1}\right) = \frac{\operatorname{count}\left(s_i^{k+1}, c_{i-1}^{k+1}\right)}{\sum_{c_{i-1} \in C_{i-1}} \operatorname{count}\left(s_i^{k+1}, c_{i-1}\right)},$$

where $R_i^k(s_i^k, c_{i-1}^k, c_i^k)$ is the immediate reward obtained by agent i at time step k, when actions c_{i-1}^k, c_i^k are actions executed by agent $i - 1$ and i. Similarly, $Q_i^{k+1}(s_i^k, c_{i-1}^k, c_i^k)$ and $Q_i^k(s_i^k, c_{i-1}^k, c_i^k)$ are the Q values for agent i at step $k + 1$ and step k, respectively. C_{i-1} is the action set of agent $i - 1$. count$(s_i^{k+1}, c_{i-1}^{k+1})$ returns the number of visits for state-action pair $(s_i^{k+1}, c_{i-1}^{k+1})$. Thus, $p(c_{i-1}^{k+1} \mid s_i^{k+1})$ is the probability for agent $i - 1$ selecting action c_{i-1}^{k+1} at state s_i^{k+1}. Models and the related symbols shown in Figure 2 will be specified in the flowing section.

4. Modified Asymmetric Cell Transmission Model

A first-order macroscopic traffic flow model named asymmetric cell transmission model (ACTM) is applied as one of the models in the Dyna-Q architecture. This model is derived from the widely used cell transmission model (CTM) [36] and has been used for ramp control problems [11, 37]. In this paper, we modify ACTM to incorporate the traffic dynamics under incident conditions.

4.1. Traffic Dynamics during the Incident. As shown in Figure 3(a), when an incident happens in the critical section, one or more lanes of the motorway will be blocked according to the incident extent. Because of the lane blockage, incident may reduce the normal road capacity and spatial storage space, which will produce a new relationship between traffic flow and road density, that is, fundamental diagram presented in Figure 3(b). As suggested by [38], additional parameters can be used to regulate fundamental diagram for incident situations. We introduce three parameters ($\lambda_1, \lambda_2, \lambda_3 \in [0, 1]$) to reflect this new dynamics. These three parameters are defined as $\lambda_1 = v_j^{\text{In}}/v_j$, $\lambda_2 = w_j^{\text{In}}/w_j$, and $\lambda_3 = d_{j,\text{main}}^{\text{In,max}}/d_{j,\text{main}}^{\text{max}}$. v_j and w_j are the free flow speed and congestion wave speed of cell j. $d_{j,\text{main}}^{\text{max}}$ is the maximum departure flow of cell j. v_j^{In}, w_j^{In}, and $d_{j,\text{main}}^{\text{In,max}}$ are these three variables during the incident. $\rho_{j,C}$ and $\rho_{j,C}^{\text{In}}$ are the critical densities for normal and incident situations. $\rho_{j,J}$ and $\rho_{j,J}^{\text{In}}$ are the jam densities for normal and incident situations.

4.2. Modified ACTM. Given three incident-related parameters, the traffic dynamics in each cell can be derived from

the fundamental diagram illustrated in Figure 3(b) and represented by the following equations.

Departure rates of the mainline and on-ramp:

$$
d_{j,\text{main}}^{k} = \min \left\{ \lambda_1 \cdot \frac{v_j}{l_j} \cdot \left(q_{j,\text{main}}^{k} + \theta_j \cdot d_{j,\text{on}}^{k} \cdot \Delta t - d_{j,\text{off}}^{k} \cdot \Delta t \right); \ \lambda_2 \cdot \frac{w_{j-1}}{l_j} \cdot \left(q_{j-1,\text{main}}^{\max} - q_{j-1,\text{main}}^{k} - \theta_j \cdot d_{j-1,\text{on}}^{k} \cdot \Delta t \right); \ \lambda_3 \right.
$$

$$
\left. \cdot d_{j,\text{main}}^{k,\max} \right\},
$$

$$
d_{j,\text{on}}^{k} =
\begin{cases}
\min \left\{ \dfrac{\left(q_{j,\text{main}}^{k} + a_{j,\text{on}}^{k} \cdot \Delta t \right)}{\Delta t}; \ \eta_j \cdot \dfrac{\left(q_{j,\text{main}}^{\max} - q_{j,\text{main}}^{k} \right)}{\Delta t}; \ \dfrac{c_i^{k}}{\Delta t} \right\}, & \text{if } j \text{ is metered on-ramp cell,} \\[4mm]
\min \left\{ \dfrac{\left(q_{j,\text{main}}^{k} + a_{j,\text{on}}^{k} \cdot \Delta t \right)}{\Delta t}; \ \eta_j \cdot \dfrac{\left(q_{j,\text{main}}^{\max} - q_{j,\text{main}}^{k} \right)}{\Delta t} \right\}, & \text{if } j \text{ is unmetered on-ramp cell.}
\end{cases}
$$

(5)

Conservation of the mainline and on-ramp:

$$
q_{j,\text{main}}^{k+1} = q_{j,\text{main}}^{k} + \Delta t
$$

$$
\cdot \left(a_{j,\text{main}}^{k} + d_{j,\text{on}}^{k} - d_{j,\text{main}}^{k} - d_{j,\text{off}}^{k} \right),
$$

$$
q_{j,\text{on}}^{k+1} = q_{j,\text{on}}^{k} + \Delta t \cdot \left(a_{j,\text{on}}^{k} - d_{j,\text{on}}^{k} \right),
$$

(6)

where $a_{j,\text{main}}^{k}$ and $d_{j,\text{main}}^{k}$ are the mainline arrival and departure rates for the cell j at step k. $a_{j,\text{on}}^{k}$ and $d_{j,\text{on}}^{k}$ are the on-ramp arrival and departure rates in cell j at step k. $d_{j,\text{off}}^{k}$ is the off-ramp departure rate for cell j at step k (if cell j is not an off-ramp cell, $d_{j,\text{off}}^{k} = 0$). $q_{j,\text{main}}^{k}$ represent the number of vehicles on the mainline of cell j at step k. $q_{j,\text{main}}^{\max}$ is the maximum number of this value limited by the mainline space of cell j. Similarly, $q_{j,\text{on}}^{k}$ and $q_{j,\text{on}}^{\max}$ denote the current (at step k) and maximum number of vehicles in the on-ramp of cell j, respectively. Δt (min) is the time duration between each two time steps. c_i^{k} is the metering rate for the on-ramp cell of the ith motorway section at step k. $\eta_j \in [0, 1]$ is the flow allocation parameter of cell j. $\theta_j \in [0, 1]$ is the flow blending parameter of traffic flow from the on-ramp to the mainline of cell j. The unit of all the arrival and departure rates is modified to veh/min in this study.

For motorway section i with J cells, the number of vehicles in the mainline can be calculated by $q_{i,\text{main}}^{k} = \sum_{j=1}^{J} q_{j,\text{main}}^{k}$, while the number of vehicles in the on-ramp of motorway section i is presented by $q_{i,\text{on}}^{k} = q_{j,\text{on}}^{k}$. In this way, the maximum number of vehicles in the mainline and on-ramp of motorway section i is presented by $q_{i,\text{main}}^{\max} = \sum_{j=1}^{J} q_{j,\text{main}}^{\max}$ and $q_{i,\text{on}}^{\max} = q_{j,\text{on}}^{\max}$.

4.3. Estimation of Arrival and Departure Rates. Arrival rates of the boundary cells in each motorway section (such as

$j + 2$, $j + 5$, and $j + 8$) and all the on-ramps, as well as the departure rates of off-ramps, are inputs of the ACTM for each planning step between two real control steps. Considering the short time of planning process (10 steps), we assume these rates can remain stable during the planning and are estimated directly from the recent flow data collected from detectors. The method described by Wang [16] is used here to do the estimation, which simply averages the most recently observed data to get the predicted flow rates. In our model, we use the flow data collected from the last N time steps ($N = 5$). Therefore, these three rates can be calculated by

$$
a_{i,\text{main}}^{k,k+1} = a_{j,\text{main}}^{k,k+1} = \frac{\sum_{n=0}^{N-1} a_{j,\text{main}}^{k-n}}{N}, \quad \text{if } j \text{ is the boundary cell,}
$$

$$
a_{i,\text{on}}^{k,k+1} = a_{j,\text{on}}^{k,k+1} = \frac{\sum_{n=0}^{N-1} a_{j,\text{on}}^{k-n}}{N}, \quad \text{if } j \text{ is the on-ramp cell,}
$$

(7)

$$
d_{j,\text{off}}^{k,k+1} = \frac{\sum_{n=0}^{N-1} d_{j,\text{off}}^{k-n}}{N}, \quad \text{if } j \text{ is the off-ramp cell,}
$$

where $a_{j,\text{main}}^{k,k+1}$ and $a_{j,\text{on}}^{k,k+1}$ are the estimated arrival rates of mainline and on-ramp of cell j for the planning step between real step k and $k+1$. $d_{j,\text{off}}^{k,k+1}$ is the estimated off-ramp departure rate of cell j. If cell j is the boundary cell of motorway section i, the arrival or departure rate of this cell is also the arrival or departure rate of motorway section i.

5. Definition of RL Elements

Except for the architecture and models defined in Section 3, three basic elements, environment state, control action, and reward function, should be specified to form a RL problem.

This section details these three elements and the relevant algorithm.

5.1. Environment State.

Environment states of a motorway section are composed of mainline states and on-ramp states. The same method mentioned in [27, 29] is used here to obtain the state space. Generally, for the mainline of motorway section i, the number of vehicles ranges from 0 to the maximum number $q_{i,\text{main}}^{\max}$ which is uniformly divided into n_i intervals. Each interval represents a state of the mainline. Therefore, each mainline section can be represented by a state set $S_{i,\text{main}}$ with n_i states. Similarly, on-ramp traffic is represented by a state set $S_{i,\text{on}}$ with m_i states according to the maximum number of vehicles $q_{i,\text{on}}^{\max}$. n_i and m_i should be adjusted for different motorway sections according to the section length. In this way, if motorway section i is the critical section, the external traffic environment is represented by

$$S_i = S_{i,\text{main}} \times S_{i,\text{on}}, \quad s_i^k \in S_i \tag{8}$$

which contains $n_i \cdot m_i$ states. At each time step, a state s_i^k will be selected from S_i as the environment state. If motorway section i is a normal section, state sets of its neighbour agent should be incorporated. Thus, traffic state is represented by

$$S_i = S_{i,\text{main}} \times S_{i,\text{on}} \times S_{i-1,\text{main}} \times S_{i-1,\text{on}}, \quad s_i^k \in S_i \tag{9}$$

which contains $n_i \cdot m_i \cdot n_{i-1} \cdot m_{i-1}$ states.

5.2. Control Action.

In a ramp control problem, the aim of the control action is to regulate the number of vehicles entering mainline in each control step. Similar to [29], we adopt flow control as the control action which can be presented by an action set $C = \{4, 6, 8, 10, 12, 14, 16, 18, 20\}$ with 9 flow rates between the minimum (4 veh/min) and maximum (20 veh/min) values.

Exploitation and exploration are two basic behaviours of the RL agent. Exploitation means the agent takes the control action that can get the most rewards from the previous experience. Exploration instead means the agent tries new actions with less rewards. In order to balance these two behaviours, we use the ε-greedy policy to select control actions [30]. Specifically, this policy takes a random action with probability ε and chooses the greedy action (with the maximum Q value) with probability $1-\varepsilon$ for each control step.

The action selection probability can be formally expressed as

$$
\widetilde{p}\left(c_i^k \mid s_i^k\right)
= \begin{cases}
1 - \varepsilon, & \text{if } c_i^k = \arg\max_{c_i^k}\left(Q^{k-1}\left(s_i^k, c_i^k\right)\right), \\
\varepsilon, & \text{otherwise.}
\end{cases} \tag{10}
$$

5.3. Reward Function.

Reward function is used to calculate the immediate reward after executing a specific action at each time step, which guides the agent to achieve its objective. Considering a common objective of traffic control system (i.e., minimising total travel time), we define our reward to guide the agent to minimise total time spent (TTS) through learning process.

TTS is defined as the total time spent by vehicles in the network during a period of time. For our case, TTS can be obtained from the following equation:

$$\text{TTS} = \Delta t \cdot \sum_{k=0}^{K}\left(q_{i,\text{main}}^k + q_{i,\text{on}}^k\right). \tag{11}$$

In the above equation, Δt is a fixed value; therefore, minimising TTS is equivalent to minimising the number of vehicles on the network $\sum_{k=0}^{K}(q_{i,\text{main}}^k + q_{i,\text{on}}^k)$. To minimise this value, the reward function defined here is composed of two negative rewards used to indicate penalties for vehicles on the mainline and on-ramp. The formal reward function at step k is defined according to two situations.

(1) Motorway Section i Is the Critical Section. Consider

$$
R_i^k\left(s_i^k, c_i^k\right)
= \begin{cases}
-\dfrac{q_{i,\text{main}}^k + q_{i,\text{on}}^k}{q_{i,\text{main}}^{\max} + q_{i,\text{on}}^{\max}}, & \text{if } q_{i,\text{main}}^k < q_{i,\text{main}}^{\max},\ q_{i,\text{on}}^k < q_{i,\text{on}}^{\max}, \\
-1, & \text{otherwise,}
\end{cases} \tag{12}
$$

where $R_i^k(s_i^k, c_i^k)$ is the immediate reward for agent i in state s_i^k when executing action c_i^k at control step k. $q_{i,\text{main}}^{\max}$ and $q_{i,\text{on}}^{\max}$ are used to normalise the number of vehicles on mainline and on-ramp, which guarantees that $R_i^k(s_i^k, c_i^k) \in [-1, 0]$.

(2) Motorway Section i Is Not the Critical Section. Here a new negative reward is introduced to maintain the system equity, that is, to make sure that the on-ramp queues and related travel times at different on-ramps should be close to each other:

$$
R_i^k\left(s_i^k, c_{i-1}^k, c_i^k\right)
= \begin{cases}
-\dfrac{q_{i,\text{main}}^k + q_{i,\text{on}}^k}{q_{i,\text{main}}^{\max} + q_{i,\text{on}}^{\max}} - \dfrac{\left|q_{i,\text{on}}^k - q_{i-1,\text{on}}^k\right|}{\max\left(q_{i,\text{on}}^{\max}, q_{i-1,\text{on}}^{\max}\right)}, & \text{if } q_{i,\text{main}}^k < q_{i,\text{main}}^{\max},\ q_{i,\text{on}}^k < q_{i,\text{on}}^{\max}, \\
-2, & \text{otherwise.}
\end{cases} \tag{13}
$$

For each agent i and episode **do**

$\quad L \longleftarrow \text{CEIL}\left(\dfrac{\text{Incident Duration}}{\Delta t}\right)$

$\quad$**IF** i is the critical section

$\quad\quad$Initialise $R_i^0(s_i, c_i)$, $Q_i^0(s_i, c_i)$

$\quad$**ELSE**

$\quad\quad$Initialise $R_i^0(s_i, c_{i-1}, c_i)$, $Q_i^0(s_i, c_{i-1}, c_i)$, $P_i^0(s_i, c_{i-1})$

$\quad$**For** each control step $k \in K$ **do**$\qquad\qquad\qquad\qquad$**(Loop I)**

$\quad\quad$(i) get detected data from each cell j: $a_{j,\text{main}}^k, d_{j,\text{main}}^k, d_{j,\text{off}}^k, a_{j,\text{on}}^k, d_{j,\text{on}}^k$

$\quad\quad$(ii) get state s_i^k through (8) and (9)

$\quad\quad$(iii) get action c_i^k by ε-greedy policy (10)

$\quad\quad$(iv) get $q_{j,\text{main}}^k, q_{j,\text{on}}^k$ through (6) and do $q_{i,\text{main}}^k \leftarrow \sum_{j=1}^J q_{j,\text{main}}^k$, $q_{i,\text{on}}^k \leftarrow q_{j,\text{on}}^k$

$\quad\quad$**IF** i is the critical section

$\quad\quad$update $R_i^k(s_i^k, c_i^k)$, $Q_i^k(s_i^k, c_i^k)$ through (2) and (12)

$\quad\quad$**ELSE** update $R_i^k(s_i^k, c_{i-1}^k, c_i^k)$, $Q_i^k(s_i^k, c_{i-1}^k, c_i^k)$, $p(c_{i-1}^k \mid s_i^k)$ through (4) and (13)

$\quad\quad$**IF** $s_i^k = s^{\text{initial}}$ and $k + 1 \geq L$ end the algorithm

$\quad\quad$**ELSE** get $a_{i,\text{main}}^{k,k+1}, a_{i,\text{on}}^{k,k+1}, d_{j,\text{off}}^{k,k+1}$ by (7) and do $l \leftarrow k$, $s_i^l \leftarrow s_i^k$, $q_{j,\text{main}}^l \leftarrow q_{j,\text{main}}^k$, $q_{j,\text{on}}^l \leftarrow q_{j,\text{on}}^k$ and start loop II

$\quad\quad\quad$**For** each planning step $l \in L$ **do**$\qquad$**(Loop II)**

$\quad\quad\quad$(i) generate flow rates for each cell j: $d_{j,\text{main}}^l d_{j,\text{on}}^l$ through (5)

$\quad\quad\quad$(ii) get the state s_i^l

$\quad\quad\quad$(iii) get $q_{j,\text{main}}^l, q_{j,\text{on}}^l$ and do $q_{i,\text{main}}^l \leftarrow \sum_{j=1}^J q_{j,\text{main}}^l$, $q_{i,\text{on}}^l \leftarrow q_{j,\text{on}}^l$

$\quad\quad\quad$(iv) get action c_i^l by ε-greedy policy

$\quad\quad\quad$**IF** i is the critical section

$\quad\quad\quad$update $R_i^l(s_i^l, c_i^l)$, $Q_i^l(s_i^l, c_i^l)$

$\quad\quad\quad$**ELSE** update $R_i^l(s_i^l, c_{i-1}^l, c_i^l)$, $Q_i^l(s_i^l, c_{i-1}^l, c_i^l)$, $p(c_{i-1}^k \mid s_i^k)$

$\quad\quad\quad$**IF** $(l = k + 9)$ or $(s_i^l = s^{\text{initial}}$ and $l + 1 \geq L)$ go back to loop I

$\quad\quad\quad$**ELSE** repeat loop II

$\quad\quad\quad$**EndFor**

$\quad$**EndFor**

EndFor

ALGORITHM 1: Algorithm for Dyna-MARL.

Compared to (12), a new term $|q_{i,\text{on}}^k - q_{i-1,\text{on}}^k|/\max(q_{i,\text{on}}^{\max}, q_{i-1,\text{on}}^{\max})$ is added into (13), which is a penalty for on-ramp queue difference in motorway section i and $i - 1$. As two adjacent agents cooperated in this situation, $R_i^k(s_i^k, c_{i-1}^k, c_i^k)$ is related to two control actions c_{i-1}^k and c_i^k. $\max(\cdot, \cdot)$ returns the maximum value of two given parameters, which is used for normalisation.

5.4. Description of the Algorithm. Based on the Dyna-Q architecture and RL elements defined in previous subsections, an algorithm Dyna-MARL is developed and described in this subsection. Two main loops corresponding to direct RL and planning shown in Figure 1 are detailed in Dyna-MARL. Between two real control steps in loop I, 10 planning steps will be run in loop II. The pseudocode of Dyna-MARL can be seen from Algorithm 1.

An episode in Dyna-MARL represents a control cycle which starts from incident occurrence and terminates when the traffic state returns to initial state s^{initial} that is the traffic state before the incident occurrence. Incident duration is assumed to be known in advance.

6. Case Study and Results

One of the metered motorway segments (southbound direction) of M6 in the UK is chosen for the case study. This segment is between junction 21A (J21A) and junction 25 (J25) with an approximate length of 12.4 km (see Figure 4). Making the noncontrolled (NC) situation as the base line, we designed a series of experiments to compare the proposed Dyna-MARL algorithm with Isolated RL (Q-learning without coordination). Experiments and relevant results are described as follows.

6.1. Partitions of the Test Segment. The test motorway segment with a three-lane mainline, three metered on-ramps, and five off-ramps is simulated by AIMSUN [39] which is a microscopic traffic simulation package. According to the detectors location and road layout, the whole segment is divided into three sections. Each section contains a metered on-ramp. Motorway section 3 is divided into 4 cells, and motorway sections 2 and 3 are both divided into 3 cells. The partitions of each section can be seen from Figure 5. According to the section length, the maximum number of
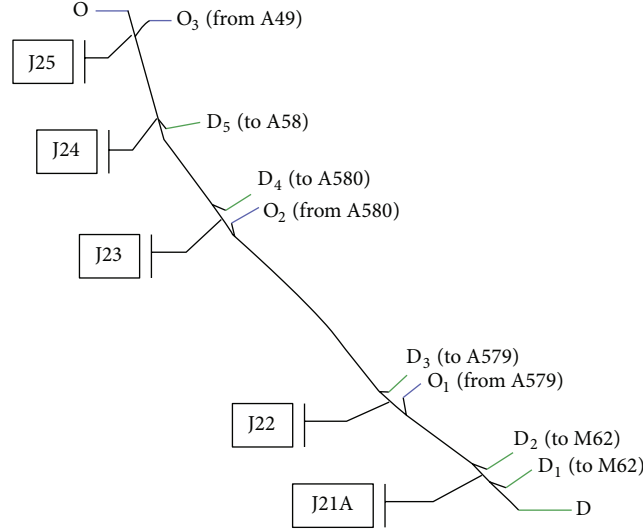
Figure 4: Test motorway segment of M6.
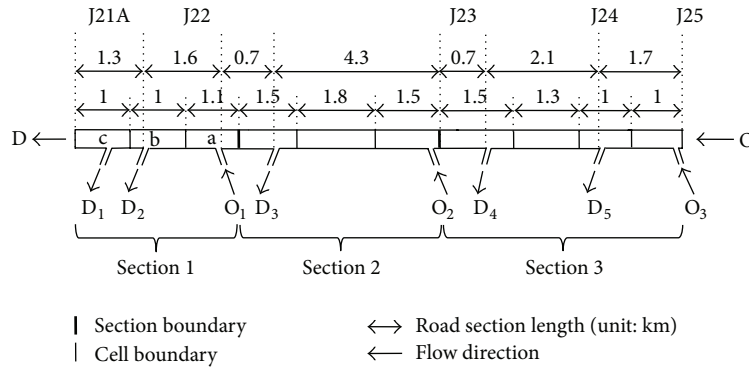
Figure 5: Partitions of test segment.

vehicles in each mainline section and on-ramps is as follows: $q_{1,\text{main}}^{\max} = 1860$, $q_{2,\text{main}}^{\max} = 2880$, $q_{3,\text{main}}^{\max} = 2880$, $q_{1,\text{on}}^{\max} = 108$, $q_{2,\text{on}}^{\max} = 90$, and $q_{3,\text{on}}^{\max} = 120$.

6.2. Real Data Source. Real detector data collected from 17 loop detectors located in the motorway segment (including both mainline and on-/off-ramps) are used for case study, which can be extracted from Traffic Information System (HATRIS) [40]. These traffic count data are averaged from April 2012 to March 2013 with 15-minute intervals. Only working day data (from Monday to Friday) are used due to the dramatic reduction of traffic load in weekends. Some of the detector data collected from mainline and three on-ramps are presented in Figure 6, from which we can see that two peak periods including AM peak period (around 07:00:00–09:00:00) and PM peak period (around 16:00:00–18:00:00) exist during the daily traffic operation.

In the test site, ramp metering only works at peak hours. Meanwhile, it is valuable to test the performance of the proposed algorithm in the high demand situation. If it can work under the high traffic load, it should be also useful for common situations. Therefore, AM peak period with heavy
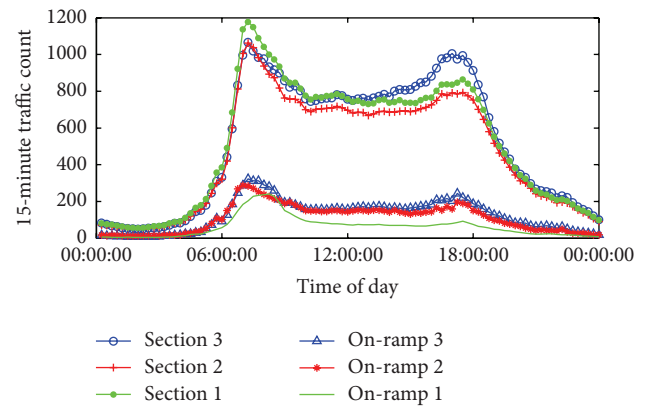
Figure 6: Real averaged traffic data.

traffic load is considered for case study. Specifically, we use the averaged traffic data during AM peak period collected from TRADS to estimate O/D (origins and destinations) matrix for the simulation. A model proposed in [41] is adopted by AIMSUN to do the estimation where the number of iterations

TABLE 1: O\D matrix estimated.

Origins\destinations	D	D_5	D_4	D_3	D_2	D_1	Totals
O	2089	375	686	728	1169	771	5818
O_3	875	65	212	193	117	46	1507
O_2	886	0	0	61	315	216	1477
O_1	824	0	0	0	292	226	1343
Totals	4675	440	898	981	1893	1258	10146

TABLE 2: Parameters for ACTM.

Parameter	$d_{\text{main}}^{\max}$	v	w	θ	η	λ_1	λ_2	λ_3
Value	6300 veh/h	107 km/h	11.6 km/h	0.5	0.16	0.55	0.9	0.6

is set as 1000 to get convergence. Table 1 shows the O/D matrix estimated from real traffic data.

6.3. Incident Scenarios. Considering the difficulty of capturing real incident data, we simulate some incident scenarios in AIMSUN. To make each ramp meter work during the incident, the incident is located near the most downstream motorway section, that is, motorway section 1. Therefore, three incident scenarios A, B, and C are designed corresponding to three different incident locations in a, b, and c (as illustrated in Figure 5), respectively.

The simulation experiment lasts for one and a half hours from 07:00:00 to 08:30:00 during AM peak period. After 30-minute normal operation (for warm-up), the incident is triggered at 07:30:00 and lasts for 30 minutes. In the preliminary experiments designed in this paper, the incident with one lane blocked is considered. Parameters introduced here can also be regulated for multiple lane-blockage situations. The incident extent is 50 meters which is assumed to be constant during the incident.

Learning-related parameters are set as typical values [30]; that is, α is 0.2, γ is 0.8, and ε is 0.1. Other parameters related to ACTM are calibrated and summarised in Table 2. All the cells have the same θ and η.

6.4. Results. The comparison of Dyna-MARL, Isolated RL, and NC is conducted from three aspects: density evolution, some general indicators, and the system equity. The experimental results are described as follows.

(1) Density Evolution. We can see from Figure 7 that four dense areas exist during the traffic operation. Three of them near on-ramp entrances (motorway length around 0.5 km, 5 km, and 10 km) are caused by heavy traffic loads from on-ramps. The dense area close to the segment end forms due to the incident.

In scenario A, incident location is close to on-ramp 1 (O_1). Without control, this incident leads to sever congestion which blocks on-ramp 1 and propagates to motorway section 2 (around 9 km in Figure 7(a)). Under this scenario, Isolated RL cannot alleviate incident-induced congestion effectively

(see Figure 7(b)). In the beginning of congestion formulation, without coordination, only the nearest ramp controller reacts to the congestion. Because of the space limit of on-ramp, one ramp controller is insufficient to dissolve this congestion that still propagates to motorway section 2. Dyna-MARL, on the other hand, coordinates all three ramp controllers and makes full use of the storage space of three on-ramps to deal with incident-induced congestion. In this way, mainline congestion can be restricted in a smaller area and will not propagate to motorway section 2 (see Figure 7(c)).

For scenarios B and C, incidents are near the motorway end and far from on-ramp 1. Without blocking on-ramp 1, incidents do not lead to sever congestion. Under such circumstances, both Isolated RL and Dyna-MARL work well on easing congestion in the mainline. As shown in Figures 7(e)–7(i), compared with the NC situation, both Isolated RL and Dyna-MARL can restrict the congestion in a small range near the on-ramp entrances.

(2) General Indicators. In this comparison, some general indicators, including total travel time (should be reduced), total throughput (should be improved), and total CO_2 emission (should be reduced), are used to show how the proposed system can benefit road users. These indicators are widely used in the transport community to test the performance of newly developed traffic control systems.

As shown in Figure 8(a), compared with the NC situation, both Isolated RL and Dyna-MARL can reduce the total travel time of road users in all three scenarios. Specifically, Isolated RL decreases total travel time by up to 6.2%, while Dyna-MARL achieves a maximum reduction of 12.2% (see Figure 8(d)). The comparison of total throughput is presented in Figure 8(b). Dyna-MARL can improve the total throughput by up to 2.3% (see Figure 8(d)) which outperforms Isolated RL in all three scenarios. In scenario B, Isolated RL even fails to improve the total throughput. For the comparison of total CO_2 emission (shown in Figure 8(c)), both Isolated RL and Dyna-MARL achieve their best performance in scenario B with a reduction of 4.7% and 4.6%, respectively. In scenarios A and C, Dyna-MARL has a much better performance than Isolated RL.

Through the above comparison, we can see that Dyna-MARL outperforms Isolated RL for almost all the scenarios and indicators.

(3) System Equity. Although the general indicators presented in comparison (2) have shown their effectiveness on testing the performance of different systems, they cannot measure the issue of system equity, which is also an important aspect of the system performance. In this paper, we only consider the spatial equity issue that is defined as a measurement of equity of user delays on different on-ramps [42]. In this study, we assume the road users from all three on-ramps have the same importance. If all users from different on-ramps can experience the similar travel time, the control system is defined as an equitable system. This term is used to measure the system equity; that is, a large queue difference leads to a highly inequitable system. In [43], the variance of travel time on different on-ramps is used as an indicator to measure

(a)

(b)

(c)

(d)

(e)

(f)
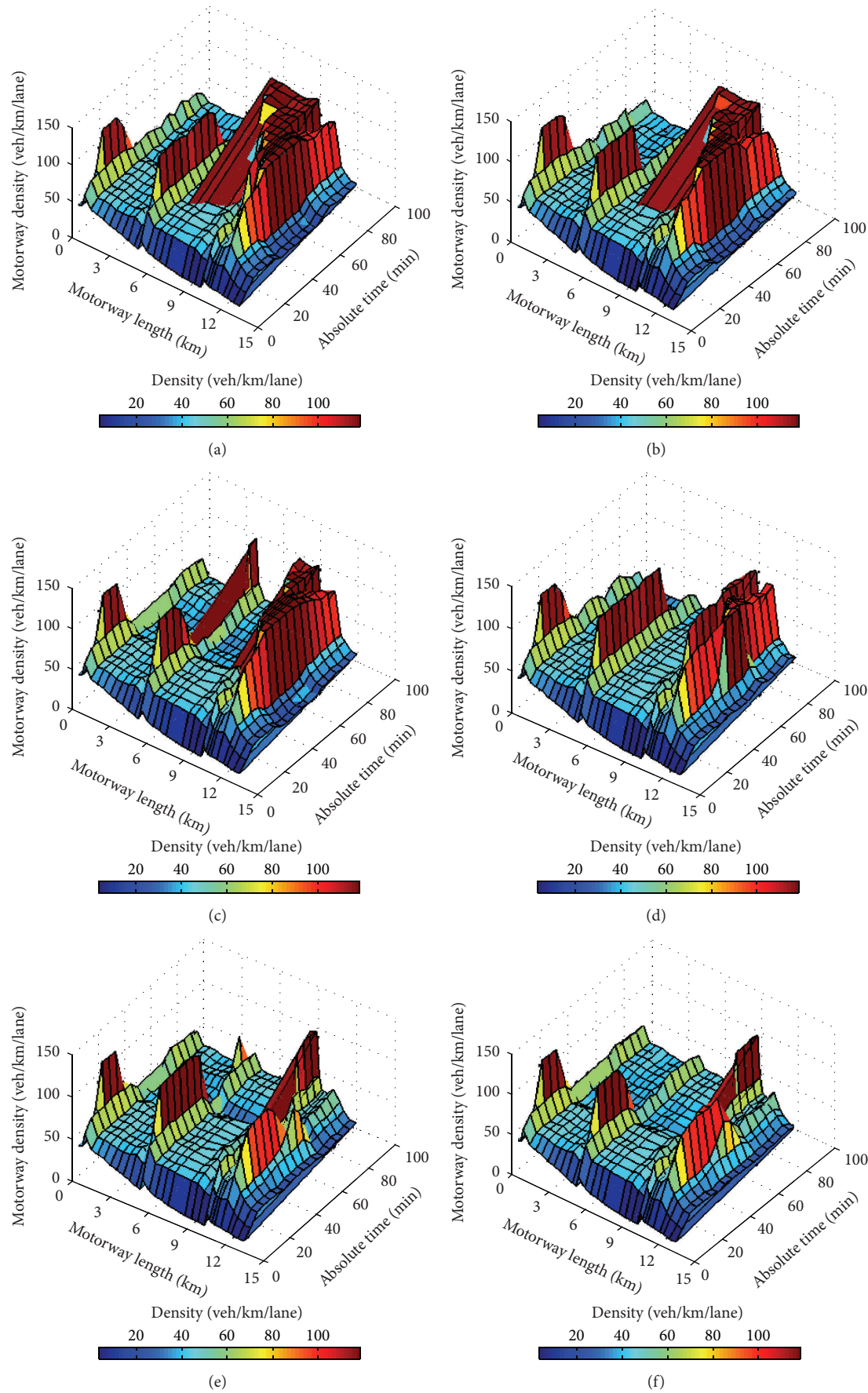
Figure 7: Continued.

(g)

(h)

(i)
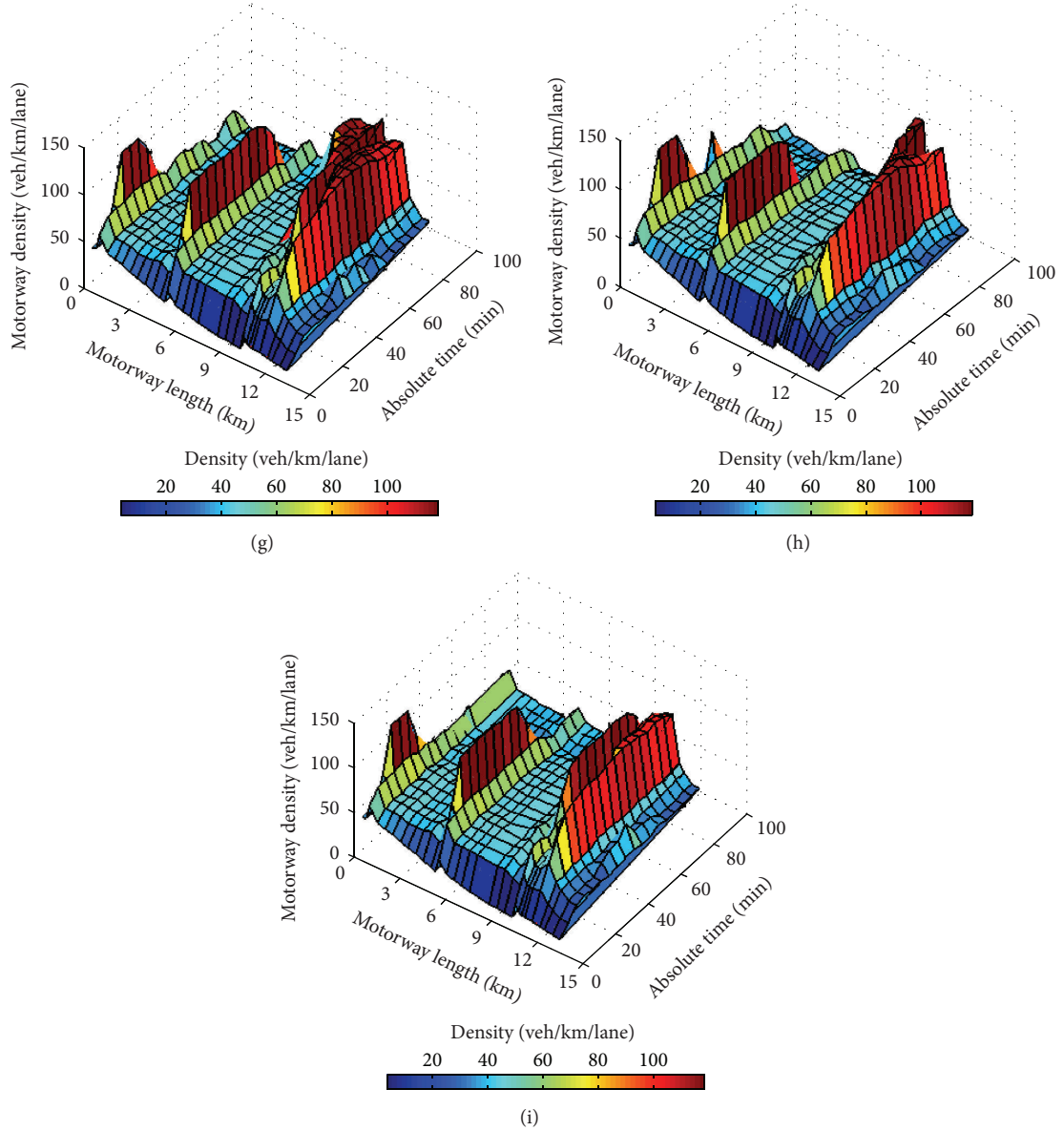
FIGURE 7: Density profiles for (a) NC in scenario A, (b) Isolated RL in scenario A, (c) Dyna-MARL in scenario A, (d) NC in scenario B, (e) Isolated RL in scenario B, (f) Dyna-MARL in scenario B, (g) NC in scenario C, (h) Isolated RL in scenario C, and (i) Dyna-MARL in scenario C.

system equity. Similar to [43], for the sake of comparison, the standard deviation is considered in our case. This indicator is defined as

$$SD(k) = \sqrt{\frac{\sum_{i=1}^{n} \left[\bar{t}^k - t_i^k\right]^2}{n}}, \tag{14}$$

where $SD(k)$ is the standard deviation of travel time of different on-ramps at time step k. t_i^k is the estimated total travel time of on-ramp i at step k. $\bar{t}^k$ is the averaged total travel time of n on-ramps at step k.

Results about the comparison of system equity can be seen from Figure 9. For the NC situation, good equity can

be maintained due to no restrictions of entering vehicles in scenarios B and C (as shown in Figures 9(b) and 9(c)). However, when one of the on-ramp entrances is blocked by the congestion in scenario A, a long queue forms and leads to imbalance and resultant inequity for users on different on-ramps (see Figure 9(a)). For controlled cases, Isolated RL performs poorly in all scenarios. This is because the ramp controller near congestion takes much more restricted measures than other controllers on the controlled traffic. Because of the coordination strategy, Dyna-MARL outperforms Isolated RL on maintaining system equity in all scenarios, especially during the incident (from 07:30:00 to 08:00:00).

(a)

(b)

(c)

(d)
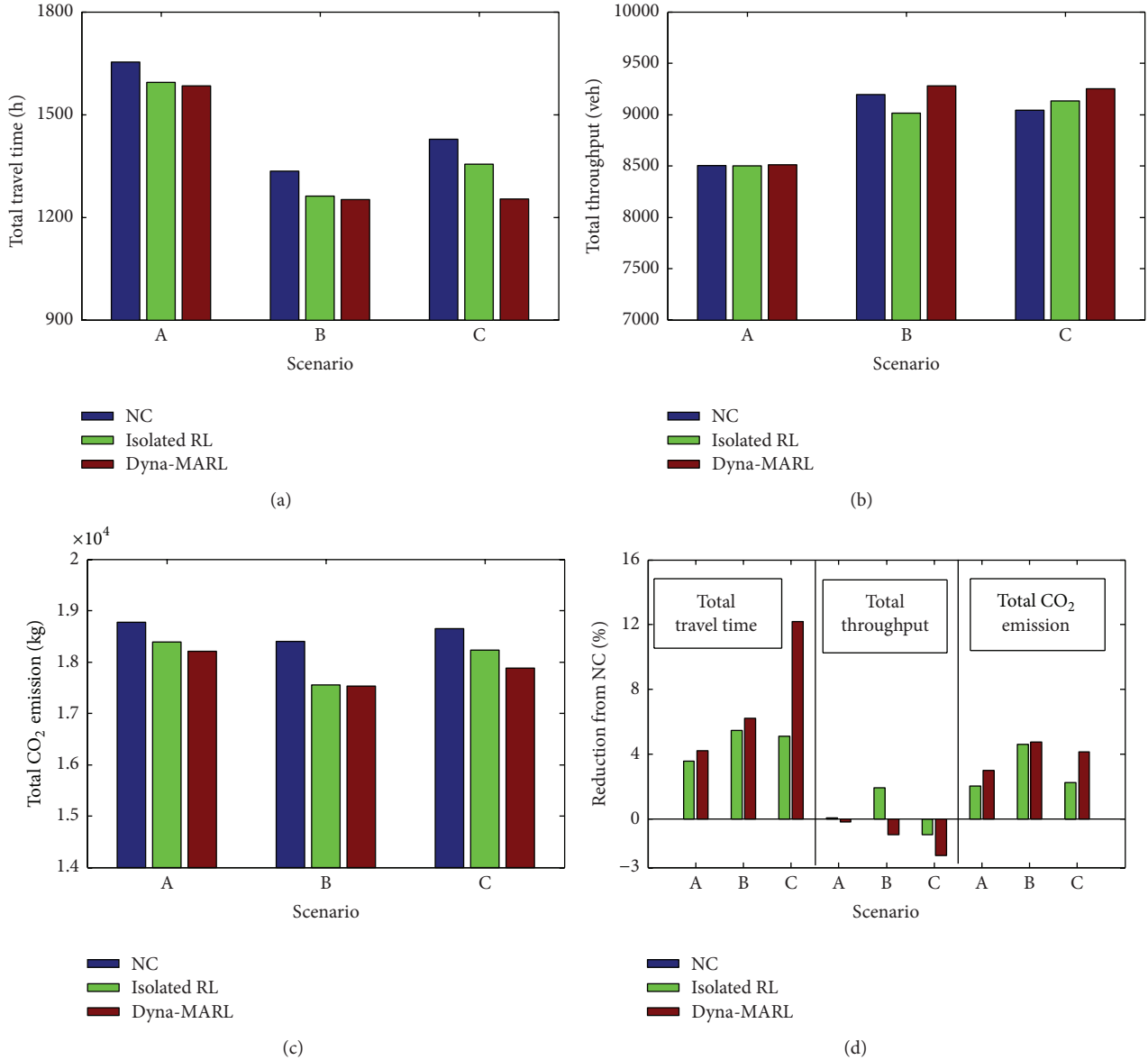
Figure 8: Comparison of general measures for different scenarios.

7. Conclusions and Future Work

A Dyna-Q based multiagent reinforcement learning method referred to as Dyna-MARL for motorway ramp control has been developed in this paper. Dyna-MARL is compared with Isolated RL (Q-learning without coordination) and noncontrolled situation under the simulation environment. Real traffic data collected from a metered motorway segment in the UK are used to form the simulation.

Through a series of simulation-based experiments, we can conclude the following: (1) Isolated RL can improve the motorway performance in terms of increasing total throughput, reducing total travel time and CO_2 emission, but this improvement is at the expense of poor system equity on different on-ramps; (2) with a suitable coordination strategy,

much higher system equity can be achieved by Dyna-MARL; (3) in addition to the system equity, Dyna-MARL outperforms Isolated RL in almost all scenarios regarding all indicators, which means Dyna-MARL can deal with the network-wide problems effectively.

Although the simulation tests have shown some positive results regarding the performance of Dyna-MARL, a simplified incident scenario with fixed duration is considered in the current work. In the practical situation, incident duration is highly unstable and affected by a number of factors, such as weather conditions, road conditions, and arriving time of the incident management team. Therefore, incident duration should be considered as an uncertainty which will be investigated in our future work.

(a)

(b)

(c)
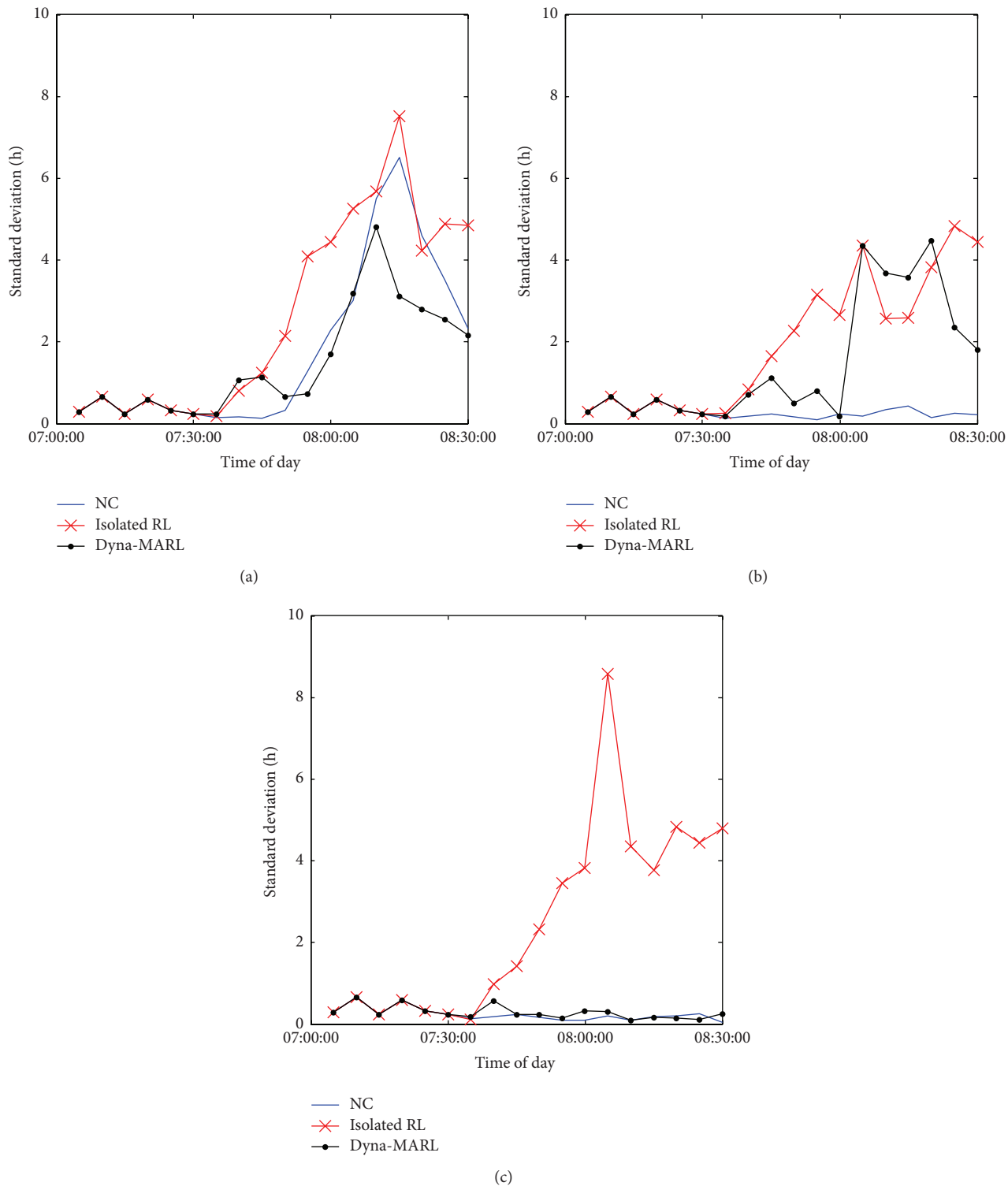
FIGURE 9: Standard deviation for different scenarios: (a) scenario A, (b) scenario B, and (c) scenario C.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: an overview," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 271–281, 2002.

[2] A. Skabardonis, P. Varaiya, and K. F. Petty, "Measuring recurrent and nonrecurrent traffic congestion," *Transportation Research Record*, vol. 1856, pp. 118–124, 2003.

[3] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville, "ALINEA: a local feedback control law for on-ramp metering," *Journal of the Transportation Research Board*, vol. 1320, pp. 58–64, 1991.

[4] E. Smaragdis and M. Papageorgiou, "Series of new local ramp metering strategies," *Transportation Research Record*, vol. 1856, pp. 74–86, 2003.

[5] L. N. Jacobson, K. C. Henry, and O. Mehyar, "Real-time metering algorithm for centralized control," *Transportation Research Record*, vol. 1232, pp. 17–26, 1989.

[6] G. Paesani, J. Kerr, P. Perovich, and F. Khosravi, "System wide adaptive ramp metering (SWARM)," in *Proceedings of the 7th ITS America Annual Meeting and Exposition: Merging the Transportation and Communications Revolutions*, Washington, DC, USA, June 1997.

[7] R. Lau, *Ramp Metering by Zone—The Minnesota Algorithm*, Minnesota Department of Transportation, 1997.

[8] H. M. Zhang and W. W. Recker, "On optimal freeway ramp control policies for congested traffic corridors," *Transportation Research Part B: Methodological*, vol. 33, no. 6, pp. 417–436, 1999.

[9] A. Kotsialos, M. Papageorgiou, and F. Middelham, "Optimal coordinated ramp metering with advanced motorway optimal control," *Transportation Research Record*, no. 1748, pp. 55–65, 2001.

[10] A. Hegyi, B. De Schutter, and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research C: Emerging Technologies*, vol. 13, no. 3, pp. 185–209, 2005.

[11] G. Gomes and R. Horowitz, "Optimal freeway ramp metering using the asymmetric cell transmission model," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 4, pp. 244–262, 2006.

[12] A. H. F. Chow and Y. Li, "Robust optimization of dynamic motorway traffic via ramp metering," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1374–1380, 2014.

[13] R. W. Hall, "Non-recurrent congestion: how big is the problem? Are traveler information systems the solution?" *Transportation Research Part C*, vol. 1, no. 1, pp. 89–103, 1993.

[14] P. Prevedouros, B. Halkias, K. Papandreou, and P. Kopelias, "Freeway incidents in the United States, United Kingdom, and Attica Tollway, Greece: characteristics, available capacity, and models," *Transportation Research Record*, vol. 2047, pp. 57–65, 2008.

[15] T. L. Greenlee and H. J. Payne, "Freeway ramp metering strategies for responding to incidents," in *Proceedings of the IEEE Conference on Decision and Control including the 16th Symposium on Adaptive Processes and a Special Symposium on Fuzzy Set Theory and Applications*, pp. 987–992, New Orleans, LA, USA, December 1977.

[16] M. H. Wang, *Optimal ramp metering policies for nonrecurring congestion with uncertain incident duration [Ph.D. thesis]*, Purdue University, West Lafayette, Ind, USA, 1994.

[17] J.-B. Sheu, "Stochastic modeling of the dynamics of incident-induced lane traffic states for incident-responsive local ramp control," *Physica A: Statistical Mechanics and its Applications*, vol. 386, no. 1, pp. 365–380, 2007.

[18] J.-B. Sheu and M.-S. Chang, "Stochastic optimal-control approach to automatic incident-responsive coordinated ramp control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 359–367, 2007.

[19] C. Jacob and B. Abdulhai, "Machine learning for multi-jurisdictional optimal traffic corridor control," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 2, pp. 53–64, 2010.

[20] M. Davarynejad, A. Hegyi, J. Vrancken, and J. van den Berg, "Motorway ramp-metering control with queuing consideration using Q-learning," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC '11)*, pp. 1652–1658, IEEE, Washington, DC, USA, October 2011.

[21] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Application of reinforcement learning with continuous state space to ramp metering in real-world conditions," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems (ITSC '12)*, pp. 1590–1595, IEEE, Anchorage, Alaska, USA, September 2012.

[22] C. Jacob and B. Abdulhai, "Automated adaptive traffic corridor control using reinforcement learning: approach and case studies," *Transportation Research Record*, vol. 1959, pp. 1–8, 2006.

[23] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Self-Learning adaptive ramp metering: analysis of design parameters on a test case in Toronto, Canada," *Transportation Research Record*, vol. 2396, pp. 10–18, 2013.

[24] X.-J. Wang, X.-M. Xi, and G.-F. Gao, "Reinforcement learning ramp metering without complete information," *Journal of Control Science and Engineering*, vol. 2012, Article ID 208456, 8 pages, 2012.

[25] A. Fares and W. Gomaa, "Multi-agent reinforcement learning control for ramp metering," in *Progress in Systems Engineering*, vol. 330 of *Advances in Intelligent Systems and Computing*, pp. 167–173, Springer, Basel, Switzerland, 2015.

[26] K. Veljanovska, K. M. Bombol, and T. Maher, "Reinforcement learning technique in multiple motorway access control strategy design," *PROMET-Traffic & Transportation*, vol. 22, no. 2, pp. 117–123, 2010.

[27] C. Lu, H. Chen, and S. Grant-Muller, "An indirect reinforcement learning approach for ramp control under incident-induced congestion," in *Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC '13)*, pp. 979–984, IEEE, The Hague, The Netherlands, October 2013.

[28] L. Buşoniu, R. Babuška, and B. De Schutter, "A comprehensive survey of multiagent reinforcement learning," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 38, no. 2, pp. 156–172, 2008.

[29] C. Lu, H. Chen, and S. Grant-Muller, "Indirect Reinforcement Learning for Incident-responsive ramp control," *Procedia—Social and Behavioral Sciences*, vol. 111, pp. 1112–1122, 2014.

[30] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.

[31] C. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[32] J. R. Kok and N. Vlassis, "Collaborative multiagent reinforcement learning by payoff propagation," *Journal of Machine Learning Research*, vol. 7, pp. 1789–1828, 2006.

[33] C. Guestrin, M. G. Lagoudakis, and R. Parr, "Coordinated reinforcement learning," in *Proceedings of the 19th International Conference on Machine Learning*, pp. 227–234, Sydney, Australia, July 2002.

[34] S. El-Tantawy, B. Abdulhai, and H. Abdelgawad, "Multiagent reinforcement learning for integrated network of adaptive traffic signal controllers (marlin-atsc): methodology and large-scale application on downtown toronto," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1140–1150, 2013.

[35] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: a survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[36] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[37] J. Haddad, M. Ramezani, and N. Geroliminis, "Cooperative traffic control of a mixed network with two urban regions and a freeway," *Transportation Research Part B: Methodological*, vol. 54, pp. 17–36, 2013.

[38] H. Mongeot and J.-B. Lesort, "Analytical expressions of incident-induced flow dynamics perturbations: using macroscopic theory and extension of Lighthill-Whitham theory," *Transportation Research Record*, vol. 1710, pp. 58–68, 2000.

[39] Transport Simulation Systems, *Aimsun User's Manual 6.1*, TTS, Barcelona, Spain, 2010.

[40] Highways England, "Hatris Homepage," 2013, https://www.hatris.co.uk/.

[41] E. Cascetta, "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator," *Transportation Research B*, vol. 18, no. 4-5, pp. 289–299, 1984.

[42] L. Zhang and D. Levinson, "Balancing efficiency and equity of ramp meters," *Journal of Transportation Engineering*, vol. 131, no. 6, pp. 477–481, 2005.

[43] A. Kotsialos and M. Papageorgiou, "Efficiency and equity properties of freeway network-wide ramp metering with AMOC," *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 6, pp. 401–420, 2004.

Research Article

The Entropy-Cost Function Evaluation Method for Unmanned Ground Vehicles

Ya-Nan Zhao, Kai-Wen Meng, and Li Gao

Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China

Correspondence should be addressed to Ya-Nan Zhao; zyn@bit.edu.cn

To improve the technology of unmanned ground vehicles, it is necessary to conduct a proper evaluation on various technologies. Previous evaluation methods are mainly based on completion of the task; this may mislead most of teams of unmanned ground vehicles using a conservative strategy during the evaluation. In this paper, a new evaluation method is proposed. Based on typical working conditions including intersection, car-following, and obstacle-avoiding, the new evaluation indicator system is established, and the entropy-cost function method is applied to the comprehensive evaluation of unmanned ground vehicles. As reported in a numerical example, the proposed evaluation method can get a quantitative result that authentically reflects the intelligent behavior level of unmanned ground vehicles.

1. Introduction

Based on the preestablished evaluation system, the technology of intelligent vehicles abroad has developed rapidly in recent years. Obviously an excellent evaluation system can guide participants to improve the performance of intelligent vehicles in evaluation/test. For example, none unmanned vehicles finished the entire race in the US intelligent vehicle competitions DARPA 2004 [1]. Guided by the rules/evaluation rules of Grand Challenge, the teams improved their unmanned vehicles and five teams accomplished the entire distance in 2005 competition [2]. In the history of technology development, evaluation system has played an important role.

The 2nd DARPA Grand Challenge simply used the "the number of finishing the races" and "the total number of gates through" to rank the teams [3]. The 3rd DARPA Grand Challenge, Urban Challenge, required that all participants must obey traffic rules [4]. "Future Challenge 2009" was held in June 2009 [5]; the race used the third-party testing to evaluate the implementation efficiency of intelligent vehicles in competitions. This attempt has promoted and facilitated the innovation and development of vehicle verification platform [6]. "Future Challenge 2012" competition was held in November 2012, Chifeng City, Inner Mongolia [7]. The race mainly tested the research level of cognitive computing about visual and auditory information [8]. U-Turn evaluation environment is accepted for more competitions in recent years, but the evaluation environment cannot take more uncertain factors into consideration [9, 10].

Most of existing foreign evaluation methods of domestic and abroad intelligent vehicle competitions used the mission-driven evaluation approach, which has an obvious shortcoming, leading many teams to adopt a conservative strategy [11]. For example, in order to avoid the violation of traffic regulations, many teams chose to park and wait [12] rather than interact with the environment as manned vehicles do. Even though they can get high score, it is a departure from the purpose of technological development [13]. As for domestic methods, the evaluation system is obviously not comprehensive; many factors are not considered in evaluation/test. For example, the papers [7–10] just focus on dynamic driving abilities and ignore static conditions, interactive abilities with environment. In the future, intelligent vehicles will interact with more people and cars in complex scenes [14–16]; thus the evaluation method needs improvement. The method must take more comprehensive evaluation factors into account and reflect the performance of unmanned vehicles objectively.

This paper concentrates on proposing a novel evaluation method for intelligent vehicles that is based on information entropy and cost function. The information entropy checks
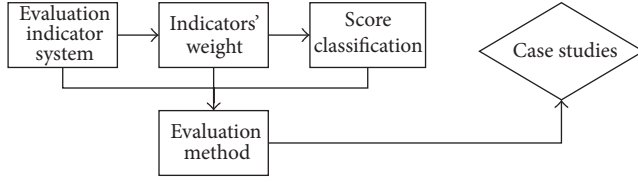
Figure 1: The logic and construction of the paper.

Table 1: The evaluation system of unmanned ground vehicles' intelligent behavior based on typical working conditions.

Evaluation objective	Typical intelligent behaviors	Evaluation auxiliary indicators
Multi-indicator evaluation system for unmanned vehicles	Intersection behavior	Parking precision Restart ability Speed capability Braking deceleration
	Obstacle-avoiding behavior	Early warning Avoidance in right angle Path replanning
	Car-following behavior	Stimulation Safe distance Speed optimization

all evaluation indicators' weight from information amount aspect to handle uncertainty problems in evaluation/test; the cost function checks the intelligent level of each of abilities to get specific evaluation score. The evaluation method also concludes a new rigorous evaluation indicator system based on typical working conditions; each typical working condition of intelligent vehicles is subdivided into different physical indicators to reflect real situation. Not only the completion time of indicators but also all the details are taken into account, even down to the completion quality and veracity of the various secondary indicators. The new evaluation method refines three aspects of evaluation process: evaluation indicator system establishment, indicators' weight arrangement, and evaluation score classification to guarantee the objectivity, comprehensiveness, and scientificalness of evaluation results. The logic and construction of the paper are shown in Figure 1.

The final evaluation results can show the intelligent level of unmanned vehicles and its weakness and then guide the participating vehicles to move in the right direction and goals of high-tech development.

2. Foundation of the Evaluation Indicator System

As described above, most of the practical evaluation activities were mission-driven, which resulted in the incomprehensive evaluation indicator system. Thus the evaluation results were partial. Most participants could get higher score just because of their own research superiorities. It is not fair and objective. That means a proper selection of evaluation indicators is an important part in the evaluation.

Evaluation of unmanned vehicle intelligent behavior is a multilevel comprehensive evaluation problem. Considering the characteristics of unmanned vehicles' data which are scattered and the advantages and disadvantages of traditional evaluation indicators now, expert opinions and analysis of typical working conditions are selected; this can not only make full use of the experts' cognitive knowledge of unmanned vehicles but also avoid missing important indicators. The result is relatively accurate with the rigorous indicators selection process (see Figure 2).

The paper mainly takes two factors into consideration: objective one and subjective one, which are given in Figure 2. Objective working conditions reflect the external environment; subjective opinions from experts reflect the internal attitude in the evaluation/test.

The typical working conditions are summarized from many intelligent vehicle competitions. In each of intelligent

vehicle competitions, the participants will encounter three main working conditions; those working conditions almost conclude all of intelligent driving behaviors in evaluation/test. According to the typical working conditions including intersection, car-following [17], and obstacle-avoiding, the paper analyzes the performance of different situations that can show the real level of unmanned vehicles' intelligence and then summarizes essential features of each typical working condition. Each typical working condition of intelligent vehicles reflects main intelligent behavior; each behavior has its tasks. Evaluation indicator structure is given in Table 1. All indicators are established through considering real situations in the driving process; the physical characteristics can be reflected in the indicator system.

The indicator called "Parking precision" tests the ability to park the vehicle at the right place and the right time when facing intersection. "Restart ability" tests the ability to identify complex intersection conditions and participate in the traffic intelligently. "Speed capability" reflects the performance on the speed controlling, acceleration, and deceleration. "Braking deceleration" tests braking quality. "Early warning" tests video detection system and the ability to identify different obstacles. "Avoidance in right angle" tests whether the vehicle can use less space to avoid the obstacles. "Path replanning" can guarantee the vehicle will not abort the original destination. "Stimulation" tests the ability to calculate the fore car's real-time speed, thus taking actions timely. "Safe distance" evaluates the stationary and the following model of the intelligent vehicle [18, 19]. "Speed optimization" tests the ability to control distance through keeping optimal speed.

3. Information Theory and Entropy for Indicators' Weight

This paper introduces knowledge of information theory to determine the weight distribution parameters. In 1948, in the paper "A Mathematical Theory of Communication," Shannon used Probability and Statistics approach to the study of communication systems, revealing that the object of communication system is information and then proposed the concept of entropy through describing the information quantitatively.
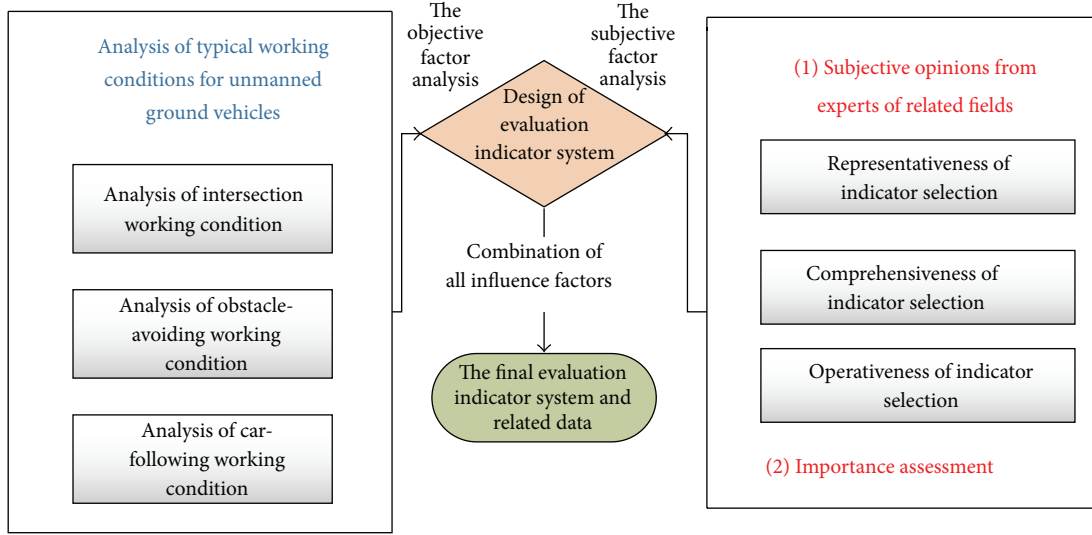
Figure 2: Foundation of the evaluation indicator system.

Based on Shannon's theory, the recipient cannot predict the message that will be received; therefore, uncertainty exists objectively because of the message's randomness. Delivery of messages is a process to eliminate the uncertainty of information, a process from uncertainty to certainty. Uncertainty in information theory is a kind of inherent property existing in working process, so the concept of entropy can be applied to the evaluation of intelligent behaviors of unmanned ground vehicles. Information entropy can determine the key indicators in terms of the amount of information to calculate the indicator weight parameters, and the entropy of information can be calculated as follows:

$$H(x_i) = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)} = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i). \quad (1)$$

3.1. Judgment Matrix. To judge all factors with 1~9 scaling method to determine each indicator's importance, experts' suggestion is summarized to construct the judgment matrix $X = (x_{ij})_{m \times n}$ of fractional values, and evaluation indicator system has k aspects behaviors, m objects, n auxiliary indicators, $1 \le i \le m$, and $1 \le j \le n$.

3.2. Matrix Data. According to the linear proportional relationship, the decision matrix $X = (x_{ij})_{m \times n}$ is reformed as standardization to get standardized matrix $Y = (y_{ij})_{m \times n}$ and then normalize $Y = (y_{ij})_{m \times n}$:

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^{m} y_{ij}}, \quad 1 \le i \le m, \ 1 \le j \le n. \quad (2)$$

3.3. Information Entropy. Calculation of j indicator's information entropy is as follows:

$$e_j = -\lambda \sum_{i=1}^{m} p_{ij} \ln p_{ij}, \quad \lambda = (\ln m)^{-1}. \quad (3)$$

3.4. Information Utility Value. The real parameters that reflect indicator weights are the values of information utility; the values can be calculated by the following formula:

$$h_j = 1 - e_j, \quad 1 \le j \le n. \quad (4)$$

3.5. Total Indicator Weight. The importance weight of all factors of this layer corresponding to the above layer of a certain factor that can be calculated using the result of all single hierarchical sorts at the same hierarchy is as follows:

$$\beta_k = (\beta_{k1}, \beta_{k2}, \ldots, \beta_{kj}, \ldots, \beta_{kn}),$$
$$\beta_{kj} = \frac{h_j}{\sum_{j=1}^{n} h_j}. \quad (5)$$

β_{kj} is objective weight of each indicator.

4. Cost Function for Score Classification

According to the 2007 DARPA Urban Challenge's low-speed, low-density traffic environment issues, scholars at Carnegie Mellon University proposed a kind of robust highway autonomous driving technology that is combined with the cost evaluation method [20]. The cost function method can also be used in Chinese intelligent vehicle competitions to quantify the indicators. The cost function method is bound to promote competition vehicles toward the direction of the minimum cost function in the process of technology development, thus leading unmanned ground vehicles to finish the task with high-quality. The technological level of Chinese unmanned vehicles on its natural environment perception and decision-making capacity is gradually improving.

Taking the indicator "Path replanning" as a research object, unmanned vehicles need to consider the process cost control that means more planning time and more useless

operating range deserve more penalty and higher process cost. The cost can be calculated by the following formula:

$$C = t * d_{\text{finish}}. \tag{6}$$

To make sure of the comprehensive consideration for all indicators and the full use of evaluation information, the paper adopts a weighted average algorithm to get final evaluation results by decomposing calculation. As for typical intelligent behaviors, the paper presents each cost function as C_1, C_2, and C_3. C_1 can be calculated as follows:

$$C_1 = \begin{cases} \sum_{j=1}^{n} c_{1j}\omega_{1j} & \text{Participator} \\ \beta C_m \ (\beta > 1, \beta = 1.5) & \text{Abstainer}. \end{cases} \tag{7}$$

c_{1j} presents the cost of n indicator of intersection behavior; ω_{1j} presents the weight of n indicator of intersection behavior; C_m presents the maximum value of the cost values;

β is the penalty factor. The final cost of unmanned vehicles can be calculated as follows:

$$C = C_1\omega_1 + C_2\omega_2 + C_3\omega_3. \tag{8}$$

5. The Entropy-Cost Function Evaluation Method

5.1. Weight Distribution of Each Indicator. According to the evaluation indicators for unmanned vehicle intelligent behavior comprehensive evaluation model, the judgment matrix of each level is constructed following the 1~9 scaling methods by expert group composed by research members in related fields. β_{kj}, calculated by formulae (2)–(5), is the weight in accordance with the overall goal layer of 10 evaluation indicators. According to the additivity of information entropy, the weight of typical working conditions' behavior can be calculated through accumulating affiliated indicators directly as follows.

Typical working conditions' behavior weight is as follows:

$$W = \begin{bmatrix} 0.417 & 0.267 & 0.317 \end{bmatrix}. \tag{9}$$

Affiliated indicators weight is as follows:

$$W = \begin{bmatrix} 0.113 & 0.110 & 0.092 & 0.102 & 0.112 & 0.067 & 0.088 & 0.112 & 0.087 & 0.116 \end{bmatrix}. \tag{10}$$

Then the level of indicators' weight coefficients is determined, and the evaluation results will be more objective because of different treatments of indicators.

5.2. Comprehensive Score Based on Cost Function. The weight of each indicator in the evaluation system has been calculated in Section 5.1, and the cost combined with weight will be the final value that represents the comprehensive score of unmanned vehicles. In the testing, the paper uses ranking method to define the basic cost value; the real performance will be quantified in the corresponding rank. Taking "Parking precision" as a research object, the cost can be represented as follows:

$$c_{11} = \begin{cases} 0 & \text{Ideal Condition} \\ 20 & \text{Score Rank 1} \\ 40 & \text{Score Rank 2} \\ 60 & \text{Score Rank 3} \\ 80 & \text{Score Rank 4} \\ 100 & \text{Eliminated District}. \end{cases} \tag{11}$$

Since the data are rough, the score can just estimate a general intelligent level of unmanned vehicles; two teams with almost same score must make further comparison about

more details. The rest of the indicators' cost value can be calculated identically, and the first class indicators' cost is

$$C_1 = \sum_{i=1}^{3} c_{1i}\omega_{1i} = c_{11}\omega_{11} + c_{12}\omega_{12} + c_{13}\omega_{13}. \tag{12}$$

6. Unmanned Vehicle Intelligent Behavior Evaluation

6.1. Each Indicator's Score. The score of each indicator is given based on the comparison between optimal performances and minimum acceptable performances from manned driving conditions roughly; the gap will be divided into several score ranks (see Figure 3).

The method for calculating score is a transition from manned driving to driverless. As unmanned vehicles' development direction is human-like capacity in driving, the method is accordingly beneficial for technical updating.

6.2. Each Indicator's Cost Value. Also taking the indicator "Path replanning" as the research object, some data in "Future Challenge 2012" competition are shown in Table 2.

Table 2 shows that car F's data are invalid because of the lane departure in this test. The score is corresponding to the planning time and operating range, optimal condition in manned driving cannot be quantified with 7–9 (s) and 5–7 (m), and minimum acceptable condition is 18–20 (s) and 15–17 (m). The threshold value of score classification is based

TABLE 2: The "Path replanning" data.

The competing teams	A	B	C	D	E	F
Planning time (s)	13	18	10	13	11	15
Operating range (m)	8.3	7.4	9.2	6.5	10.9	None
Cost value	40	60	20	20	60	100

TABLE 3: The data and score of team A in the competition.

Indicators	Z_{11}	Z_{12}	Z_{13}	Z_{14}	Z_{21}	Z_{22}	Z_{23}	Z_{31}	Z_{32}	Z_{33}
Data-1	0.7	4.5	59	12.3	30	1.3	13	7	3.6	36
Unit-1	m	s	h	m	s	m	s	s	M	km/h
Data-2	13	1.5		2	89.3	53	8.3			18
Unit-2	s	m/s^2		m/s^2	%	°	m			s
Cost	20	20	40	20	20	60	40	80	20	40
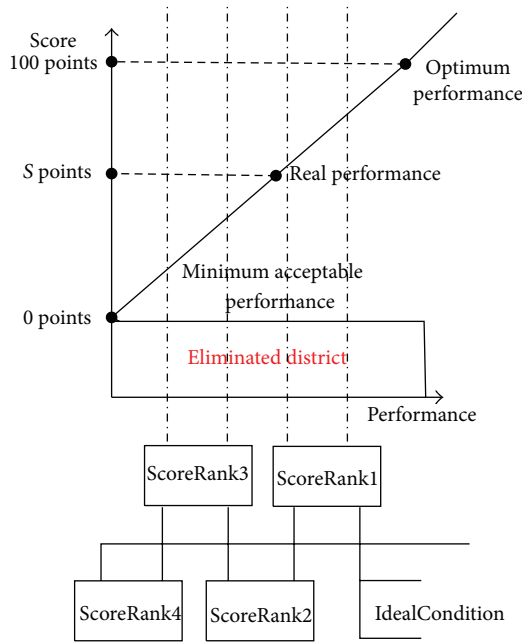
FIGURE 3: Calculation of each indicator's score.

on human experience, so the data can only be processed roughly to get the score-rank of performance.

Analyzing the data of 10 indicators of team A in the competition, "Parking precision" is corresponding to parking time and the distance between the front of the vehicle and the stop line; "Restart ability," restart time and acceleration; "Speed capability," horsepower; "Braking deceleration," deceleration and braking distance; "Early warning," lead time and accuracy rate; "Avoidance in right angle," offset distance and maximum offset angle; "Stimulation," time gap in gear shift; "Safe distance," reasonable distance; "Speed optimization," optimal speed and adjustment time. The indicators' data and the corresponding score are listed in Table 3.

6.3. Total Cost Value. Calculating the data from Table 3 to get A's total cost value as described in Section 5.2, the cost values of three typical intelligent behaviors are

$$C_1 = 24.41;$$

$$C_2 = 36.63; \tag{13}$$

$$C_3 = 48.39.$$

The total cost is $C = 35.30$.

The result shows that team A is at 2nd (20–40) level in the competition.

7. Conclusion

In this paper, the evaluation of unmanned ground vehicles is studied. Based on the typical working conditions of unmanned ground vehicles, a multilevel indicators evaluation system is established. Because the uncertainty is intrinsic property of each evaluation process, information entropy is applied to quantify the weight of each indicator, and each factor matches different weight coefficients to highlight the importance of the evaluation factor. Then entropy-cost function evaluation method is proposed to evaluate team A's unmanned vehicle in "Future Challenge 2012" competition. From the quantitative results, the teams can learn the vehicle's intelligent level generally and find their technical shortcomings in some specific indicators; thus the team will get the right development direction.

The first issue to be developed is more detailed quantification of indicators' score. The technology development is from manned driving to driverless, but manned driving behaviors are also difficult to quantify. Therefore, the cost function should be reformed for more precise cost value through both qualitative and quantitative manned driving empirical data.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] R. Behringer, S. Sundareswaran, B. Gregory et al., "The DARPA a grand challenge—development of an autonomous vehicle," in *Proceedings of the IEEE Intelligent Vehicles Symposium*, pp. 226–231, Parma, Italy, June 2004.

[2] DARPA, http://www.darpa.mil/.

[3] E. Krotkov, S. Fish, L. Jackel, B. McBride, M. Perschbacher, and J. Pippine, "The DARPA PerceptOR evaluation experiments," *Autonomous Robots*, vol. 22, no. 1, pp. 19–35, 2007.

[4] C. Urmson, J. Anhalt, D. Bagnell et al., "Autonomous driving in urban environments: boss and the urban challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 425–466, 2008.

[5] Future Challenge, http://ccvai.xjtu.edu.cn/news.

[6] G. Xiong, X. Zhao, H. Liu et al., "Research on the quantitative evaluation system for unmanned ground vehicles," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '10)*, pp. 523–527, IEEE, San Diego, Calif, USA, June 2010.

[7] X. Gao, *Research on Information Processing of Intelligent Behavior and Evaluation of Unmanned Vehicle Based on Typical Working Condition*, Beijing Institute of Technology, 2012.

[8] S. X. Liu, *Research on Evaluation of Unmanned Vehicle Intelligent U-Turn Behavior*, Beijing Institute of Technology, Beijing, China, 2014.

[9] X. Zhang, Y. N. Zhao, L. Gao, and D. H. Hao, "Evaluation framework and method of the intelligent behaviors of unmanned ground vehicles based on AHP scheme," *Applied Mechanics and Materials*, vol. 721, pp. 476–480, 2015.

[10] X. Zhang, *Research on Evaluation of Intelligent Behaviors through Specials Areas for Unmanned Ground Vehicles*, Beijing Institute of Technology, 2015.

[11] J. Wei and J. M. Dolan, "A robust autonomous freeway driving algorithm," in *Proceedings of the IEEE 2009 Intelligent Vehicles Symposium*, pp. 1015–1020, Xi'an, China, June 2009.

[12] M. Bertozzi, A. Broggi, and A. Fascioli, "Vision-based intelligent vehicles: state of the art and perspectives," *Robotics and Autonomous Systems*, vol. 32, no. 1, pp. 1–16, 2000.

[13] J. Leonard, J. How, S. Teller et al., "A perception-driven autonomous urban vehicle," *Journal of Field Robotics*, vol. 25, no. 10, pp. 727–774, 2008.

[14] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[15] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Optimal dimensioning and power management of a fuel cell/battery hybrid bus via convex programming," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 1, pp. 457–468, 2015.

[16] X. Hu, L. Johannesson, N. Murgovski, and B. Egardt, "Longevity-conscious dimensioning and power management of the hybrid energy storage system in a fuel cell hybrid electric bus," *Applied Energy*, vol. 137, pp. 913–924, 2015.

[17] S. Li, K. Li, J. Wang, X. Lian, H. Ukawa, and D. Bai, "Modeling of heavy-duty truck driver's car-following characteristics and its comparative verification," *International Journal of Automotive Technology*, vol. 11, no. 1, pp. 81–87, 2010.

[18] B. Ran, H. Tan, J. Feng, W. Wang, Y. Cheng, and P. Jin, "Estimating missing traffic volume using low multilinear rank tensor completion," *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 2015.

[19] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li, "A tensor-based method for missing traffic data completion," *Transportation Research Part C: Emerging Technologies*, vol. 28, pp. 15–27, 2013.

[20] J. Wei, J. M. Dolan, and B. Litkouhi, "A prediction- and cost function-based algorithm for robust autonomous freeway driving," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV '10)*, pp. 512–517, IEEE, San Diego, Calif, USA, June 2010.

Research Article

Traffic Management as a Service: The Traffic Flow Pattern Classification Problem

Carlos T. Calafate,[1] David Soler,[2] Juan-Carlos Cano,[1] and Pietro Manzoni[1]

[1]*Department of Computer Engineering (DISCA), Universitat Politècnica de València, 46022 Valencia, Spain*
[2]*Institute for Pure and Applied Mathematics (IUMPA), Universitat Politècnica de València, 46022 Valencia, Spain*

Correspondence should be addressed to Carlos T. Calafate; calafate@disca.upv.es

Intelligent Transportation System (ITS) technologies can be implemented to reduce both fuel consumption and the associated emission of greenhouse gases. However, such systems require intelligent and effective route planning solutions to reduce travel time and promote stable traveling speeds. To achieve such goal these systems should account for both estimated and real-time traffic congestion states, but obtaining reliable traffic congestion estimations for all the streets/avenues in a city for the different times of the day, for every day in a year, is a complex task. Modeling such a tremendous amount of data can be time-consuming and, additionally, centralized computation of optimal routes based on such time-dependencies has very high data processing requirements. In this paper we approach this problem through a heuristic to considerably reduce the modeling effort while maintaining the benefits of time-dependent traffic congestion modeling. In particular, we propose grouping streets by taking into account real traces describing the daily traffic pattern. The effectiveness of this heuristic is assessed for the city of Valencia, Spain, and the results obtained show that it is possible to reduce the required number of daily traffic flow patterns by a factor of 4210 while maintaining the essence of time-dependent modeling requirements.

1. Introduction

In densely populated urban areas, traffic-related problems, such as air quality, noise, vibration, and accidents, are critical issues for management authorities. In terms of solutions to make traffic flow more efficient or to reduce it, especially in downtowns, authorities develop initiatives to promote the use of public transportation, forbid access to the most polluting vehicles, alternate the days of downtown access according to the vehicles' plate number, charge drivers for access, and so forth. In addition to these initiatives, traffic engineers analyze the traffic flow in our cities taking into account important factors like the adequate street directions to minimize travel times, influence of traffic lights synchronization and placement in traffic congestion, fuel consumption and CO_2 emissions, traffic noise modeling [1–6], and so forth.

Particularly, in the field of fuel consumption and exhaust pollutant, Intelligent Transportation Systems (ITS) have recently emerged as a powerful ally in order to improve traffic flows [7]. Moreover, the massive adoption of smartphones and the ever increasing efforts to achieve smartphone-vehicle integration [8, 9] pave the way towards novel traffic management solutions where real-time interaction between drivers and traffic management authorities becomes possible. Such interaction provides mutual benefits since traffic authorities are able to have real-time feedback about traffic congestion states at different parts of a city, while drivers are also able to have more information, aiding them in the decision process of finding the optimal route.

In this paper we present a novel platform for centralized traffic management in urban environments which attempts to avoid known problems associated with current route planning solutions based on fixed path costs. The proposed solution takes into account the historical data about traffic patterns in order to provide time-dependent route recommendations to drivers traveling through dense traffic areas. As a first approach to deploy this solution, we propose using existing traffic measurements based on induction loop detections [10] in order to obtain all the required time-dependent traffic flow models. We focus on the specific case of the city

of Valencia, Spain, to gain further insight into the problem. Based on the results obtained, we propose a heuristic to address the problem efficiently by grouping elements with a similar behavior, and we assess the effectiveness of the proposed heuristic in terms of the number of interpolation functions required. We show that it is possible to reduce the required number of interpolations functions describing daily traffic patterns by a factor of 4210, which significantly reduces the problem complexity.

The paper is organized as follows: in the next section we introduce some related works. In Section 3 we present the proposed traffic management platform. Section 4 describes the time-dependent traffic analysis problem and provides an overview of the traffic patterns for the city of Valencia, Spain. Section 5 describes the selected heuristic to the modeling problem, along with the results achieved. Section 6 then presents the overall aggregation gains, detailing the origin of those gains. Finally, in Section 7 we conclude the paper.

2. Related Works

After several decades of research, the existing traffic engineering literature is quite broad and extensive. Recently, some solutions have emerged that rely on mobile devices to monitor the traffic in real time, for example, the Mobile Millennium [11] project. Such information can be used for administrative purposes, for example, to visually analyze the traffic conditions, but, in addition, it can also be useful to optimize the routes taken by vehicles, as shown analytically by Kim et al. [12].

Among these proposals we can find TrafficView [13], which defines a framework to gather and disseminate information about the vehicles on the road. With such a system, drivers will be provided with road traffic information that helps driving in adverse situations such as foggy weather or finding an optimal route in a long trip. Work and Bayen [14] highlight the potential of mobile devices to provide real-time traffic information for the entire transportation network, providing some case studies. Claudel et al. [15] emphasize how mobile devices may allow obtaining more reliable estimations about the time required to traverse specific routes. Leontiadis et al. [16] propose an opportunistic traffic management system where vehicles share traffic information in an ad hoc manner, allowing them to dynamically reroute based on individually collected traffic information. Recently, solutions such as EcoTrec [2] introduced a VANET-based ecofriendly routing algorithm for vehicular traffic which considers road characteristics and traffic conditions to improve the fuel savings of vehicles, thereby reducing gas emissions.

Moreover, when attempting to solve the vehicle route planning problem in the most accurate way, we must take into account the traffic variability throughout the day, as well as other situations that take place in real life when driving a vehicle [17, 18]. For instance, it is quite clear that, on large metropolitan areas, the cost of traversing certain arteries, especially large avenues, heavily depends on the time of day, being critical at peak traffic hours [19]. However, it has been proved that integrating time-dependencies in route optimization algorithms significantly increases their complexity [20, 21].

To tackle this increase of complexity, we present in this paper an approach to significantly reduce the amount of data that our platform will need to find the time-dependent shortest routes. Specifically, we detail how to aggregate large amounts of historical traffic flow data into the most meaningful set of information to properly describe traffic flow variations throughout the day on the different streets and avenues of a city.

To this aim, we will use a clustering technique. Cluster analysis is an unsupervised learning technique used for the classification of data. Data elements are partitioned into groups called clusters that represent proximate collections of data elements based on a distance or dissimilarity function. There exist two main clustering methods. The hierarchical methods basically start with each member of the set in a cluster of its own and fuse nearest clusters until there are k remaining. The partitioning methods start by building a set of k representative objects and cluster around those, iterating until (locally) optimal clustering is found. See, for example, the classical book by Kaufman and Rousseeuw [22] and Xu and Wunsch II [23].

Clustering techniques have been already used in the last years as part of ITS solutions in order to provide real insights into traffic management policies. For briefness, we only refer to some of these works. We recommend consulting Guardiola et al. [24] for further information on the topic.

For example, Wang et al. [25] present a dynamic traffic prediction model that deals with traffic flow data to convert them into traffic status. In this model, two data mining techniques, the clustering analysis and the classification analysis, are applied to historical traffic flow data. Caceres et al. [26] present a methodology for estimating traffic flows using road features as clustering variables, so that it can be applied to any road section, even without detector data. More recently, Yildirimoglu and Geroliminis [27] partition the historical data set from loop detectors on Californian freeways in clusters with similar characteristics based on the traffic patterns observed on the roadway. The building block of their methodology is the development of stochastic congestion maps, which identify the probability that a space-time domain is congested. Finally, Guardiola et al. [24] present a new methodology for analyzing the daily traffic flow profile using Functional Data Analysis. They claim that their methodology allows a maximum exploitation of the recorded historical data and results in the detection of changes in the flow pattern, which would otherwise be difficult to detect via classical statistical methods.

3. Traffic Management as a Service

Current vehicle navigation systems are typically based on locally stored static information from which routes are calculated. Among such systems we can find commercial applications like TomTom (http://www.tomtom.com/) or Garmin (http://www.garmin.com/). There are also free tools, like

Google Maps Navigator and OsmAnd (http://osmand.net/) that operate in a similar manner. The main drawbacks of navigation systems based on static information are the inability to adapt to traffic congestion states or unexpected events, like accidents or other problems on the road, which cause travel times to be much higher than expected.

More sophisticated route navigation solutions update route information in real time, based on reported traffic conditions. As an example, the TomTom navigation software has been enhanced to support client-server interaction in order to inform clients about alternative routes when atypical traffic delays are detected.

In this paper we will address the specific problem of traffic congestion in urban environments. Instead of accidents and other conditions causing atypical delays, we will focus on predicting daily traffic flow patterns for a specific urban environment, detailing how it is possible to reduce travel times based on historical information about the traffic density distribution throughout the day.

The proposed traffic management platform is named ABATIS: *Automatic Balancing of Traffic through the Integration of Smartphones with vehicles*. The main novelty of ABATIS as a route planning system is providing time-dependent route recommendations based on traffic congestion history. Specifically, it offers client-server interaction, where the route selection process is performed at the route server (see Figure 1) based on real-time information stored in the route database and historical data. The traffic analysis and visualization server allows making traffic congestion forecasts based on historical data while also allowing traffic management authorities to check the traffic conditions in real time.

Clients contribute to improving the route database information by providing real-time feedback about traffic congestion conditions, which allows maintaining both a real-time map of traffic fluidity in a city and accurate historical data of traffic behavior. This approach supports global traffic load balancing and event-based management (e.g., reducing traffic congestion in the route of an ambulance).

This strategy, although offering significantly better routes, has a higher cost since the estimated time for traversing each path segment will no longer be a fixed value based on segment length and speed limit, but instead it will vary dynamically along the day. In order to achieve time-dependent costs for the different streets and avenues in a city, ABATIS will use existing historical data about traffic logs in a city to estimate travel times. Since such logs provide per-hour congestion measurements for all induction loop detectors in a city for a whole year, they must be properly summarized and synthesized by the traffic analysis server to allow seamlessly integrating such information in the route server. Thus, in the remainder of the paper, we will focus on the traffic analysis component, proposing a heuristic able to reduce the complexity of the problem by converting huge amounts of historical data about traffic intensity into a small but representative set of daily patterns able to describe the expectable traffic behavior in the city along the day.
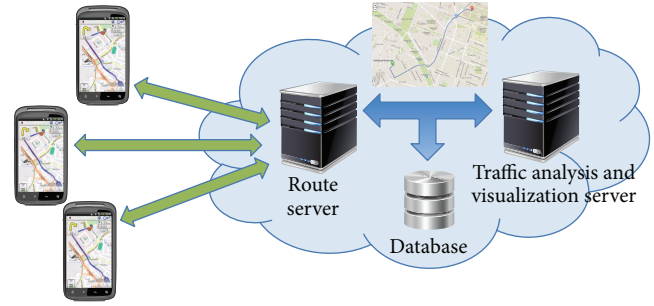
FIGURE 1: ABATIS traffic management architecture.

4. Flow Pattern Classification Problem

Attempting to model the daily traffic flow pattern of hundreds of streets/avenues for every day of the year would lead to hundreds of thousands of interpolation functions able to provide a smooth description of per-street traffic flow variations throughout the day, based on several million input values (assuming a per-hour granularity). Such modeling effort for a single city can be considered excessive and, in addition, causes route recommendation tasks at the server to have an extremely high computational cost. Nevertheless, when attempting to provide an accurate characterization of path segment costs in a specific urban environment, it quickly becomes clear that (i), from a yearly perspective, seasonal differences are expectable as, for example, more people use their vehicles during cold weather seasons than during the warm and hot seasons where, for example, bicycles or public transport can become a more attractive alternative; (ii), from a weekly perspective, labor days are characterized by mobility patterns and traffic congestion states that drastically differ from the behavior during weekends and holidays; (iii), from an hourly perspective, different hours of the day are associated with different congestion levels (e.g., day versus night); and finally (iv), from a spatial perspective, different streets/avenues have different traffic levels at any time of the day, requiring independent modeling.

Taking the aforementioned factors into consideration, in this section we will take an in-depth look into traffic behavior when focusing on a medium-size European city like Valencia, Spain, which is the third largest metropolitan area in Spain with about 1.77 million inhabitants. Detailed trace files containing the amount of traffic flowing in each of the streets/avenues each hour for a full year (2013) were provided to us by Valencia's City Hall Traffic Department, in particular, data concerning the 421 most relevant streets/avenues (those monitored by traffic services through induction loop detectors).

Our goal is to obtain insight into the traffic flow, detecting traffic patterns according to the day of the week, hour, and type of street. Based on the traffic patterns detected, we will propose a heuristic in order to simplify the number of models required while maintaining most of the time-dependent modeling effectiveness. Although we use the city of Valencia as the target of our analysis, the modeling methodology followed is quite general, being applicable to other cities as well.
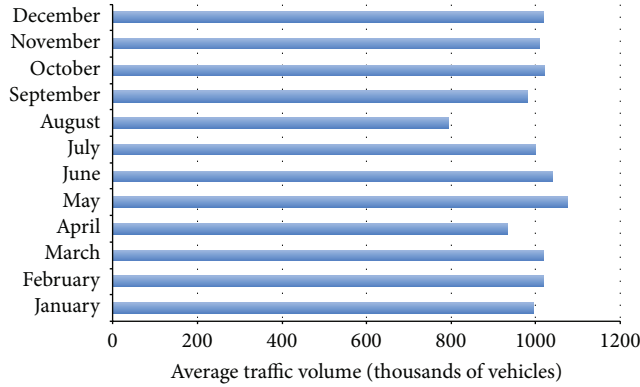
FIGURE 2: Average traffic volume in Valencia per month.
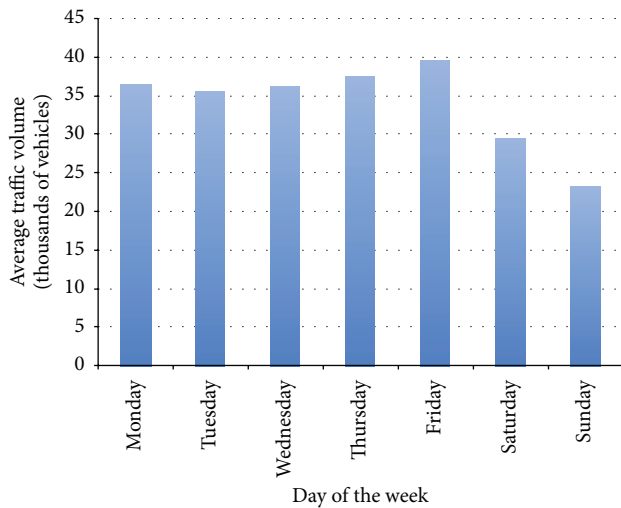
FIGURE 3: Average traffic volume in Valencia for the different days of the week.

We start by analyzing the monthly traffic, assessing whether we can detect significant seasonal differences. As shown in Figure 2, there are minor fluctuations in terms of overall traffic on a monthly basis. It quickly becomes evident that holiday periods, like August and also Easter (in April), have a clear and expectable impact on the overall traffic volume. For the remaining months of the year the values can be considered relatively similar, having a mean value of about 1 million vehicles.

For the analysis that follows we picked a month with an average overall traffic volume close to the mean; specifically, we selected November, which has no holiday periods. Focusing on the traffic pattern variation throughout the week, Figure 3 shows that there are very significant differences between the days of the week, especially between the weekend and weekdays. Also, we can observe an overall increasing trend from Monday to Friday, with Friday being the weekday with higher traffic volume.

In addition to the differences in terms of daily traffic volume, there are also clear differences in terms of the daily traffic pattern itself. For instance, Figure 4 shows that on

Mondays the traffic follows a typical pattern where the peak hour is between 8 and 9 a.m., when most people go to work. Another peak occurs between 2 and 3 p.m., which denotes mobility from people working in the afternoon. Finally, a last traffic peak is detected between 6 and 8 p.m., when workers return to their homes. Other weekdays follow a similar pattern.

A totally different pattern is detected, for example, on a Sunday. Compared to weekdays we find that (i) work-related traffic peaks are no longer present; (ii) the total traffic volume is significantly lower; and (iii) the peak hours differ. In particular, peak hours are now related to mobility towards food courts at lunch time (between 1 and 2 p.m.) and mobility from relax areas to homes (between 6 and 8 p.m.).

When focusing on the traffic distribution throughout a city, it is well known that main streets and avenues will experience a much higher traffic load than secondary and isolated ones. Discriminating between them is a relevant issue since some streets barely experience any traffic load increase during peak hours, meaning that travel times are not affected by congestion in the same way as the main arteries of the city.

To be able to discriminate between the streets of Valencia based on traffic flow, we first obtained the peak traffic intensity per street during November, and we then obtained the cumulative distribution for these values (see Figure 5).

We observe that 30.3% of all streets have a traffic intensity lower than 690 vehicles/hour during peak hours, which according to [28] means that these low traffic intensity streets will not experience traffic congestion even at peak hours, and so they can be discarded from our time-dependent modeling efforts. Additionally, we observe that the number of streets/avenues with very high traffic volumes (more than 10.000 vehicles during the peak hour) is rather limited (about 10%). Thus, the majority of the streets in a city will experience moderate traffic volumes, and the global peak hour behavior will not cause any noticeable effect on these streets. To confirm this observation, Figure 6 shows the traffic load per hour in two different streets for the same day. Notice that although both share quite similar values for peak traffic intensity, the daily traffic patterns significantly differ that the peaks in one pattern often match valleys in the other pattern.

Observing the daily traffic pattern in Figure 6(a), we find that it closely matches the traffic pattern of a typical Monday, as shown in Figure 4(a); on the contrary, Figure 6(b) shows a quite different traffic pattern. Hence, it becomes necessary to discriminate between the different streets based on their daily traffic pattern. To achieve this goal, we will apply a clustering technique in order to automatically classify streets according to their daily traffic pattern.

5. Clustering Heuristic

In this section we propose a heuristic to simplify traffic modeling for the city of Valencia by taking into consideration the results presented in the previous section.

The proposed heuristic aggregates into a single pattern all those daily traffic patterns having a common behavior. This is made possible by making the obtained time-dependent

(a) Monday

(b) Sunday
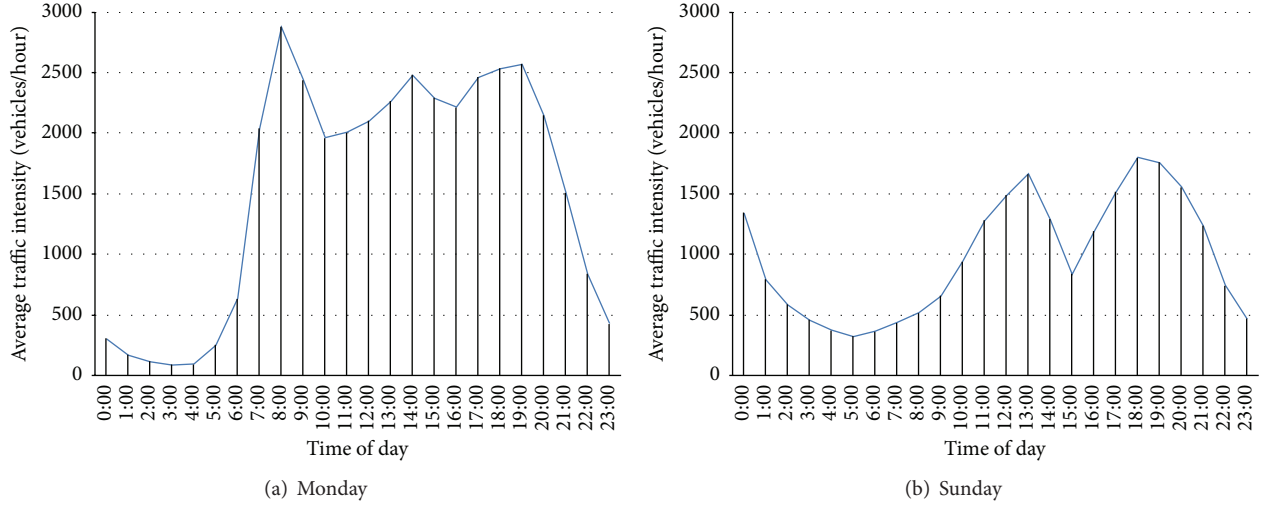
FIGURE 4: Average daily behavior for different days of the week.
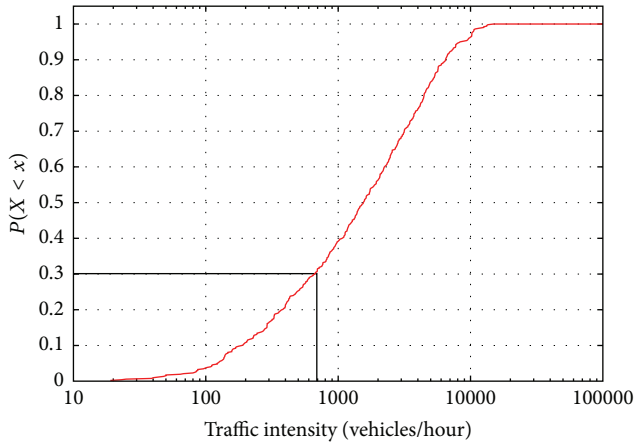
FIGURE 5: Cumulative distribution for traffic intensity using the monthly peak hours.

models independent of the actual number of vehicles in each street through normalization using the mean daily value.

To this aim, we use Mathematica 9.0.1 [29], which is a widely recognized tool to solve mathematical problems, especially in engineering. This tool provides function *FindClusters*, which returns the number of clusters as well as the elements on each cluster. This function has several options and suboptions. In fact, we can choose between a hierarchical method or a partitioning method. The partitioning method it uses is based on the Partitioning Around Medoids (PAM) algorithm [22], which seeks to find k representative objects called medoids from the data set such that the sum of the dissimilarities within a cluster are minimized. A medoid can be defined as that object of a cluster whose average dissimilarity to all the objects in the cluster is minimal. After finding the set of medoids, each object of the data set is assigned to the nearest medoid.

We have chosen the partitioning method of *FindClusters* for two reasons. The first one is that this method is the default

option, and the second and most important one is that the PAM algorithm is the one used by reference authors on the topic such as Guardiola et al. (see [24]), who claim that the choice of PAM is due in part to the large number of statistics it provides for thorough analysis of the resultant clusters.

At this point, we want to stress the fact that while [24] (and also [27]) try to cluster different days corresponding to the same section of a freeway, the aim of our procedure is quite different; particularly, we attempt to cluster different streets corresponding to the same day. Moreover, as far as we know, the clustering distance that we will use here has not been used in any previous paper on ITS.

Finally, note that although we have not made use of them, function *FindClusters* has suboptions in order, for instance, to fine-tune the number of clusters. Probably the best known suboption to do this is the silhouette statistic [22], but according to [23] there is no criterion providing evidence about its superiority compared to others in the general case of adjusting the number of clusters. In addition, notice that two properties that define a good heuristic and that we have taken into account to our aim are low time overhead and simplicity of its steps.

Below we describe the five steps followed to reduce the number of independent daily patterns to be modeled: (i) select the appropriate clustering metric, (ii) find the optimal number of clusters per day of the week, (iii) determine how representative mean days are, (iv) group days of the week with similar characteristics, and (v) group clusters with similar daily patterns.

5.1. Selection of a Clustering Metric for Per-Hour Street Behavior. If for each street (or street segment) we have the number of cars that traverse it every hour, we can represent each street by a point $x = (x_1, x_2, \ldots, x_{24})$ in $\mathbb{R}^{24}$, where x_i is the number of cars traversing the street at hour i. Suppose we have two streets $x = (x_1, x_2, \ldots, x_{24})$ and $y = (y_1, y_2, \ldots, y_{24})$. By default, the distance used to form clusters is the Euclidean

(a) Street following the expected pattern

(b) Street not following the expected pattern
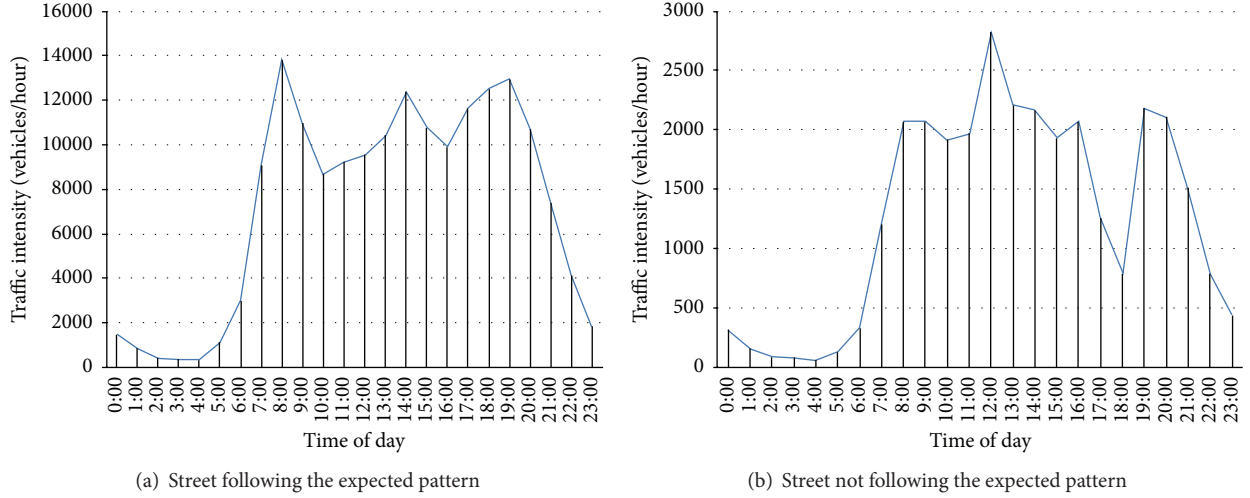
FIGURE 6: Daily traffic intensity pattern for streets with different characteristics.

distance, $\sqrt{\sum_{i=1}^{24}(x_i - y_i)^2}$. If the Euclidean distance between two points is relatively small, both streets will belong to the same cluster. However, if we attempt to classify streets taking into account the traffic variability as a function of the time of day, we believe that this distance is not adequate. Let us take a small illustrative example in this regard. Suppose that we only consider six consecutive hours for four different streets and that their respective points are $a = (12, 9, 10, 9, 8, 11)$, $b = (24, 20, 22, 20, 17, 21)$, $c = (6, 16, 20, 25, 17, 7)$, and $d = (15, 35, 44, 48, 34, 17)$.

Streets a and b have a similar behavior: the relative number of vehicles traversing them every hour is more or less the same, within certain bounds. Although the actual number of vehicles differs greatly from one street to another, both streets should be in the same group encompassing all those streets where there is little traffic variability, where vehicle speeds can be considered mostly constant over the considered period.

With respect to streets c and d, central hours are peak periods where we have about three times the traffic volume compared to edge values. Although the number of vehicles differs greatly from one street to another, they should belong to the same group characterized by a single peak corresponding to hours in the mid-range and with much lower values on the edges.

However, if we classify the four streets using the Euclidean distance, the result is quite predictable: $\{a, c\}$ and $\{b, d\}$. In this example the Euclidean distance has created two clusters grouping the two streets with less traffic and the two streets with high traffic volume. To address this problem, we believe that the distance metric that best fits our objective is the correlation distance, defined as $1 - |r_{xy}|$, where r_{xy} is the correlation coefficient:

$$r_{xy} = \frac{\sum_{i=1}^{24}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sqrt{\sum_{i=1}^{24}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{24}(y_i - \overline{y})^2}}. \tag{1}$$

Recall that $|r_{xy}|$ is always less than or equal to 1 and that values close to 1 indicate that variables x and y have a direct linear relationship, meaning that the graphical representation of the 24 points (x_i, y_i) is approximately a straight line. Therefore, the higher the correlation between points x and y is, the closer to zero $1 - |r_{xy}|$ becomes, and so the probability of belonging to the same cluster will increase. If we classify the four streets according to correlation distance, the result obtained is the desired one: $\{a, b\}$ and $\{c, d\}$.

On the other hand, it is easy to see that the correlation distance is the same if we work with the coordinates (x_i, y_i) or with coordinates $(x_i/\sum_{j=1}^{24} x_j, y_i/\sum_{j=1}^{24} y_j)$, taking into account that, to compare streets considering traffic variability throughout the day, it also seems useful to compare the percentage of the daily traffic passing on every street for each hour. This way, it does not matter whether we compare both streets considering the number of cars per hour or the percentage of traffic per hour: the classification using the correlation distance will generate the same clusters. This is obviously not true when adopting Euclidean distances.

5.2. Finding the Optimal Number of Clusters for Each Day of the Week. Using the correlation distance defined previously, in this section we will determine the optimal number of clusters for the 292 streets in Valencia considered by the City Hall as representative in terms of traffic flow for every day of the week. Subsequently, to reduce the overall number of clusters, we will attempt to join the different days in a week whenever the same number of clusters are detected.

Therefore, for our analysis, we apply the *FindClusters* function to each of the 28 days of November studied enabling the correlation distance option. For each day, the function will cluster the 292 points in $\mathbb{R}^{24}$ corresponding to the streets taken for our study.

In the analysis that follows we work with the percentage of vehicles traversing each street every hour with respect to the overall daily value. As referred in the previous section, the actual number of vehicles *per se* is not relevant to

TABLE 1: Number of clusters obtained and associated statistics.

	Mo	Tu	We	Th	Fr	Sa	Su
A: Week 1	3	3	1	2	1	2	3
B: Week 2	1	2	4	1	3	2	4
C: Week 3	5	3	1	4	2	2	3
D: Week 4	3	1	1	3	3	2	1
E: mean(A, B, C, D)	3	2.25	1.75	2.5	2.25	2	2.75
F: median(A, B, C, D)	3	2.5	1	2.5	2.5	2	3
G: average day	4	2	2	4	3	2	2
H: round(E) == G	False	True	True	False	False	True	False
I: mean(E, F, G)	3.3(3)	—	—	3	2.58(3)	—	2.58(3)
Number of clusters	**3**	**2**	**2**	**3**	**3**	**2**	**3**

TABLE 2: Percentages of matching for the different clusters compared to the average day clusters.

	Mo	Tu	We	Th	Fr	Sa	Su
Number of clusters	**3**	**2**	**2**	**3**	**3**	**2**	**3**
	83.11	92.31	84.42	30.86	70.15	91.98	73.72
Week 1	66.67	84.56	62.32	59.32	68.75	71.43	66.67
	81.33			56.99	90.43		35.82
	60.14	89.74	80.52	81.48	70.15	96.26	74.36
Week 2	55.07	58.09	59.42	43.22	60.64	81.90	69.56
	80.00			58.06	51.56		73.13
	62.84	84.62	80.52	58.02	70.15	88.77	51.28
Week 3	69.57	32.35	91.30	75.42	74.47	84.76	47.83
	84.00			31.18	35.94		89.55
	81.76	96.15	74.68	62.96	86.57	97.87	82.05
Week 4	88.41	84.56	82.61	74.58	65.96	53.30	59.42
	76.00			65.59	56.25		58.21
Average	**73.63**	**78.68**	**77.05**	**58.56**	**69.18**	**86.21**	**66.70**

our purposes, and the correlation distance metric adopted provides the same output on both cases.

Since our study period encompasses 4 weeks, we create an "average day" for each day of the week, which is calculated for each street by averaging the number of vehicles traversing it each hour. Such "average day" attempts to filter out the peculiarities of a specific day, obtaining a representative trend instead.

Table 1 shows the results obtained, where the last row shows the cluster allocation for each day of the week. To attain those values, we first apply function *FindClusters* to different weeks (A–D) and to the "average days" (G). In addition, we calculate the mean (E) and the median (F) for the cluster groups corresponding to the different weeks. If this mean value (E) is rounded to a number that matches the number of clusters for the average day (G), then we define such value as the number of clusters for that day of the week. Otherwise, we obtain the average of the mean (E), median (F), and average day (G) to obtain a value (I) that when rounded defines the number of clusters to be used. We find that the proposed

number of clusters matches the rounded mean (E) except for a minor change in one day.

5.3. Determining Cluster Matching on a Per-Day Basis. Once the number of clusters for each day of the week was defined, the next step was to validate that cluster elements for each day of the week resembled the cluster elements obtained for the average day. If a good degree of matching is obtained, then the conclusions associated with streets in that cluster are valid; otherwise, we could be considering that streets belong to a group with a specific behavior, when in fact their behavior significantly differs.

For our endeavor we apply the *FindClusters* function to the 35 days (28 real days plus 7 average days), but this time fixing the number of clusters defined a priori, as obtained in the previous section. Afterwards, for each of the four weeks under analysis, we compare the clusters obtained against the average day of the week, determining the percentage of streets that both clusters have in common. These results are presented in Table 2.

Table 3: Percentages of cluster matching for average days of the week with same number of assigned clusters. Valid combinations are shown in boldface.

Combinations	Degree of matching (%)	Average matching (%)
Monday-Thursday	57.43	
	8.69	48.29
	66.67	
Monday-Friday	77.70	
	72.46	68.84
	48.00	
Monday-Sunday	58.78	
	27.54	43.49
	28.00	
Thursday-Friday	37.04	
	71.19	59.25
	63.44	
Thursday-Sunday	30.86	
	55.93	41.78
	33.33	
Friday-Sunday	61.94	
	36.17	51.37
	51.56	
Tuesday-Wednesday	**91.67**	
	91.91	**91.78**
Tuesday-Saturday	71.15	
	44.11	58.56
Wednesday-Saturday	69.48	
	42.03	56.51

We find that the average degree of matching for all the days of the week is 72.71%. Globally, we find that this value is quite acceptable and that differences appearing on specific days are expectable since traffic patterns may suffer some changes depending on weather, special events, or other conditions.

5.4. Grouping Days of the Week with Similar Cluster Characteristics. The next step of our clustering procedure was to assess the feasibility of grouping those days of the week having the same number of clusters. With this purpose we tested all combinations and calculated the percentage of cluster matching for each pair of mean days of the week. The results are shown in Table 3.

All combinations show an average degree of matching below 70%, except for the Tuesday-Wednesday combination which is close to 92%. Thus, we agree that these two weekdays can be combined as if they were a single day since similar patterns are obtained in terms of traffic variability throughout the day. Data shown earlier in Figure 3 also emphasize this similitude.

To confirm that the grouping did not have a negative impact on the error associated with specific days, we now proceed to compare the degree of matching for the different clusters against the average day, the crossed average day, and the proposed union of both days. These results are shown in Table 4.

We find that the differences between the three cases are quite low. Specifically, the impact of grouping these two days into one is of only 1.6%, which is quite acceptable. The results using cross averages also strengthen the point of unifying these two days. As a result, by accounting for the number of clusters of each average day and by merging Tuesday and Wednesday into a single day, we obtain a total of 16 different traffic patterns.

5.5. Grouping Clusters with Similar Daily Patterns. In this section we present the normalized traffic patterns corresponding to the 16 clusters created: 3 for Monday, 2 for Tuesday/Wednesday, 3 for Thursday, 3 for Friday, 2 for Saturday, and 3 for Sunday.

As shown in Figure 7, there are some pattern similitudes between the first weekdays (Monday versus Tuesday/Wednesday), between the last weekdays (Thursday versus Friday), and between weekend days (Saturday versus Sunday). However, this initial insight obtained visually must be confirmed through statistical evidence. With this purpose we picked the clusters for those days which visually show some similitude and calculated the correlation between the daily patterns associated with each cluster for relevant time ranges. The results of these analyses are presented in Table 5.

When comparing the daily pattern for the clusters of Monday against Tuesday/Wednesday (see Table 5(a)), we find that there is a high correlation (>92%) between the patterns corresponding to the first 2 clusters of each of these days. Thus, a single model will suffice when attempting to represent the daily pattern for these clusters that only a different model is required for Monday's Cluster number 3.

When comparing Thursday against Friday, we find that only Cluster number 2 for Thursday and Cluster number 1 for Friday present a high correlation (~94%).

Finally, when comparing Saturday against Sunday, we find that Cluster number 1 and Cluster number 3 present a good degree of matching (~94%), and these two clusters can also be represented through same daily pattern.

6. Generalization and Benefits of the Proposed Model

In this section we assess the benefits of our model in terms of the minimum number of patterns required to adequately describe traffic intensity throughout the day for the city of Valencia. Then, we detail how these different models obtained can be integrated in our traffic management platform to predict route costs. Finally we summarize our proposal by presenting the proposed heuristic in pseudocode format to allow generalizing the proposed procedure to any target city.

6.1. Aggregation Gains Achieved. Below we discuss the different aggregation techniques that integrate our heuristic and the previous analysis.

Table 4: Percentages of matching for the different clusters against the average day, the crossed average day, and the proposed union of both days.

	Original average days		Crossed average days		Union of average days	
	Tu	We	Tu	We	Tu	We
Week 1	92.31	84.42	90.26	86.54	86.83	86.23
	84.56	62.32	81.16	65.44	84.00	69.60
Week 2	89.74	80.52	89.61	78.21	85.63	77.25
	58.09	59.42	57.25	58.09	56.80	60.00
Week 3	84.62	80.52	86.36	78.85	82.04	72.46
	32.35	91.30	34.06	90.44	30.40	88.00
Week 4	96.15	74.68	95.45	73.08	92.22	68.26
	84.56	82.61	82.61	81.62	86.40	80.00
Average	**78.68**	**77.05**	**77.82**	**76.71**	**77.14**	**75.34**

Table 5: Correlation between clusters (period between 7 a.m. and 9 p.m.).

(a) Monday and Tuesday/Wednesday

		Tuesday/Wednesday	
		Cluster number 1	Cluster number 2
Monday	Cluster number 1	**0.9221668**	0.578729
	Cluster number 2	0.6229643	**0.9422671**
	Cluster number 3	0.5900097	0.7910942

(b) Thursday and Friday

		Friday		
		Cluster number 1	Cluster number 2	Cluster number 3
Thursday	Cluster number 1	0.6741969	0.2552095	0.7292981
	Cluster number 2	**0.9393144**	0.6691599	0.6666628
	Cluster number 3	0.7247197	0.7841533	0.8645128

(c) Saturday and Sunday

		Sunday		
		Cluster number 1	Cluster number 2	Cluster number 3
Saturday	Cluster number 1	0.8859214	0.8585393	**0.9368844**
	Cluster number 2	0.8948805	0.8840648	0.7977545

Yearly Analysis. The monthly behavior results shown before allow assuming that traffic volumes throughout the year are mostly constant, except for vacation periods like summer and festivities lasting for long periods (e.g., Easter), meaning that partitioning weeks into three groups (typical week, relevant holiday period, and summer holidays) seems appropriate.

Monthly Analysis. Results have shown that, for the same type of period, data is consistent across weeks, which allows clustering the different days of a month in a single *average representative week*.

Traffic Intensity Analysis. Concerning traffic congestion for the different streets and avenues of a city, our heuristic assumes that only a subset of these streets/avenues actually face significant congestion problems deserving time-dependent modeling, while for the rest, the use of traditional fixed-cost approaches suffices. Based on the thresholds defined in [28] for class IV (urban) arterial types, we consider that only those streets with a peak traffic value surpassing 690 vehicles per hour are actually experiencing congestion-related traffic delays. This way, the target number of streets/avenues can be reduced from 421 (total number of streets being monitored by traffic services) to 292 (number of streets with a relevant traffic load).

Clustering Analysis. Focusing on the street/avenue subset significantly affected by congestion, the clustering analysis showed that a small number of groups can be created, where for each group all streets/avenues follow very similar traffic congestion patterns. Thus, the target number of models required can be reduced from 292 per 7 days in a week to a total of 18, and this value can be further reduced to 16 by noticing the similarity between Tuesday and Wednesday.

Daily Pattern Analysis. An analysis of the daily patterns associated with the different clusters defined for the different days of the week has shown that some of these clusters have
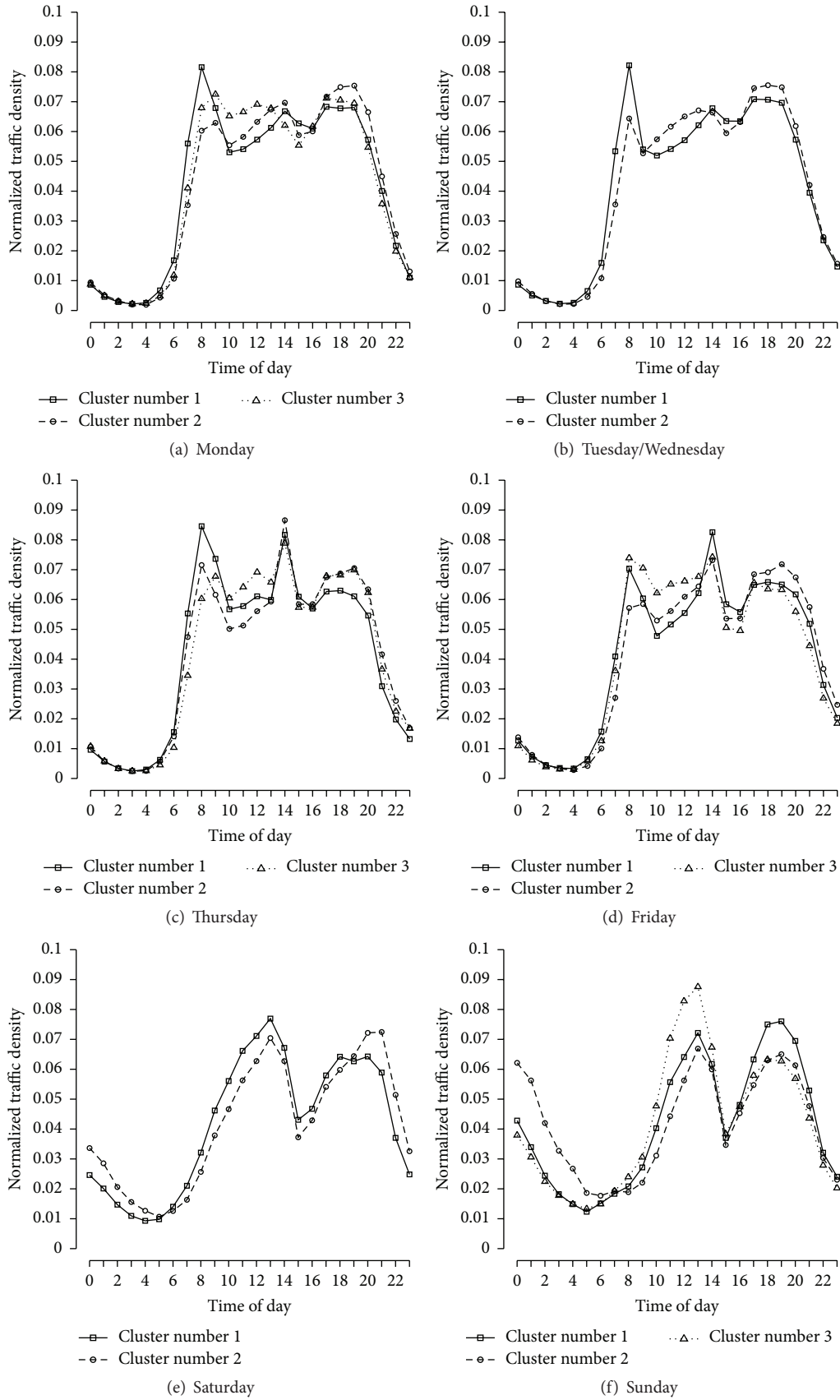
(a) Monday

(b) Tuesday/Wednesday

(c) Thursday

(d) Friday

(e) Saturday

(f) Sunday

FIGURE 7: Cluster description for the different average days considered.

TABLE 6: Benefits of the proposed heuristic in terms of aggregation gain.

Target heuristic	Number of elements	Aggregation gains	Independent modeling domains
Monthly patterns per year	12	12 : 3	3
Daily patterns per month		30 : 7	
Traffic intensity analysis		421 : 292	
Street clustering	421×30	$(292 \times 7) : 18$	12
Similar days clustering		18 : 16	
Daily pattern analysis		16 : 12	
Total	**151,560**	**4210**	**36**

a common behavior. This means, in general, that the same group of streets behaves similarly across different days, which allows slightly reducing the number of patterns from 16 to 12.

Based on aforementioned aggregation proposals for the city of Valencia, in Table 6 we detail the benefits obtained in terms of model simplification. As can be observed, street clustering is the key element when reducing the number of separate modeling domains required to characterize the traffic behavior throughout the year. In particular, aggregation based on the clustering analysis is the most critical one, allowing for substantially reducing the number of interpolation functions required. The second most relevant aggregation gain is associated with yearly and week behavior, based on segregating work periods from short/long holiday periods and by finding that we have the same behavior across the different weeks. Eliminating secondary streets that experience fluid traffic throughout the whole year also provides some contribution in terms of aggregation gain by eliminating the need for modeling their traffic throughout the day. Finally, the daily pattern analysis across clusters has further helped reducing the number of models required.

Overall, the proposed heuristic allows reducing the required number of interpolation functions for the city of Valencia by a factor of 4210 while maintaining the essence of time-dependent modeling requirements. Such a significant reduction certainly simplifies the integration of these models in our ABATIS platform and allows accelerating the associated calculations. This way, route decisions are taken in a centralized route server based on traffic states prediction throughout the day and for the different streets/avenues of a city, thus providing the most time-efficient routes.

6.2. Applicability of the Model in the Context of ABATIS.

The relationship between traffic flow levels and average travel speed is a well-known topic in traffic flow theory [30]. As shown in Figure 8, this relationship can be closely approximated through a parabolic behavior represented through the following expression, obtained by interpolating points $(0, 0)$, $(s_f/2, v_m)$, and $(s_f, 0)$:

$$v(s) = \frac{4 \cdot v_m}{s_f^2} \cdot s \cdot (s_f - s).$$ (2)

As expected, average travel speed starts to decay when traffic density per lane increases beyond a certain threshold
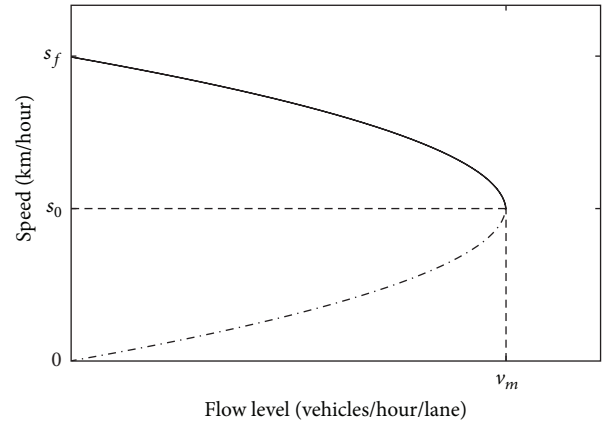
FIGURE 8: Relationship between vehicle flow level and vehicle speed.

and becomes close to zero when approaching the maximum road capacity.

Since our models required a normalization of traffic levels of each street in order to perform model aggregation for similar patterns, given a street and an instant of time a vehicle is expected to enter the street, we show below the four steps involved in calculating the travel time for that street starting at the given instant of time. Note that, for simplicity, we do not put to the variables the subindexes corresponding to the given street and instant of time.

(i) Find the normalized traffic intensity (pattern) n at the time the vehicle is expected to enter the target street, using the daily pattern for the target street.

(ii) Obtain the expected traffic flow level v for that street and instant of time by denormalizing the obtained value using the mean traffic volume $\bar{v}$ for the target street:

$$v = n \cdot \bar{v}.$$ (3)

(iii) Based on the average free-flow speed s_f and the maximum flow v_m for the target street (provided by authorities), the expected travel speed s_v can be obtained based on the predicted traffic flow level v. Specifically and taking the behaviour of Figure 8 corresponding

```
input: 3D array of traffic density per street, per hour, per day
output: pattern-dependant cluster classification
BEGIN
for each street in All_streets do {
    if (peak_traffic_intensity in P < 690 veh/h) then
        remove street from All_streets
}
for each Week_day in WEEK_DAY do {
    average_Week_day = get_average_pattern(Week_day)
    clusters[] = FindClusters(Week_day, average_Week_day)
    mean_clusters = get_average(clusters[Week_day])
    median_clusters = get_median(clusters[Week_day])
    if (clusters[average_Week_day] == round(mean_clusters)) then
        num_clusters[week_day] = clusters[average_Week_day]
    else
        num_clusters[week_day] = round(get_average(mean_clusters,
                median_clusters, clusters[average_Week_day]))
}
for all week_day pairs (w_i, w_j)
where num_clusters[w_i] == num_clusters[w_j] do {
    if (Matching(cluster_elements(w_i), cluster_elements(w_j)) > 90%)
        then pattern[w_i] = pattern[w_j]
}
for all week_day pairs (w_i, w_j) with different pattern do {
    for all clusters c_i in w_i and c_j in w_j do {
        if (correlation(average_street(c_i), average_street(c_j)) > 0.9)
            then
                pattern[c_i] = pattern[c_j]
    }
}
RETURN cluster pattern classification
END algorithm
```

ALGORITHM 1: Cluster patterns.

to *below flow saturation levels* as reference (solid line section), s_v can be approximated as follows:

$$s_v = \frac{s_f}{2} \cdot \left(1 + \sqrt{1 - \frac{v}{v_m}} \right). \qquad (4)$$

(iv) Calculate the travel time t for the target street with length L using the expected travel speed:

$$t = \frac{L}{s_v}. \qquad (5)$$

Notice that, since the ABATIS platform is able to offer, among others, Traffic Management as a Service, it is able to serve optimal routes to clients. Currently, route costs are calculated using free-flow speeds. Thus, the proposed models can be integrated in the route calculation engine so that optimality conditions now account for the updated path costs using our predictive model. In addition, if the current status of the traffic flow is available in the future, it can be combined with the predicted value to further improve path cost accuracy.

6.3. Pseudocode for the Proposed Heuristic. Let P represent the time period under analysis and let *Week_day* represent

the set of days in P corresponding to a particular day of the week. *WEEK_DAY* is a superset containing all *Week_day* sets and *All_streets* represents the set containing all the streets for the target city.

Algorithm 1 shows the pseudocode that allows applying the proposed heuristic in a systematic manner, thereby making it applicable to any target city.

7. Conclusions

Traffic management has evolved substantially in the last decades. Nowadays, traffic engineers require effective solutions to help them improve the traffic flow in cities, while minimizing travel times and tackling traffic-related problems such as CO_2 emissions, noise, and accidents.

In this paper we define a procedure to obtain reliable traffic congestion estimations for all the streets/avenues in a city for the different times of the day and for every day in a year. Considering the modeling effort required, we proposed a heuristic that allows reducing the number of required interpolation functions characterizing daily traffic patterns.

By specifically addressing the city of Valencia, we made a detailed analysis of traffic behavior on the different streets/ avenues of the city to determine (i) the behavior along the

year, (ii) which days of the week show a similar pattern, (iii) which streets/avenues experience more traffic congestion, and (iv) how streets can be grouped into clusters based on their daily traffic pattern. The results of our analysis show that it is possible to model the traffic behavior in the city by aggregating elements with a similar behavior in the same interpolation function. This way, we will be able to account for the travel time variations along the main paths of a city, providing users with both optimized and accurate travel plans, while reducing the modeling complexity.

As future work we will develop a smartphone application that interacts with the ABATIS platform in order to obtain the most efficient routes, and we will implement a route planning algorithm that allows selecting these best paths while accounting for time-dependencies, FIFO restrictions, turn penalties, and so forth.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] S. Salcedo-Sanz, D. Manjarrés, Á. Pastor-Sánchez, J. Del Ser, J. A. Portilla-Figueras, and S. Gil-López, "One-way urban traffic reconfiguration using a multi-objective harmony search approach," *Expert Systems with Applications*, vol. 40, no. 9, pp. 3341–3350, 2013.

[2] R. Doolan and G.-M. Muntean, "VANET-enabled eco-friendly road characteristics-aware routing for vehicular traffic," in *Proceedings of the IEEE 77th Vehicular Technology Conference (VTC Spring '13)*, pp. 1–5, Dresden, Germany, June 2013.

[3] A. Corti, V. Manzoni, and S. M. Savaresi, "Simulation of the impact of traffic lights placement on vehicle's energy consumption and CO_2 emissions," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation Systems (ITSC '12)*, pp. 620–625, Anchorage, Alaska, USA, September 2012.

[4] Y. Xu, Y. Zhang, and M. Liu, "Multiagent based decentralized traffic light control for large urban transportation system," *Mathematical Problems in Engineering*, vol. 2014, Article ID 104349, 13 pages, 2014.

[5] F. Pop and C. Dobre, "An efficient pagerank approach for urban traffic optimization," *Mathematical Problems in Engineering*, vol. 2012, Article ID 465613, 9 pages, 2012.

[6] V. Nedic, S. Cvetanovic, D. Despotovic, M. Despotovic, and S. Babic, "Data mining with various optimization methods," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3993–3999, 2014.

[7] M. K. Nasir, R. Md Noor, M. A. Kalam, and B. M. Masum, "Reduction of fuel consumption and exhaust pollutant using

[8] G. Alli, L. Baresi, A. Bianchessi et al., "Green move: towards next generation sustainable smartphone-based vehicle sharing," in *Proceedings of the Sustainable Internet and ICT for Sustainability (SustainIT '12)*, pp. 1–5, IEEE, Pisa, Italy, October 2012.

[9] J. Zaldivar, C. T. Calafate, J. C. Cano, and P. Manzoni, "Providing accident detection in vehicular networks through OBD-II devices and android-based smartphones," in *Proceedings of the 36th Annual IEEE Conference on Local Computer Networks (LCN '11)*, pp. 813–819, IEEE, Bonn, Germany, October 2011.

[10] M. Westerman, R. Litjens, and J. Linnartz, "Integration of probe vehicle and induction loop data: estimation of travel times and automatic incident detection," Tech. Rep., 1996.

[11] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobson, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment," *Transportation Research, Part C: Emerging Technologies*, vol. 18, no. 4, pp. 568–583, 2010.

[12] S. Kim, M. E. Lewis, and C. C. White III, "Optimal vehicle routing with real-time traffic information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 178–188, 2005.

[13] T. Nadeem, S. Dashtinezhad, C. Liao, and L. Iftode, "Trafficview: traffic data dissemination using car-to-car communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 8, no. 3, pp. 6–19, 2004.

[14] D. Work and A. Bayen, "Impacts of the mobile internet on transportation cyberphysical systems: traffic monitoring using smartphones," in *Proceedings of the National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, & Rail*, pp. 18–20, Washington, DC, USA, November 2008.

[15] C. Claudel, A. Hofleitner, N. Mignerey, and A. Bayen, "Guaranteed bounds on highway travel times using probe and fixed data," in *Proceedings of the 88th TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2009.

[16] I. Leontiadis, G. Marfia, D. MacK, G. Pau, C. Mascolo, and M. Gerla, "On the effectiveness of an opportunistic traffic management system for vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1537–1548, 2011.

[17] C. Qian, C.-Y. Chan, and K.-L. Yung, "Reaching a destination earlier by starting later: revisited," *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 5, pp. 641–647, 2011.

[18] O. Bräysy, E. Martínez, Y. Nagata, and D. Soler, "The mixed capacitated general routing problem with turn penalties," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12954–12966, 2011.

[19] A. L. Kok, E. W. Hans, and J. M. J. Schutten, "Vehicle routing under time-dependent travel times: the impact of congestion avoidance," *Computers and Operations Research*, vol. 39, no. 5, pp. 910–918, 2012.

[20] F. Dehne, M. T. Omran, and J.-R. Sack, "Shortest paths in time-dependent fifo networks," *Algorithmica*, vol. 62, no. 1-2, pp. 416–435, 2012.

[21] D. Soler, J. Albiach, and E. Martínez, "A way to optimally solve a time-dependent vehicle routing problem with time windows," *Operations Research Letters*, vol. 37, no. 1, pp. 37–42, 2009.

[22] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York, NY, USA, 1990.

[23] R. Xu and D. Wunsch II, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[24] I. G. Guardiola, T. Leon, and F. Mallor, "A functional approach to monitor and recognize patterns of daily traffic profiles," *Transportation Research, Part B: Methodological*, vol. 65, pp. 119–136, 2014.

[25] Y. Wang, Y. Chen, M. Qin, and Y. Zhu, "Dynamic traffic prediction based on traffic flow mining," in *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA 0'6)*, vol. 2, pp. 6078–6081, Dalian, China, June 2006.

[26] N. Caceres, L. M. Romero, and F. G. Benitez, "Estimating traffic flow profiles according to a relative attractiveness factor," *Procedia—Social and Behavioral Sciences*, vol. 54, pp. 1115–1124, 2012.

[27] M. Yildirimoglu and N. Geroliminis, "Experienced travel time prediction for congested freeways," *Transportation Research Part B: Methodological*, vol. 53, pp. 45–63, 2013.

[28] Transportation Research Board, *Highway Capacity Manual*, Transportation Research Board, 2010.

[29] Wolfram, Mathematica, 2015, http://www.wolfram.com/mathematica/.

[30] Federal Highway Administration, "Revised monograph on traffic flow theory," Tech. Rep., Federal Highway Administration, Washington, DC, USA, 2001.

Research Article

Robustness of SOC Estimation Algorithms for EV Lithium-Ion Batteries against Modeling Errors and Measurement Noise

Xue Li,[1,2] Jiuchun Jiang,[1,2] Caiping Zhang,[1,2] Le Yi Wang,[3] and Linfeng Zheng[1,2]

[1]*National Active Distribution Network Technology Research Center (NANTEC), Beijing Jiaotong University, Beijing 100044, China*
[2]*Collaborative Innovation Center of Electric Vehicles in Beijing, Beijing Jiaotong University, Beijing 100044, China*
[3]*Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA*

Correspondence should be addressed to Jiuchun Jiang; jcjiang@bjtu.edu.cn

State of charge (SOC) is one of the most important parameters in battery management system (BMS). There are numerous algorithms for SOC estimation, mostly of model-based observer/filter types such as Kalman filters, closed-loop observers, and robust observers. Modeling errors and measurement noises have critical impact on accuracy of SOC estimation in these algorithms. This paper is a comparative study of robustness of SOC estimation algorithms against modeling errors and measurement noises. By using a typical battery platform for vehicle applications with sensor noise and battery aging characterization, three popular and representative SOC estimation methods (extended Kalman filter, PI-controlled observer, and H_∞ observer) are compared on such robustness. The simulation and experimental results demonstrate that deterioration of SOC estimation accuracy under modeling errors resulted from aging and larger measurement noise, which is quantitatively characterized. The findings of this paper provide useful information on the following aspects: (1) how SOC estimation accuracy depends on modeling reliability and voltage measurement accuracy; (2) pros and cons of typical SOC estimators in their robustness and reliability; (3) guidelines for requirements on battery system identification and sensor selections.

1. Introduction

Electric vehicles (EVs), including hybrid electric vehicles (HEVs), battery electric vehicles (BEVs), and plug-in hybrid electric vehicles (PHEVs), have become a critical driving force for green economy and attracted great research effort recently. An appropriate battery management system (BMS) is indispensable for safe, reliable, and efficient operations of EV battery systems [1]. The state of charge (SOC) is one of the most important state variables in BMS. Failure to estimate SOC accurately may cause overdischarging or overcharging, resulting in decreased battery longevity and even causing dangerous accidents [2].

There are many methods to estimate the SOC, with their own pros and cons. The Coulomb integral method [3] is easy to implement, but it needs the prior knowledge of the initial SOC and suffers from accumulated errors from measurement noise and bias. The open circuit voltage (OCV) method is a more reliable approach for SOC estimation. There

is a monotonic relationship between the OCV and SOC. However, this relationship is accurate only at a steady-state after several hours of open circuit condition. As a result, the OCV method cannot be used reliably for online SOC estimation [4]. The Kalman filter [5, 6] and extended Kalman filter (EKF) [7–12] have the appealing property of minimizing the mean-square estimation errors when the state and output measurement noises are additive, independent, zero mean, and Gaussian. On the other hand, they are susceptible to modeling errors and noise feature variations. The nonlinear observer method [13] employs a feedback mechanism to correct SOC estimation errors. Although this method works well under noise-free environment, its feedback gain must be carefully designed to achieve noise attenuation and robustness, which is highly challenging.

The accuracy of the model parameters is one of the main reasons affecting the SOC estimation accuracy. In [14], the researchers analyze the effects of the model parameters on the SOC estimation accuracy, when the model parameters drift

due to battery aging. Various degrees of impact of different model parameters are established, leading to some parameter updating guidelines to focus on high-impact parameters so that computational complexity can be reduced.

The paper is a comparative study of several typical SOC estimation algorithms on their robustness against modeling errors and measurement noises. We focus on variations of model parameters caused by battery aging. Our simulation results demonstrate that the Ohmic resistance R_O, polarization resistance R_P, and the open circuit voltage OCV are the key parameters affecting SOC estimation accuracy. However, the polarization capacitor C_P which is an important parameter only influences the dynamic response characteristics of SOC estimation and does not have noticeable effects on the steady-state accuracy of SOC estimation. Within the extended Kalman filter algorithm, H_∞ observer, and PI observer studied in this paper, our results indicate that the robustness and the estimation accuracy of the three methods against modeling errors and measurement noises are similar. However, in the view of application and SOC accuracy, the PI observer has advantages over the H_∞ observer and the EKF algorithm to be applied in BMS.

This paper reveals that SOC estimation accuracy depends critically on voltage measurement errors. While random noises in voltage measurements can be effectively filtered out, any bias or persistent errors will cause substantial deterioration on SOC estimation accuracy, which is a major reason for many algorithms to fail. Since voltage measurement accuracy varies substantially among BMS manufacturers, careful examination and enhancement of robustness of algorithms by design improvement and online parameter estimation are of essential importance. This paper utilizes some common scenarios of battery aging and parameter variations to study this issue and provide some related guidelines on how to select a robust method which has a strong tolerance towards voltage measurement errors.

The main contributions of this paper are in the following aspects which are essential for BMS design: (1) a clear analysis of the influence of each model parameter on the SOC estimation precision; (2) comparison of the robustness of various SOC estimation algorithms against model errors; (3) establishment of the quantitative relationship between measurement noise and SOC estimation accuracy.

The remainder of this paper is organized as follows. The battery model is introduced in Section 2. The three observer-based algorithms under study are described. Estimation accuracy of these algorithms is evaluated in Section 3 and their robustness is quantitatively compared. Section 4 investigates adaptability of the three algorithms against system uncertainties. Finally, some conclusions are drawn in Section 5.

2. Battery Model

Lithium-ion battery is a complex, nonlinear electrochemical system. It is difficult to find a very accurate model to describe the complex changes in its charging and discharging processes. Extensive research on batteries has generated many battery models [15–19]. In these battery models, the equivalent circuit model is used commonly, including the
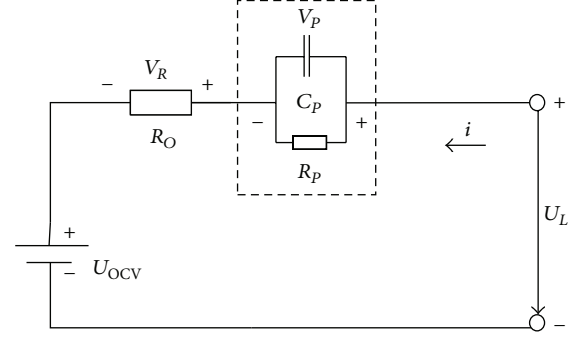
Figure 1: First-order RC model.

Rint model, the first-order RC model, and the second-order RC model [9, 20]. In general, an accurate battery model is essential for precise battery state estimation. However, high-fidelity battery models need more complex structures, more parameters, and carry high computational complexity. Therefore, it is necessary to find a compromise between accuracy of SOC estimation and simplicity of the model. In this paper, a first-order RC model is employed, shown in Figure 1, in which the battery terminal voltage error is within ±20 mV, meeting the requirements of the estimation accuracy.

One of the most important state variables in BMS is state of charge (SOC), which is defined as

$$\text{SOC}(t) = \text{SOC}_0 + \frac{1}{Q} \int_0^t \eta I(\tau) \, d\tau, \tag{1}$$

where $\text{SOC}(t)$ is the SOC at time t, SOC_0 the initial value, Q the battery nominal capacity, $I(\tau)$ the current at time τ, and η the coulomb efficiency. The coulomb efficiency can be considered to be 1 [21, 22]. The influence of the self-discharge on battery SOC estimation can be neglected.

According to Kirchhoff's current and voltage laws, it is easy to obtain the following mathematical relationships:

$$U_L = U_{\text{OCV}} + I R_O + V_P, \tag{2}$$

$$\dot{V}_P = -\frac{1}{R_P C_P} V_P + \frac{1}{C_P} I, \tag{3}$$

where R_O is the Ohmic resistance, R_P the polarization resistance, C_P the polarization capacitor, U_{OCV} the open circuit voltage, U_L the terminal voltage, V_R the voltage across R_O, and V_P the polarization voltage.

According to (1), it can be converted into the derivative equation as follows:

$$\dot{\text{SOC}} = \frac{I}{Q}. \tag{4}$$

The relationship between the SOC and OCV is nonlinear. In this paper, this function is represented by piecewise linear segments,

$$U_{\text{OCV}_i} = k_i \text{SOC}_i + b_i = f(\text{SOC}), \tag{5}$$

where k_i is the slope of the ith line segment and b_i is the intercept. Their values are listed as in Table 1.

TABLE 1: The values of k_i and b_i.

i	1	2	3	4	5	6	7	8	9	10
SOC$_i$	0–7	7–12	12–17	17–22	22–27	27–31	31–36	36–41	41–46	46–51
k_i	6.48	1.75	0.60	0.64	0.52	0.49	0.60	0.59	0.44	0.41
b_i	3.20	3.53	3.67	3.66	3.69	3.69	3.66	3.67	3.73	3.74
i	11	12	13	14	15	16	17	18	19	20
SOC$_i$	51–56	56–61	61–66	66–71	71–76	76–80	80–85	85–90	90–95	95–100
k_i	0.40	0.32	0.29	0.28	0.39	0.50	0.37	0.38	0.52	1.05
b_i	3.75	3.79	3.81	3.82	3.74	3.66	3.76	3.75	3.62	3.12
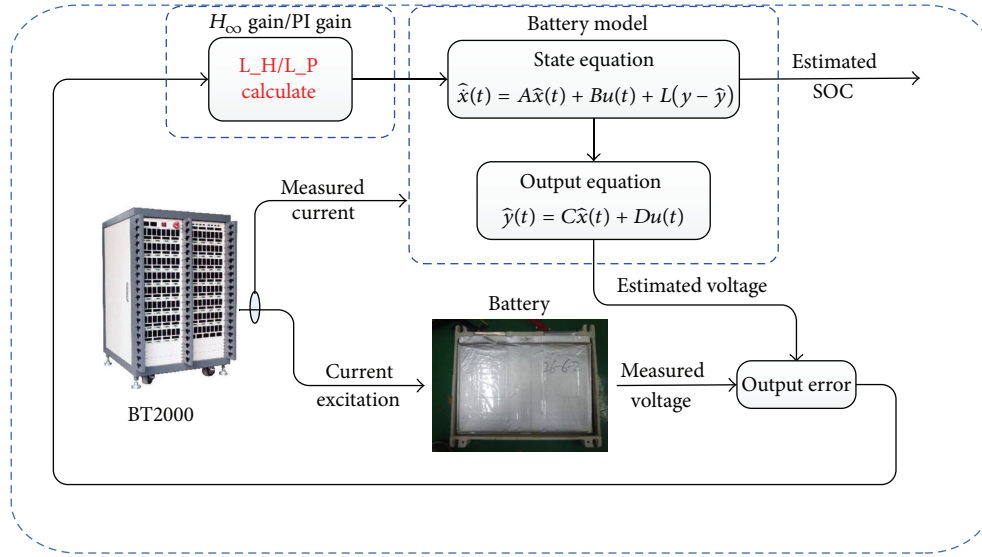
FIGURE 2: The general structure of the closed-loop observers.

If the state of the battery is defined as $x = [V_P, \text{SOC}]^T$, then the state equations of the battery in each segment are linear

$$\dot{x}(t) = Ax(t) + Bu(t),$$
$$y(t) = Cx(t) + Du(t), \tag{6}$$

where $A = \begin{bmatrix} -1/R_P C_P & 0 \\ 0 & 0 \end{bmatrix}$, $B = [1/C_P \ 1/Q]^T$, $C = [1 \ k_i]$, $D = R_O$, $u(t) = I(t)$, and $y(t) = U_L(t)$.

3. SOC Estimation Algorithms and Their Robustness

In this paper, three closed-loop observers are evaluated on their accuracy and robustness in SOC estimation. The key control principle of the closed-loop observers is to use the difference between the measured terminal voltage and the estimated value as the input to the feedback module with a gain matrix to update the polarization voltage V_P and SOC. The general structure of the closed-loop observers is shown in Figure 2.

3.1. The H_∞ Observer. The H_∞ control theory was initiated by Zames in his seminal paper [23]. Since then extensive

theoretical development, efficient solutions using frequency domain methods, state space models, numerical algorithms, and software packages have resulted in a rich treatise in this field [24–26]. Numerous successful applications have also been documented. In particular, numerical solutions to standard H_∞ observers can be found by using the Robust Control Tool Box and LMI model in MATLAB.

The main advantages of the H_∞ observer are as follows: (1) it is designed to attenuate disturbances of broader types than Kalman filters and Wiener filters which target Gaussian white noises; (2) it is robust against unstructured model uncertainty. However, as a worst-case robust design approach, it may be conservative, namely, nonoptimal, if the noise spectrum is actually known.

Consider a generic nonlinear battery system described by

$$\dot{x}(t) = Ax(t) + Bu(t) + F\lambda,$$
$$y(t) = Cx(t) + Du(t) + E(t) + G\lambda, \tag{7}$$

where A, B, C, D, E, F, and G are coefficient matrices, which depend on the actual battery system, $x(t)$ is the state, and $y(t)$ is the output. Consider $E(t) = b_i(t)$, $\lambda = [\omega \ v]^T$, $\omega = \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}$, $F = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and $G = [0 \ 0 \ 1]$.

The structure of the observer is

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + L(y - \hat{y}), \quad (8)$$

$$\hat{y}(t) = C\hat{x}(t) + Du(t) + E(t), \quad (9)$$

where $\hat{x}(t)$ and $\hat{y}(t)$ are the estimates for $x(t)$ and $y(t)$ and L is the gain vector of the observer.

Define the state estimation error $e(t) = x(t) - \hat{x}(t)$. Then the error dynamics is

$$\dot{e}(t) = \dot{x}(t) - \dot{\hat{x}}(t) = (A - LC)e(t) + (F - LG). \quad (10)$$

The goal of the observer design is disturbance attenuation: for a given (acceptable) sensitivity coefficient $\gamma > 0$, design the observer gain L such that the error system (10) is stable and that the following inequality is met under the zero initial condition:

$$\|e(t)\| \leq \gamma \|\lambda(t)\|. \quad (11)$$

The gain $L = P^{-1}X$ may be numerically solved by using the LMI approach [27]

$$\min(\gamma^2)$$

$$P > 0 \quad (12)$$

$$\begin{bmatrix} A^T P - C^T X^T + PA - XC + I & PF - XG \\ (PF - XG)^T & -\gamma^2 \end{bmatrix} < 0,$$

where $P = P^T$, $X = PL$, for which the LMI Toolbox in MATLAB can be used. The derivation of (12) is similar to the derivation in [13].

Reference [13] has verified that the battery state space model (8) is observable. As a result, the LMI approach is applicable to design the H_∞ observer to estimate the SOC of the battery system.

3.2. The PI Observer.

The proportional control law is the simplest most common control law. However, it carries steady-state error which limits its applications alone. The integral controller is not only related to the size of the input bias but also related to the existence of time deviation. As long as the bias exists, the output will continue to accumulate until the bias is zero; it will stop accumulating. Therefore, the integral control can eliminate steady-state error. Although the integral control can eliminate residual error, it slows down the control action, and as such it has detrimental effect on stability and transient performance. By combining these two control actions, the proportional and integral (PI) control inherits the advantages of both.

The observability and stability of the PI observer estimating the battery SOC are proved in [28]. Therefore, the PI observer can be utilized to estimate the battery SOC.

The relationship of the input and the output in the PI observer is

$$\varphi(t) = K_P \left[e(t) + \frac{1}{T_i} \int e(t)\, dt \right]$$

$$= K_P e(t) + K_i \int e(t)\, dt, \quad (13)$$

where K_p is the proportional gain of the observer, T_i the time constant of integration, K_i the integral gain of the observer, $\varphi(t)$ the output of the feedback system in the observer, and $e(t)$ the error between the estimated voltage and measured value, which is the input to the feedback loop in the observer.

The Ziegler-Nichols tuning method [29] is a heuristic method of tuning a PID controller. It was developed by Nichols and Ziegler. It is performed by setting the integral gain K_i to zero. The proportional gain K_p is then increased (from zero) until it reaches the ultimate gain K_u, at which the output of the control loop oscillates with a constant amplitude. The ultimate gain K_u and the oscillation period T_u are used to set the proportional gain K_p and the integral gain K_i. According to [29], the proportional gain K_p and the integral gain K_i in the simulation are set as follows:

$$K_p = 1.5,$$

$$K_i = 0.3. \quad (14)$$

The PI observer takes the advantage of the proportional control to generate control action immediately and that of the integral control to eliminate residual error. The control parameters must be properly designed to achieve a desirable balance between dynamic quality and steady-state performance of the observer.

3.3. The Extended Kalman Filter.

In 1960, Kalman published his famous paper describing a recursive solution to optimal discrete-time linear filtering problems under additive and independent Gaussian noise [30]. A Kalman filter estimates the state of a dynamic system with a linear process model and measurement model [31]. Its extension to nonlinear systems employs local linearization, leading to the extended Kalman filter (EKF) [32].

In its application to the battery systems considered in this paper, the discrete-time nonlinear system with additive noise is given by

$$\begin{bmatrix} V_P(k+1) \\ \text{SOC}(k+1) \end{bmatrix}$$

$$= \begin{bmatrix} \exp\left(-\dfrac{\Delta t}{R_P C_P}\right) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_P(k) \\ \text{SOC}(k) \end{bmatrix} \quad (15)$$

$$+ \begin{bmatrix} R_P\left(1 - \exp\left(-\dfrac{\Delta t}{R_P C_P}\right)\right) \\ \dfrac{\Delta t}{Q} \end{bmatrix} I(k) + \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix}.$$

By substituting (5) into (2), the terminal voltage can be expressed as

$$U_L(k) = k_i \text{SOC}(k) + I(k) R_O + V_P(k) + b_i + \nu(k). \quad (16)$$

Therefore, coefficient matrices can be derived as

$$A = \begin{bmatrix} \exp\left(-\dfrac{\Delta t}{R_P C_P}\right) & 0 \\ 0 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} R_P \left(1 - \exp\left(-\dfrac{\Delta t}{R_P C_P}\right)\right) \\ \dfrac{\Delta t}{Q} \end{bmatrix}, \quad (17)$$

$$C = \begin{bmatrix} 1 & k_i \end{bmatrix},$$

$$D = R_O.$$

The EKF algorithm involves the following steps:

(i) Prediction update:

$$\widehat{x}_k^- = A\widehat{x}_{k-1} + BI_{k-1},$$
$$P_k^- = AP_{k-1}A^T + \varepsilon. \quad (18)$$

(ii) Measurement update:

$$\widehat{x}_k = \widehat{x}_k^- + K_k \left(U_L(k) - \widehat{U}_L(k)\right),$$

$$K_k = P_k^- C^T \left[CP_k^- C^T + \delta\right]^{-1}, \quad (19)$$

$$P_k = \left[1 - K_k C\right] P_k^-,$$

where ε and δ are the variances of the noises $\omega(k)$ and $v(k)$, respectively, and $\widehat{x}_k^-$ is the updated state estimate from the previous estimate $\widehat{x}_{k-1}$.

Implementation of the EKF is depicted by the flowchart in Figure 3.

3.4. Experiments. In this paper, one battery with a nominal capacity of 92 Ah, whose anode is lithium manganese oxide and whose cathode is graphite, is used in our experiments to verify parameter identification and SOC estimation accuracy. All experiments are accomplished on the battery testing platform, shown in Figure 4, which includes the Arbin Testing System, thermostat, PC, BMS, and a high precision multimeter. The charge and discharge tests are finished by the Arbin Instrument BT2000 battery testing system, whose maximum voltage and charge/discharge current are 5 V and 400 A, respectively, in which the current can be set to the low range (−1 A∼1 A), the middle range (−50 A∼50 A), or the high range (−400 A∼400 A), according to the required maximum testing current. The controllable temperature range of the thermostat is −373.15 K∼233.15 K. The BMS is manufactured by Huizhou Epower Electronic Co., Ltd. The digital multimeter with a 6.5-digit resolution has a precision of 0.1 mV, so the voltage measured by this device is considered as true values. The noise is acquired by subtracting the BMS measured value from the digital multimeter measured value, whose statistical distribution is shown in Figure 5. The statistical characteristics of the terminal voltage measurement noise of the BMS are specified in Table 2.

TABLE 2: Statistical properties of terminal voltage and current measurement noise.

	Mean	Variance
The terminal voltage measurement noise of the BMS	1.3×10^{-3} V	4.1368×10^{-7} V^2

The tests are composed of two parts. One part is parameter identification experiment and the other part is SOC estimation accuracy verification experiment. A 1/3 C constant current is used to charge the battery to 4.2 V, and the capacity C in this paper refers to the maximum available capacity of the battery in current state of health if not figured out specifically. These experiments were first performed when the battery was brand-new. When the capacity of the battery was reduced to 74.5 Ah, the experiments were repeated to analyze the impact of the model parameter variations caused by battery aging. Data points are acquired at 1 Hz during the tests.

The battery model is built on the MATLAB/Simulink platform. The measured data including the voltage and current from the test bench are used as the input information to the three algorithms to estimate the SOC value. The SOC estimation accuracies are compared with the current integration values which serve as the SOC reference values.

The current integration method has two disadvantages: (a) it needs to know the initial SOC value in advance; (b) there is an accumulated error caused by the current measurement that is not accurate.

First, in this paper, the battery is discharged entirely in advance, so the initial SOC value is 0. Therefore, the initial SOC value is known. Second, the data used in the paper is obtained by the Arbin Instrument BT2000 battery testing system, and the precision of the current is 0.1% of the measuring range. The current range in experiments is set to the middle range (−50 A∼50 A), so the precision of the current sampling is 0.1 A. When the battery is charged in 1/3 C current rate, the maximum SOC accumulated error in one full charging is (0.1 A ∗ 3 h)/92 Ah ∗ 100% = 0.32%, which can be neglected. Therefore, the SOC value calculated by the current integration method in instrument can serve as the SOC reference values in this paper, which overcomes the shortcomings of the current integration method effectively.

3.5. Verification. The SOC is estimated from experimental data by using the three estimation methods, respectively. Figure 6(a) shows the results of SOC estimation from the H_∞ observer; Figure 6(b) shows the results of SOC estimation from the PI observer; and Figure 6(c) shows the results of SOC estimation from the EKF algorithm. The left y-axis is the SOC value, including the estimated value and experimental data, and the right y-axis is the SOC estimation error, and the x-axis is the time. All of three observers demonstrate good SOC estimation accuracy and convergence to the true SOC value, with a very short response time.
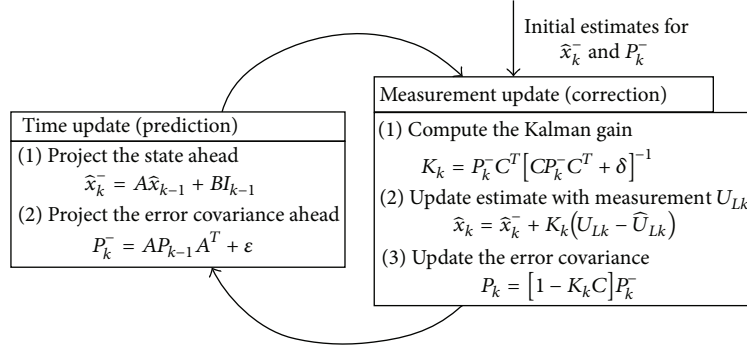
FIGURE 3: The operation of EKF.
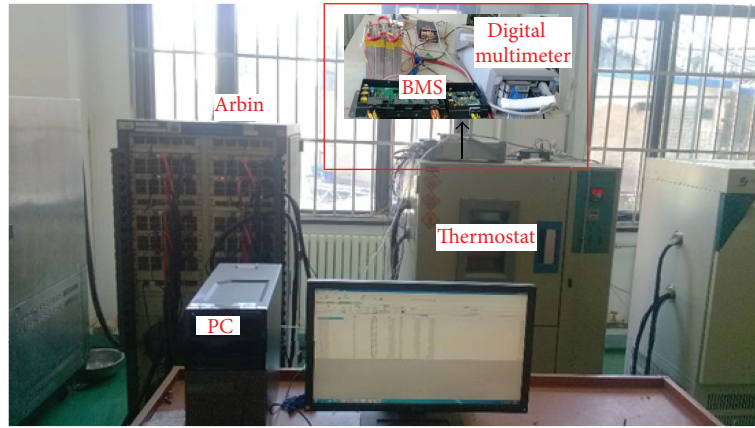
FIGURE 4: The battery testing platform.
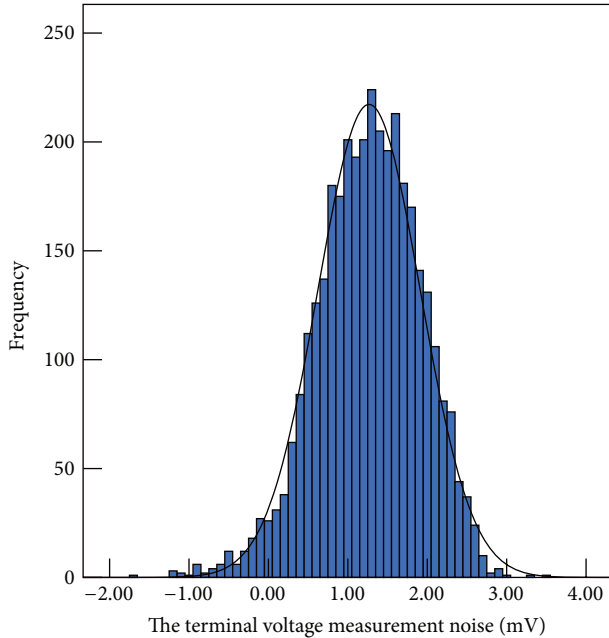
FIGURE 5: The statistical properties of the terminal voltage measurement.

While the SOC estimation errors of the three estimation methods have the same trends, there are important differences among them, as indicated by Figure 6(d). In Figure 6(d), it is apparent that the EKF algorithm has the largest transient volatility, which is partially due to the initial value of P_k^-, ε, and δ. After SOC estimation reaches steady-state, the H_∞ observer has the largest SOC estimation errors, with the upper limit 1.67% and lower limit −0.79% implying that the H_∞ observer is least accurate. This may be attributed to the fact that the H_∞ observer is a conservative estimation method, which does not attenuate noise optimally.

The steady-state SOC estimation errors of the three estimation methods can maintain between the 2% band and −2% band when the SOC initial error is 20%, see Figure 6(d).

4. Robust Analysis of the Algorithms against System Uncertainties

In this section, the adaptability of the noise characteristics is discussed in detail. The noise is divided into two parts. One is the modelling error due to parameter changes caused by battery aging and the other is the terminal voltage measurement noise caused by BMS sampling accuracy. The

(a) H_∞ observer

(b) PI observer

(c) EKF algorithm

(d) Three estimation methods
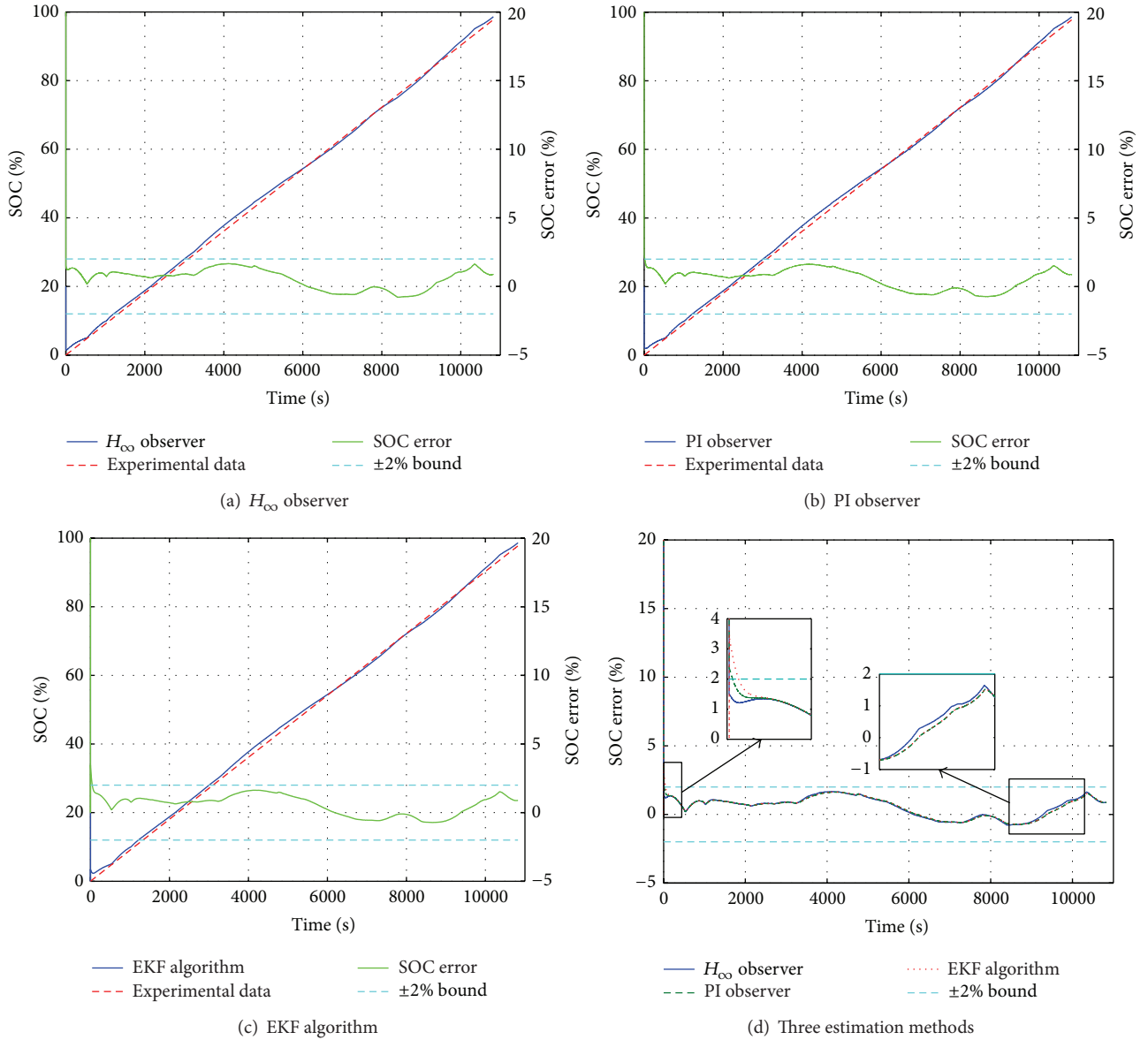
FIGURE 6: The SOC estimation results when initial SOC error is 20%.

adaptability of the model noise is discussed in Section 4.1, and the adaptability of the measurement noise is analyzed in Section 4.2.

4.1. The Model Parameter Perturbation. When the capacity of the battery with nominal capacity 92 Ah declines to 74.5 Ah, other parameters of the battery will change too. If the parameters for the new battery of capacity 92 Ah are used to estimate the SOC of the old battery of capacity 74.5 Ah, the SOC estimation accuracy of the three estimation methods will be affected significantly. Therefore, the impacts of the inaccurate battery parameters caused by battery aging to the SOC estimation accuracy are of essential importance.

In order to analyze the effects of parameter variations on SOC estimation accuracy, four cases are considered in our simulation.

Case 1. It is the process of estimating the SOC of the old battery using the parameters of the new battery.

Figure 7 shows the results of the SOC estimation in Case 1 using three algorithms. Using the H_∞ observer, the maximum value of the SOC estimation error is 6.39%, and the minimum value of the SOC estimation error is 0.4%. Using the PI observer, the maximum value of the SOC estimation error is 6.33%, and the minimum value of the SOC estimation error is 0.42%. Using the EKF algorithm, the maximum value of the SOC estimation error is 6.33%, and the minimum value of the SOC estimation error is 1.54%.

Case 2. It is the process of estimating the SOC of the old battery by updating the battery's parameters used in Case 1 to the old battery's parameters except the SOC-OCV curve.

(a) SOC curves with three algorithms and experiment

(b) SOC estimation error curves
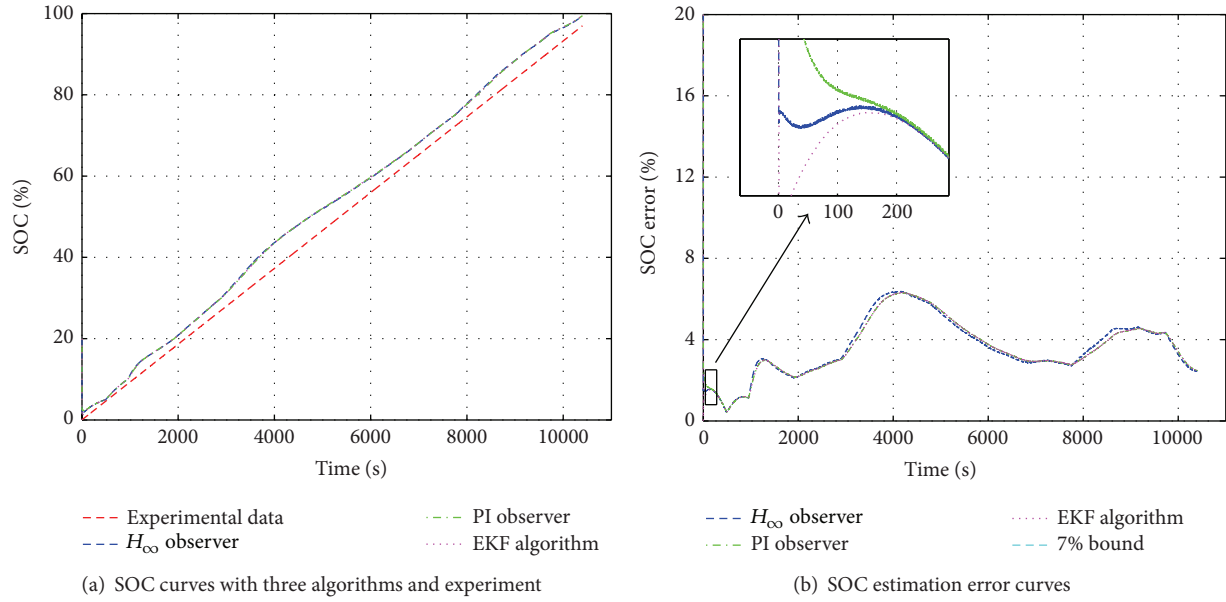
FIGURE 7: All parameters are the new battery's ones.

(a) SOC curves with three algorithms and experiment

(b) SOC estimation error curves
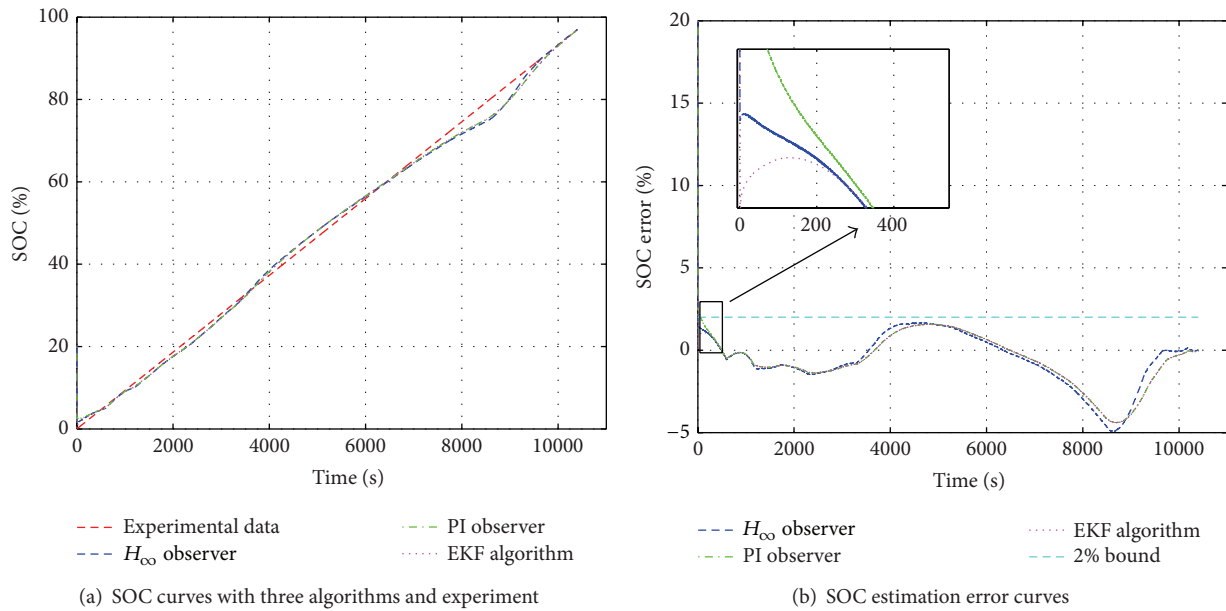
FIGURE 8: Only the OCV is the new battery's one.

Figure 8 shows the results of the SOC estimation in Case 2 using three algorithms. Using the H_∞ observer, the maximum value of the SOC estimation error is 1.69%, and the minimum value of the SOC estimation error is −4.96%. Using the PI observer, the maximum value of the SOC estimation error is 1.59%, and the minimum value of the SOC estimation error is 1.59%, and the minimum value of the SOC estimation error is −4.41%.

Case 3. It is the process of estimating the SOC of the old battery by updating only the SOC-OCV curve to the old

battery's parameters, while the other parameters are the same as in Case 1.

Figure 9 shows the results of the SOC estimation in Case 3 using three algorithms. Using the H_∞ observer, the maximum value of the SOC estimation error is 8.54%, and the minimum value of the SOC estimation error is 0.38%. Using the PI observer, the maximum value of the SOC estimation error is 8.38%, and the minimum value of the SOC estimation error is 0.40%. Using the EKF algorithm, the maximum value of the SOC estimation error is 8.38%, and the minimum value of the SOC estimation error is 1.40%.

(a) SOC curves with three algorithms and experiment

(b) SOC estimation error curves
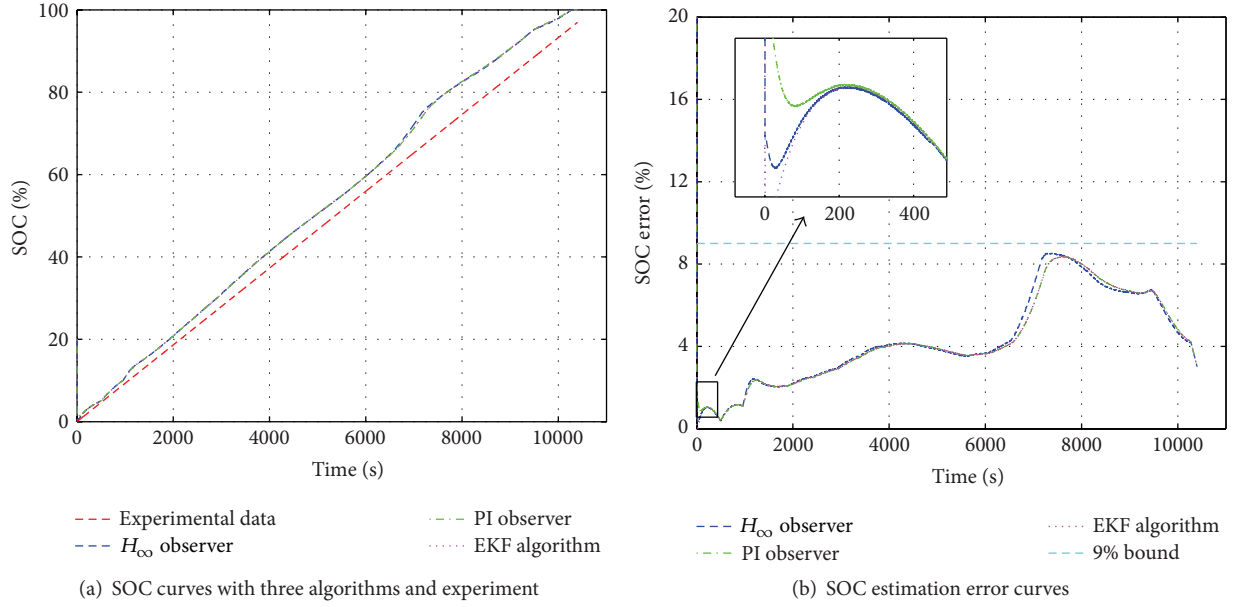
Figure 9: Only the OCV is the old battery's one.

(a) SOC curves with three algorithms and experiment

(b) SOC estimation error curves
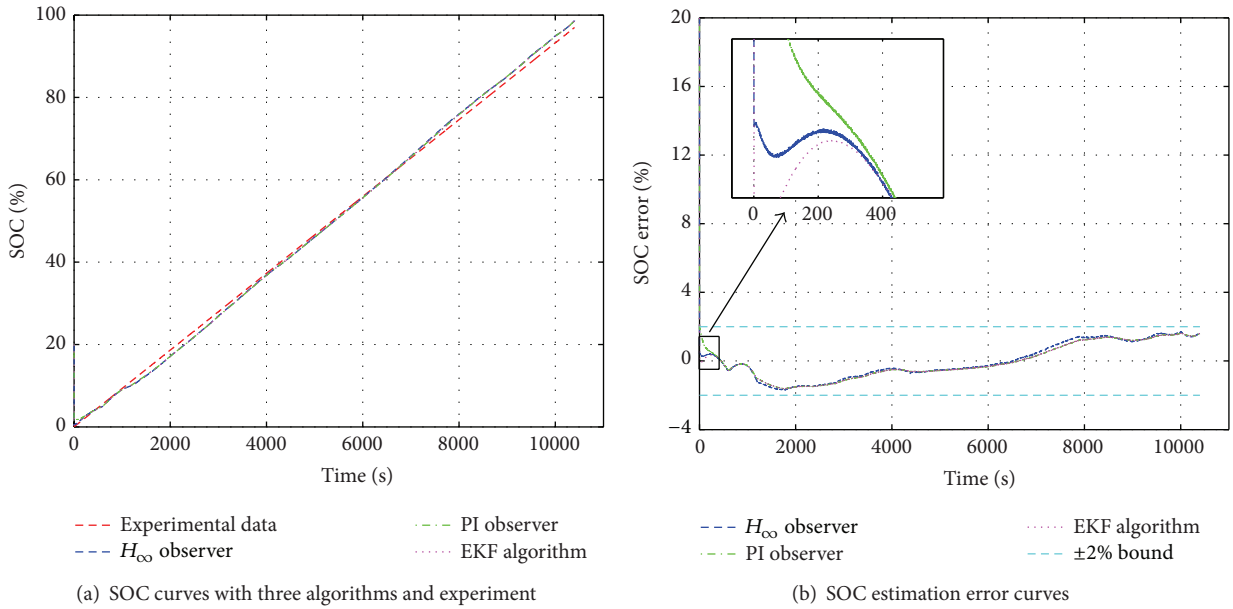
Figure 10: All parameters are the old battery's ones.

Case 4. It is the process of estimating the SOC of the old battery by updating all the parameters to the old battery's parameters.

Figure 10 shows the results of the SOC estimation in Case 4 using three algorithms. Using the H_∞ observer, the maximum value of the SOC estimation error is 1.72%, and the minimum value of the SOC estimation error is −1.69%. Using the PI observer, the maximum value of the SOC estimation error is 1.64%, and the minimum value of the SOC estimation error is −1.63%. Using the EKF algorithm, the maximum

value of the SOC estimation error is 1.63%, and the minimum value of the SOC estimation error is −1.62%.

Discussions. Note that

$$y_{est} = U_{OCV}(SOC) + V_P + IR_O, \quad (20)$$

$$\Delta y = y_{exp} - y_{est}, \quad (21)$$

$$\Delta SOC = \int_0^t \left(\frac{1}{Q}I + L_2 \Delta y \right) dt, \quad (22)$$

TABLE 3: The parameters changes.

	The maximum	The minimum	The average
$\Delta R_O = R_{O\_old} - R_{O_{new}} = f(\text{SOC})$	0.87 mΩ,	0.20 mΩ,	0.28 mΩ,
$\Delta R_P = R_{P\_old} - R_{P_{new}} = f(\text{SOC})$	1.6 mΩ,	0.31 mΩ,	0.64 mΩ,
$\Delta C_P = C_{P\_old} - C_{P_{new}} = f(\text{SOC})$	5730 F	−14968 F	−604 F
$\Delta\text{OCV} = \text{OCV}_{old} - \text{OCV}_{new} = f(\text{SOC})$	106 mV	−14.3 mV	2.5 mV

where y_{exp} is the measured terminal voltage and y_{est} is the estimated terminal voltage.

From (21), since the measured terminal voltage y_{exp} is known, the estimated terminal voltage y_{est} determining the terminal voltage error Δy affects the accuracy and convergence time of the three SOC estimation methods. The estimated terminal voltage y_{est} includes open circuit voltage $U_{OCV}(\text{SOC})$, polarization voltage V_P, and IR_O, as (20) shows. Therefore, they have obvious effects on estimation accuracy and robustness of the three SOC estimation methods, as (22) shows, demonstrated by Figures 7–10.

The parameters changes from the new battery to the old battery's parameters are shown as in Table 3.

When the battery is aging from 92 Ah to 74.5 Ah, during a complete charging process in 1/3 C rate, the average polarization voltage increase is 16 mV, and the average Ohmic voltage increase is 7 mV. However, the average OCV decrease is 2.5 mV.

When we use the parameters of the old battery to estimate the SOC of the old battery, the SOC estimation error is only caused by the battery model error. The terminal voltage error caused by the battery model is considered to be Δ, and the SOC estimation result is shown in Figure 10.

However, using the parameters of the new battery to estimate the SOC of the old battery, which means that the parameters are not updated, the estimated value of the terminal voltage y_{est} is $(\Delta + 20.5)$ mV smaller than the measured terminal voltage value. This causes a great SOC estimation error, whose maximum $E1\_max$ is between 2% and 8%, as shown in Figures 7 and 10. By updating the Ohmic resistance and the polarization resistance to the true values of the old battery, the estimated value of the terminal voltage y_{est} is $(\Delta - 2.5)$ mV smaller than the measured terminal voltage value. This causes only a little SOC estimation error, whose minimum $E2\_min$ is less than −2%, as shown in Figures 8 and 10. If the OCV is updated, but not other parameters, to the value of the old battery, the estimated value of the terminal voltage y_{est} is $(\Delta+23)$ mV smaller than the measured terminal voltage value. Because $(\Delta+23)$ mV is larger than $(\Delta+20.5)$ mV, estimation errors by using the updated OCV of old battery are worst among all cases, whose maximum $E3\_max$ is more than 8%, larger than $E1\_max$, as shown in Figures 9 and 10. The clear comparison is shown in Table 4.

In our recent studies [33], it is shown that SOC estimation accuracy is dependent on the curve of SOC-OCV significantly, if the SOC-OCV curve varies substantially. In this study during the progress of the battery aging, the SOC-OCV curve does not change much. The experimental results are included in Figure 11. As a result, the impact of aging on the
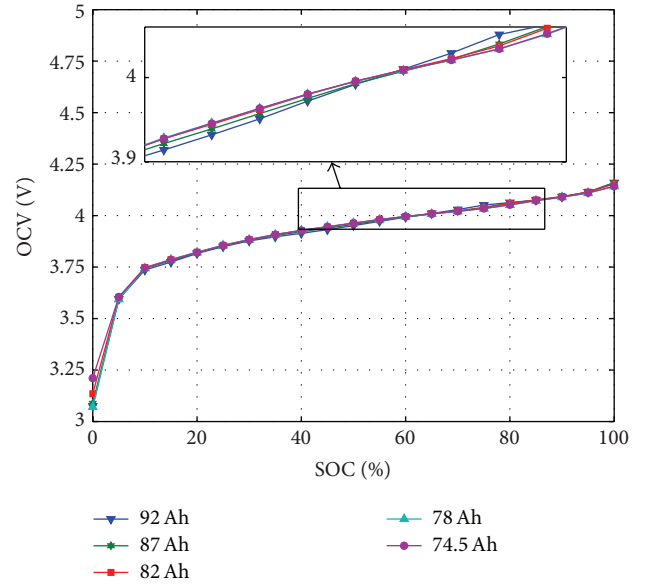
FIGURE 11: OCV-SOC curves at different aging states.

OCV has negligible effect on SOC estimation accuracy, as analyzed in Section 4.1. In contrast, effect of aging on other model parameters plays much more prominent roles in SOC estimation accuracy.

In addition, we may conclude from comparing Figures 7, 8, 9, and 10 that the SOC estimation error when using the parameters of the new battery to estimate the SOC of the old battery is not a linear superposition of the SOC estimation error caused by the changes of the resistance and capacitance and the changes of the curve of SOC-OCV. This is clearly indicated in Figure 12, taking the SOC estimation results of the PI observer as an example.

4.2. The Terminal Voltage Measurement Errors. In order to find the relationship between the SOC estimation accuracy and the statistical characteristics of the terminal voltage measurement noise, we further add measurement noises with different means and variances to the experimental data of the terminal voltage in Section 3.5. There are different SOC estimation results using different estimation methods.

(i) The H_∞ Observer. When the mean of the measurement noise changes from −6 mV to 4 mV and standard deviation of the measurement noise increases from 0 to 10 mV, the SOC accuracy is depicted in Figure 13.

TABLE 4: Terminal voltage error and SOC estimation error.

Cases estimating SOC of the old battery	$\Delta y = y_{exp} - y_{est}$	SOC estimation error
Case 1: all parameters are the new battery's ones	$(\Delta + 22.5)$ mV	$2\% < E1\_\max < 8\%$
Case 2: only the OCV is the new battery's one	$(\Delta - 2.5)$ mV	$-5\% < E2\_\min < -2\%$
Case 3: only the OCV is the old battery's one	$(\Delta + 23)$ mV	$E1\_\max < 8\% < E3\_\max$
Case 4: all parameters are the old battery's ones	Δ mV	$-2\% < E4 < 2\%$

SOC error in case 3 − SOC error in case 4
SOC error in case 2 − SOC error in case 4
SOC error in case 1 − SOC error in case 4
Green line + blue line
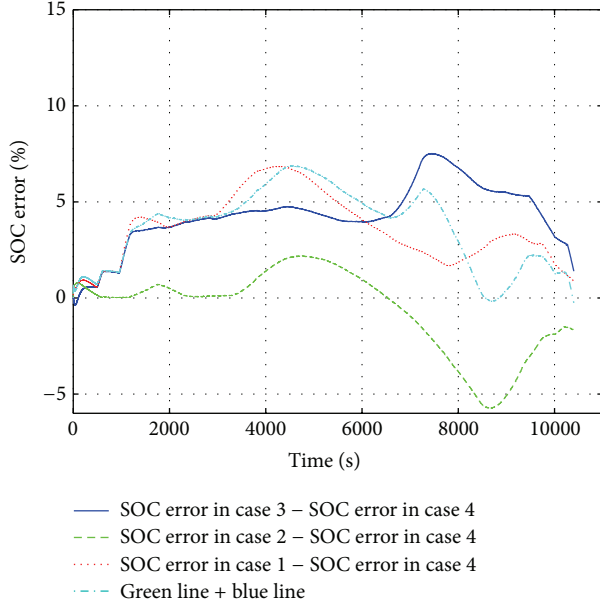
FIGURE 12: The comparison of the SOC estimation errors.

SOC error

FIGURE 13: The SOC estimation error of the H_∞ observer when statistical characteristics of the noise change.

If the SOC estimation error is required to be less than 2%, the mean and standard deviation of the measurement noise must be confined from −3.172 mV to 1.172 mV and 0 to 3.43 mV, respectively.

(ii) The PI Observer. When the mean of the measurement noise changes from −6 mV to 4 mV and standard deviation of the measurement noise increases from 0 to 10 mV, the SOC accuracy is shown in Figure 14.

If the SOC estimation error is required to be less than 2%, the mean and standard deviation of the measurement noise must be confined from −3.36 mV to 1.172 mV and 0 to 5.15 mV, respectively.

(iii) The EKF Algorithm. When the mean of the measurement noise changes from −6 mV to 4 mV and standard deviation of the measurement noise increases from 0 to 10 mV, the SOC accuracy is illustrated in Figure 15.

If the SOC estimation error is required to be less than 2%, the mean and standard deviation of the measurement noise must be confined from −3.36 mV to 1.3 mV, 0 to 3.535 mV, respectively.

(iv) Discussions. The principle of the H_∞ observer is similar to the Luenberger observer, whose feedback gain is a proportional gain. It controls quite aggressively with an expected fast

response. However, it shows certain levels of overreaction, causing nonsmooth contours of SOC estimation errors. In contrast the PI observer with its integral part acts as a signal smoother (a low-pass filter type), so the contours of its SOC estimation errors are smoother than that of the H_∞ observer. Note that the gain of the EKF algorithm is updated timely, so the contours of its SOC estimation errors are the smoothest.

To capture transient errors more concretely, in Figures 13–15, we define the area enclosed by the x-axis and the contour line as the total absolute error. The total absolute error ratio is then calculated as the total absolute error divided by the area which is equal to the product of the maximum value of x-axis and maximum value of y-axis in the figure. The ratios of the three observers are shown in Figure 16.

Under the same SOC estimation accuracy, the adaptation range against the measurement noise of the PI observer is the biggest and that of the H_∞ observer is the smallest, as Figure 16 shows. It is worth noting that Figures 13, 14, and 15 are not symmetrical about the mean, result from the total effect of the terminal voltage noise mean and the inaccurate parameters on the SOC estimation accuracy.

5. Conclusions

In this paper, several typical SOC estimation algorithms including the H_∞ observer, PI observer, and extended
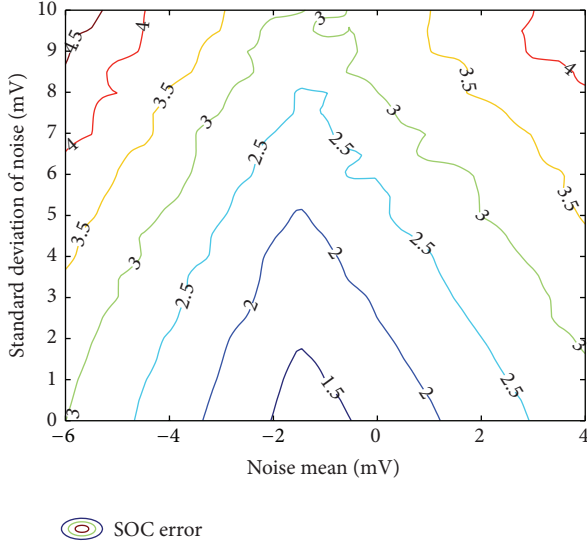
FIGURE 14: The SOC estimation error of the PI observer when statistical characteristics of the noise change.

FIGURE 16: The total absolute error ratio of the measurement noise.

FIGURE 15: The SOC estimation error of the EKF observer when statistical characteristics of the noise change.

Kalman filter are applied to estimate SOC. By considering four categories of variations of model parameters caused by battery aging and studying the influence of each category on the SOC estimation precision, we compare the algorithms in terms of their robustness against modeling errors. In addition, their tolerance to voltage measurement errors is quantitatively evaluated.

The robustness and the estimation accuracy of the three methods against modeling errors and measurement noises are similar. However, the H_∞ observer needs to know the gain L, which needs to calculate the Linear Matrix Inequalities in advance; the EKF algorithm needs to know the distribution of the measurement noise, which is difficult to be obtained

in fact. Compared to these two methods, the PI observer can acquire the proportional gain K_p and the integral gain K_i depending on the experiences. Therefore, in the view of application and SOC accuracy, the PI observer has advantages over the H_∞ observer and the EKF algorithm to be applied in BMS. Through simulation results we reach the following conclusions:

(1) The Ohmic resistance R_O, polarization resistance R_P, and the open circuit voltage OCV are the key parameters affecting SOC estimation accuracy. However, the polarization capacitor C_P, which is an important parameter, only influences the dynamic response characteristics of SOC estimation but does not have noticeable effects on the steady-state accuracy of SOC estimation.

(2) Under the same SOC estimation accuracy and the robustness against modeling errors and measurement noises, the PI observer has advantages over the H_∞ observer and the EKF algorithm to be applied in BMS.

(3) The relationship between SOC estimation accuracy and voltage measurement errors has been resolved, and some related guidelines on how to select a robust method which has a strong tolerance against voltage measurement errors are provided.

There are several important related topics that are not covered in this paper. First, optimal design of PI observers requires essential statistical information on measurement noises and individualized models. Learning algorithms for noise characterizations and parameter estimation can lead to adaptive PI observers with improved SOC estimation accuracy. Furthermore, implementation of SOC estimators in battery management systems on electric vehicles encounters hardware and computational complexity constraints. For

example, sensor precision levels and sampling rates will limit data flow rates and reliability. Analysis of the influence of synchronous sampling, sampling rates, asynchronous sampling, and quantization on SOC estimation accuracy will be pursued in our future studies.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] K. W. E. Cheng, B. P. Divakar, H. Wu, K. Ding, and H. F. Ho, "Battery-management system (BMS) and SOC development for electrical vehicles," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 1, pp. 76–88, 2011.

[2] V. Pop, H. J. Bergveld, D. Danilov, P. P. L. Regtien, and P. H. L. Notten, *Battery Management Systems: Accurate State-of-Charge Indication for Battery-Powered Applications*, Springer, New York, NY, USA, 2008.

[3] S. Piller, M. Perrin, and A. Jossen, "Methods for state-of-charge determination and their applications," *Journal of Power Sources*, vol. 96, no. 1, pp. 113–120, 2001.

[4] S. Lee, J. Kim, J. Lee, and B. H. Cho, "State-of-charge and capacity estimation of lithium-ion battery using a new open-circuit voltage versus state-of-charge," *Journal of Power Sources*, vol. 185, no. 2, pp. 1367–1373, 2008.

[5] J. Xu, C. C. Mi, B. Cao, and J. Cao, "A new method to estimate the state of charge of lithium-ion batteries based on the battery impedance model," *Journal of Power Sources*, vol. 233, pp. 277–284, 2013.

[6] M. Mastali, J. Vazquez-Arenas, R. Fraser, M. Fowler, S. Afshar, and M. Stevens, "Battery state of the charge estimation using Kalman filtering," *Journal of Power Sources*, vol. 239, pp. 294–307, 2013.

[7] C. P. Zhang, J. C. Jiang, W. G. Zhang, and S. M. Sharkh, "Estimation of state of charge of lithium-ion batteries used in HEV using robust extended Kalman filtering," *Energies*, vol. 5, no. 4, pp. 1098–1115, 2012.

[8] Z. Chen, Y. Fu, and C. C. Mi, "State of charge estimation of lithium-ion batteries in electric drive vehicles using extended Kalman filtering," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 3, pp. 1020–1030, 2013.

[9] R. Xiong, H. He, F. Sun, and K. Zhao, "Evaluation on state of charge estimation of batteries with adaptive extended kalman filter by experiment approach," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 108–117, 2013.

[10] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation," *Journal of Power Sources*, vol. 134, no. 2, pp. 277–292, 2004.

[11] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: part 2. Modeling and identification," *Journal of Power Sources*, vol. 134, no. 2, pp. 262–276, 2004.

[12] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs. Part 1. Background," *Journal of Power Sources*, vol. 134, no. 2, pp. 252–261, 2004.

[13] F. Zhang, G. Liu, L. Fang, and H. Wang, "Estimation of battery state of charge with H_∞ observer: applied to a robot for inspecting power transmission lines," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 2, pp. 1086–1095, 2012.

[14] X. Hu, S. Li, H. Peng, and F. Sun, "Robustness analysis of State-of-Charge estimation methods for two types of Li-ion batteries," *Journal of Power Sources*, vol. 217, pp. 209–219, 2012.

[15] M. Chen and G. A. Rincón-Mora, "Accurate electrical battery model capable of predicting runtime and I-V performance," *IEEE Transactions on Energy Conversion*, vol. 21, no. 2, pp. 504–511, 2006.

[16] L. Gao, S. Liu, and R. A. Dougal, "Dynamic lithium-ion battery model for system simulation," *IEEE Transactions on Components and Packaging Technologies*, vol. 25, no. 3, pp. 495–505, 2002.

[17] V. H. Johnson, "Battery performance models in ADVISOR," *Journal of Power Sources*, vol. 110, no. 2, pp. 321–329, 2002.

[18] T.-S. Dao, C. P. Vyasarayani, and J. McPhee, "Simplification and order reduction of lithium-ion battery model based on porous-electrode theory," *Journal of Power Sources*, vol. 198, pp. 329–337, 2012.

[19] A. Szumanowski and Y. Chang, "Battery management system based on battery nonlinear dynamics modeling," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 3, pp. 1425–1432, 2008.

[20] H. He, R. Xiong, X. Zhang, F. Sun, and J. Fan, "State-of-charge estimation of the lithium-ion battery using an adaptive extended kalman filter based on an improved Thevenin model," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 4, pp. 1461–1469, 2011.

[21] G. L. Wu, R. G. Lu, C. Zhu, and C. C. Chan, "An improved Ampere-hour method for battery state of charge estimation based on temperature, coulomb efficiency model and capacity loss model," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC '10)*, pp. 1–4, Lille, France, September 2010.

[22] M.-H. Chang, H.-P. Huang, and S.-W. Chang, "A new state of charge estimation method for LiFePO$_4$ battery packs used in robots," *Energies*, vol. 6, no. 4, pp. 2007–2030, 2013.

[23] G. Zames, "Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 301–320, 1981.

[24] J. Huang and C.-F. Lin, "Numerical approach to computing nonlinear H_∞ control laws," *Journal of Guidance, Control, and Dynamics*, vol. 18, no. 5, pp. 989–996, 1995.

[25] J. C. Doyle, K. Glover, P. P. Khargonekar, and B. A. Francis, "State-space solutions to standard H_2 and H_∞ control problems," *IEEE Transactions on Automatic Control*, vol. 34, no. 8, pp. 831–847, 1989.

[26] J. W. Helton and A. Sideris, "Frequency response algorithms for H_∞ optimization with time domain constraints," *IEEE Transactions on Automatic Control*, vol. 34, no. 4, pp. 427–434, 1989.

[27] K. Zhou and J. Doyle, *Essentials of Robust Control*, Prentice Hall, Hoboken, NJ, USA, 1998.

[28] J. Xu, C. C. Mi, B. Cao, J. Deng, Z. Chen, and S. Li, "The state of charge estimation of lithium-ion batteries based on a proportional-integral observer," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 4, pp. 1614–1621, 2014.

[29] N. B. Nichols and J. G. Ziegler, "Optimum settings for automatic controllers," *Journal of Dynamic Systems, Measurement, and Control*, vol. 115, no. 2, pp. 220–222, 1993.

[30] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[31] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter*, Artech House, Boston, Mass, USA, 2004.

[32] M. Hoshiya and E. Saito, "Structural identification by extended Kalman filter," *Journal of Engineering Mechanics*, vol. 110, no. 12, pp. 1757–1770, 1984.

[33] C. P. Zhang, L. Y. Wang, X. Li et al., "Robust and adaptive estimation of state of charge for lithium-ion batteries," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 4948–4957, 2015.

Research Article

A Hybrid MMC Topology with dc Fault Ride-Through Capability for MTDC Transmission System

Xinhan Meng,[1] Ke-Jun Li,[1] Zhuodi Wang,[1] Wenning Yan,[1] and Jianguo Zhao[2]

[1]*School of Electrical Engineering, Shandong University, Jinan 250061, China*
[2]*State Grid of China Technology College, Jinan 250002, China*

Correspondence should be addressed to Ke-Jun Li; lkjun@sdu.edu.cn

This paper proposes a hybrid modular multilevel converter (MMC) topology based on mismatched-cascade mechanism. The blocking conditions of different submodule (SM) structures under dc fault are analyzed and a series double submodule is presented. With series-double submodules and mismatched-cascade submodules, the proposed hybrid MMC can ride-through the dc side short-circuit fault and provide an output voltage with the feature of low harmonic content. This hybrid MMC topology can be used in the VSC based multiterminal dc (VSC-MTDC) transmission system. The dc fault ride-through properties of the new structure and the total harmonic distortion (THD) are analyzed compared with the previous full-bridge and clamp-double architectures. An appropriate fault blocking procedure is presented, and a typical four-terminal dc transmission simulation system is given in the power system simulation software. Finally, simulation of steady-state and dc bipolar short-circuit fault verifies that the MTDC system based on this new hybrid MMC topology is stabilized and can block the dc fault and return the nonfault parts to normal.

1. Introduction

With the development of power electronic technology, the voltage source converter (VSC) based on full-controllable electric semiconductor device is widely applied to high voltage direct current transmission field. Compared to the traditional HVDC, VSC based high voltage dc (VSC-HVDC) transmission system has the advantages such as flexible power control, reactive power compensation, supplying power to passive network, and forming multiterminal dc network [1–5]. According to the different structures of voltage source converter, VSC-HVDC can be divided into two kinds: the low level traditional VSC-HVDC and modular multilevel converter based high voltage dc (MMC-HVDC) transmission system. MMC-HVDC is superior to the low level VSC-HVDC in the following aspects: it has lower switching frequency, lower switching loss, and higher scalability; it does not need to switch the serial IGBTs at the same time and can be applied in high voltage occasion. Consequently, it has been rapidly developed for the last few years [6–8].

The present research of VSC-HVDC is mostly focused on half-bridge MMC (HBMMC) and its control strategies.

However, half-bridge MMC cannot clear the fault current when dc fault occurs because of the freewheeling diode [9, 10]. At the same time, high power dc current breaker for HVDC applications is not sufficiently mature and cost-effective [11]. So, when dc fault occurs, the common method is to turn off the whole dc system with ac circuit breakers [12]. This approach costs lots of time and reduces the reliability of VSC-HVDC system. To avoid dc fault, cable with low failure rate is used as transmission lines, but this increases the engineering cost and is easily restricted by the working environment. Thus, the present VSC-HVDC technology could hardly be used in long distance or multiterminal dc transmission system [13, 14].

To overcome the shortcomings of traditional MMC, different topologies have been addressed by many scholars. References [15, 16] propose an MMC-HVDC system based on full-bridge MMC (FBMMC) topology. FBMMC can block the fault current when dc fault occurs. However, as too many IGBTs are needed, under the same dc voltage and power level, FBMMC's engineering investment and operation cost is high, which limits its application in engineering practice.

In order to reduce the IGBT used quantity and make the converter capable of blocking dc fault, a new kind of clamp-double MMC (CDMMC) is proposed [14, 17, 18]. When a fault occurs on the dc side, CDMMC turns off IGBTs immediately and utilizes the diode reversed-phase blocking ability to complete the fault handling process. CDMMC needs less semiconductors than FBMMC and also has the ability to block dc fault. However, due to the characteristics of parallel structure, the equivalent capacitance in a bridge arm shows two kinds of states according to the different flows of short-circuit current. So it requires longer time to cut off the short-circuit current, and its dc fault blocking ability is inferior compared with FBMMC [14, 19].

The contribution of this paper is to analyze the equivalent states of various MMC topologies under the dc short-circuit fault and propose an improved MMC topology to improve its performance. Based on the mismatched-cascade mechanism and the principle of dc fault blocking, a hybrid MMC topology which has dc fault ride-through capability and is very suitable for MTDC system is presented. According to the "handshaking method" of MTDC system [20, 21], the process of clearing dc fault and recovering nonfault lines is explained in detail. Finally, a typical four-terminal dc system is introduced and a simulation model is built to verify the system characteristics under the bipolar short-circuit fault which is the most serious dc fault.

This paper is organized as follows: after introduction, the dc fault blocking analysis, which includes analyzing the fault blocking principle and the current paths of different submodules under blocking states, is explained in detail in Section 2. In Section 3, a new topology of hybrid MMC based on the mismatched-cascade mechanism is introduced. In Section 4, a four-terminal dc simulation model is built to explain the application of the new hybrid MMC in MTDC transmission system, including the process to remove the fault lines and recover the nonfault lines under the dc fault. This new hybrid MMC used in MTDC system is tested with the steady-state and dc bipolar short-circuit fault simulations in Section 5. The conclusion of this paper is made in Section 6.

2. Fault Blocking Ability Analysis

2.1. Fault Blocking Principle. An MMC topology consists of two arms per each phase where each arm is comprised of n series-connected submodules and a series-connected inductor. These submodules of each bridge arm can be replaced with an ideal voltage source. After the treatment of presenting network parameter by per-unit value normalization, the equivalent circuit under dc fault state is shown in Figure 1.

Considering the upper and lower loops, the dynamics of the converter can be described as follows:

$$u_k - L_T \frac{di_k}{dt} = -L \frac{di_{kp}}{dt} - Ri_{kp} - u_{kp} + u_{dp}$$
$$u_k - L_T \frac{di_k}{dt} = L \frac{di_{kn}}{dt} + Ri_{kn} + u_{kn} + u_{dn}$$
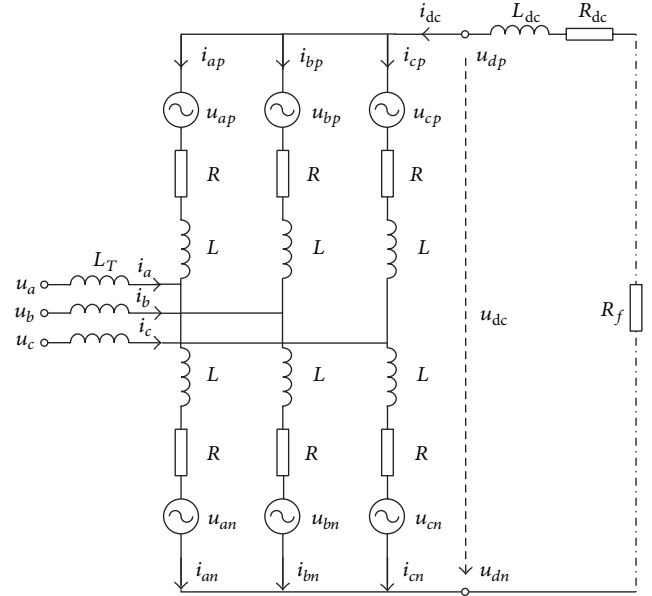
FIGURE 1: Equivalent circuit under dc fault state.

$$i_k = i_{kn} - i_{kp}$$
$$i_{dc} = \sum_k i_{kp} = \sum_k i_{kn},$$
(1)

where $k = a, b, c$ and L, R, and L_T represent the equivalent values of the inductance, resistance, and transformer leakage inductance.

The dynamics of the dc fault current i_{dc} and voltage u_{dc} are expressed as

$$i_{dc} = C_{eq} \frac{d \sum_k u_{kp}}{dt} = C_{eq} \frac{d \sum_k u_{kn}}{dt}$$
$$u_{dc} = u_{dp} - u_{dn} = -L_{dc} \frac{di_{dc}}{dt} - \left(R_{dc} + R_f\right) i_{dc},$$
(2)

where L_{dc} and R_{dc} represent the equivalent values of the dc side inductance and resistance, R_f represents the equivalent value of the short-circuit fault resistance, and C_{eq} represents the equivalent value of the capacitance under the fault state.

Combining these aforementioned equations, the following equation can be obtained:

$$\frac{d^2 i_{dc}}{dt^2} + \frac{R_e}{L_e} \frac{di_{dc}}{dt} + \frac{1}{L_e C_e} i_{dc} = 0,$$
(3)

where $L_e = 2L/3 + L_{dc}$, $R_e = 2R/3 + R_{dc} + R_f$, and $C_e = 3C_{eq}$.

Thus, the fault equivalent circuit is a second-order oscil-lated discharging circuit. When dc fault occurs, the dc side fault current can be determined by the derivative of (3) at $t = 0$, which is expressed by

$$i_{dc} = e^{-t/\tau} \left[-\frac{I_{dc0} \omega_0}{\omega} \sin(\omega t - \alpha) - \frac{U_{dc0}}{\omega L_e} \sin(\omega t) \right],$$
(4)

(a)

(b)

Figure 2: Schematic and block operation mode of FBSM: (a) schematic of FBSM; (b) block operation mode of FBSM.

where I_{dc0} and U_{dc0} are the initial values of i_{dc} and u_{dc}. $\tau = 2L_e/R_e$, $\omega = \sqrt{1/(L_eC_e) - [R_e/(2L_e)]^2}$, $\omega_0 = \sqrt{1/(L_eC_e)}$, and $\alpha = \arctan(\omega\tau)$.

The short-circuit fault current can be calculated by (4). The discharge process before converter blocking is an oscillated discharge procedure which has already known the initial conditions. The equivalent value of the capacitance changes after the converter blocking. If and only if the equivalent capacitor voltage under any circuit state is larger than the ac line voltage amplitude, the short-circuit current can be reduced to zero by the antiparallel diodes. Then, the dc fault would be blocked.

2.2. SM Circuit Topologies under Blocking States

2.2.1. Full-Bridge Submodule.
The full-bridge submodule is as shown in Figure 2(a). The control system sends blocking signals to all the IGBTs on the bridge arms when dc fault is found. The path of the short-circuit current is shown in Figure 2(b). The SM's equivalent capacitor voltage after blocking is in the opposite direction with the ac side voltage. Thus, the short-circuit current is blocked from feeding into the converter. However, because a full-bridge submodule needs four IGBTs, the number of required components is doubled to that of the half-bridge submodule needs. So, the full-bridge submodule based MMC is not cost-effective.

2.2.2. Clamp-Double Submodule.
The clamp-double submodule, as shown in Figure 3(a), is constituted by two

equivalent half-bridge SMs and an IGBT with freewheeling diode. When the dc side fault occurs, the CDSM is switched to the block mode, as shown in Figure 3(b). The equivalent capacitor voltage can stop the fault current from feeding into the converter. However, the equivalent capacitor voltages of a SM are different when the flow of the fault current i_{sm} changes. If $i_{sm} > 0$, the voltage is $2U_c$; otherwise, it turns to be U_c. Correspondingly, the equivalent capacitances are $C_0/2$ and $2C_0$ for each of them separately. Thus, during the dc fault periods, the system charges the SM capacitors repeatedly. This leads to the delay of fault blocking time.

2.2.3. Series-Double Submodule.
As the HBSM has the advantages of having simple structure and mature control strategy and being economical, reliable, and efficient, the series-double submodule (SDSM) presented is based on two HBSMs, as shown in Figure 4(a). Between the two HBSMs, there is an IGBT with freewheeling diode to isolate them. SDSM features are economical, reliable, and efficient as HBSM and can block dc fault. Under steady-state operations, the middle IGBT is turned on, and the SDSM is equivalent to two HBSMs in series. At this point, the SDSM's control strategy is simple as the HBSM's. When the dc side fault happens, the control system sends blocking signals to all the IGBTs. As shown in Figure 4(b), no matter how the current flow direction is, the equivalent capacitor voltage of SDSM is $2U_c$, and the voltage direction is opposite to the alternating voltage direction at any moment. So the short-circuit current from the ac system is blocked, and the SDSM based MMC has the dc fault ride-through capability.
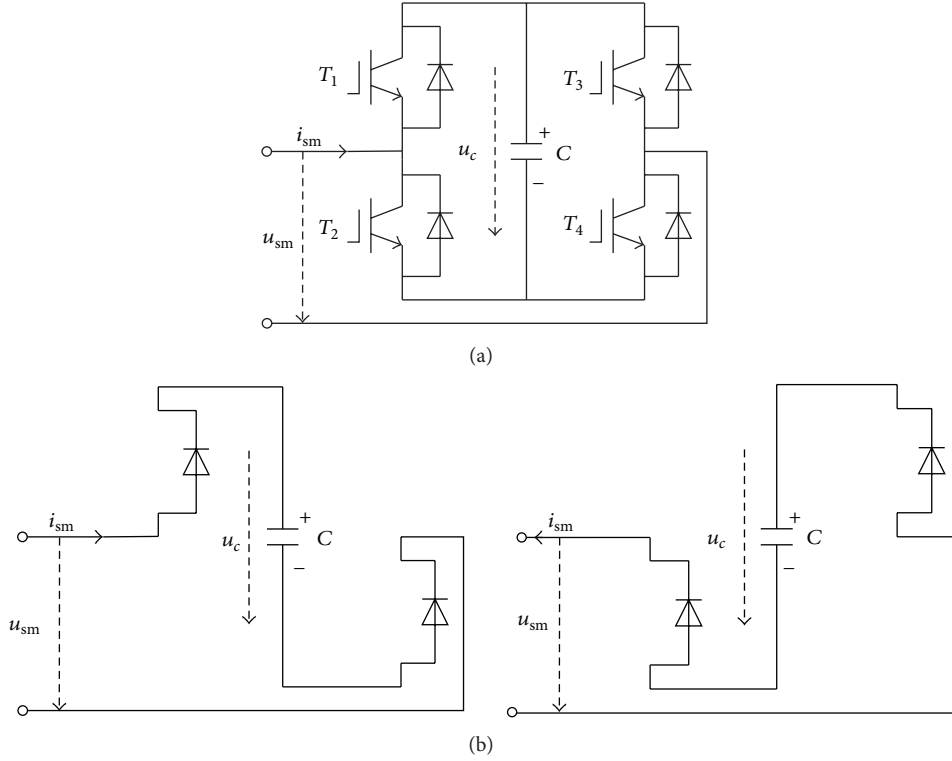
(a)

(b)
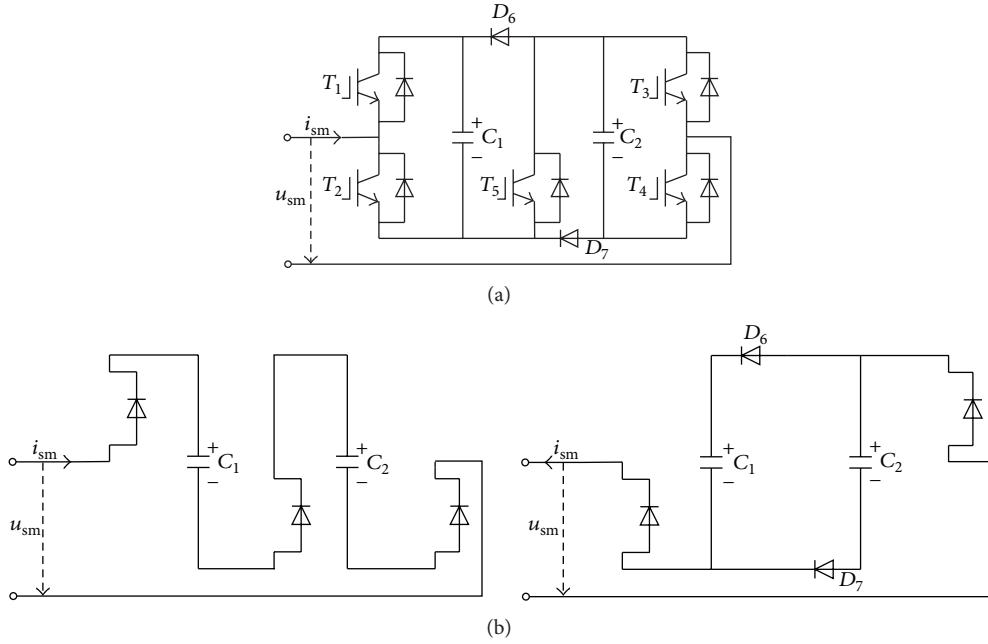
FIGURE 3: Schematic and block operation mode of CDSM: (a) schematic of CDSM; (b) block operation mode of CDSM.

(a)   (b)
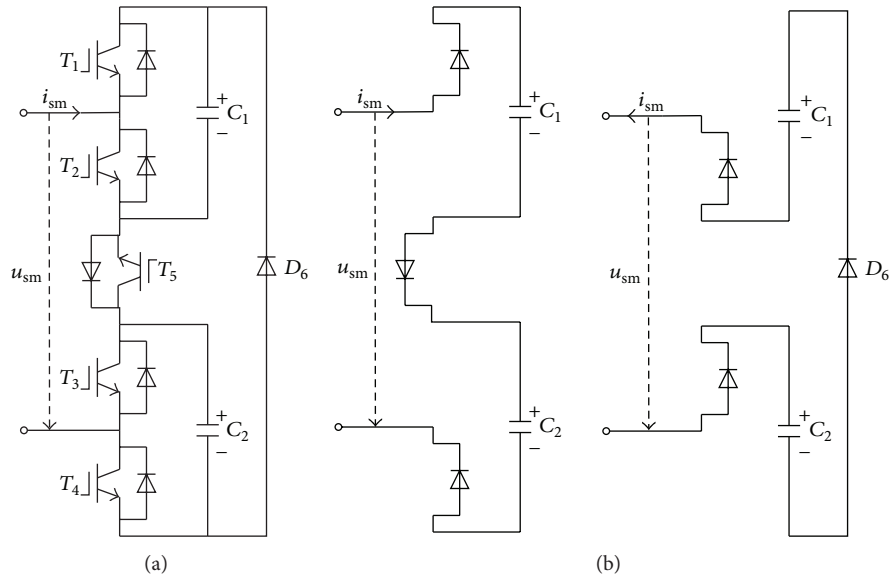
FIGURE 4: Schematic and block operation mode of SDSM: (a) schematic of SDSM; (b) block operation mode of SDSM.

As shown in Table 1, SDMMC needs less semiconductors and has the same dc fault blocking capability compared to FBMMC. As for CDMMC, due to the characteristics of parallel structure, the equivalent capacitance in a bridge arm shows two kinds of states according to the different flows of short-circuit current, as shown in Figure 3(b). So it requires longer time to cut off the short-circuit current, and its dc fault blocking ability is inferior compared with FBMMC and SDMMC. The comparison of dc fault blocking time of various MMC configurations is shown in Figure 5. In consideration of the number of semiconductors used and power losses,

SDMMC is better than FBMMC. Power losses are represented by the estimated power losses, which is the ratio of converter's losses (primarily the IGBT switching losses) and the rated capacity of MMC. Among these, the switching losses are calculated by the on-off time and the single switch loss of IGBT. The estimated power losses in Table 1 are under the circumstances when N is equal to 10. So SDSM is a kind of good submodule topology applied into MMC structure with dc fault ride-through capability.

Also, SDSM has a weakness as well as these advantages above. When all the IGBTs are blocked, the semiconductors

TABLE 1: Comparison of the MMC configurations with various SMs.

Features	Half-bridge MMC	Full-bridge MMC	Clamp-double MMC	Series-double MMC
SM circuit	HBSM	FBSM	CDSM	SDSM
dc fault blocking capability	×	√	√	√
Blocking time	×	18 ms	24 ms	18 ms
Number of SMs per arm	$2N$	$2N$	N	N
Number of IGBTs per arm	$4N$	$8N$	$5N$	$5N$
Number of diodes per arm	$4N$	$8N$	$7N$	$6N$
Number of capacitors per arm	$2N$	$2N$	$2N$	$2N$
Voltage levels	$2N + 1$	$2N + 1$	$2N + 1$	$2N + 1$
Estimated power losses	0.69%	0.96%	0.83%	0.83%
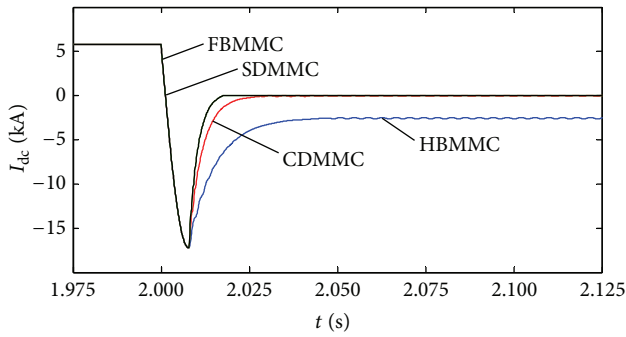
FIGURE 5: dc fault blocking time of various MMC configurations.

T_5 and D_6 need to support doubling the blocking voltage (i.e., $2U_c$). So the related semiconductors (T_5 and D_6) need a special design to overcome this larger voltage than other ones. A common method is to be realized by series connection of two semiconductors. In this instance, the number of IGBTs per arm needed by SDMMC is increased from $5N$ to $6N$. It is a little more than CDMMC needed, but still less than FBMMC used. So the power losses produced by SDMMC (T_5 is replaced by the series connection of two semiconductors) are less than FBMMC but are a little more than CDMMC. However, as shown in Figure 5 and Table 1, SDMMC is better than CDMMC in dc blocking time. It will save 25% time than the CDMMC to block the dc fault. Consequently, in consideration of the double blocking voltage problem, the proposed SDMMC still has its special advantages compared with other MMC configurations.

3. Topology of Mismatched-Cascade Mechanism Based Hybrid MMC

3.1. Application of Mismatched-Cascade Mechanism. Mismatched-cascade mechanism can be used to improve the quality of step wave. Firstly, it produces a small range of dislocations to the basic waveform. Secondly, it stacks these mismatched waveforms to produce a more meticulous waveform. Mismatched-cascade mechanism based MMC is constituted by adding a mismatched-cascade submodule (MCSM) to the former structure. The MCSM is formed by connecting some half-bridge units in series, as shown in Figure 6.

m represents the number of half-bridge units in an MCSM. U_{mc} represents the nominal voltage of the half-bridge unit capacitor, and it must be restricted as

$$U_{mc} = \frac{U_c}{m + 1}. \tag{5}$$

When $n = 4$, $m = 2$, for example, then $U_{mc} = U_c/3$; the output voltage waveform in the effect of MCSM and the switching mode of SMs are shown separately in Figure 7 and Table 2. As the upper and lower bridges are symmetrical, the output voltage waveform only considers the condition of $u_a > 0$.

When the number of half-bridge units in upper bridge's MCSM is $m_p = 1$, correspondingly, the number of half-bridge units in lower bridge's MCSM is $m_n = 1$; the output voltage waveform remains precisely unchanged. However, when m_p increases and m_n decreases, on the contrary, the output voltage levels can produce small dislocations, as shown in Figure 7. The FFT analysis demonstrates that the output voltage has a better quality (THD is changed from about 12% to 4%) after the MCSMs are added to the MMC structure.

After the MCSM applied, the number of output voltage levels is changed from $(n + 1)$ to N_{out}:

$$N_{out} = (n + 1)(m + 1). \tag{6}$$

When $n = 20$, $m = 2$, for example, the number of output voltage levels will be changed from 21 to 63 after the MCSM applied. And the voltage total harmonic distortion (THD) under different modulation ratio k is shown in Figure 8.

Thus, it can be observed that only an MCSM is needed to increase the number of output voltage levels several times over, and the structure of MMC is not changed sharply. This will reduce the harmonic content of output waveform largely.

3.2. Topology of Hybrid MMC. Based on the fault blocking principle shown in Section 2 and the mismatched-cascade mechanism shown above, a hybrid MMC topology is presented here. As stated above, HBSM has the advantages of having simple structure and mature control strategy and being economical, reliable, and efficient. SDSM has the dc fault ride-through capability. And MCSM can optimize

TABLE 2: Switching mode of SMs under different output states.

Number of SMs switched		Number of HBSMs switched in MCSM		Output voltage
Upper bridge	Lower bridge	Upper bridge	Lower bridge	
2	2	1	1	0
2	2	0	2	U_{mc}
1	3	2	0	$2U_{mc}$
1	3	1	1	$3U_{mc}$
1	3	0	2	$4U_{mc}$
0	4	2	0	$5U_{mc}$
0	4	1	1	$6U_{mc}$
0	4	0	2	$7U_{mc}$
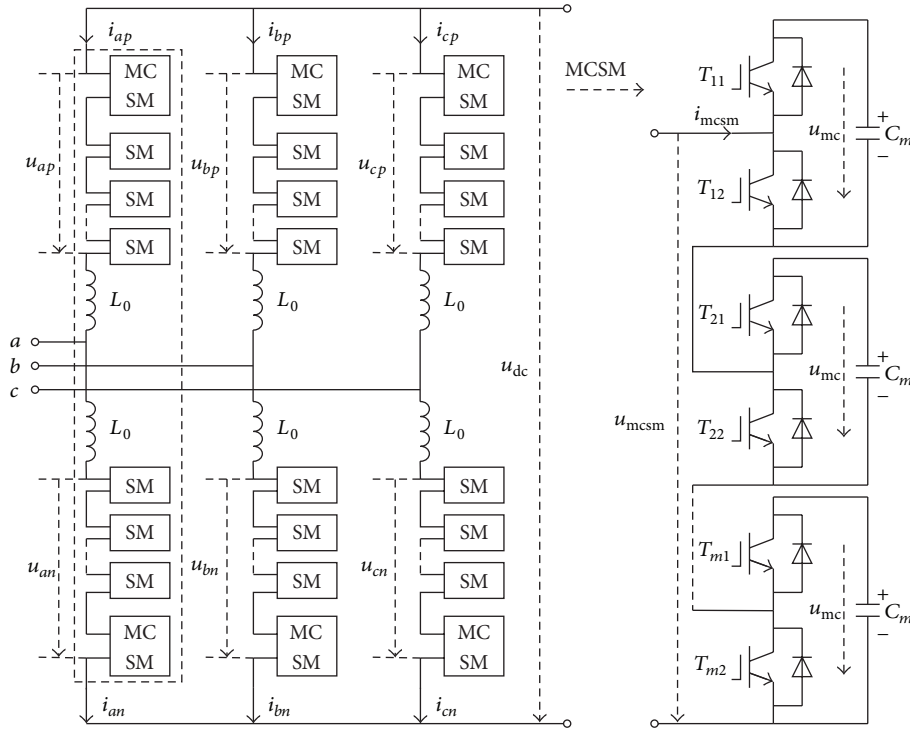
FIGURE 6: Topology structure of mismatched-cascade based MMC.
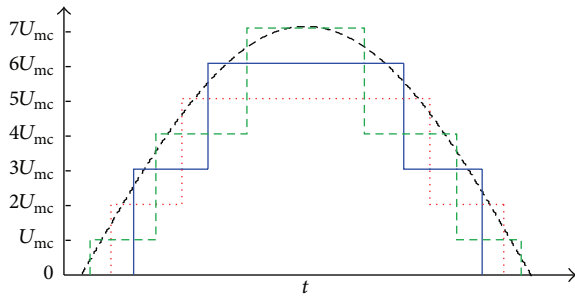
FIGURE 7: Schematic diagram of voltage waveform in the effect of MCSM.

the quality of output voltage waveform. So it is considered to connect them to build a hybrid MMC which has all these advantages. In the hybrid MMC, N_S represents the minimum number of SDSMs which a bridge arm needed. Considering the constraint conditions of dc fault blocking, the equivalent capacitor voltage must be larger than the reversed alternating voltage at any moment. So N_S must follow the principle as follows:

$$N_S \geq \frac{\sqrt{3}U_m}{4U_c}, \tag{7}$$

where U_m represents the maximum value of ac side phase voltage and U_c represents the voltage of a single capacitor in SDSM. On equal conditions, the reverse voltage is larger when N_S increases, and so the blocking time needed is shorter. Thus, the hybrid MMC needs more time to block the dc fault than the MMC made up by SDSM only. To balance dc fault blocking capacity and economical efficiency, under the premise of meeting the requirement of fault blocking time, the number of SDSMs will be decreased as much as possible.

TABLE 3: dc fault clearing time in different combinations of N_H and N_S.

Scheme	N_H	N_S	dc fault blocking time/ms	Number of IGBTs used in a single-bridge arm	THD	Efficiency of each scheme	Estimated power losses
1	12	4	25	44	0.04	0.477	0.79%
2	8	6	16	46	0.04	0.457	0.79%
3	4	8	12	48	0.04	0.438	0.79%
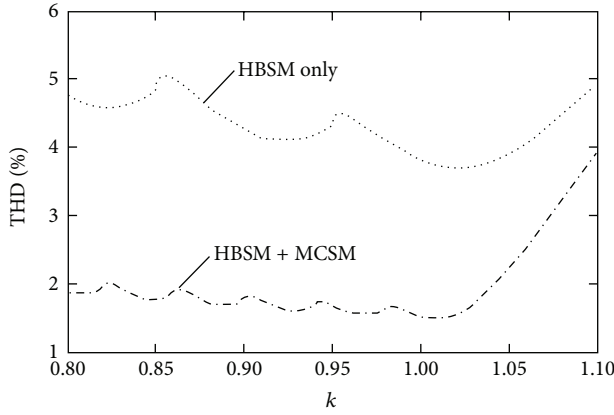
FIGURE 8: The voltage THD changed after the MCSM applied.

Here N_H is defined to represent the number of HBSMs on a bridge arm, and N_S is the number of SDSMs used. When the total voltage level number N_P is fixed, N_H and N_S should follow the principle as follows:

$$N_H + 2N_S + 1 = N_P. \tag{8}$$

To determine the appropriate proportional relation of the number of HBSMs and SDSMs, a simulation model for hybrid MMC was built in PSCAD/EMTDC, and the dc fault blocking time in different combinations of N_H and N_S was measured as shown in Table 3.

As the SDSM is constituted by two HBSMs, and the middle IGBT is turned on under steady-state operations, the SDSM is equivalent to two HBSMs in series. So the THDs and the estimated power losses of schemes 1~3 are almost equal. The efficiency of each scheme means each IGBT's contribution to the total output levels. It is the ratio of the output levels and the number of IGBTs used in a single-bridge arm. From Table 3, it can be observed that scheme 1 has the highest efficiency among all the three schemes since the blocking time is the longest too. However, as all the blocking time is less than 30 ms which is short enough to block the fault, scheme 1 is able to meet the dc fault blocking time requirement when the target is not too strict. Under such circumstances, the required number of IGBTs is suitable, and the hybrid MMC based transmission project is economical.

When MCSM is added, the total voltage level number would be changed from N_P to N'_P. N_M is defined to represent the number of half-bridge units in an MCSM. Then, N'_P should be processed as

$$N'_P = (N_H + 2N_S + 1) \times (N_M + 1). \tag{9}$$

This means that the output voltage level number is increased several times over when adding MCSM to the hybrid MMC.

4. Application of Hybrid MMC in MTDC Transmission System

The multiterminal direct current (MTDC) transmission system can be divided into series, parallel, and hybrid forms, by the different connection modes. Parallel connection mode is the most popular one owing to its characteristics of simplicity, flexibility, and expandability [5]. In this section, a typical parallel four-terminal direct current transmission system is utilized to analyze the procedure of the hybrid MMC-MTDC transmission system blocking dc side fault.

Reference [20] proposes a "handshaking method" and its procedure to deal with the dc fault in two levels' VSC-MTDC transmission system. Using this method, the voltage stability and power recovery after failure can be insured. However, as circuit breakers are needed to cut off the short-circuit current from ac side, the time required to block the fault and recover the rest is too long. Meanwhile, the hybrid MMC presented has the dc fault ride-through capability by itself, so the hybrid MMC based MTDC transmission system can process the dc fault much faster.

As is shown in Figure 9, hybrid MMC is applied in this four-terminal MTDC. $S_1 \sim S_8$ are the disconnecting switches in each dc transmission line. Based on this system, the process of the presented hybrid MMC-MTDC removing the fault lines and recovering the nonfault lines under dc side fault is as follows.

Step 1. Each converter monitors the dc side bus voltage and current in real time. Using the dc fault detection method [20, 21], and combined with the system parameters, the control system is able to acquire when the dc fault occurs and which kind of it.

Step 2. After the dc fault is detected, according to the changes of current, voltage, and power flow at the dc side bus of each MMC, the potential fault lines can be detected [22].

Step 3. The disconnecting switches in the potential fault lines are marked, so that the fault lines can be isolated after the dc fault blocking.

Figure 9: Structure of four-terminal MMC-MTDC system.

Step 4. Blocking signals are sent to the SMs which have the dc fault ride-through capability in the hybrid MMC. At the same time, all the by-pass switches paralleled with other SMs are turned on to protect the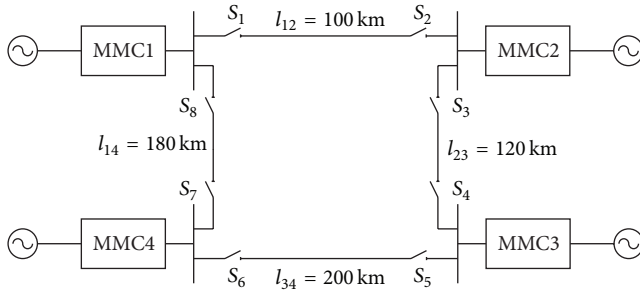 semiconductor devices in them. Considering the delay time of detecting fault and blocking IGBTs, the time needed for MMC to deal with the dc fault is within 5~10 ms.

Step 5. The disconnecting switches marked in Step 3 are turned off. According to the simulation results and the dc fault clearing time listed in Table 3, the switches should be disconnected at about 30 ms after the converters receive blocking signals.

Step 6. The control system sends unblocking signals to the converters to unblock all the IGBTs in MMC and turns off the by-pass switches in Step 4.

Step 7. The converters are set to run in constant dc voltage control mode until the dc bus voltage increases to its rated value. At the moment, the disconnecting switches in the potential fault lines measure the voltages on their both sides. If the voltage difference is less than a certain value, the switch would close itself.

Step 8. After the switches in the nonfault lines closed, all of the converters are set to their original modes. And then, the fault line is isolated and the nonfault lines are returned to service. The hybrid MMC-MTDC transmission system reaches a new steady state under the presupposed operating conditions.

5. Simulation Analysis

To verify the reliability of the represented hybrid MMC-MTDC transmission system and its capability to block dc fault, a four-terminal MTDC simulation model is built in PSCAD/EMTDC. Where MMC1 uses the constant dc voltage and ac voltage control mode, MMC2 adopts the dc voltage drop and constant ac voltage control mode; both MMC3 and MMC4 take the constant active and reactive power control mode. Each bridge arm in hybrid MMC is consisted by 12 HBSMs, 4 SDSMs, and 1 MCSM, and an MCSM is made up by 4 half-bridge units. According to (9), the maximum number

of output voltage levels is changed from 21 to 105 after MCSM added. It reduces the harmonic content of output waveform immensely.

The parameters of this hybrid MMC-HVDC simulation system are listed in Table 4. The power direction flows into the converter are set as positive. The converter transformers adopt the YNd11 connection scheme. The converters are connected by high voltage direct current cable, and the unit of distance impedance is 0.01 Ω/km.

5.1. Steady-State Simulation and Analysis. When the hybrid MMC-MTDC transmission system is under the steady-state operation, the active power of MMC3 is changed from 50 MW to 80 MW and the active power of MMC4 is changed from 100 MW to 50 MW at 0.6 seconds. The active power of MMC3 is changed from 80 MW to 100 MW and the active power of MMC4 is turned from 50 MW to −50 MW at 0.8 seconds. The reactive power of MMC4 is turned from −50 Mvar to 50 Mvar at 1.0 second. The active power of MMC3 and MMC4 is resumed to 80 MW and 100 MW at 1.2 seconds. The simulation waveforms are shown in Figure 10.

As shown in Figure 10, since MMC1 adopts the constant dc voltage control mode and the power change is within the adjustable range, the active power imbalance is compensated by MMC1 at the time of 0.6 seconds. The active power of MMC4 is changed significantly at the time of 0.8 seconds and the power flow turns to the opposite direction. According to the dc voltage control mode of MMC2, both MMC1 and MMC2 are needed to compensate the active power and stabilize the dc voltage. Meanwhile, the dc voltage of MMC4 is changed from slightly larger to less than 400 kV, so that MMC4 can absorb the active power from the system. At the time of 1.0 second, the reactive power of MMC4 changes from −50 Mvar to 50 Mvar and the dc voltage of MMC4 tends towards stability after a slight oscillation. During this process, the other converters' dc voltages always keep stable. It means that the change of reactive power rarely has influence on the stability of the dc system. At the time of 1.2 seconds, the power of MMC3 and MMC4 recovered, so MMC1 and MMC2 changed their output to maintain the system stability. During the whole process, after the changes of parameters, the system restores to a steady state within a short time (less than 50 ms), and the fluctuation is within ±1%. All of these indexes meet the requirements of the system stability.

5.2. Simulation of the DC Bipolar Short-Circuit Fault. To verify the capability of the new hybrid MMC-MTDC under dc side fault and analysis of the characteristic of nonfault lines' recovery, a dc fault in lines 1-2 close to MMC1 is set at the time of 2 seconds. The dc fault can be divided into three kinds: monopolar grounding fault, tripping fault, and bipolar short-circuit fault. In all the three kinds of dc faults, the bipolar short-circuit fault causes the most serious effect [23]. So, to make the focal points stand out, only bipolar short-circuit fault is simulated and analyzed in this section. Combined with the control strategy presented in Section 4, a hybrid MMC based four-terminal dc transmission system is built in PSCAD/EMTDC, and the waveforms of fault state simulation are shown in Figure 11.

TABLE 4: Parameters of the hybrid MMC-HVDC simulation system.

Parameters	MMC1	MMC2	MMC3	MMC4
Rated capacity of MMC	400 MVA	300 MVA	100 MVA	100 MVA
ac bus voltage	220 kV	220 kV	110 kV	110 kV
dc bus voltage		±200 kV		
Rated capacity of transformer	400 MVA	320 MVA	120 MVA	120 MVA
Ratio of transformer		220 kV/230 kV	110 kV/230 kV	
Leakage reactance		0.1 pu		
Bridge arm reactance	32 mH	55 mH	165 mH	165 mH
HBSMs' number		12		
SDSMs' number		4		
MCSMs' number		1		
SM capacitance	3400 μF	2550 μF	850 μF	850 μF

(a)

(b)

(c)
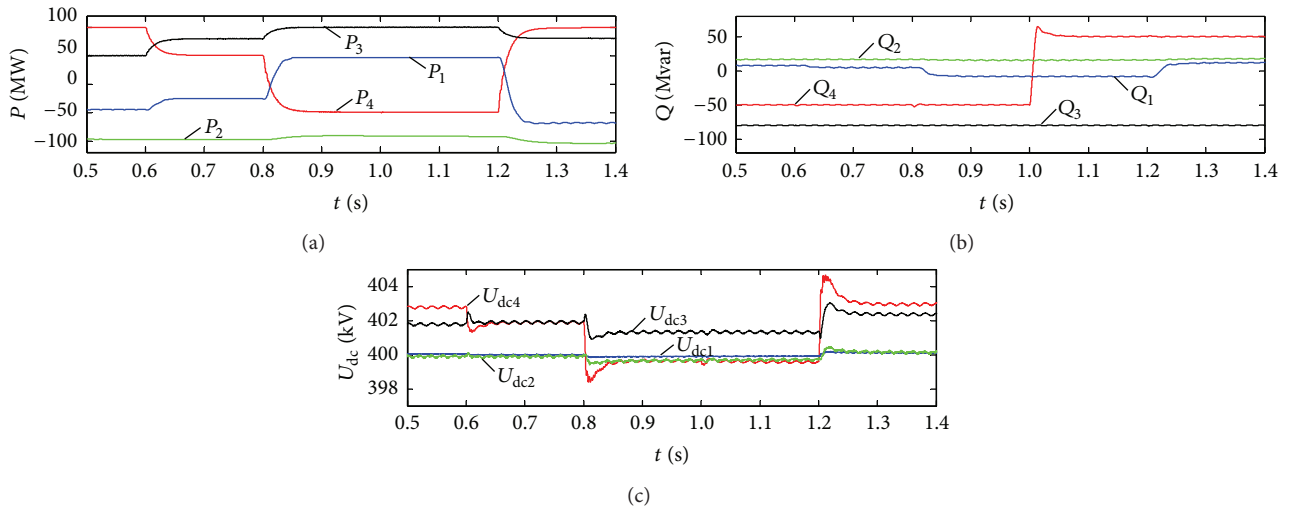
FIGURE 10: Waveforms of steady-state simulation: (a) active power variations; (b) reactive power variations; (c) dc voltage variations.

As shown in Figures 11(a) and 11(b), after the dc fault occurs, the hybrid MMC blocks itself; then the active power and reactive power reduce to zero within 35~40 ms. When the converter is unblocked, the dc voltage recovers to 400 kV; then the active power and reactive power turn to their new steady states. Due to the fact that the fault lines have been isolated and that the system network parameters have been changed, the transmission power of MMC1 and MMC2 changes from the original state to a new steady state.

As shown in Figures 11(c) and 11(d), after the dc fault occurs, the fault current increases sharply and the dc voltage falls off due to the capacitor discharge and the ac side through fault current. Then, the control system sends blocking signals to block the hybrid MMC. After blocking, the capacitor voltage stops the ac current from flowing into the dc side. Then, the short-circuit current and dc voltage go down to zero, respectively. After that, the disconnecting switches are turned off, and unblocking signals are sent to the converters. At the same time, all the converters are set to run in constant dc voltage control mode until the dc bus voltages increase to their rated values. At this moment, the disconnecting switches in the nonfault lines detect the voltage differences

are less than the set value and close themselves. At last, all the converters are set to their original modes and i_{dc} rises to a new steady value.

As shown in Figures 11(e) and 11(f), the SMs' voltages stay near the rated voltage value 20 kV. So, after the fault line is removed, there is no need to charge the capacitors in the SMs again, and then the recovery time will not be too long.

As shown in Figures 11(g) and 11(h), the ac side current decreases to zero during the blocking time and recovers to the steady value when the system is put into operation again. During this process, there is no off-limit condition, and the ac voltage variation is less than 5%. So, all the results indicate that the dc fault can hardly influence the ac system and the new hybrid MMC-MTDC transmission system is stable during the fault process.

6. Conclusions

In this paper, the mismatched-cascade mechanism and dc fault blocking principle are implemented in a hybrid MMC topology. Firstly, the dc fault blocking principle and different topologies of submodules are presented to analyse the fault
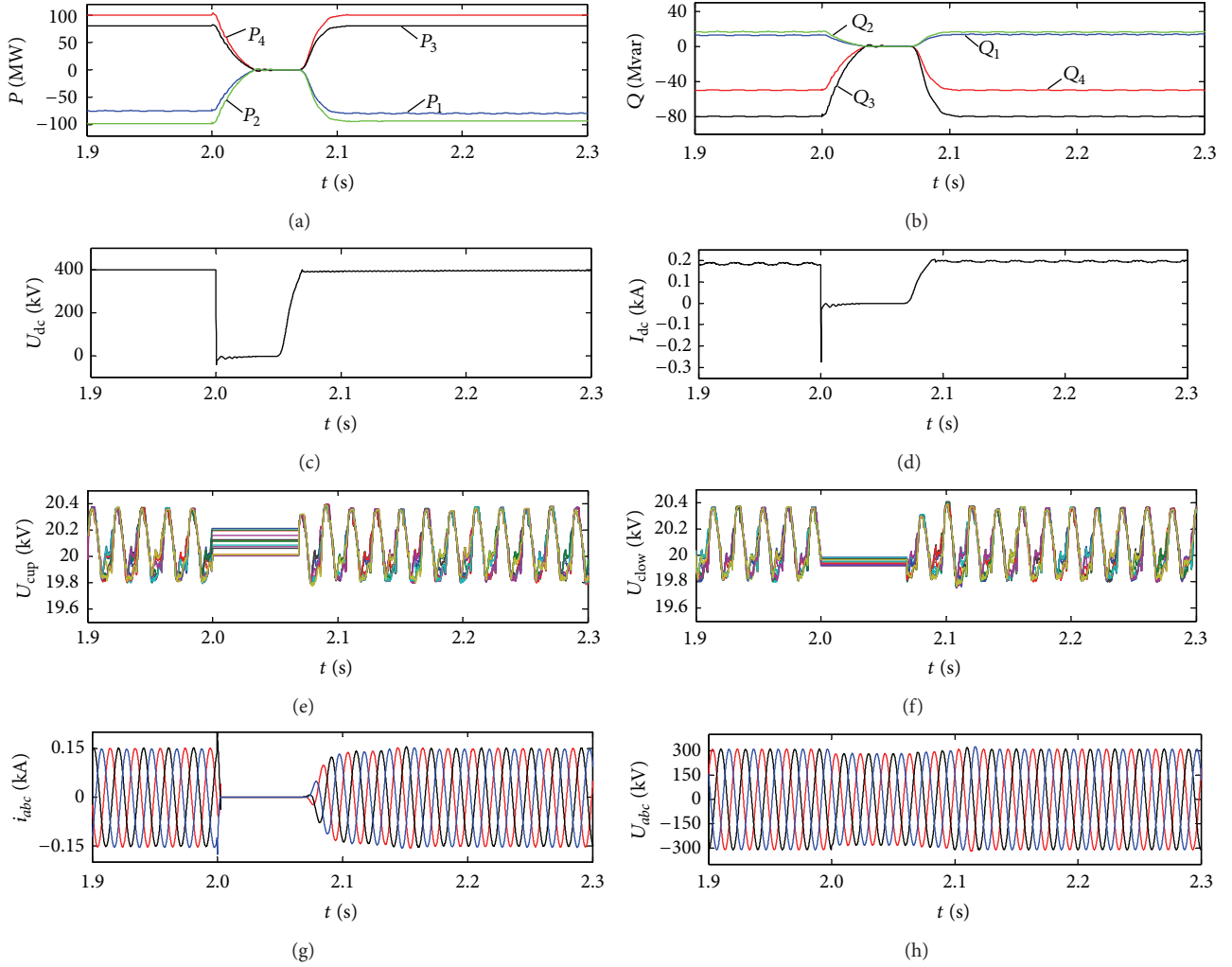
FIGURE 11: Waveforms of fault state simulation: (a) and (b) active and reactive power variations; (c) and (d) dc voltage and current of MMC1; (e) and (f) upper and lower bridge SM capacitor voltages; (g) and (h) ac side three-phase voltage and current of MMC1.

ride-through capability. A new kind of series-double sub-module is proposed. Based on this submodule, the MMC can block dc fault and the number of semiconductor devices is reduced. A new hybrid MMC topology constituted by half-bridge SM, series-double SM, and mismatched-cascade SM is proposed. The effect of this kind of MMC to block dc fault is verified by simulation in PSCAD/EMTDC. Then, an MTDC transmission system based on this hybrid MMC and its process to remove the fault lines and recover the nonfault lines are presented. Finally, a four-terminal dc simulation model is built. Based on this model, both steady state and fault state are analyzed by simulations. According to test results, it is shown that dc system based on this new kind of hybrid MMC can block the dc fault accurately and quickly. And the nonfault parts can be restored to normal operation. With the fault blocking ability and feasible control strategy, this hybrid MMC presented can be applied into connecting the renewable source to power grid and forming multiterminal dc network.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the republication of this paper.

References

[1] N. Flourentzou, V. G. Agelidis, and G. D. Demetriades, "VSC-based HVDC power transmission systems: an overview," *IEEE Transactions on Power Electronics*, vol. 24, no. 3, pp. 592–602, 2009.

[2] Y. Wei, Q. He, Y. Sun, Y. Sun, and C. Ji, "Improved power flow algorithm for VSC-HVDC system based on high-order Newton-type method," *Mathematical Problems in Engineering*, vol. 2013, Article ID 235316, 10 pages, 2013.

[3] N. R. Chaudhuri, R. Majumder, B. Chaudhuri, and J. Pan, "Stability analysis of VSC MTDC grids connected to multimachine AC systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2774–2784, 2011.

[4] S. Cole, J. Beerten, and R. Belmans, "Generalized dynamic VSC MTDC model for power system stability studies," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1655–1662, 2010.

[5] O. Gomis-Bellmunt, J. Liang, J. Ekanayake, R. King, and N. Jenkins, "Topologies of multiterminal HVDC-VSC transmission for large offshore wind farms," *Electric Power Systems Research*, vol. 81, no. 2, pp. 271–281, 2011.

[6] R. Marquardt, "Modular multilevel converter: an universal concept for HVDC-networks and extended dc-bus-applications," in *Proceedings of the International Power Electronics Conference*, pp. 502–507, Sapporo, Japan, June 2010.

[7] J. Dorn, H. Huang, and D. Retzmann, "A new multilevel voltage-sourced converter topology for HVDC applications," in *Proceedings of the Conference International des Grands Reseaux Electriques*, pp. 1–8, Paris, France, 2008.

[8] R. José, L. G. Franquelo, K. Samir et al., "Multilevel converters: an enabling technology for high-poer applications," *Proceedings of the IEEE*, vol. 97, no. 11, pp. 1786–1817, 2009.

[9] P. Bordignon, M. Marchesoni, G. Parodi, and L. Vaccaro, "Modular multilevel converter in HVDC systems under fault conditions," in *Proceedings of the 15th European Conference on Power Electronics and Applications*, pp. 1–10, Lille, France, September 2013.

[10] X. Chen, C. Zhao, and C. Cao, "Research on the fault characteristics of HVDC based on modular multilevel converter," in *Proceedings of the IEEE Electrical Power and Energy Conference (EPEC '11)*, pp. 91–96, Winnipeg, MB, Canada, October 2011.

[11] C. M. Franck, "HVDC circuit breakers: a review identifying future research needs," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 998–1007, 2011.

[12] X. Li, Q. Song, W. Liu, H. Rao, S. Xu, and L. Li, "Protection of nonpermanent faults on DC overhead lines in MMC-based HVDC systems," *IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 483–490, 2013.

[13] F. Chang, Z. Yang, Y. Wang, F. Lin, and S. Liu, "Fault characteristics and control strategies of multiterminal high voltage direct current transmission based on modular multilevel converter," *Mathematical Problems in Engineering*. In press.

[14] T. Modeer, H.-P. Nee, and S. Norrga, "Loss comparison of different sub-module implementations for modular multilevel converters in HVDC applications," in *Proceedings of the 14th European Conference on Power Electronics and Applications (EPE '11)*, pp. 1–7, Birmingham, UK, September 2011.

[15] C. Zhao, J. Xu, and T. Li, "DC faults ride-through capability analysis of full-bridge MMC-MTDC system," *Science China Technological Sciences*, vol. 56, no. 1, pp. 253–261, 2013.

[16] R. Marquardt, "Modular multilevel converter topologies with DC-short circuit current limitation," in *Proceedings of the 8th IEEE International Conference on Power Electronics and ECCE Asia (ICPE & ECCE '11)*, pp. 1425–1431, IEEE, Jeju, Republic of Korea, May-June 2011.

[17] D. R. Trainer, C. C. Davidson, C. D. M. Oates, N. M. Macleod, D. R. Critchley, and R. W. Crookes, "A new hybrid voltage sourced converter for HVDC power transmission," in *Proceedings of the Conference International des Grands Reseaux Electriques*, pp. 1–12, Paris, France, 2010.

[18] M. M. C. Merlin, T. C. Green, P. D. Mitcheson, D. R. Trainer, D. R. Critchley, and R. W. Crookes, "A new hybrid multi-level voltage-source converter with DC fault blocking capability," in *Proceedings of the 9th IET International Conference on AC and DC Power Transmission (ACDC '10)*, pp. 1–5, London, UK, October 2010.

[19] J. Qin, M. Saeedifard, A. Rockhill, and R. Zhou, "Hybrid design of modular multilevel converters for HVDC systems based on various submodule circuits," *IEEE Transactions on Power Delivery*, vol. 30, no. 1, pp. 385–394, 2015.

[20] L. Tang and B.-T. Ooi, "Locating and isolating DC faults in multi-terminal DC systems," *IEEE Transactions on Power Delivery*, vol. 22, no. 3, pp. 1877–1884, 2007.

[21] J. Yang, J. E. Fletcher, and J. O'Reilly, "Short-circuit and ground fault analyses and location in VSC-based DC network cables," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 10, pp. 3827–3837, 2012.

[22] L. Tang and B.-T. Ooi, "Protection of VSC-multi-terminal HVDC against DC faults," in *Proceedings of the IEEE 33rd Annual Power Electronics Specialists Conference (PESC '02)*, vol. 2, pp. 719–724, June 2002.

[23] D. Schmitt, Y. Wang, T. Weyh, and R. Marquardt, "DC-side fault current management in extended multiterminal-HVDC-grids," in *Proceedings of the 9th International Multi-Conference on IEEE Systems, Signals and Devices*, pp. 1–5, Chemnitz, Germany, March 2012.

Research Article

An Improved Macro Model of Traffic Flow with the Consideration of Ramps and Numerical Tests

Zhongke Shi, Wenhuan Ai, and Dawei Liu

College of Automation, Northwestern Polytechnical University, No. 127 Youyi Road (West), Beilin, Xi'an, Shaanxi 710072, China

Correspondence should be addressed to Zhongke Shi; shizknwpu@126.com

We present an improved macro model for traffic flow based on the existing models. The equilibrium point equation of the model is obtained. The stop-and-go traffic phenomenon is described in phase plane and the relationship between traffic jams and system instability is clearly shown in the phase plane diagrams. Using the improved model, some traffic phenomena on a highway with ramps are found in this paper. The numerical simulation is carried out to investigate various nonlinear traffic phenomena with a single ramp generated by different initial densities and vehicle generation rates. According to the actual road sections of Xi'an-Baoji highways, the situations of morning peak with several ramps are also analyzed. All these results are consistent with real traffic, which shows that the improved model is reasonable.

1. Introduction

In recent years, traffic jams has become more and more serious. They do not only cause a large number of costs but they also have a negative impact on the environment and energy sustainability. Therefore, researchers have made many efforts to develop transportation electrification to alleviate the impact. Hu et al. [1, 2] analyzed the energy efficiencies of a series plug-in hybrid electric bus with different energy management strategies and battery sizes and comparatively examined three different electrochemical energy storage systems for a hybrid bus powertrain. Recently, they [3] also investigated the optimal component sizing and power management of a fuel cell/battery hybrid bus. Sun et al. [4] presented a traffic data-enabled predictive energy management framework for plug-in hybrid electric vehicles. Furthermore, they [5] studied the velocity predictors for predictive energy management in hybrid electric vehicles. However, many other physicists and engineers have tried to develop traffic models with the aim of optimizing traffic flow. During the past decades, lots of traffic models have been constructed to replicate the formation mechanism and inherent law of the traffic phenomena. In microscopic view, the traffic flow system was regarded as a complex self-driven many-particle system composed of a large number of vehicles. The

microscopic traffic flow models investigated the dynamical behavior of a single vehicle and the interactions between the vehicles, so they can also describe the traffic phenomena of the whole system. Among them, the car-following model is a favorable type of traffic models describing the driver's following behavior in view of the stimulus from its preceding vehicle. On the macroscopic view, physicists paid close attention on the collective behavior of traffic. Due to the analogy of vehicle stream with gas stream or fluid stream, large numbers of the gas kinetic models or fluid-dynamic models have been developed to approximately describe the traffic phenomena. Based on these models, researchers can use a lot of system simulation methods to analyze the traffic phenomena.

The traffic flow on a highway with the ramp has been studied for decades through observation and modeling. Lee et al. [6] studied the presence of the external vehicle flux through ramps and found a new kind of traffic phenomenon, called "recurring humps" (RH). In this state, the density and the flow oscillated periodically and the oscillations concentrated around the ramp. Gupta and Katiyar [7] studied the phase transition on a highway in a modified anisotropic continuum model with an on-ramp. Huang [8] observed an interesting phase: jam-max.-free when the on-ramp is placed before the off-ramp. He also demonstrated that the bulk properties on the roadway are totally controlled by the ramp flow through

the boundaries. Tang et al. [9] indicated that ramps often have different effects on the main road traffic during the morning rush period and the evening rush period and that the effects are related to the initial status of the main road traffic flow.

However, these models cannot completely describe the various complex phenomena resulted by different input and output conditions on ramps. In particular, the phenomena of fixed vehicle generation rate but increasing initial homogeneous density with a single ramp, the situation of morning peak, and the congested traffic stream with several ramps are rarely studied in the past. In this paper, we present an improved macro model for traffic flow to analyze these phenomena on a highway with a single ramp and multiple ramps. Moreover, we introduce a completely different method to describe traffic phenomena in the phase plane diagrams from a stability perspective. The variable substitution is adopted in the models and the traffic congestion corresponds to the unstable system in phase plane. So the traffic flow problems can be converted into the system stability problems.

The remainder of the paper is organized as follows. In Section 2, we present an improved continuum model based on the existing traffic flow model. In Section 3, we deduce the equilibrium point equation of the model. In Section 4, we analyze the well-known stop-and-go waves using the phase plane diagrams based on the improved model and compare them with the traditional temporal evolution of vehicle density. In Section 5, we use the improved model to describe various nonlinear phenomena on a highway with a single ramp. In Section 6, the traffic flow on a highway with multiple ramps is also studied and the actual traffic phenomena of Xi'an-Baoji highways are discussed. We conclude the paper in Section 7.

2. Models

The macroscopic traffic flow models consider vehicles as interacting particles and consider traffic flow as a one-dimensional compressible flow of these particles. The study of macroscopic traffic flow models began with the LWR model proposed by Lighthill and Whitham [10] and Richards [11]. To overcome the shortage of the LWR model, Payne [12] developed a higher order model by using a dynamic equation for the mean velocity. Hereafter, many researchers presented a great number of models based on Payne's model [13–15]. However, these models fail to describe the property that the characteristic speeds are always less than or equal to the macroscopic flow speed. Later, Zhang [16] proposed a macroscopic traffic flow model which overcomes the backward travel problem. Gupta and Katiyar [17] also developed an anisotropic continuum model which is referred to as GK model. Although these models can describe many complex traffic phenomena, they cannot be used to directly explore the effects of ramps since they do not consider this factor. So far, some theoretical models have been developed to study the effects of ramps [6–9]. However, these models cannot completely describe the various complex phenomena resulted by different input and output conditions on ramps. In particular, the phenomena of fixed vehicle generation rate but increasing initial homogeneous density with a single ramp, the situation of morning peak, and the congested traffic stream with

several ramps are rarely studied in the past. In this paper, we present an improved macro model for traffic flow on a highway with ramps based on the GK model as follows:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho v)}{\partial x} = s(x, t)$$

$$\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = \frac{1}{\tau} \cdot \left[V_e(\rho) - v \right] + \frac{1}{\tau}$$

$$\cdot V_e'(\rho) \left[\frac{1}{2\rho} \frac{\partial \rho}{\partial x} + \frac{1}{6\rho^2} \frac{\partial^2 \rho}{\partial x^2} - \frac{1}{2\rho^3} \left(\frac{\partial \rho}{\partial x} \right)^2 \right] \quad (1)$$

$$- 2\beta c(\rho) \frac{\partial v}{\partial x},$$

where ρ is the density; v is the velocity; x and t represent space and time, respectively; τ is the driver's reaction time; β is a nonnegative dimensionless parameter; $V_e[\rho(x,t)]$ is the optimal velocity function and has the following form [18]:

$$V_e[\rho]$$

$$= v_f \left\{ \left[1 + \exp\left(\frac{\rho/\rho_m - 0.25}{0.06} \right) \right]^{-1} - 3.72 \times 10^{-6} \right\}. \quad (2)$$

Consider $V_e'(\rho) = dV_e(\rho)/d\rho$, v_f is the free-flow speed, ρ_m is the maximum or jam density, and $c(\rho) < 0$ is the traffic sound speed given by

$$c^2(\rho) = - \frac{aV_e'(\rho)}{2}. \quad (3)$$

Consider $s(x, t)$ is the flow generation rate. For simplicity, we here adopt the definition of flow generation rate in Jiang et al. [19]; that is

$$s(x, t) = \begin{cases} \dfrac{q_{\text{ramp}}}{L_{\text{ramp}}} & \forall x \in \Omega_{\text{ramp}}, \\ 0 & \text{else}, \end{cases} \quad (4)$$

where Ω_{ramp} is the region of the ramp, L_{ramp} is the length of the ramp, and q_{ramp} is the total ramp flow. We here define q_{ramp} as follows:

$$q_{\text{ramp}} = q_{\text{ramp}}^{\text{on}} - q_{\text{ramp}}^{\text{off}}, \quad (5)$$

where $q_{\text{ramp}}^{\text{on}}$ is the input flow of the on-ramp and $q_{\text{ramp}}^{\text{off}}$ is the output flow of the off-ramp.

Furthermore, we employ a simple transformation as follows:

$$\sigma = \frac{1}{v}$$

$$\eta = \frac{1}{\rho_m - \rho}. \quad (6)$$

Substituting the variables into (1), we have a new traffic flow model as follows:

$$\frac{\partial \eta}{\partial t} = - \frac{1}{\sigma} \frac{\partial \eta}{\partial x} + \frac{\rho_m \eta^2 - \eta}{\sigma^2} \frac{\partial \sigma}{\partial x} + \eta^2 \cdot s(x, t)$$

$$\frac{\partial \sigma}{\partial t} = \left(2\beta \sqrt{-\frac{v'_e(\eta)}{2\tau}} - \frac{1}{\sigma} \right) \frac{\partial \sigma}{\partial x} - \frac{\sigma^2 v'_e(\eta)}{\tau}$$

$$\cdot \left\{ \frac{1}{2\eta(\rho_m\eta - 1)} \cdot \frac{\partial \eta}{\partial x} + \frac{1}{6(\rho_m\eta - 1)^2} \cdot \frac{\partial^2 \eta}{\partial x^2} \right.$$

$$\left. - \left(\frac{1}{3\eta(\rho_m\eta - 1)^2} + \frac{1}{2\eta(\rho_m\eta - 1)^3} \right) \left(\frac{\partial \eta}{\partial x} \right)^2 \right\}$$

$$- \frac{\sigma^2 v_e(\eta)}{\tau} + \frac{\sigma}{\tau}. \tag{7}$$

Similarly, substituting the variables into (2), the equilibrium velocity $v_e(\eta)$ is as follows:

$$V_e(\eta)$$
$$= v_f \left\{ \left[1 + \exp\left(\frac{0.75 - 1/\eta\rho_m}{0.06} \right) \right]^{-1} - 3.72 \times 10^{-6} \right\}. \tag{8}$$

According to the variable substitution $\sigma = 1/v$, we can see that as long as the traffic becomes congested and the vehicles velocity goes to zero, the state variable σ will approach infinity. Likewise, from the variable substitution, $\eta = 1/\rho_m - \rho$, we can see that if the vehicle density becomes saturated, the state variable η will approach infinity in the same way. So we can use the phase plane diagrams about the variable η or σ to describe clearly the relationship between traffic jams and system instability. As long as the traffic has a very small density fluctuation, the value of η and σ will change sharply. Moreover, as long as there is traffic jam formation, the value of η and σ will approach infinity. The more the value of η and σ increases, the greater the fluctuation of the vehicle density is and the more unstable the traffic system is. On the contrary, the system becomes more stable, so the problem of traffic flow could be converted into that of system stability. We can describe all kinds of nonlinear traffic phenomena with the phase plane diagrams and determine whether there will be traffic congestion or other abnormal phenomena from a global stability point. It may be possible to apply some mathematical tools such as branch and bound to the nonlinear stability analysis of traffic system. We can find the equilibrium solutions and some bifurcations of the new model to regulate the stability of traffic system in the future work.

3. The Equilibrium Point Equation Analysis

When the traffic system reaches equilibrium state, the density and velocity of the whole road will not change with time. Moreover, when the traffic system reaches some special equilibrium points, the density and velocity of the whole road will not change with time and displacement at the same time. In order to find these equilibrium solutions of the new model, the equilibrium point equation of the system is analyzed firstly.

When system (7) does not change with time, we have

$$\frac{\partial \eta}{\partial t} = 0,$$
$$\frac{\partial \sigma}{\partial t} = 0. \tag{9}$$

It is assumed that the input flow of the on-ramp is equal to the output flow of the off-ramp. So we have $s(x, t) = 0$. By substituting (9) into (7), the equilibrium points satisfy the formula as follows:

$$\left(\eta - \rho_m\eta^2 \right) \frac{\partial \sigma}{\partial x} + \sigma \frac{\partial \eta}{\partial x} = 0, \tag{10a}$$

$$\left(\frac{1}{\sigma} - 2\beta \sqrt{-\frac{v'_e(\eta)}{2\tau}} \right) \frac{\partial \sigma}{\partial x} + \frac{\sigma^2 v'_e(\eta)}{\tau} \left\{ \frac{1}{2\eta(\rho_m\eta - 1)} \right.$$

$$\cdot \frac{\partial \eta}{\partial x} + \frac{1}{6(\rho_m\eta - 1)^2} \cdot \frac{\partial^2 \eta}{\partial x^2}$$

$$\left. - \left(\frac{1}{3\eta(\rho_m\eta - 1)^2} + \frac{1}{2\eta(\rho_m\eta - 1)^3} \right) \left(\frac{\partial \eta}{\partial x} \right)^2 \right\}$$

$$+ \frac{\sigma^2 v_e(\eta)}{\tau} - \frac{\sigma}{\tau} = 0. \tag{10b}$$

Equation (10a) can be rewritten as

$$\frac{1}{\sigma} \frac{\partial \sigma}{\partial x} = \frac{1}{\eta(\rho_m\eta - 1)} \frac{\partial \eta}{\partial x}. \tag{11}$$

The right hand side of (11) can be written as

$$\frac{1}{\eta(\rho_m\eta - 1)} \frac{\partial \eta}{\partial x} = \frac{\partial f(\eta)}{\partial \eta} \frac{\partial \eta}{\partial x} = \left(\frac{a}{\eta} + \frac{b}{\rho_m\eta - 1} \right) \frac{\partial \eta}{\partial x}$$

$$= \left(\frac{\rho_m}{\rho_m\eta - 1} - \frac{1}{\eta} \right) \frac{\partial \eta}{\partial x}. \tag{12}$$

So,

$$f(\eta) = \ln(\rho_m\eta - 1) - \ln\eta = \ln\left(\frac{\rho_m\eta - 1}{\eta} \right)$$

$$= \ln\left(\rho_m - \frac{1}{\eta} \right). \tag{13}$$

By integrating (11) at both ends, we have

$$\ln\sigma = f(\eta) + \text{const} = \ln k_0 \left(\rho_m - \frac{1}{\eta} \right). \tag{14}$$

k_0 is a nonzero constant and (14) can be rewritten as

$$\sigma = k_0 \left(\rho_m - \frac{1}{\eta} \right). \tag{15}$$

By substituting (15) and (11) into (10b), we obtain

$$\frac{k_0^2 \eta V_e'(\eta)(\rho_m \eta - 1)}{6\tau} \frac{\partial^2 \eta}{\partial x^2} + \frac{k_0^2 V_e'(\eta)}{\tau} \left[\frac{1}{3}(1 - \rho_m \eta) - \frac{1}{2} \right]$$

$$\cdot \left(\frac{\partial \eta}{\partial x} \right)^2 + \left[\eta^2 - 2\beta k_0 \eta (\rho_m \eta - 1) \sqrt{-\frac{V_e'(\eta)}{2\tau}} \right.$$

$$\left. + \frac{k_0^2 V_e'(\eta)(\rho_m \eta - 1)^2}{2\tau} \right] \frac{\partial \eta}{\partial x}$$

$$+ \frac{k_0 \eta (\rho_m \eta - 1)^2}{\tau} \left[k_0 (\rho_m \eta - 1) V_e(\eta) - \eta \right] = 0. \tag{16}$$

In summary, the system equilibrium points satisfy the following equations:

$$\sigma = k_0 \left(\rho_m - \frac{1}{\eta} \right)$$

$$\frac{k_0^2 \eta V_e'(\eta)(\rho_m \eta - 1)}{6\tau} \frac{\partial^2 \eta}{\partial x^2} + \frac{k_0^2 V_e'(\eta)}{\tau} \left[\frac{1}{3}(1 - \rho_m \eta) - \frac{1}{2} \right]$$

$$\cdot \left(\frac{\partial \eta}{\partial x} \right)^2 + \left[\eta^2 - 2\beta k_0 \eta (\rho_m \eta - 1) \sqrt{-\frac{V_e'(\eta)}{2\tau}} \right.$$

$$\left. + \frac{k_0^2 V_e'(\eta)(\rho_m \eta - 1)^2}{2\tau} \right] \frac{\partial \eta}{\partial x} \tag{17}$$

$$+ \frac{k_0 \eta (\rho_m \eta - 1)^2}{\tau} \left[k_0 (\rho_m \eta - 1) V_e(\eta) - \eta \right] = 0$$

$$V_e(\eta) = v_f \left\{ \left[1 + \exp\left(12.5 - \frac{1}{0.06\rho_m \eta} \right) \right]^{-1} - 3.72 \right.$$

$$\left. \times 10^{-6} \right\}.$$

The meaning of every parameter is the same as above. If the initial values of η and σ are given by the solution of (17), the density and velocity of the whole road will not change with time. At the same time, we can see from (15) that the product of density and velocity is equal to $1/k_0$.

Furthermore, when system (17) also does not change with displacement, we obtain the equilibrium points equation as follows:

$$\sigma = k_0 \left(\rho_m - \frac{1}{\eta} \right), \tag{18a}$$

$$\frac{k_0 \eta (\rho_m \eta - 1)^2}{\tau} \left[k_0 (\rho_m \eta - 1) V_e(\eta) - \eta \right] = 0. \tag{18b}$$

Next, we analyze the solution of (18b). η cannot be zero according to $\eta = 1/(\rho_m - \rho)$. If $\rho_m \eta - 1 = 0$, then ρ is equal to 0. In this case, η is trivial equilibrium point and has no practical significance. So we only need to investigate the following equation:

$$k_0 (\rho_m \eta - 1) V_e(\eta) - \eta = 0. \tag{19}$$

Equation (19) can be written as

$$V_e(\eta) = \frac{\eta}{k_0 (\rho_m \eta - 1)}. \tag{20}$$

At the same time, (18a) can be written as $\sigma = k_0(\rho_m \eta - 1)/\eta$. So we can see from (20) that $V_e(\eta) = 1/\sigma = v$.

Therefore, we may conclude that if the value of initial density is set as a random constant in the reasonable range of traffic flow and the initial velocity is given as the equilibrium velocity which is corresponding to the initial density, the density and velocity of the whole road will not change with time and displacement. These conclusions are also consistent with the phenomena observed in realistic traffic flow.

4. The Stop-and-Go Traffic Phenomena on the Phase Plane

The stop-and-go traffic phenomena are international well-known nonlinear phenomena. Traditional researches on it mainly focused on using the figures of temporal development of density through the original traffic flow models. The new model mentioned above can also describe it through the phase plane diagrams from a system stability perspective. The comparisons and discussions between the two methods by numerical experiments were given as follows. Here we assume that the input flow of the on-ramp is equal to the output flow of the off-ramp.

The stop-and-go phenomena can be observed in the amplification of a small disturbance. In this section, we simulate the stop-and-go phenomena with respect to an amplified localized perturbation in an initial homogeneous condition. The following initial variation of the average density ρ_0 is used as in [20]:

$$\rho(x, 0) = \rho_0 + \Delta\rho_0 \left\{ \cosh^{-2} \left[\frac{160}{L} \left(x - \frac{5L}{16} \right) \right] \right.$$

$$\left. - \frac{1}{4} \cosh^{-2} \left[\frac{40}{L} \left(x - \frac{11L}{32} \right) \right] \right\} \quad x \in [0, L] \tag{21}$$

$$v(x, 0) = V(\rho(x, 0)) \quad x \in [0, L],$$

where ρ_0 is the initial vehicle density, $\Delta\rho_0 = 0.01$ veh/m is the amplitude of localized perturbation, and $L = 32.2$ km is the length of road section under consideration. The dynamic approximate boundary condition was given by

$$\rho(1, t) = \rho(2, t),$$
$$\rho(L, t) = \rho(L - 1, t),$$
$$v(1, t) = v(2, t),$$
$$v(L, t) = v(L - 1, t). \tag{22}$$

For computational purpose, the space domain was divided into equal intervals of length of 100 m and time

interval was chosen as 1 s. The related parameters of our model were as follows:

$$\beta = 2,$$

$$\tau = 14\,\text{s},$$

$$v_f = 30\,\text{m/s},$$

$$\rho_m = 0.2\,\text{veh/m},$$ (23)

$$\rho_0 = 0.052\,\text{veh/m},$$

$$\Delta\rho_0 = 0.01\,\text{veh/m}.$$

The critical density values of the GK model corresponding to the parameters above were 0.037 veh/m and 0.091 veh/m, which can easily be found out by the stability condition [17]. The traffic flow will be unstable between these critical densities. The small disturbance in these initial homogeneous conditions will be amplified, and the stop-and-go phenomena will occur.

Traditionally, people used the temporal evolution of vehicle density or velocity to describe the stop-and-go waves, such as in [17, 21–24]. In particular, [17] employed the GK model to analyze it. The variation range of vehicle density is 0–0.25 veh/m and velocity is 0–30 m/s. So there are rather limited changes in the diagrams about these variables. When the traffic becomes congested, the vehicle density and velocity both tend to a specific value. We cannot see significant changes from the traditional temporal development of density or velocity. However, through our variable substitutions, the state variable η and σ both tend to infinity. As long as the traffic has a small fluctuation, the value of η or σ will change sharply. Moreover, as long as there is traffic jam formation, the value of η or σ will approach infinity. Using the new model by such variable substitution, we can describe clearly the relationship between traffic jams and system instability in the phase plane. The numerical solution of η and σ can be obtained by applying the finite difference method on the new model. Then we analyze the stop-and-go phenomena with four phase plane diagrams. The coordinate systems of them are $(\eta, \partial\eta/\partial t)$, $(\eta, \partial\eta/\partial x)$, $(\sigma, \partial\sigma/\partial t)$, and $(\sigma, \partial\sigma/\partial x)$, respectively. Through the four graphs the variation of density or velocity with time or sections can be investigated more clearly. Thus we can completely convert the fluctuations of traffic flow into the stability analysis charts.

Figure 1 shows the unstable traffic situation with small perturbations divergence when the initial density was set to 0.052 veh/m. Figure 1(a) is the temporal evolution of vehicle density. Since the value of initial density we set was in the unstable range, the amplitude of the initial small perturbations grows in time, leading to traffic instability. A complex localized structure consisting of two or more clusters forms. This situation corresponds to stop-and-go traffic.

A cross-sectional analysis was made along a time axis of Figure 1(a) to observe the variation of density with time when the section is fixed. Since the state variable η increases strictly monotonously with the density, we draw the change curves of η with time on each road section in Figure 1(b). Similarly,

since σ is inversely proportional to the vehicle velocity, we draw the change curves of σ with time on each road section in Figure 1(d). If the traffic approaches congestion, the density will approach the jam density and the velocity will be close to zero. The state variables η and σ will tend to infinity and the system will become unstable.

If the curves in Figures 1(b) and 1(d) are drawn one by one in the order of road section, we can find the curve of each downstream section moved toward the outer ring of upstream section. It shows that density fluctuations were gradually amplified toward the upstream section and the average velocity gradually reduced. The initial small perturbations spread upstream and were divergent, leading to the whole traffic system instability.

The phase plane diagrams emphasized the instable situation of the traffic system we are chiefly concerned about. It can be seen that the value of η and σ will change sharply in Figures 1(b) and 1(d) even when there is a very small fluctuation in Figure 1(a). When the vehicle density approached congestion, the value of η and σ will get larger. The corresponding curves of such traffic jams accounted for a large proportion in the graphs, while most small amplitude density fluctuations accounted for a quite small percentage and they are just centered in a small area near the initial value.

Then we made a cross-sectional analysis along a displacement axis of Figure 1(a) to observe the variation of density with displacement when time is fixed. Figures 1(c) and 1(e) are, respectively, the change curves of η and σ per second on the whole road. If these curves are drawn one by one in the order of time, we can see all the curves change with time from the inner circle to the outer ring and many of them tend to infinity. It also demonstrates that density fluctuations of the whole road are gradually amplified with time. The amplitude of the initial small perturbations grows in time. The whole traffic system is unstable. Compared with the temporal evolution of density, the phase plane diagrams can more clearly reflect the density variation of current time and the next time. So through the phase plane diagrams we can convert directly the traffic jam phenomenon into the curves of instability system. The result is more obvious when the traffic system is more instable.

Figure 2 is the change curves of density and η on the ninetieth road section which changes along with time. It can be seen from Figure 2(a) that there were many density fluctuations at this road section in the first ten minutes and this situation corresponds to stop-and-go traffic.

It reflects the density fluctuations outstandingly in Figure 2(b). We find out the starting point of the curve to observe its trajectory. It moves from the innermost point which is labeled as 6.6225 to the outermost circle and then turns toward the small circles inside. A circle of the phase plane diagram corresponds to a density fluctuation of the density curve chart. The variation of the circle ring in horizontal direction corresponds to the amplitude of a density fluctuation and in vertical direction it corresponds to the change rate of a density fluctuation. So the phase plane diagrams clearly reflect the size and speed of the density fluctuations. These results are consistent with the stop-and-go traffic phenomena described by the density curve

(a)

(b)

(c)

(d)

(e)
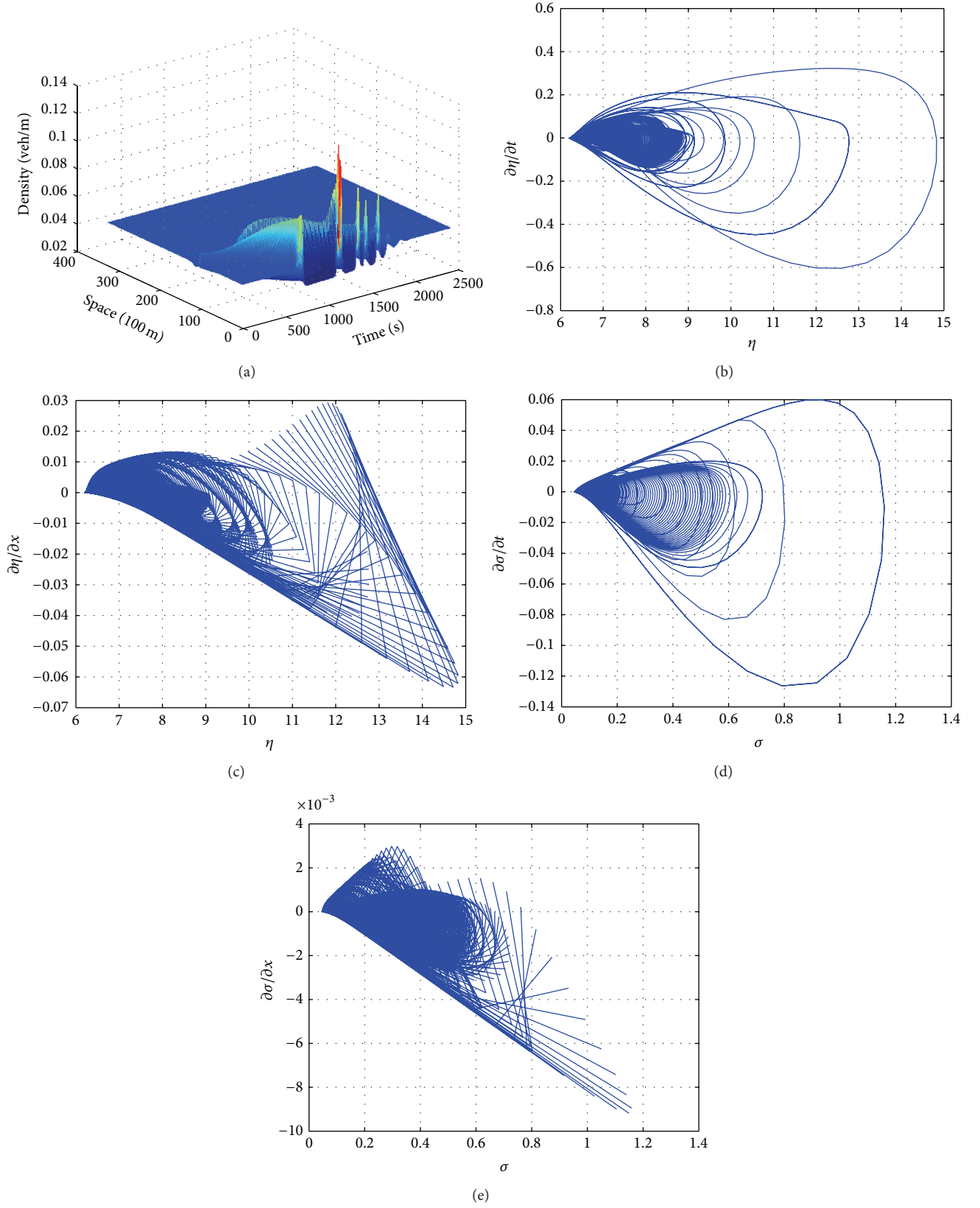
Figure 1: Evolution of small perturbation under the initial density 0.052 veh/m. (a) The temporal evolution of vehicle density. (b) The phase plane diagram of $(\eta, \partial\eta/\partial t)$. (c) The phase plane diagram of $(\eta, \partial\eta/\partial x)$. (d) The phase plane diagram of $(\sigma, \partial\sigma/\partial t)$. (e) The phase plane diagram of $(\sigma, \partial\sigma/\partial x)$.

(a)

(b)
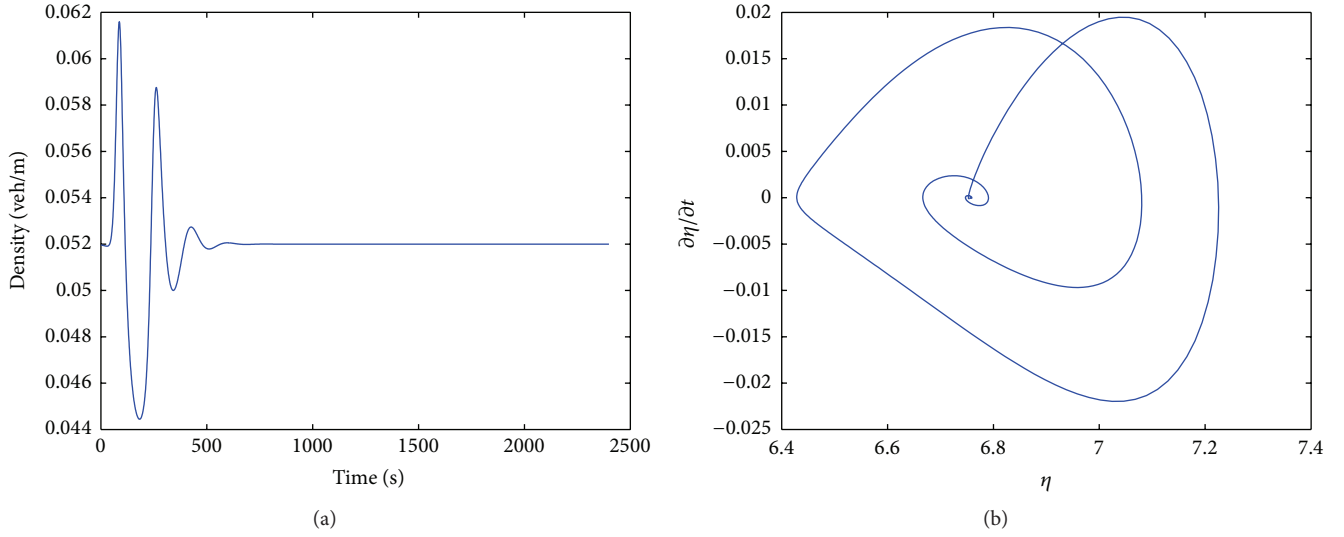
FIGURE 2: (a) The density-time variation curve on the ninetieth road section. (b) The phase plane diagram of $(\eta, \partial\eta/\partial t)$ on the ninetieth road section.

The 50th section
The 60th section
The 70th section

The 80th section
The 90th section

(a)

The 50th section
The 60th section
The 70th section

The 80th section
The 90th section

(b)
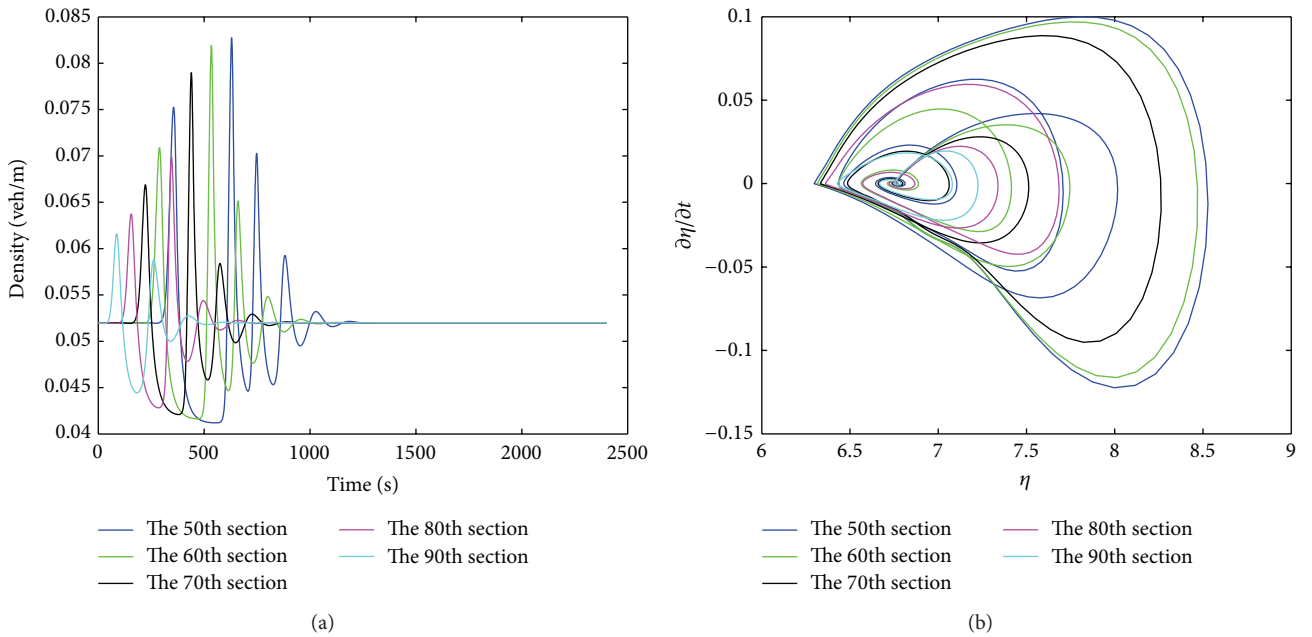
FIGURE 3: (a) Density-time variation curves on the ninetieth, the eightieth, the seventieth, the sixtieth, and the fiftieth road sections. (b) The phase plane diagram of $(\eta, \partial\eta/\partial t)$ on the ninetieth, the eightieth, the seventieth, the sixtieth, and the fiftieth road sections.

chart. Moreover, the phase plane diagrams mainly reflect the unstable traffic conditions we are chiefly concerned with.

Figure 3 is the change curves of density and η in the first 40 minutes on the ninetieth, the eightieth, the seventieth, the sixtieth, and the fiftieth road sections which change along with time. As you can see from Figure 3(a), the amplitude of the density fluctuation on the 90th section is small, as shown by the cyan line. It is growing toward the upper section of the road because the initial small perturbations spread upstream and are divergent. In Figure 3(b), the variation of η on the 90th section is small, as shown by the cyan

circle inside. It grows toward the upper sections and reaches the largest on the 50th section as shown by the blue circle outside. So it also shows the initial perturbation is amplified upstream. Although there are very small density fluctuations in Figure 3(a), the variation range of η is big in Figure 3(b). The phase plane diagram mainly reflects the density fluctuations.

Figure 4(a) is the density variation curves of the 1100th, the 1101st, the 1102nd, the 1103rd, the 1104th, the 1105th, and the 1106th seconds of the whole road section. The curves almost coincide together and we cannot see the

(a)

(b)
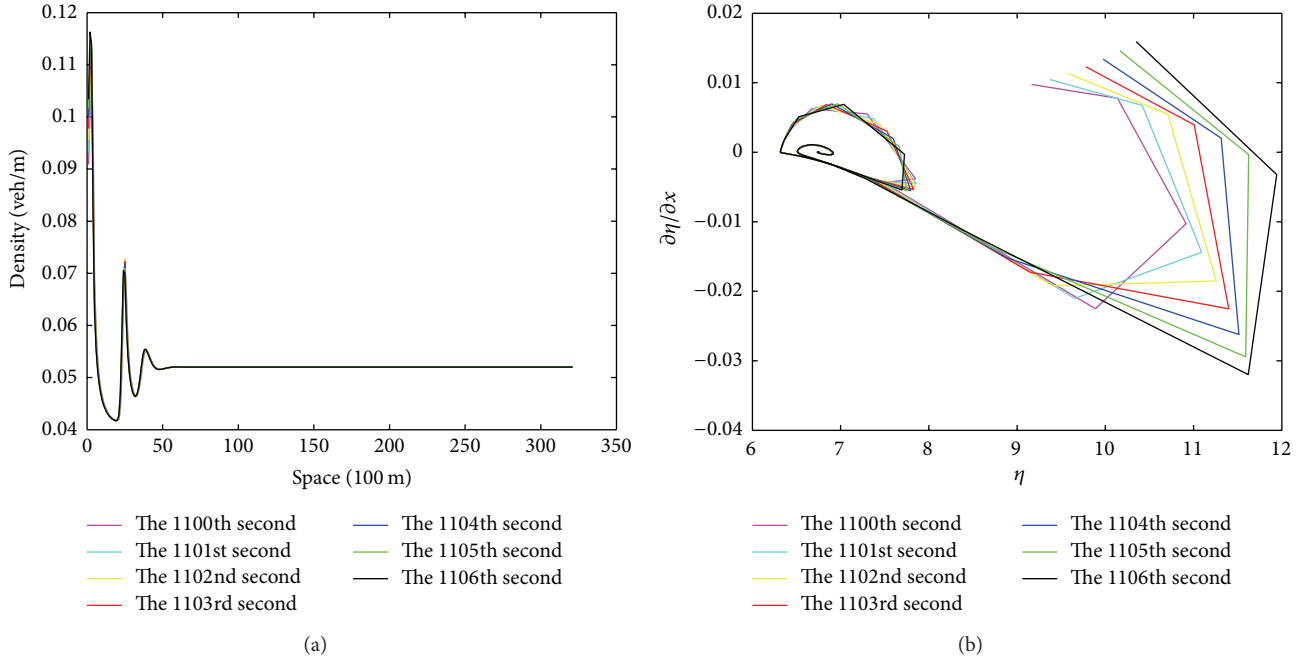
FIGURE 4: (a) Density-space variation curves on the 1100th, the 1101st, the 1102nd, the 1103rd, the 1104th, the 1105th, and the 1106th seconds of the whole road section. (b) The phase plane diagrams of $(\eta, \partial\eta/\partial x)$ on the 1100th, the 1101st, the 1102nd, the 1103rd, the 1104th, the 1105th, and the 1106th seconds of the whole road section.

obvious changes. Figure 4(b) is the phase plane diagrams corresponding to these times and it can clearly show that the curves enlarge per second from inside to outside; that means the amplitude of the density on the whole road grows in time and the traffic system is divergent.

Figure 5(a) is the velocity-space variation curves of the 679th, 680th, the 681st, the 682nd, the 683rd, the 684th, and the 685th seconds on the whole road section. All of the curves almost coincide together and we cannot see the obvious changes. However, the corresponding phase plane diagram in Figure 5(b) obviously shows that the inner circle of the current time turns to the outer ring of the next second and all of the curves tend to infinity. So it clearly reflects unstable traffic phenomena.

Comparing the phase plane diagrams with the temporal evolution of density shows that the stop-and-go traffic phenomena described by the new model are consistent with that described by the original model. But, unlike the previous temporal evolution of density, the phase plane diagram focuses on the density fluctuation we mainly care for and does not highlight the most homogeneous state of stable traffic flow. Moreover, it describes the variation of density or velocity with time or sections more clearly.

5. The Traffic Phenomena on a Highway with a Single Ramp

Although some empirical studies have been conducted to evaluate traffic data on highways with ramps, they cannot completely describe the various complex phenomena of fixed vehicle generation rate but increasing initial homogeneous density with a single ramp and the situation of morning peak. As it is very difficult to investigate the traffic phenomena induced by ramps, we use simulations by our model to describe the effects that ramps have on a main road. We take the test road section as 32.2 km long and set a ramp in the middle of the road section. We assume that the number of vehicles through an on-ramp is 36 veh more than that through an off-ramp every meter per hour. The length of the ramp is 100 m. So the fixed vehicle generation rate is 0.0001 veh/m/s. Other parameter values are the same as in Section 4. The results are shown in Figure 6.

It is clear from Figure 6(a) that since the vehicle generation rate is a small constant and the initial density of the main road is also lower than the down-critical unstable density, the vehicles coming from the on-ramp can drive quickly downstream and will not have any effect on the upstream traffic. That is to say, vehicles upstream can keep their speeds even when they drive past the ramp. So the density increments just appear from the on-ramp and reduce downstream gradually. The ramp has a small effect on the main road at this situation and this phenomenon will seldom appear because the main road density is relatively high.

In Figure 6(b), since the initial density is just above the down-critical unstable density, the ramp can disturb the stability of the main road traffic. A small quantity of vehicles coming from the ramp can be seen as a small localized perturbation on the initial homogeneous traffic flow. The amplitude of perturbation grows in time and eventually forms the stop-and-go traffic. So the ramp produces stop-and-go traffic when the main road density is between two critical values. The fluctuation amplitude of traffic flow is much larger

(a)

(b)
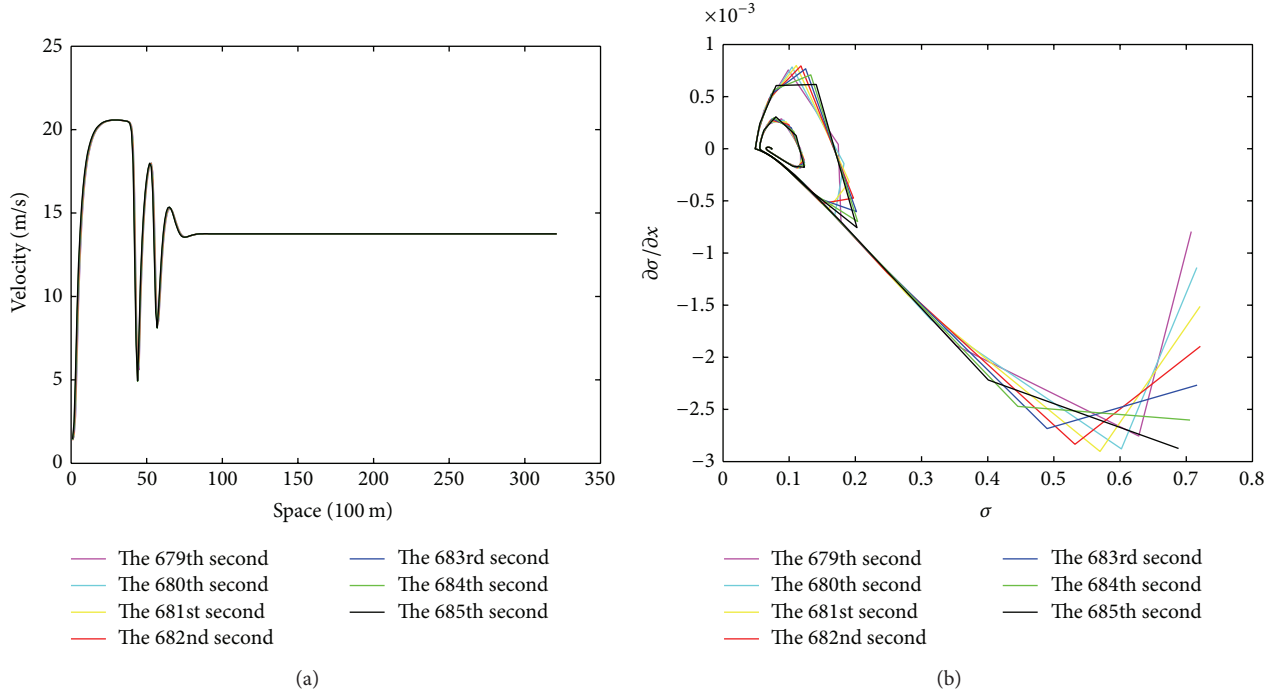
FIGURE 5: (a) Velocity-space variation curves on the 679th, the 680th, the 681st, the 682nd, the 683rd, the 684th, and the 685th seconds of the whole road section. (b) The phase plane diagrams of $(\sigma, \partial\sigma/\partial x)$ on the 679th, the 680th, the 681st, the 682nd, the 683rd, the 684th, and the 685th seconds of the whole road section.

than the vehicle generation rate on ramp. These phenomena often appear during the rush hours. In this case, the inflow of the ramp should be reduced to improve the main road traffic.

Figure 6(c) shows that when the initial density becomes greater than the up-critical density, a stable regime of the model is reached again and the perturbation is dissipated. Most of the cars which entered from the ramp accumulate on the ramp road section and the ramp becomes jammed very quickly. It shows the vehicles queue near the on-ramp when the main road density is relatively high. The ramp should be closed at this situation.

The phenomena described above are the situation of fixed vehicle generation rate but increasing initial homogeneous density. Next, we will analyze the phenomena of fixed initial homogeneous density but changing vehicle generation rate. The morning peak is a common phenomenon in traffic flow. In order to simulate it approximately, we set the value of vehicle generation rate on the ramp to change with a sine wave and the maximal amplitude of it is 36 veh/m/h. The number of vehicles which entered the ramp is increasing at the beginning of the peak hours in the morning and again is decreasing gradually after the peak hours. The initial uniform density is set to 0.027 veh/m and other parameter values are the same as above. The temporal evolution of vehicle density is shown in Figure 7.

It is clear from Figure 7 that the density is increased first and gradually decreased with the morning peak on the ramp section which we set in the middle of the road. The density increment of the ramp section spreads gradually downstream. As the initial density of the main road is very

low and the vehicle generation rate is also small, the vehicles coming from the on-ramp can drive quickly downstream.

6. The Traffic Phenomena on a Highway with Multiple Ramps

This section primarily analyzes traffic phenomena on a highway with multiple ramps by using the real traffic data. According to the new model and phase plane diagrams, several situations of morning traffic in Xi'an-Baoji Highway (China) were simulated. The total length in Xi'an to Baoji freeway is 187.404 kilometers and the station distances of all road sections are shown in Table 1.

There are ramps on each toll station according to the practical distribution. To simulate the morning traffic, we can assume the values of vehicle generation rate on every ramp all change with a sine wave and the maximal amplitude of them are all 0.001 veh/m/s. Firstly, we set the initial density of the whole road at 0.025 veh/m which is in the stable range of our model. For computational purpose, the space domain was divided into equal intervals of length of 200 m and the time interval was chosen as 1 s. Other parameter values used were as follows:

$$\beta = 6,$$
$$v_f = 125 \, \text{m/s},$$
$$\tau = 14 \, \text{s},$$
$$\rho_m = 0.25 \, \text{veh/m}.$$
$$(24)$$

(a) $\rho_0 = 0.027$ veh/m

(b) $\rho_0 = 0.047$ veh/m

(c) $\rho_0 = 0.127$ veh/m
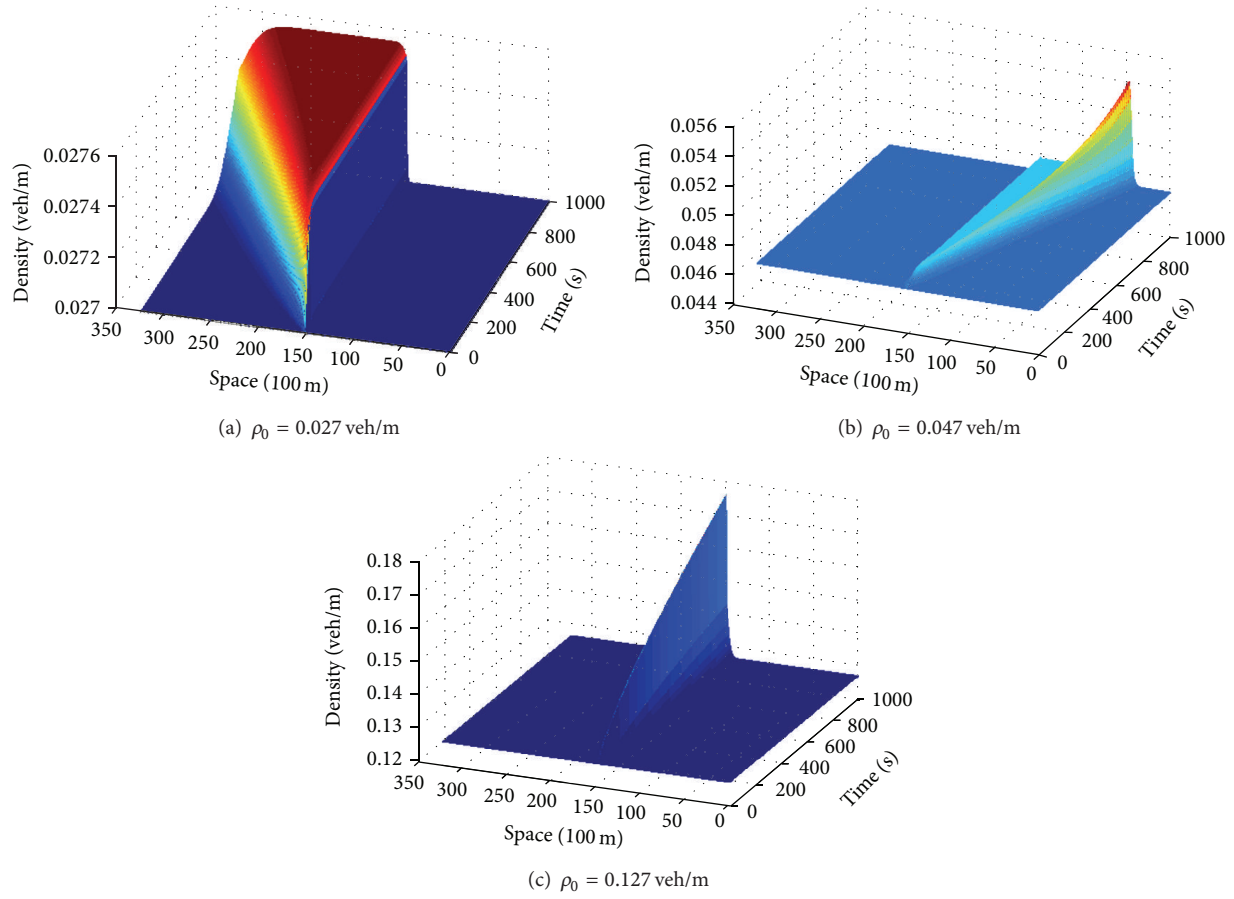
Figure 6: The density temporal evolution with initial homogeneous traffic and fixed vehicle generation rate.
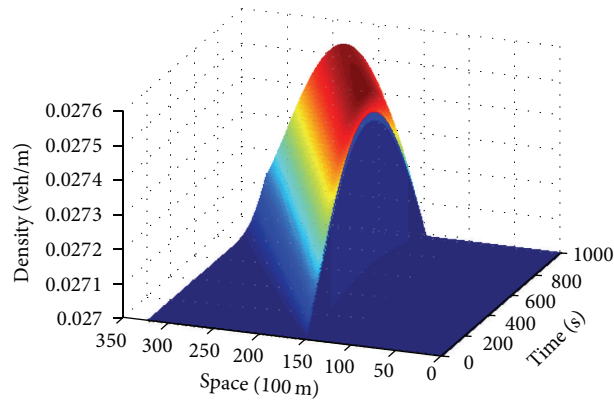
Figure 7: The density temporal evolution with initial homogeneous traffic of amplitude $\rho_0 = 0.027$ veh/m to simulate the morning peak.

Table 1: The general situation of Xi'an-Baoji highway.

Toll stations or control points	Sanqiao	Xianyang	Xian Yang Xi	Xingping	Wugong	Yangling	Jiang Zhang	Changxing	Meixian	Cai Jia Po	Guozhen	Baoji
The station distances (KM)	Starting point	9.698	6.092	13.594	27.419	11.83	13.08	10.515	11.725	9.855	21.855	17.815
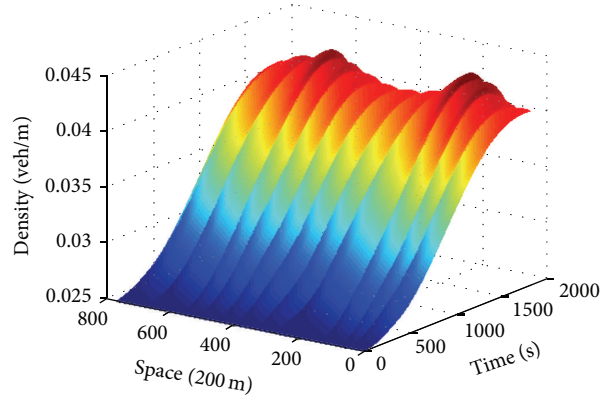
FIGURE 8: The density temporal evolution on the Xi'an-Baoji highway with initial homogeneous traffic of amplitude $\rho_0 = 0.025$ veh/m.

(a)

(b)
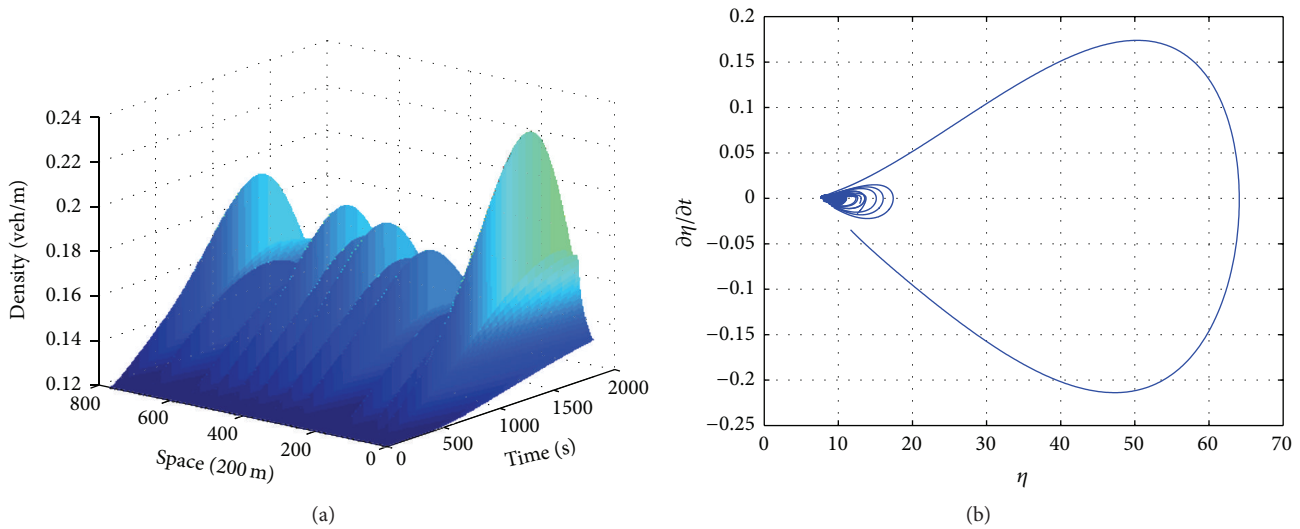
FIGURE 9: The density temporal evolution and phase plane diagrams on the Xi'an-Baoji highway with initial homogeneous traffic of amplitude $\rho_0 = 0.12$ veh/m. (a) The temporal evolution of vehicle density. (b) The phase plane diagram of $(\eta, \partial \eta / \partial t)$.

The meanings of these parameters are the same as in Section 2. The result illustrated by the temporal evolution of density was shown in Figure 8.

Since we assumed that the number of vehicles which entered the ramp was always greater than that exited the ramp at rush hours, the values of vehicle generation rate on every ramp were positive and this could eventually cause the density of the whole road to increase continuously. However, the densities were far less than the jam density and the traffic did not reach the congestion state because of the small initial homogeneous density.

To further simulate the morning traffic approaching the value of jam density on the Xi'an-Baoji highway, we increased the initial density of 0.12 veh/m and remained the value of vehicle generation rate; the temporal evolution of vehicle density and phase plane diagrams can be compared in Figure 9.

In Figure 9(a), since the initial density is increased and the vehicles continually enter from the ramp, the density of the whole road increases greatly and some road sections

approach the congestion eventually. Figure 9(b) is the combination of variation curves of η on each road section during the first 30 minutes. Since the number of the vehicles which entered the ramp increases first and then decreases before and after the morning peak, the trajectory of η also increases first and then decreases. So there are lots of irregular cycles in the figure corresponding to different density fluctuations. When the amplitude value of the density fluctuation is small, the cycle radius is small. However, the radius of cycle becomes very large when the density reaches the jam density. Although such cycles are not many, they account for very big proportion in the graph.

If we continually increased the initial density of 0.127 veh/m and other conditions were the same as above, the density of the whole road will increase continually with time and reach the jam density eventually. When the sum of the traffic demand of the upstream section and the traffic flow expected into highway is greater than the traffic capacity of the downstream section, ramp metering can be applied usually to restrict the number of the vehicles which entered

(a)

(b)
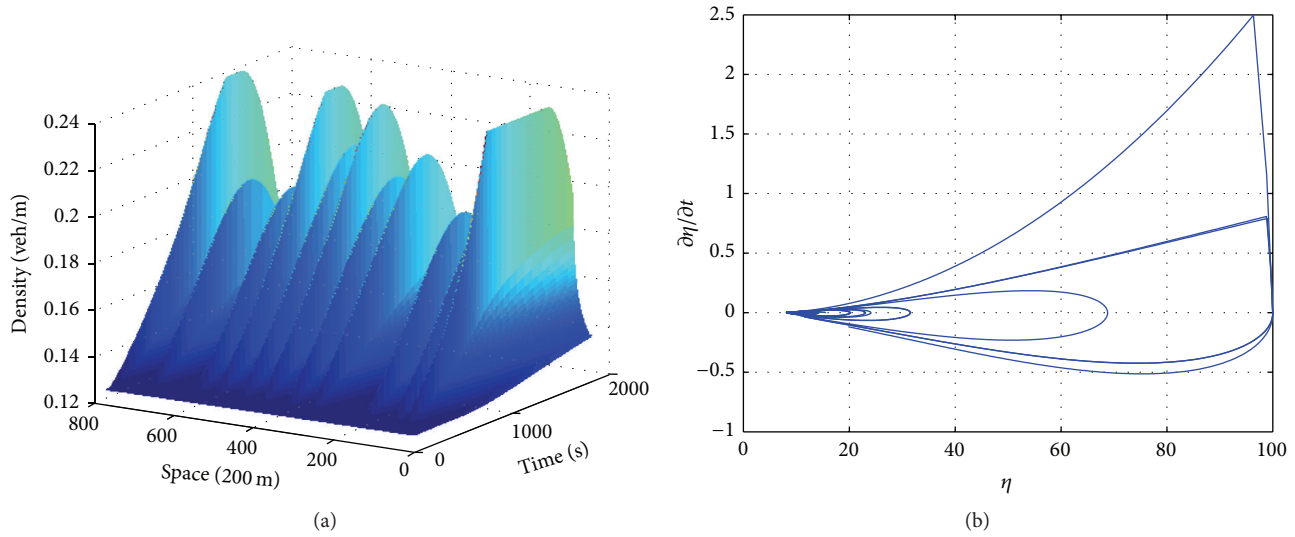
FIGURE 10: The density temporal evolution and phase plane diagrams on the Xi'an-Baoji highway with initial homogeneous traffic of amplitude $\rho_0 = 0.127$ veh/m. (a) The temporal evolution of vehicle density. (b) The phase plane diagram of $(\eta, \partial\eta/\partial t)$.

the ramp and excess vehicles will wait in the queue. So we let $\rho = \rho_m - 10^{-2}$ when $\rho \geq \rho_m - 10^{-2}$ and the results were shown in Figure 10.

It is clear from Figure 10(a) that there are three congested roads and the congestion spreads upstream gradually. However, the phase plane diagrams mainly described the instability of the system. When the vehicle density becomes saturated, the state variable η increases greatly. So we can see three curves changing greatly from Figure 10(b). But most small amplitude traffic flow fluctuations account for a small proportion in the graphs and they are just centered in a small area near the initial value.

7. Conclusions

The existing models cannot fully describe the phenomena raised by different input and output on ramps and there are few studies on the traffic phenomena using the phase plane diagrams. In this paper, we present an improved macro model for traffic flow on a highway with ramps based on the existing models. Furthermore, two new variable substitutions are adopted to extend the range of the variable from a specific value to infinity, so the model is transformed into a new model which is suitable for the stability analysis in phase plane. The problem of traffic flow could be converted into that of system stability. When the input flow of the on-ramp is equal to the output flow of the off-ramp, the equilibrium point equation of the model is calculated and the stop-and-go traffic is described in phase plane. The results show that the traffic phenomena described by the new model are consistent with those of real traffic, which shows that the new model is reasonable. When the stop-and-go waves appear and the traffic flow fluctuations tend to be unstable, the curves in phase plane diagrams are divergent and many of them tend to infinity. The phase plane analysis highlights the unstable traffic phenomena we are chiefly concerned about and describes the variation of density or velocity with time or sections more

clearly. When the input flow of the on-ramp is unequal to the output flow of the off-ramp, some traffic phenomena on a highway with ramps are found by our model. The numerical simulation shows that the model can reproduce some complex phenomena of fixed vehicle generation rate but increasing initial homogeneous density with a single ramp and the situation of morning peak. According to the actual road sections of Xi'an-Baoji highways, the traffic phenomena on a highway with multiple ramps are also analyzed. As all these numerical results are consistent with general traffic situations, the new model is reasonable. Moreover, the phase plane diagrams highlight the instability of the system. As long as the traffic becomes congested, the curves will be divergent and approach infinity in phase plane. However, the continuously changing vehicle generation rate and the varying initial density can pose more complex traffic phenomena. Our ongoing research will study various phenomena induced by the ramps, develop a ramp model that takes more factors into consideration, such as the road conditions, and examine the validation of the model using actual traffic data collected from the field.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] X. S. Hu, N. Murgovski, L. Johannesson, and B. Egardt, "Energy efficiency analysis of a series plug-in hybrid electric bus with

different energy management strategies and battery sizes," *Applied Energy*, vol. 111, pp. 1001–1009, 2013.

[2] X. S. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[3] X. S. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Optimal dimensioning and power management of a fuel cell/battery hybrid bus via convex programming," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 1, pp. 457–468, 2015.

[4] C. Sun, S. J. Moura, X. S. Hu, J. K. Hedrick, and F. C. Sun, "Dynamic traffic feedback data enabled energy management in plug-in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1075–1086, 2015.

[5] C. Sun, X. S. Hu, S. J. Moura, and F. C. Sun, "Velocity predictors for predictive energy management in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1197–1204, 2015.

[6] H. Y. Lee, H.-W. Lee, and D. Kim, "Origin of synchronized traffic flow on highways and its dynamic phase transitions," *Physical Review Letters*, vol. 81, no. 5, pp. 1130–1133, 1998.

[7] A. K. Gupta and V. K. Katiyar, "Phase transition of traffic states with on-ramp," *Physica A: Statistical Mechanics and Its Applications*, vol. 371, no. 2, pp. 674–682, 2006.

[8] D.-W. Huang, "Ramp-induced transitions in traffic dynamics," *Physical Review E*, vol. 73, no. 1, Article ID 016123, 2006.

[9] T.-Q. Tang, H.-J. Huang, S. C. Wong, Z.-Y. Gao, and Y. Zhang, "A new macro model for traffic flow on a highway with ramps and numerical tests," *Communications in Theoretical Physics*, vol. 51, no. 1, pp. 71–78, 2009.

[10] M. J. Lighthill and G. B. Whitham, "On kinematic waves: II. A theory of traffic flow on long crowed roads," *Proceedings of the Royal Society of London*, vol. 229, pp. 317–345, 1955.

[11] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, pp. 42–51, 1956.

[12] H. J. Payne, *Models of Freeway Traffic and Control*, vol. 1 of *Simulation Council Proceedings: Mathematical Models of Public Systems*, Simulation Councils, La Jolla, Calif, USA, 1971.

[13] W. F. Phillips, "A kinetic model for traffic flow with continuum implications," *Transportation Planning and Technology*, vol. 5, no. 3, pp. 131–138, 1979.

[14] P. Ross, "Traffic dynamics," *Transportation Research. Part B. Methodological*, vol. 22, no. 6, pp. 421–435, 1988.

[15] P. G. Michalopoulos, P. Yi, and A. S. Lyrintzis, "Continuum modelling of traffic dynamics for congested freeways," *Transportation Research Part B*, vol. 27, no. 4, pp. 315–332, 1993.

[16] H. M. Zhang, "Anisotropic property revisited—does it hold in multi-lane traffic?" *Transportation Research Part B: Methodological*, vol. 37, no. 6, pp. 561–577, 2003.

[17] A. K. Gupta and V. K. Katiyar, "Analyses of shock waves and jams in traffic flow," *Journal of Physics A. Mathematical and General*, vol. 38, no. 19, pp. 4069–4083, 2005.

[18] B. S. Kerner and P. Konhäuser, "Cluster effect in initially homogeneous traffic flow," *Physical Review E*, vol. 48, no. 4, pp. 2335–2338, 1993.

[19] R. Jiang, M.-B. Hu, B. Jia, R. Wang, and Q.-S. Wu, "Spatiotemporal congested traffic patterns in macroscopic version of the Kerner-Klenov speed adaptation model," *Physics Letters A*, vol. 365, no. 1-2, pp. 6–9, 2007.

[20] R. Jiang, Q.-S. Wu, and Z.-J. Zhu, "A new continuum model for traffic flow and numerical tests," *Transportation Research Part B: Methodological*, vol. 36, no. 5, pp. 405–419, 2002.

[21] L. Yu, T. Li, and Z.-K. Shi, "The effect of diffusion in a new viscous continuum traffic model," *Physics Letters, Section A: General, Atomic and Solid State Physics*, vol. 374, no. 23, pp. 2346–2355, 2010.

[22] H. X. Ge and X. L. Han, "Density viscous continuum traffic flow model," *Physica A*, vol. 371, no. 2, pp. 667–673, 2006.

[23] A. K. Gupta and S. Sharma, "Analysis of the wave properties of a new two-lane continuum model with the coupling effect," *Chinese Physics B*, vol. 21, no. 1, Article ID 015201, 2012.

[24] L. Yu and Z.-K. Shi, "Density wave in a new anisotropic continuum model for traffic flow," *International Journal of Modern Physics C*, vol. 20, no. 11, pp. 1849–1859, 2009.

Research Article

Temporal-Spatial Analysis of Traffic Congestion Based on Modified CTM

Chenglong Chu,[1] Na Xie,[2] Xiqun Chen,[3] Yuxin Wu,[4] and Xiaoxiao Sun[5]

[1]*School of Civil Engineering, Tsinghua University, Beijing 100084, China*
[2]*School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China*
[3]*College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China*
[4]*School of Accountancy, Central University of Finance and Economics, Beijing 100081, China*
[5]*School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100081, China*

Correspondence should be addressed to Na Xie; xiena@cufe.edu.cn

A modified cell transmission model (CTM) is proposed to depict the temporal-spatial evolution of traffic congestion on urban freeways. Specifically, drivers' adaptive behaviors and the corresponding influence on traffic flows are emphasized. Two piecewise linear regression models are proposed to describe the relationship of flow and density (occupancy). Several types of cellular connections are designed to depict urban rapid roads with on/off-ramps and junctions. Based on the data collected on freeway of Queen Elizabeth, Ontario, Canada, we show that the new model provides a relatively higher accuracy of temporal-spatial evolution of traffic congestions.

1. Introduction

It is well known that traffic flow dynamics estimation is important to transport engineering. To get an accurate estimation, most existing models are based on macroscopic continuum traffic flow modeling. The study of macroscopic continuum traffic flow model began with the Lighthill-Whitham-Richards (LWR) model independently proposed by Lighthill and Whitham [1] and Richards [2]. The LWR model formulated a hyperbolic partial differential equation (PDE) to describe the temporal-spatial evolution of traffic flow on a homogeneous highway.

Cell transmission model (CTM) is one of convergent numerical approximations to LWR model and was developed by adopting a simplified trapezoidal relationship between traffic flow and traffic density to solve the scalar kinematic wave model [3, 4]. It assumes a constant free flow speed, v_f, when density is lower than the critical traffic density and a constant backward shockwave speed, w, at a higher density.

Although the linear structure of CTM has the advantage of simplifying traffic control analysis, design, and data-estimation, the assumption of a simplified trapezoidal

fundamental diagram seems too strong for general applications. Some more complex models (e.g., Switching Mode Model (SMM)) derived from CTM were presented to reflect the temporal-spatial dynamic features of traffic congestion [5]. Alternatively, the investigation of oscillating congested traffic on freeways based on CTM, Lo et al. [6], leads to a dynamic traffic control formulation designed as dynamic intersection signal control optimization.

It was gradually found that the simple first-order continuum model does not have the ability to explain the amplification of small disturbances on heavy traffic, because no stability condition can be derived from the model [7]. In addition to the simple continuum model, Payne [8] introduced a higher-order model that is derived from car-following model to overcome the shortcomings of LWR model. Based on Payne's model, the deterministic and stochastic finite difference equations of validated high-order model were employed to describe freeway traffic flow dynamics, respectively [9–12]. Another higher-order continuum model that includes a dynamics equation developed from car-following theory in addition to the conservation equation was introduced
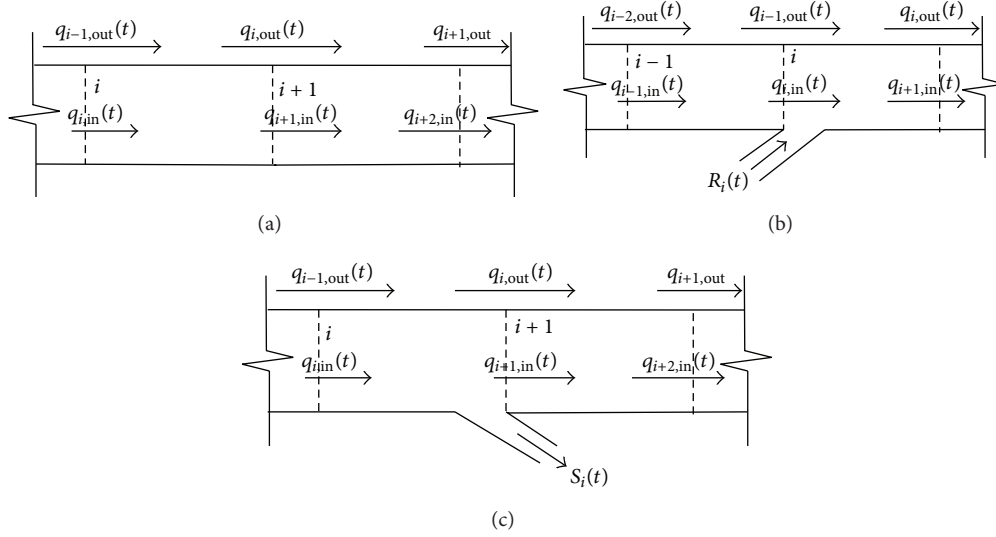
Figure 1: Three typical cell combinations of freeways: (a) with no ramps; (b) with one on-ramp; (c) with one off-ramp.

by Jiang et al. in [7]. This model overcomes the backward travel problem that exists in many other higher-order continuum models and can explain nonlinear dynamical spatial-temporal phenomena on freeways like shock waves, rarefaction waves, stop-and-go waves, and local cluster effects.

The uncertainty of traffic flow also attracts increasing concerns. Alecsandru [13] developed a stochastic compositional model of the evolution of traffic flows on freeways to extend the Daganzo cell transmission model by defining sending and receiving functions explicitly as random variables and by also specifying the dynamics of the average speed in each cell. Differently, an entropy solution was proposed for the LWR model with a flow density relationship of being piecewise quadratic, continuous, and concave, but not differentiable at the junction points where two quadratic polynomials meet [14]. Several alternative models were also developed in [15, 16].

In this paper, a modified cell transmission model (CTM) is proposed to depict the temporal-spatial evolution of traffic congestion on urban freeways. Specifically, drivers' adaptive behaviors and the corresponding influence on traffic flows are emphasized. Two piecewise linear regression models are used to describe the relationship of flow and density (occupancy). Several types of cellular connections are designed to depict urban rapid roads with on/off-ramps and junctions.

The rest of this paper is organized as follows. In Section 2, a new formation of CTM is presented. Section 3 discusses two critical issues when linear initialization of density and local perturbation analysis are conducted to reflect the dynamic and nonequilibrium properties of traffic flow. Section 4 addresses the two applications of spatial-temporal evolution analysis based on the proposed method with real detective data on unban freeways. Finally, main conclusions are summarized in Section 5.

2. An Improved Spatial-Temporal Finite Difference Model

The LWR model and the selected higher-order continuum model can be written in two PDEs as follows:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v}{\partial x} = g(x, t), \tag{1a}$$

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{v_e(\rho) - v}{T} - c(\rho, v)\frac{\partial p(\rho, v)}{\partial x}, \tag{1b}$$

where t and x denote time and distance from the origin, $\rho(x, t)$ is the traffic density, $v(x, t)$ is the space mean speed, $g(x, t)$ is net-merging traffic flow, T is the relaxation time, $v_e(\rho)$ is the steady-state speed-density relationship implied by car-following model, and $p(\rho, v)$ denotes the traffic pressure from upstream, and $c(\rho, v)$ means the coefficient of traffic pressure.

2.1. Finite Difference of the First-Order Equation. For the first-order model, CTM is an efficient finite difference solution scheme for realistic freeways and urban road networks. Figure 1 demonstrates three typical cell combinations of freeways, considering on/off-ramps.

For the basic segments without on/off-ramps demonstrated in Figure 1(a), the conservation equation can be derived from LWR model as follows:

$$\begin{aligned}
\rho_i(t+1) &= \rho_i(t) + \frac{T_s}{\lambda_i l_i}\left[q_{i-1,i}(t) - q_{i,i+1}(t)\right] \\
&= \rho_i(t) + \frac{T_s}{\lambda_i l_i}\left[q_{i,\text{in}}(t) - q_{i,\text{out}}(t)\right],
\end{aligned} \tag{2}$$

where i denotes the cell number, T_s is the time interval, l_i is the length of cell i, λ_i is the lane number of cell i, $\rho_i(t)$ means

the vehicle density of cell i at the time of t. And $q_{i,\text{in}}(t)$ and $q_{i,\text{out}}(t)$, respectively, denote the traffic flow entering and leaving cell i during the tth time interval, $T_s[t, t+1]$, including flows along the mainline and the on- and off-ramps.

In order to refine the transmission rules, it is assumed that two adjacent cells can be divided into one of five connection modes: (1) "free flow-free flow" (FF), (2) "congestion-congestion" (CC), (3) "congestion free flow" (CF), (4) "free flow congestion 1" (FC1), in which the wave front is moving downstream, and (5) "free flow congestion 2" (FC2), in which the wave front is moving downstream [5]. The SMM is composed of several linear models; straightforward linear techniques for model analysis and control design can be applied to the individual linear subsystems. The wave propagation direction is usually difficult to detect exactly. So,

four distinguishable states are formulated in this paper as follows:

FF:

$$q_{i-1,\text{out}}(t)$$
$$= \min\left\{q_{m,i-1}, \rho_{i-1}(t)\,v_{f,i-1}, \frac{\rho_{i-1}(t)\,\lambda_i l_{i-1}}{T_s}\right\}. \tag{3a}$$

CF:

$$q_{i-1,\text{out}}(t) = \min\left\{q_{m,i-1}, \frac{\rho_{i-1}(t)\,\lambda_i l_{i-1}}{T_s}\right\}. \tag{3b}$$

FC/CC:

$$q_{i-1,\text{out}}(t) = \max\left\{\min\left\{q_{m,i-1}, [\rho_{J,i}(t) - \rho_i(t)]\,w_i - R_i(t), \frac{\rho_{i-1}(t)\,\lambda_i l_{i-1}}{T_s}\right\}, 0\right\}. \tag{3c}$$

According to the conservation law at the boundary of cell $i-1$ and cell i, we get

$$q_{i,\text{in}}(t) = q_{i-1,\text{out}}(t) + R_i(t), \tag{4}$$

where $R_i(t)$ is the on-ramp flow and $\rho_{J,i}(t)$ is the jam density of cell i.

In this paper, we extend the above simple state switch transmission rule to depict more complex cloverleaf junctions. As shown in Figure 2, each ramp is considered as a cell and the other types of junctions can be formulated in the simplified expression. The traffic flow transmission matrix can be denoted as the structure of (5). Here, we mainly consider the properties of cell $i_k + 2$ ($k = 1, 2, 3, 4$) and ramp-cells by the following rules:

$$\left\{\begin{array}{cccccc}
\text{Cell} & i_k & i_k+1 & i_k+1 & i_k+1 & i_k+1 \\
i_k & 0 & q_{i_k,i_k+1} & & & \\
i_k+1 & & 0 & q_{i_k,i_k+1} & & \\
i_k+2 & & & 0 & q_{i_k,i_k+1} & \\
i_k+3 & & & & 0 & q_{i_k,i_k+1} \\
i_k+4 & & & & & 0
\end{array}\right\},$$
$$k \in \{1, 2, 3, 4\}, \tag{5}$$

$$\left\{\begin{array}{ccccc}
\text{Cell} & i_{k+1}+2 & i_{k-1}+4 & R_{ri_k} & R_{li_k} \\
i_k & 0 & 0 & q_{i_k, R_{ri_k}} & 0 \\
i_k+2 & 0 & 0 & 0 & q_{i_k+2, R_{li_k}} \\
R_{ri_k} & 0 & q_{R_{ri_k}, i_{k-1}+4} & 0 & 0 \\
R_{li_k} & q_{R_{li_k}, i_{k+1}+2} & 0 & 0 & 0
\end{array}\right\},$$
$$k \in \{1, 2, 3, 4\}, \; i_0 = i_4, \; i_5 = i_1,$$

where

$$q_{i_k+1,\text{in}}(t) = q_{i_k, i_k+1}(t),$$
$$q_{i_k+2,\text{in}}(t) = q_{i_k+1, i_k+2}(t) + q_{R_{li_{k-1}}, i_k+2}(t),$$
$$q_{i_k+3,\text{in}}(t) = q_{i_k+2, i_k+3}(t),$$
$$q_{i_k+4,\text{in}}(t) = q_{i_k+3, i_k+4}(t) + q_{R_{ri_{k+1}}, i_k+4}(t), \tag{6}$$
$$q_{R_{ri_k},\text{in}}(t) = q_{i_k, R_{ri_k}}(t),$$
$$q_{R_{li_k},\text{in}}(t) = q_{i_k+2, R_{li_k}}(t).$$

Here, R_{ri_k} and R_{li_k} represent the right/left turning ramp-cell, respectively.

2.2. Finite Difference of the Higher-Order Equation. In this paper, we applied the PDE model proposed in [7]. This PDE model replaces the density gradient term with the speed gradient term to consider drivers' anticipation as follows:

$$\frac{dv(x,t)}{dt} = \frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{v_e(\rho) - v}{T} + w\frac{\partial v}{\partial x}, \tag{7}$$

where w represents the propagation speed of the disturbance. In this model, the characteristic speed is equal to or smaller than the macroscopic flow velocity. In reality, traffic density $\rho(x,t)$ is hard to investigate or estimated directly (Figure 8); the most popular method is to monitor the time occupancy of vehicles passing detection stations. A number of literatures indicate that the time occupancy at x is proportional to traffic density around x.

Assuming there exist no merging or diverging in the section, say $g(x,t) = 0$, and applying the finite difference method to discretize (1a) and (7), we obtain the following difference equations based on point detectors like inductive

(a)

(b)

(c)
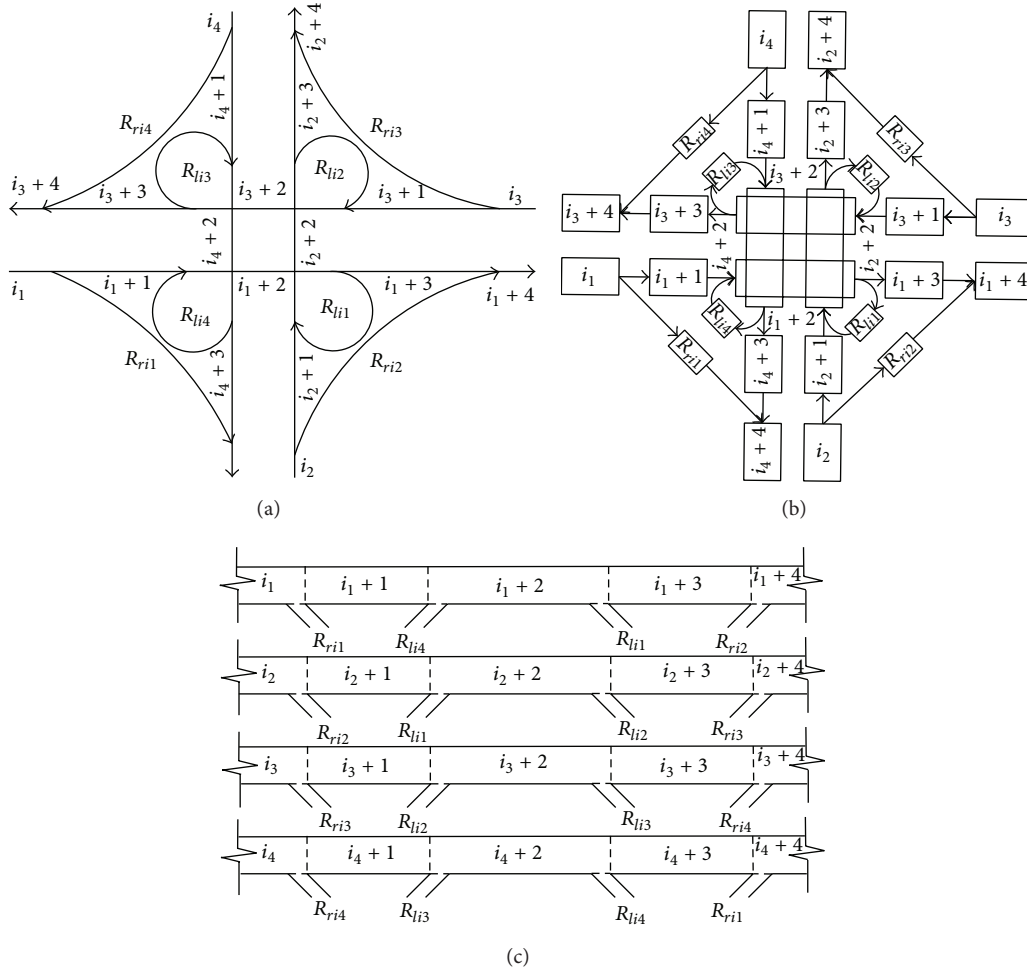
FIGURE 2: Cell representation of a cloverleaf junction: (a) the layout of junction; (b) the corresponding division of cells; (c) the corresponding CTM cells.

loops, microwave detectors, and radar sensors. The fundamental equation of traffic flow is

$$q_i(t) = \frac{\lambda_i o_i(t) v_i(t)}{g_i(t)} + \zeta_i^q(t), \qquad (8)$$

where $o_i(t)$ denotes the time occupancy at the upstream boundary of cell i, $g_i(t)$ represents the effective vehicle length which depends on both the actual vehicle length, which varies by lane and over the course of a day, and the loop's electrical circuit, and $\zeta_i^q(t)$ is the zero-mean white noise acting on the approximation to reflect the measurement inaccuracies and fluctuations in the traffic flow.

During one time interval, the drivers in cell i will adapt their speed to the local traffic density and are thus impacted by the traffic state in front of them. We suppose the adaptive behavior is influenced by the composition of traffic density of these two adjacent cells; it is a weighted average of the density in cell i and in cell $i + 1$:

$$\overline{o}_i(t) = f_o[\alpha_i(t), o_i(t), o_{i+1}(t)]. \qquad (9)$$

To simplify the combinational equations, we assume that this adaptation is linear as

$$\overline{o}_i(t) = \alpha_i o_i(t) + (1 - \alpha_i) o_{i+1}(t), \qquad (10)$$

where the parameter $0 < \alpha_i < 1$ weights how far ahead the drivers look and how well they anticipate the traffic density downstream. The drivers adapt the local traffic density more habitually; α_i is closer to 1.

Since each drivers' aggressiveness is random, the anticipant average space speed of cell i is relative to the speed of cell i and cell $i + 1$; the weighted factor $0 < \beta_i < 1$ expresses the relative impact on the calculation of the speed of convective and inertial behavior:

$$\overline{v}_i(t) = f_v[\beta_i(t), v_i(t), v_{i+1}(t)]. \qquad (11)$$

Similarly, we assume that

$$\overline{v}_i(t) = \beta_i v_i(t) + (1 - \beta_i) v_{i+1}(t). \qquad (12)$$

Because most of drivers tend to adapt their behavior whose function reflects an deterministic statistical relationship between the traffic flow and density or occupancy; the anticipant traffic flow $y_{i,\text{out}}(t)$ that passes through the boundary of cell i and cell $i + 1$ is approximately equal to $o_i(t)\overline{v}_i(t)$, when the stationary equivalent traffic flow is taken into consideration.

Meanwhile, the transmission flow is also constrained by the current sending capacity of cell i, that is, $o_i(t)l_i/T_s$, and by the receiving capacity of cell $i + 1$, that is, $[o_m - o_i(t)]l_{i+1}/T_s$, where o_m is the maximum occupancy; then it yields

$$
\begin{aligned}
&y_{i,\text{out}}(t) \\
&= \min \left\{ \frac{o_i(t)\,l_i}{T_s}, \frac{[o_m - o_{i+1}(t)]\,l_{i+1}}{T_s}, \gamma_i o_i(t)\,\overline{v}_i(t) \right. \\
&\left. + (1 - \gamma_i)\,Q_e\left[\overline{o}_i(t)\right] \right\} + \zeta_i^y(t),
\end{aligned} \quad \text{(A)}
$$

$$
v_i(t+1) = \begin{cases} v_i(t) + \dfrac{T_s}{l_i}\left[w_i - v_i(t)\right]\left[v_{i+1}(t) - v_i(t)\right] + \dfrac{T_s}{T}\left[V_e\left(o_i(t)\right) - v_i(t)\right] + \zeta_i^v(t), & \text{if } v_i(t) < w_i \\[2mm] v_i(t) + \dfrac{T_s}{l_i}\left[w_i - v_i(t)\right]\left[v_i(t) - v_{i-1}(t)\right] + \dfrac{T_s}{T}\left[V_e\left(o_i(t)\right) - v_i(t)\right] + \zeta_i^v(t), & \text{if } v_i(t) \geq w_i, \end{cases} \quad (14)
$$

where w_i represents the propagation speed of the disturbance that is the backward speed of traffic shockwave and $\zeta_i^v(t)$ is a zero-mean white noise reflecting the fluctuations in the drivers' speed.

3. Simulation Results

In the following experiments, we apply the initial stationary equivalent traffic speed and occupancy relationship function as (15) and the corresponding traffic flow and time occupancy relationship function as (16) given in [10]:

$$
V_e\left[\overline{o}_i(t)\right] = v_f \exp\left[-\frac{1}{\alpha}\left(\frac{\overline{o}_i(t)}{o_{c,i}(t)}\right)^{\alpha}\right], \quad (15)
$$

$$
\begin{aligned}
Q_e\left[\overline{o}_i(t)\right] &= \overline{o}_i(t)\,V_e\left[\overline{o}_i(t)\right] \\
&= v_f \overline{o}_i(t) \exp\left[-\frac{1}{\alpha_m}\left(\frac{\overline{o}_i(t)}{o_{c,i}(t)}\right)^{\alpha_m}\right].
\end{aligned} \quad (16)
$$

In the simulation of traffic density and speed evolution of a segment of urban freeways with periodical boundary conditions, the following constants that keep the same in all cells are applied. Geometric conditions and temporal-spatial discretization are

$$
L = 10\,\text{km},
$$

$$
\Delta L = 0.1\,\text{km},
$$

$$
T_s = 1\,\text{s},
$$

where $Q_e[\overline{o}_i(t)]$ is the stationary equivalent traffic flow and time occupancy relationship function, $Q_e[\overline{o}_i(t)] = Q_{e1}[\overline{o}_i(t)]$, if $\overline{o}_i(t) < o_{c,i}(t)$; $Q_e[\overline{o}_i(t)] = Q_{e1}[\overline{o}_i(t)]$, otherwise. $o_{c,i}(t)$ is the critical time occupancy of cell i, and the coefficient $0 < \gamma_i < 1$ expresses the relative impact on the calculation of the equivalent traffic flow of convective and macroscopic stable behavior and $\zeta_i^y(t)$ is a zero-mean white noise that describes the uncertainty of traffic transmission between two adjacent cells.

Thus, we can formulate the evolution of time occupancy under the assumption that $g_i(t)$ is approximately equal to $g_{i+1}(t)$ by the following finite difference equation:

$$
o_i(t+1) = o_i(t) + \frac{T_s}{\lambda_i l_i}\left[y_{i,\text{in}}(t) - y_{i,\text{out}}(t)\right], \quad (13)
$$

where $y_{i,\text{in}}(t)$ is equal to $y_{i,\text{out}}(t)$ when there exist no on/off-ramps. The formulas of conditions with merging or diverging areas can be obtained similarly as

$$
T = 10\,\text{s},
$$

$$
g(x,t) = 6.67\,\text{m}.
$$

$$
(17)
$$

Coefficients of fundamental diagram are

$$
v_f = 80\,\text{km/h},
$$

$$
w = 30\,\text{km/h},
$$

$$
o_{c,i}(t) = 20\%,
$$

$$
Q_m = 1800\,\text{veh/h/lane}.
$$

$$
(18)
$$

Parameters of drivers' adaptive behavior and standard deviations of Gaussian noise are

$$
\alpha_i(t) = 0.5,
$$

$$
\beta_i(t) = 0.5,
$$

$$
\gamma_i(t) = 0.5,
$$

$$
\sigma\left(\zeta_i^v(t)\right) = \sigma\left(\zeta_i^q(t)\right) = 0.
$$

$$
(19)
$$

Generally, the adaptive time can be chosen as 0 to 1. For numerical tests shown in this section, we just choose adaptive coefficients as 0.5 to show the influence of drivers' adaptive behavior.

3.1. Experiment 1: Linear Initialization of Density. In Experiment 1, we test the influence of linear initialization of density (Figure 5). During the whole process of simulation, lane
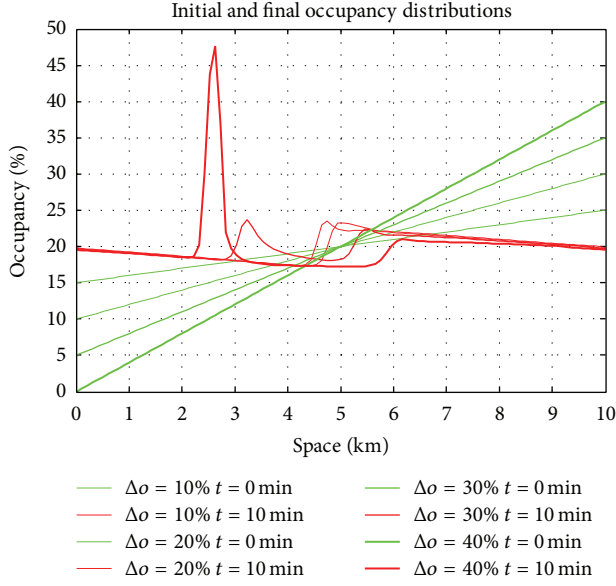
FIGURE 3: The initial and finial spatial distributions of occupancy for Experiment 1.

number and speed limit (equal to free-flow speed v_f) are not changed. Four typical experiments are conducted with the same critical occupancy of 20%; the boundary occupancy conditions are [15%, 25%], [10%, 30%], [5%, 35%], and [0%, 40%], respectively. Therefore, the initial differences between critical occupancy are 5%, 10%, 15%, and 20%.

The results show that a larger initial swing leads to a more obvious backward moving jam that is observed in Figure 3. Waves moving in the opposite directions meet at around 3.5 min and evolve into the backward moving jam with a larger density and a smaller fluctuation that propagates downstream especially obtained in the last experiment. Similarly, the evolution pattern of traffic speed shows the wave fluctuation phenomena of a moving jam in Figure 4.

When the initial density swing is small, the two reverse waves meet and the gap arises but no moving jam is achieved because of the lack of enough density difference between adjacent cells where the fluctuations cross. The scatter plot of fundamental diagram is illustrated in Figure 6; the widest line represents the equivalent functions of traffic speed/occupancy and traffic flow/occupancy with exponential expression showed in (16). Points of four different colors are four experimental results, respectively. It indicates that a linear initialization with a larger swing is more nonstable and the results scatter in a larger area in the fundamental diagram. The numerical results indicate that a larger initial difference of density generally triggers a larger fluctuation of traffic flow.

3.2. Experiment 2: Local Cluster Effect. The local cluster effect corresponds to the stop-and-go phenomena observed in the field due to a small disturbance [7]. In this subsection, we simulate the local cluster effect with respect to a localized perturbation in an initial homogeneous condition.

The following initial variation of the average occupancy o_0 is used as in [17]:

$$o(x, 0) = o_0 + \Delta o_0 \left\{ \cosh^{-2} \left[\frac{160}{L} \left(x - \frac{3L}{8} \right) \right] \right.$$
$$\left. - \frac{1}{4} \cosh^{-2} \left[\frac{40}{L} \left(x - \frac{13L}{32} \right) \right] \right\}, \tag{20}$$

where $v(x, 0) = V_e[o(x, 0)]$; the other parameters are the same as the simulation of linear initialization and the second term acts as a local perturbation to the unstable constant distribution; then the combinatorial difference methods are used to show the evolution pattern when $o(x, 0)$ increases.

Figure 7 shows the initial local perturbations with $\Delta o_0 = 5\%$ eventually evolve into local cluster effects when initial distribution of average time occupancy is limited in a specific extent and periodic boundary conditions are applied. The observed local cluster effects are also consistent with the diverse nonlinear dynamics phenomena observed in realistic traffic flow.

4. Model Validation and Application

The real data testing was conducted with traffic measurements collected from a 4.7 km, 3-lane stretch of Queen Elizabeth Way, Canada. As shown in Figure 9, it involves two on-ramps and an off-ramp at stations 50 and 53; another five loop detector stations were installed along this freeway and three at the on/off-ramps. The test stretch is subdivided into 8 segments, each with an approximate length of 650 m. Furthermore, a more refined cell division (where the size is only about 125 m) is conducted to explore the detailed spatial-temporal evolution of traffic jam.

The individual vehicle data were recorded by the detectors from 6 am to 10 am on December 15, 1998. All data were converted into aggregated traffic measurements of flow and space mean speed with a temporal-resolution-level 20 seconds. Only traffic flow and occupancy data are used for validation.

For the sake of simple and convenient computation, we use piecewise linear assumption of fundamental diagram in this section:

$$Q_e[o(x, t)] = v_f(x, t) o(x, t)$$
$$\text{if } o(x, t) < o_c(x, t); \tag{A-1}$$

$$Q_e[o(x, t)] = w(x, t) [o_J(x, t) - o(x, t)],$$
$$\text{otherwise.} \tag{A-2}$$

Figure 10 shows the fundamental diagram defined in this paper, where points A and B are the critical milestone of the formation and dissipation of traffic congestion; the corresponding occupancies are k_q and k_d where "q" means queuing and "d" means dissipating. Regressive results of the real traffic data show the occupancy gap is accepted that midpoint of line AB is the inflexion of traffic states. As a result, the assumption of critical occupancy is approximately to k_c. So, we determine that we meet a congested state if and only
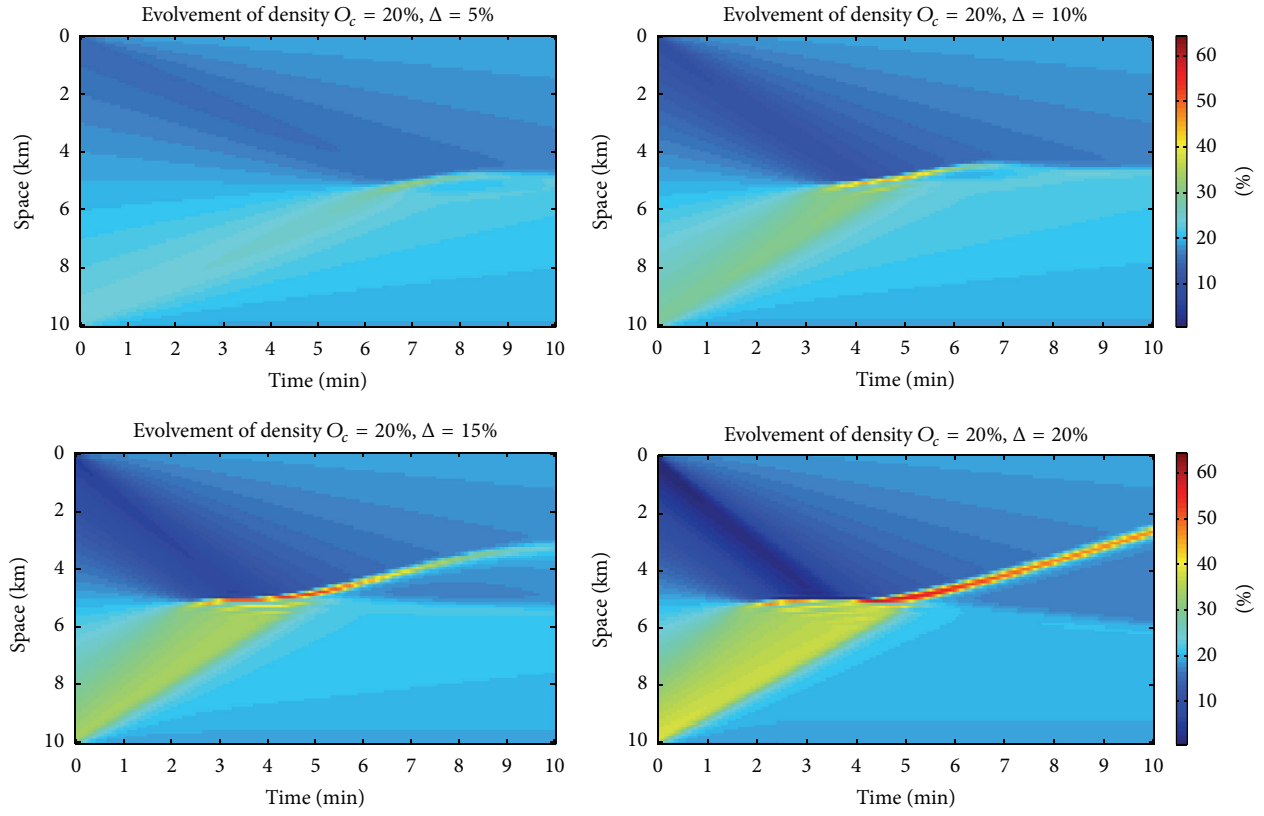
FIGURE 4: Occupancy evolution under linear initialization for Experiment 1.
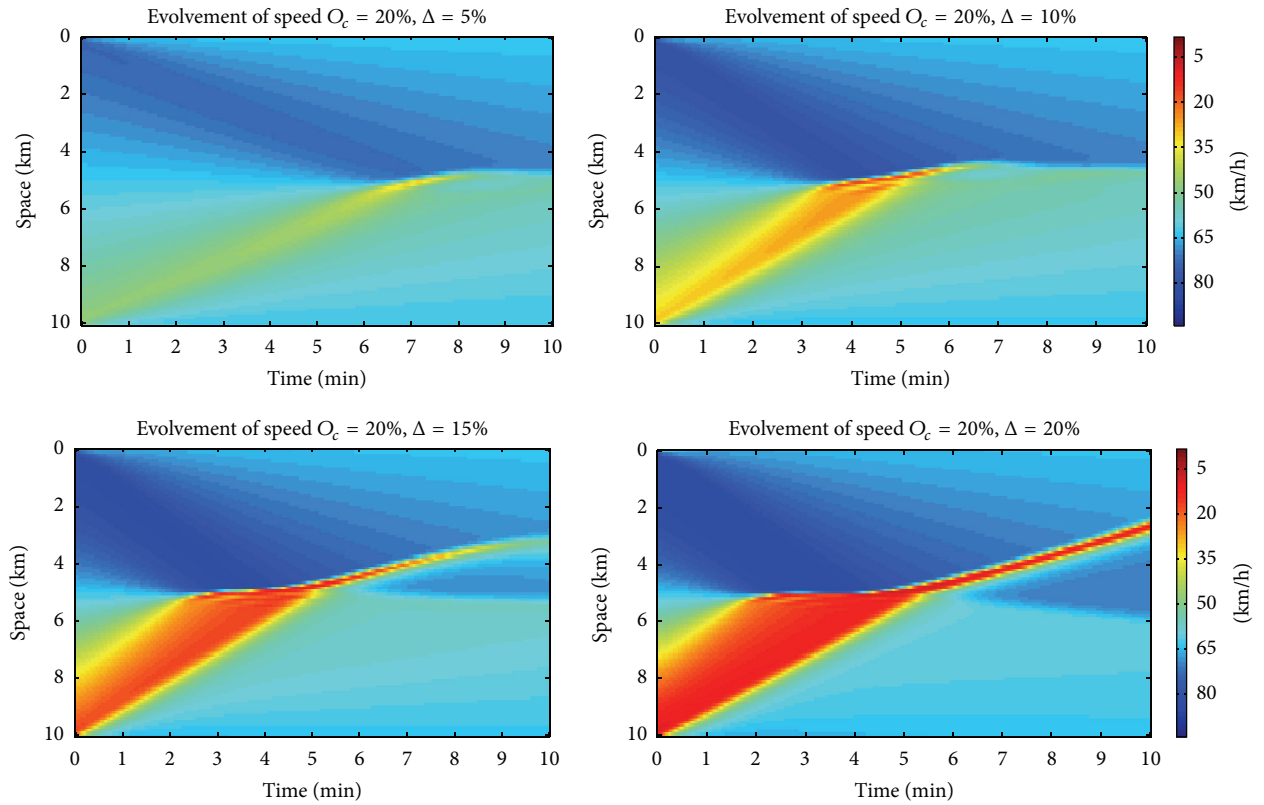
FIGURE 5: Speed evolution under linear initialization for Experiment 1.
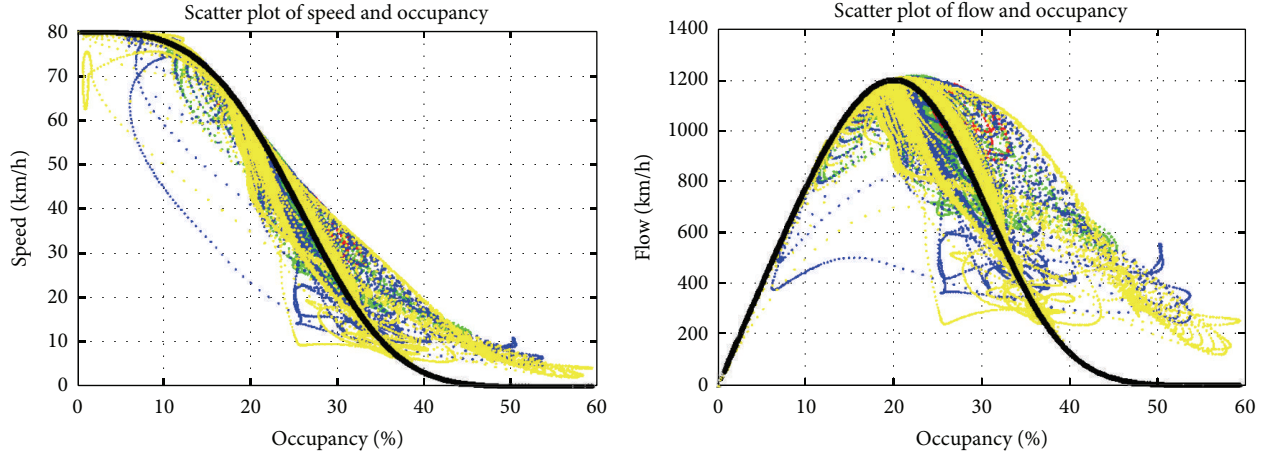
FIGURE 6: Fundamental diagram obtained for Experiment 1.
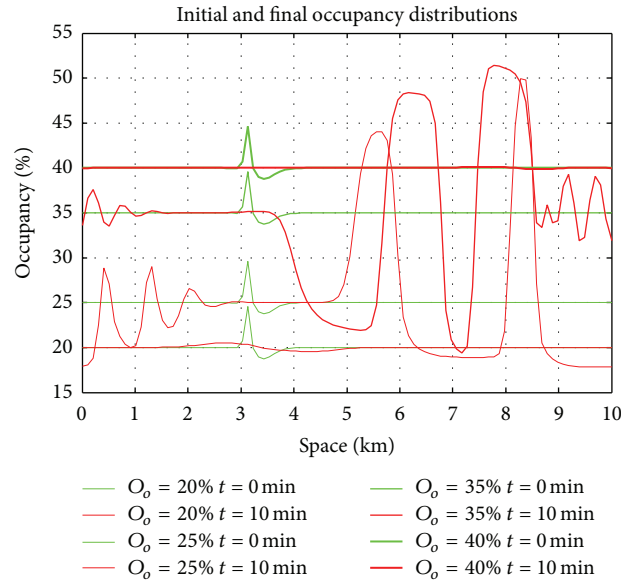
FIGURE 7: The initial and finial spatial distributions of occupancy for Experiment 2.

if the $o(x,t) > o_c$(or k_c) and $q(x,t)/o(x,t) < [q_q(x,t) - q_d(x,t)]/[o_q(x,t) - o_d(x,t)]$.

It is supposed that the traffic data collected by station 49~52 is unknown, and only date from stations 48 and 53 and the on/off-ramps is used to validate the finite different method discussed in this paper. The comparative results of measured and simulated occupancy data are shown in Figure 11. The agreement between two kinds of data suggests that this finite difference method is well applied to reflect the traffic congestion and the influences of on/off-ramp.

Tests also show that the simulated results will significantly deviate from the empirical measurements, if no drivers' anticipation term is introduced in the traffic flow model.

5. Conclusive Remarks

An improved spatial-temporal finite difference method based on LWR and a higher-order car-following model is proposed. A compositional model combining cell transmission model (CTM) and drivers' adaptive car-following behaviors is proposed for the spatial-temporal evolution of traffic flows on freeways.

Numerical tests verify that the model is able to simulate complex traffic phenomena observed in the field such as shock waves, stop-and-go waves, and local cluster effects. Empirical traffic data are also used to validate the finite difference method proposed in this paper. The results are
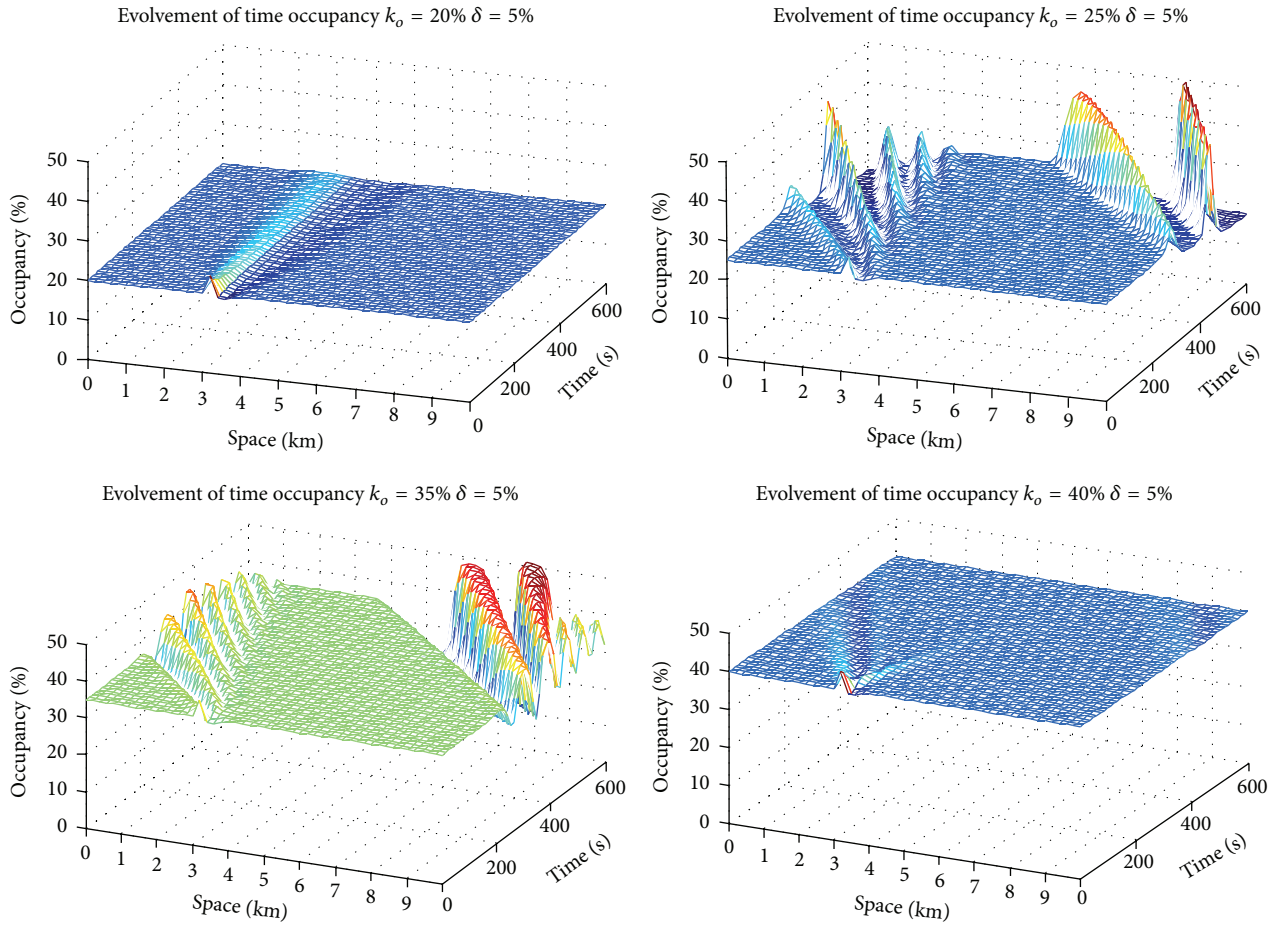
Figure 8: Traffic density evolution under small local disturbance for Experiment 2.
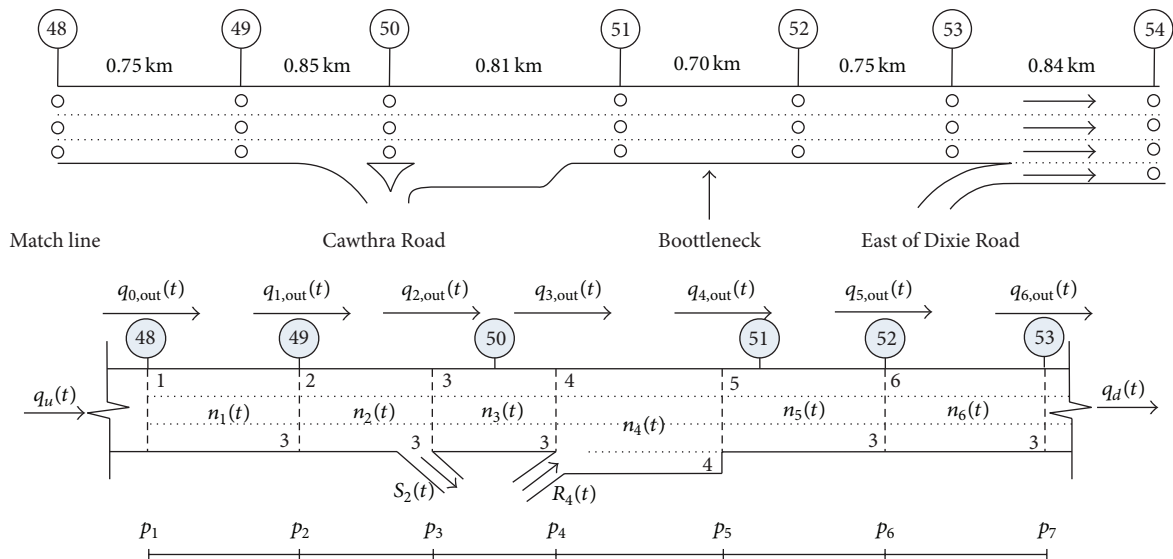
Figure 9: The layout and the corresponding cell representation of Queen Elizabeth Way, Canada.
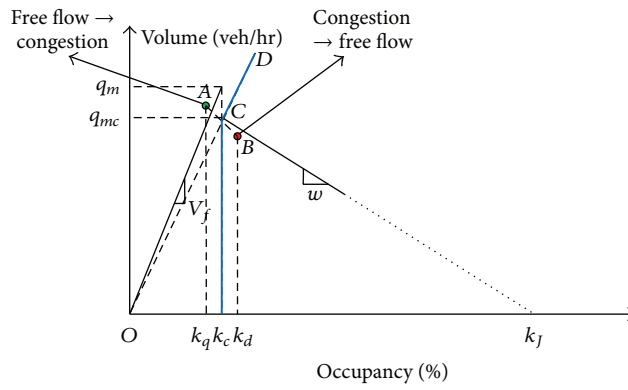
FIGURE 10: The fundamental diagram defined to depict asymmetric formation and dissipation of traffic congestion.
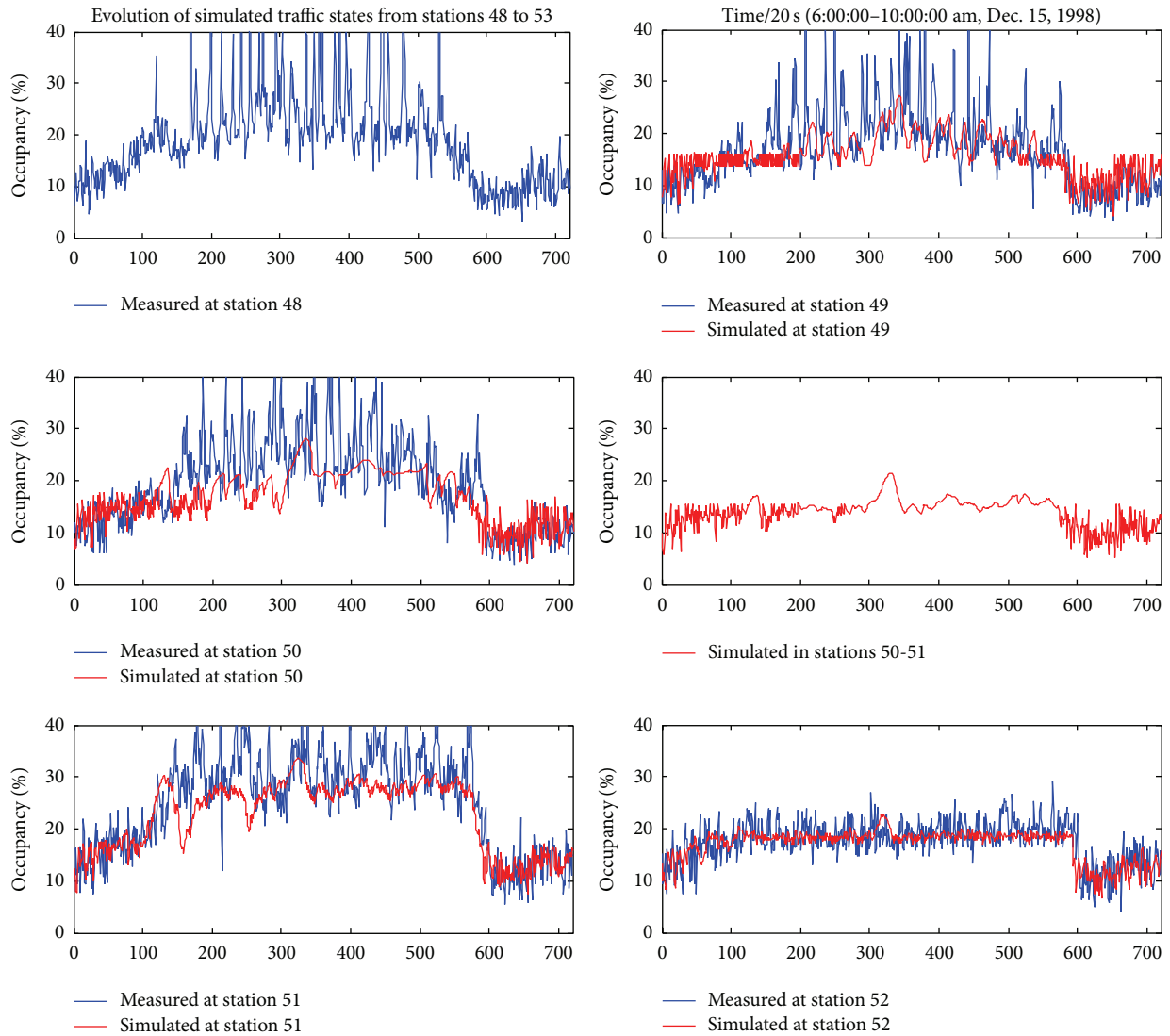
FIGURE 11: Spatial-temporal evolution of empirical and simulated occupancy.

consistent with the nonlinear dynamic properties measured in practice.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[2] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, pp. 42–51, 1956.

[3] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[4] C. F. Daganzo, "A finite difference approximation of the kinematic wave model of traffic flow," *Transportation Research Part B*, vol. 29, no. 4, pp. 261–276, 1995.

[5] R. Horowitz, X. T. Sun, L. Muñoz et al., "Design, field implementation and evaluation of adaptive ramp metering algorithms: final report," California PATH Research Report UCB-ITS-PRR-2006-21, 2006.

[6] H. K. Lo, E. Chang, and Y. C. Chan, "Dynamic network traffic control," *Transportation Research Part A: Policy and Practice*, vol. 35, no. 8, pp. 721–744, 2001.

[7] R. Jiang, Q.-S. Wu, and Z.-J. Zhu, "A new continuum model for traffic flow and numerical tests," *Transportation Research Part B: Methodological*, vol. 36, no. 5, pp. 405–419, 2002.

[8] H. J. Payne, "Models of freeway traffic and control," in *Mathematical Models of Public Systems*, G. A. Bekey, Ed., vol. 1 of *Simulation Council Proceedings Series*, pp. 51–61, 1971.

[9] A. S. Willsky, E. Y. Chow, S. B. Gershwin, C. S. Greene, A. L. Kurkjian, and P. K. Houpt, "Dynamic model-based techniques for the detection of incidents on freeways," *IEEE Transactions on Automatic Control*, vol. 25, no. 3, pp. 347–360, 1980.

[10] M. Papageorgiou, J.-M. Blosseville, and H. Hadj-Salem, "Modelling and real-time control of traffic flow on the southern part of Boulevard Périphérique in Paris: Part I: modelling," *Transportation Research Part A: General*, vol. 24, no. 5, pp. 345–359, 1990.

[11] A. Kotsialos, M. Papageorgiou, C. Diakaki, Y. Pavlis, and F. Middelham, "Traffic flow modeling of large-scale motorway networks using the macroscopic modeling tool METANET," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 4, pp. 282–292, 2002.

[12] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: adaptive capabilities and real data testing," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, pp. 1340–1358, 2008.

[13] C. D. Alecsandru, *A stochastic mesoscopic cell-transmission model for operational analysis of large-scale transportation networks [Ph.D. thesis]*, Louisiana State University, Baton Rouge, La, USA, 2006.

[14] Y. D. Lu, S. C. Wong, M. P. Zhang, C.-W. Shu, and W. Q. Chen, "Explicit construction of entropy solutions for the Lighthill-Whitham-Richards traffic flow model with a piecewise quadratic flow–density relationship," *Transportation Research Part B: Methodological*, vol. 42, no. 4, pp. 355–372, 2008.

[15] T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-Payoong, "Short-term traffic state prediction based on temporal-spatial correlation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1242–1254, 2013.

[16] R. X. Zhong, A. Sumalee, T. L. Pan, and W. H. K. Lam, "Stochastic cell transmission model for traffic network with demand and supply uncertainties," *Transportmetrica A: Transport Science*, vol. 9, no. 7, pp. 567–602, 2013.

[17] M. Herrmann and B. S. Kerner, "Local cluster effect in different traffic flow models," *Physica A: Statistical Mechanics and Its Applications*, vol. 255, no. 1-2, pp. 163–188, 1998.

Research Article

Lithium-Ion Battery Cell-Balancing Algorithm for Battery Management System Based on Real-Time Outlier Detection

Changhao Piao,[1,2] Zhaoguang Wang,[1] Ju Cao,[1] Wei Zhang,[2] and Sheng Lu[1]

[1]*Institute of Pattern Recognition and Applications, Chong Qing University of Posts and Telecommunications, Chongqing 400065, China*
[2]*Mechanical Engineering, INHA University, Incheon 400072, Republic of Korea*

Correspondence should be addressed to Sheng Lu; lusheng@cqupt.edu.cn

A novel cell-balancing algorithm which was used for cell balancing of battery management system (BMS) was proposed in this paper. Cell balancing algorithm is a key technology for lithium-ion battery pack in the electric vehicle field. The distance-based outlier detection algorithm adopted two characteristic parameters (voltage and state of charge) to calculate each cell's abnormal value and then identified the unbalanced cells. The abnormal and normal type of battery cells were acquired by online clustering strategy and bleeding circuits ($R = 33$ ohm) were used to balance the abnormal cells. The simulation results showed that with the proposed balancing algorithm, the usable capacity of the battery pack increased by 0.614 Ah (9.5%) compared to that without balancing.

1. Introduction

Electric vehicles (EV) are widely viewed as an important transitional technology for energy-saving and environmentally sustainable transportation [1]. As the new traction battery packs, critical energy sources of EV, lithium-ion (Li-ion) battery pack is drawing a vast amount of attention for its excellent advantages such as compact volume, large capacity, lower weight, and higher safety [2–4]. Single battery cells are serially connected to a battery stack to achieve higher capacity and voltage. However, the charging process has to stop as soon as one cell is completely charged and the discharging process has to stop as soon as one cell is completely discharged [5]. The capacity of the whole battery pack is thus limited by the unbalanced cells required to be balanced (also called abnormal cells in this paper) in the pack which can reduce the usable capacity of the battery pack, decrease the energy usage efficiencies, and shorten the lifetime of battery pack. Therefore, battery cell balancing that is one basic function of BMS is necessary for battery pack in EV [6–9].

Two algorithms are commonly used for cell balancing: voltage-based balancing algorithm and state of charge-based balancing algorithm. The voltage-based balancing is that when the difference between one cell voltage and the mean value of cell voltages is larger than the threshold V_{th}, the cell is probably considered to be an abnormal cell [10–12]. This method is simple and easy operating while the external voltage of the cell is affected by its internal state and environment. On the other hand, some researchers pointed that state of charge (SOC) can reflect the capacity of the battery pack in essence and proposed the SOC-based balancing algorithm which controls the range of the SOC smaller than the threshold SOC_{th} [13, 14]. However, SOC that is affected by battery model, self-discharge, temperature, and other factors can only be calculated by voltage or current indirectly and it is still difficult to get the accurate SOC of each cell. Unfortunately, there are still no observations at present about applying outlier detection algorithm to cell balancing.

Outlier detection algorithm which is an important branch of data mining is applied in many different domains [15–17]. This paper innovatively proposed to use the algorithm to identify the abnormal cells. The algorithm chooses characteristic parameters of battery cells and develops a flexible distance function to get outliers (viz. the unbalanced cells)

effectively [18, 19]. After getting the accurate category of normal and abnormal cells through clustering method, the abnormal cells are balanced by passive balancing circuit. Outlier detection algorithm can recognize the abnormal battery cell accurately and improve the performance of battery pack, such as increasing the usable energy and extending lifetime.

The research work is organized as follows. Section 2 describes the detailed processes of the proposed balancing algorithm. The simulation models and test cycle are described in Section 3. In the final section, conclusions and final remarks are given.

2. Balancing Algorithm Based on Outlier Detection

2.1. Outlier Detection Algorithm. As shown in Figure 1, the outlier detection balancing algorithm includes two modules: the unbalanced cells recognition module and the balancing control module. The former module gets normal and abnormal cells by outlier detection algorithm while the latter balances the abnormal cells and gives feedback to the former. There are N battery cells in the power battery pack. And the characteristic parameters of the cells are provided to the unbalanced cell recognition module [20].

The Li-ion battery's input-current I and output-SOC, current I, and terminal voltage U are shown in Figure 2 [21]. The characteristic parameters of voltage and SOC are used to calculate each cell's outlier value.

2.2. Unbalanced Cell Recognition. Abnormal cells are picked up as outlier point by the outlier detection method in unbalanced cells recognition module. First, z-score standardized method is used to preprocess the attribute of battery. Second, the outlier detection method based on the distance of multidimensional attribute is adopted to calculate each cell's outlier value which is the summation of distances from one cell to the others. Third, the battery pack will be balanced if the abnormality range is not less than the threshold signed as VOA1, otherwise the unbalanced cells will be obtained by the dynamic cluster method. Finally, passive equalization is applied to the abnormal battery cells. Figure 3 shows the progress of recognizing abnormal cells.

Input. The number of battery cells N and the initial threshold of abnormality range VOA1 are input.

Step 1. If the attribute values of the battery cells are equal, the process ends and the pack is considered balanced; otherwise it goes to Step 2.

Step 2. Z-score standardized method is adopted to preprocess the characteristic parameters for eliminating the influence of units. Namely, use the formula (1) to preprocess the Voltage U and the SOC as follows:

$$Z_{ij} = \frac{\text{Cell}_{ij} - \overline{\text{Cell}}_j}{\delta_j},$$
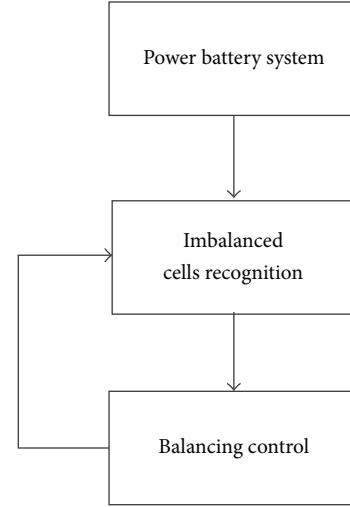
FIGURE 1: The design of the outlier detection balancing algorithm.
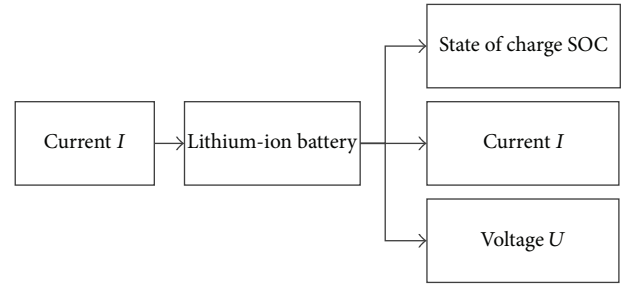
FIGURE 2: Input and output parameters of lithium-ion battery model.

$$\overline{\text{Cell}}_j = \frac{\sum_{i=1}^{n} \text{Cell}_{ij}}{n},$$

$$\delta_j = \sqrt{\frac{\sum_{i=1}^{n} \left(\text{Cell}_{ij} - \overline{\text{Cell}}_j\right)}{n-1}},$$

(1)

where Z_{ij} ($i = 1, \ldots, 40$ and $j = 1, 2$) is the standardization form of jth characteristic parameter of the ith cell (i.e., Z_{12} denotes the standard SOC of the first cell); Cell_{ij} represents the original value of jth characteristic parameter of the ith cell (i.e., Cell_{21} denotes the original voltage of the second cell); $\overline{\text{Cell}}_j$ describes the mean of jth parameters; δ_j denotes the standard deviation of the voltage or SOC when j equals to 1 or 2, respectively; n is the number of cells.

Then, Euclid-distance is used to calculate the abnormal value of each cell in the pack. The calculation formula [22] is defined as follows:

$$D_2\left(Z_m, Z_n\right) = \sqrt{\left|Z_{m1} - Z_{n1}\right|^2 + \left|Z_{m2} - Z_{n2}\right|^2},$$

$$W\left(Z_m\right) = \sum_{m=1}^{n} \sum_{j=1}^{2} D_2\left(Z_m, Z_n\right),$$

(2)
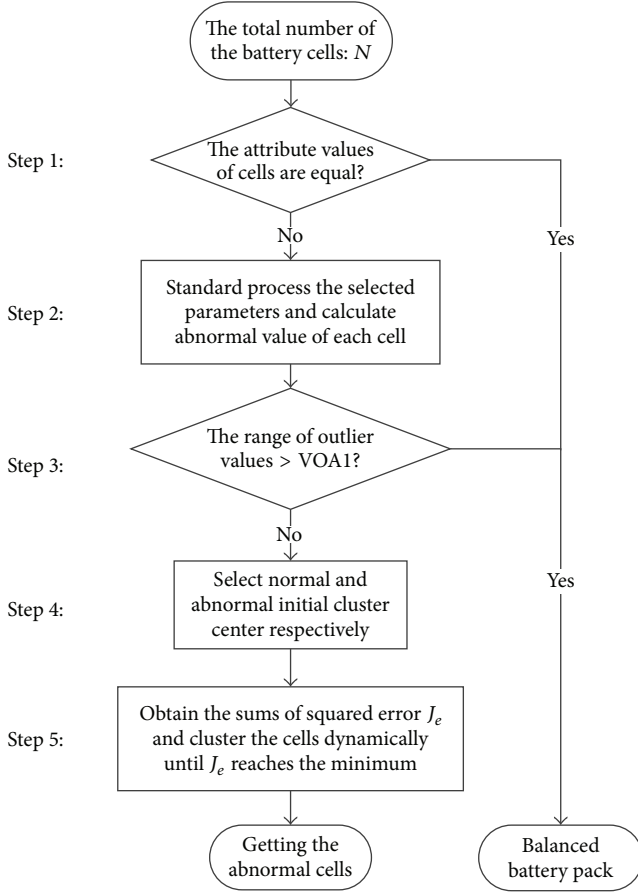
FIGURE 3: The process of acquiring abnormal battery cells.

where $D_2(Z_m, Z_n)$ ($m = 1, \ldots, 40$, $n = 1, \ldots, 40$, $m \neq n$) represents the Euclid-distance between the mth cell and the nth cell; Z_m represents the mth cell which has two attributes (i.e., $Z_m = (Z_{m1}, Z_{m2})$); $W(Z_m)$ denotes the summation of Euclid-distance betweenthe mth cell and the others. The smaller the $W(Z_m)$ is, the more normal the mth cell is. On the contrary, the mth battery is probably abnormal.

Step 3. If the range of outlier values (the difference outlier value between cell with the lowest and that with the highest outlier value) of the cells is smaller than the threshold VOA1, the process ends and the pack is considered balanced; otherwise it goes to Step 4. VOA1 is defined and updated by the formulation as follows:

$$\text{VOA1} = \frac{\sum_{m=1}^{n} W(Z_m)}{n}. \tag{3}$$

Step 4. Set the cell with the lowest and that with the highest outlier value as initial clustering centroids.

Step 5. The other cells are assigned to their nearest cluster centroid, all at once, followed by recalculation of cluster centroid. Then the other cells are individually reassigned if doing so will reduce the sums of squared error, and cluster centroids are recomputed after each reassignment

[23]. The process of obtaining the minimum sums of squared error J_e is formulated as follows:

$$J_m = \sum_{m=1}^{N_m} \left\| Z_m - C_j \right\|^2, \quad Z_m \in S_j,$$

$$J_e = \sum_{j=1}^{k} J_m = \sum_{j=1}^{k} \sum_{m=1}^{N_m} \left\| Z_m - C_j \right\|^2, \quad Z_m \in S_j, \tag{4}$$

where J_m denotes the squared error of mth cell; J_e is the sums of all the squared errors; C_j is initial cluster centroid; S_j is the normal category or the abnormal category.

When J_e converges to a global minimum, the process jumps to the next step.

Output. The battery pack is balanced or unbalanced.

By this time, the unbalanced cells are recognized by the outlier detection algorithm and can be balanced with the passive balancing circuit that will be described in detail in next section.

2.3. Balancing Control. At present, the balancing circuit can be divided into two main groups [24]: passive balancing circuit and active cell balancing circuit. Typical passive cell balancing circuit also named shunt method uses switches to control balancing. Specifically, shunt method is designed to use a resistor to discharge the unbalanced cell detected by outlier detection algorithm. With active cell balancing circuit, charge can be transferred between the cells in battery pack by a capacitor or an inductor. Very little energy would be wasted in this case compared to the passive balancing method. However, more switches and associated components are needed in the active balancing circuit. And these additional components may lead to higher cost and unreliability. Passive balancing circuit has already been used in many applications for its simple structure and reliability. Hence, passive cell balancing circuit is applied in this paper. As shown in Figure 4, every battery has a balancing circuit which comprises a resistor and a switch in series.

3. Simulation Experiments

In order to compare the efficiency of the outlier detection balancing algorithm with the traditional balancing algorithms in detail, the constant-current charging-discharging (CCCD) model and the software-in-the-loop platform (SILP) model for the BMS were established in Sections 3.1 and 3.2, respectively. The simulations were conducted on an Intel 2.3 GHz Windows platform with 4 GB RAM, and implemented in Matlab/Simulink. The battery pack is modeled in Simulink, using the electric drives library. As can be seen in Figure 5(a), the battery pack model consists of five Li-ion batteries connected in series and each battery is made up of eight cells which are also connected in series (see Figure 5(b)). The rated capacity and nominal voltage of cell are 6.5 Ah and 3.6 V, respectively, and the values of other parameters are shown in Table 1.

TABLE 1: The parameters of battery cell.

Parameters	Value	Unit
Nominal voltage	3.6	V
Rated capacity	6.5	Ah
Maximum capacity	6.5	Ah
Fully charged voltage	4.2	V
Nominal discharge current	2.82	A
Internal resistance	0.005	Ohm
Capacity at nominal voltage	5.87	Ah
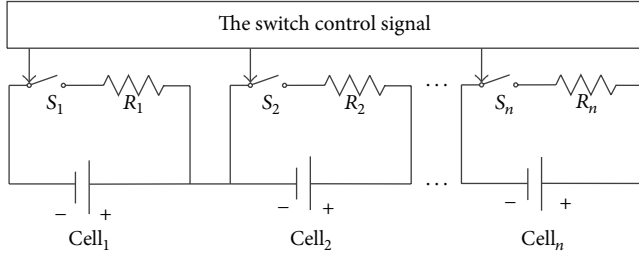
FIGURE 4: Passive cell balancing circuit.

And the SOC for a fully charged cell model is 100% and for an empty cell model is 0%. The SOC is calculated based on Coulomb-counting as

$$\text{SOC} = 100\left(1 - \frac{1}{Q}\int_0^t i(t)\,dt\right), \tag{5}$$

where Q is the rated capacity; t is the charging or discharging time; $i(t)$ is the charging or discharging current.

Here are several assumptions of the battery cell model in Simulink [21]:

(a) The parameters of the model are deduced from discharging characteristics and assumed to be the same for charging.

(b) The internal resistance is supposed to be constant during the charging and the discharging cycles and does not vary with the different amplitude of the current.

(c) The self-discharge of the battery is not represented and the battery has no memory effect.

(d) The model does not take the temperature into account.

3.1. Constant Current Charging-Discharging Model and Simulation

3.1.1. CCCD Model and Test Condition. As shown in Figure 6, the CCCD model provides a 6.5-A (1C) constant charging and discharge current for the battery pack. In a serially connected battery pack, discharging or charging progress has to be stopped immediately as soon as one of the terminal cell voltages falls below discharging voltage limit (DVL) or exceeds charging voltage limit (CVL) [2]. The values of DVL

and CVL of the cell modeled in this paper are 3.749 V and 4.2 V, and the SOC correspondingly reaches 30% and almost 100%.

Figure 7 shows the simulated SOC, current, and voltage of 10th cell during one CCCD cycle. The cell is charged by a 6.5-A constant current until voltage reaches CVL (4.2 V) and the corresponding SOC is almost 100%. It is discharged by the same current until voltage reaches DVL (3.749 V) and the corresponding SOC is 30%. The initial SOC and voltages of the cells in the battery pack is that one cell's SOC and voltage are 45.06% and 3.794 V while other cells are 35.06% and 3.769 V, respectively. The initial value of threshold VOA1 of the proposed method is 12.33.

3.1.2. Simulation Results and Analysis. The simulation results during whole CCCD test cycle are shown in Table 2 for different balancing scenarios (no balancing, voltage-based balancing, SOC-based balancing and outlier detection balancing). That the usable capacity calculated over balancing process of abnormal cell decreased by 0.584 Ah is the precondition for comparing the simulation results for different balancing scenarios.

With no balancing, the unbalanced cell (10th cell) in the battery pack cannot be completely discharged before charging and the normal cells cannot be completely charged before discharging during the whole test cycle, as detailed in Figure 8. Hence, the amount of usable energy of the pack decreased at the end of the charging process. With voltage-based balancing, the total voltage (charging cut-off) increased to 163.017 V and the SOC range of cells decreased to 1.030%, but the frequency of balancing switch on/off reaches as high as 352. With outlier detection balancing, the frequency of the switching on/off was significantly reduced from 352 to 2. And the voltage variance (charging cut-off) and the SOC variance were reduced to 0.008 and 0.158, respectively, when compared with voltage-based and SOC-based balancing algorithm. Additionally, the charging time (after CCCD testing, balancing algorithm off) of unbalanced battery pack is 2091 s and the charging capacity is 3.775 Ah while balanced battery pack (same condition with unbalanced pack testing) is 2413 s and 4.389 Ah, respectively.

The definitions of several evaluation standards in Table 2 are as follows.

Testing Time. Time of the whole simulation process.

Balancing Time. Sum of the balancing time.

Frequency of Switch on and off. Sum of the balancing circuit switch on/off times.

Usable Capacity Decrease. This can be calculated by the formula (6):

$$C_i = \int_0^t i_{\text{equ}}(t)\,dt,$$
$$i_{\text{equ}}(t) = \frac{V_i(t)}{R}, \tag{6}$$

(a) (b)
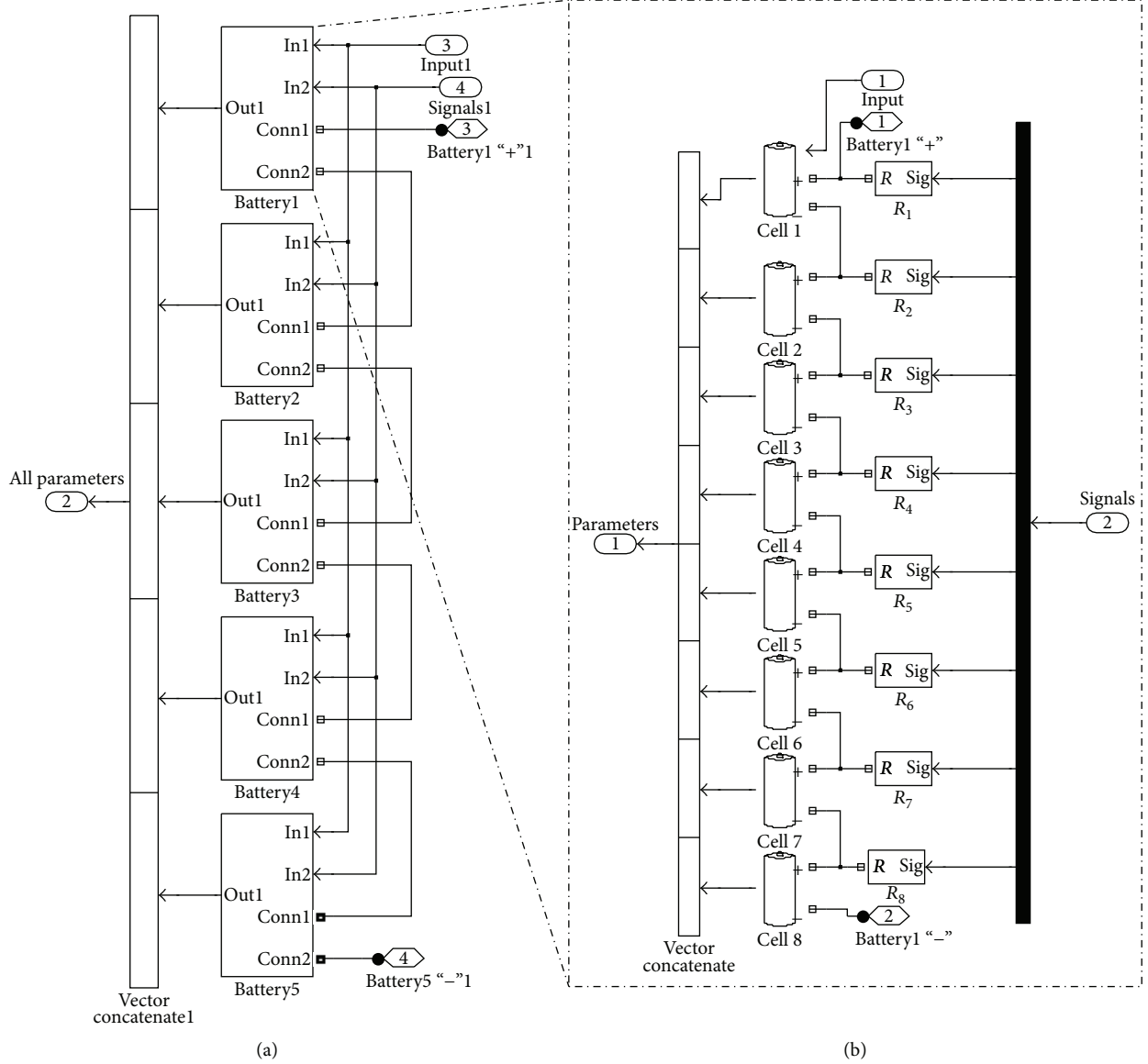
FIGURE 5: The schematic diagram of simulation model of pack (a) and battery (b).

where C_i is the decreased usable capacity of ith cell during balancing process; t represents the balancing time; $i_{equ}(t)$ is the balancing current; $V_i(t)$ denotes the voltage of battery at t time; R (33 ohm) represents the resistance in cell balancing circuit.

As illustrated by Figure 9, when the voltage-based balancing algorithm determined on and off the 10th cell as abnormal cell (upper plot), the outlier detection algorithm constantly and accurately did that (lower plot). Meanwhile, the testing time of the simulation decreased to 23037 s when implemented outlier detection equalization algorithm on CCCD model. The control signal "1" represents opening the balancing circuit and "0" means shutting it down.

The process that the battery pack transferred from unbalanced state to balanced state using different balancing

algorithms under CCCD cycle is shown as Figure 10. And Figure 11 shows the position variation of the abnormal cell (10th) detected by the proposed algorithm after balanced. The abnormal cell was closer to the other cells after it was balanced by outlier detection balancing algorithm.

3.2. Software-in-the-Loop Platform Model of BMS and Simulation

3.2.1. SILP Model of BMS and ECE + EUDC Test Condition.
The software-in-the-loop platform (SILP) model of BMS for electric vehicles gives a new idea to test the validity and reliability for BMS and power battery in different properties [20, 25]. Generally speaking, the model can also be used to test the feasibility and effectiveness of balance algorithms in
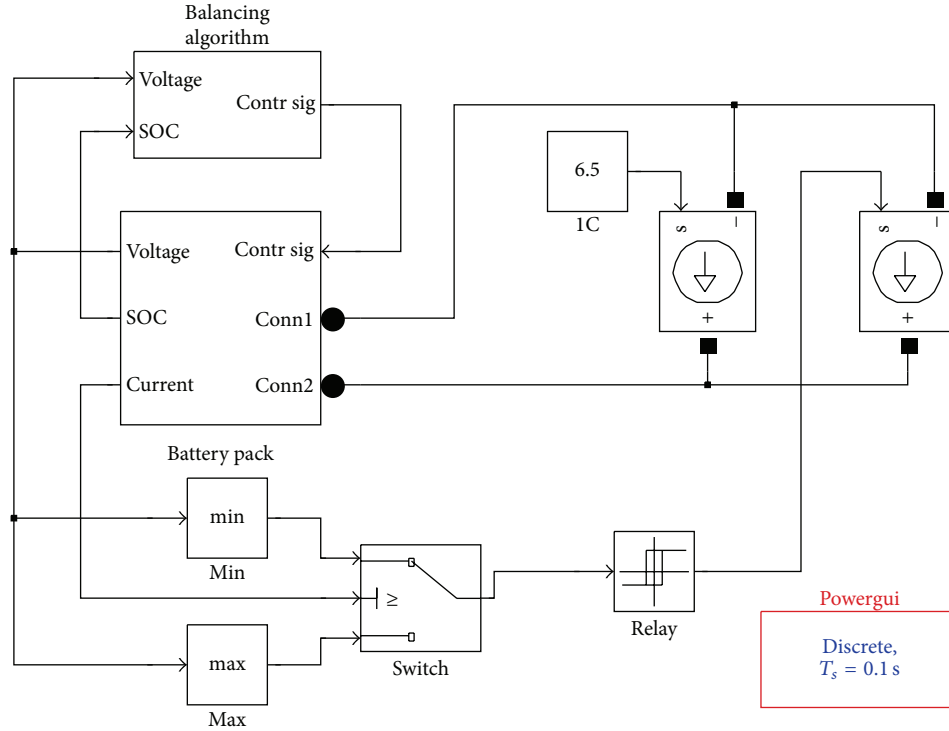
FIGURE 6: The constant current charge-discharge model.

TABLE 2: Performance comparisons of 3 algorithms under CCCD condition.

Evaluation standard	Balancing algorithm			
	No balancing	Voltage-based balancing	SOC-based balancing	Outlier detection balancing
Testing time (S)	—	37171	29048	23037
Balancing time (S)	—	17681	17783	17775
Usable capacity decrease (AH)	—	0.584	0.584	0.584
Frequency of switch on and off	—	352	2	2
Charging cut-off				
Total voltage (V)	161.214	163.017	162.939	165.738
Voltage range (V)	0.160	0.058	0.059	0.055
Voltage variance	0.025	0.009	0.009	0.008
Discharging cut-off				
Total voltage (V)	149.994	149.965	154.481	149.965
Voltage range (V)	0.034	0.005	0.005	0.001
Voltage variance	0.005	0.001	0.001	0
Charging time (S)	2091	2413	2413	2413
SOC range of cells (%)	10%	1.030%	1.025%	0.999%
SOC variance of cells	1.581	0.162	0.162	0.158

the early stage of design, greatly improving the reliability of the algorithms. In addition, the SILP model can simulate more different working conditions and the simulation results have more significance in actual engineering when compared with the CCCD model. The SILP model for BMS mainly includes driving cycle model, driver model, vehicle control unit model, battery management system software model,

battery model, power system model, and wheel model. The whole virtual environment model is shown in Figure 12.

The ECE + EUDC test cycle is used for EU type approval testing of emissions and fuel consumption from light duty vehicles. As Figure 13 shows, ECE + EUDC (bottom chart) cycle includes four ECE (upper left chart) segments repeated without interruption, followed by one EUDC (upper right
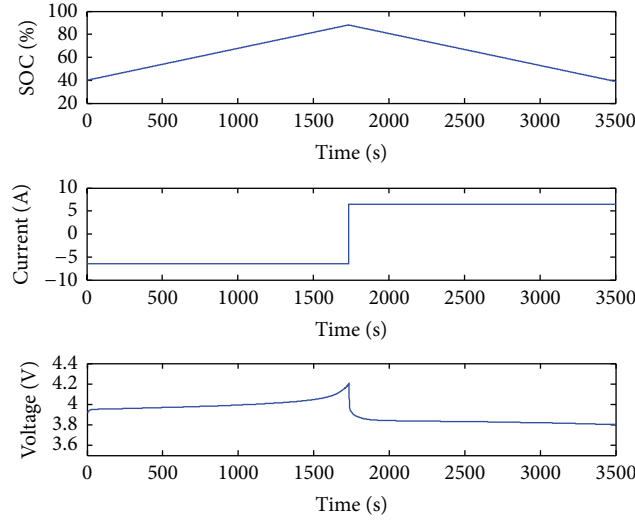
FIGURE 7: The SOC, current, and voltage of 10th cell during one 1C CCCD test cycle.

(a)

(b)
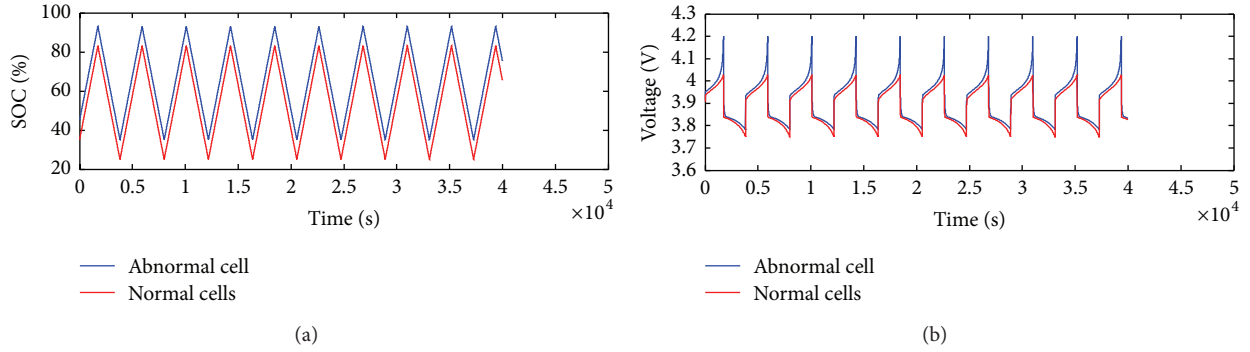
FIGURE 8: SOC (a) and voltage (b) of the unbalanced pack.

(a)

(b)
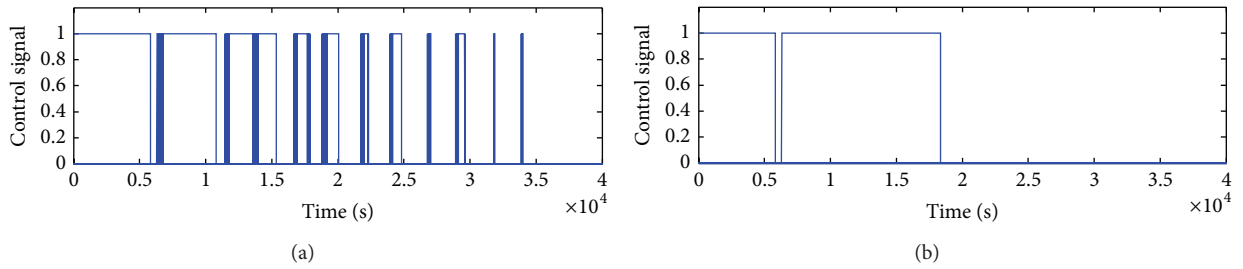
FIGURE 9: Balancing control signal of voltage-based (a) and outlier detection (b) balancing.

chart) segment. According to Figure 14, the battery pack is discharged with two ECE + EUDC cycle until 2400 s and charged by the generator in the next 822 s since the SOC of the pack reduced to 30%. Simulation test starts in the situation that there is one cell's SOC and voltage value are 64.83% and 4.140 V, and the others are 54.83% and 4.063 V, respectively.

3.2.2. Simulation Result and Analysis. The simulation results during the whole ECE + EUDC test cycle are shown in Table 3 for different balancing scenarios (no balancing, voltage-based balancing, SOC-based balancing, and outlier detection

balancing algorithm). That the usable capacity calculated over balancing process of abnormal cell decreased by 0.584Ah is the precondition for comparing the simulation results for different balancing scenarios.

With no balancing, the unbalanced cell in the battery pack could not be completely discharged before charging and the normal cells could not be completely charged before discharging during the whole test cycle, as detailed in Figure 15. Hence, the amount of usable energy of the pack decreased at the end of the charging process. With voltage-based balancing, the total voltage (charging cut-off) increased to 164.738 V

(a)

(b)
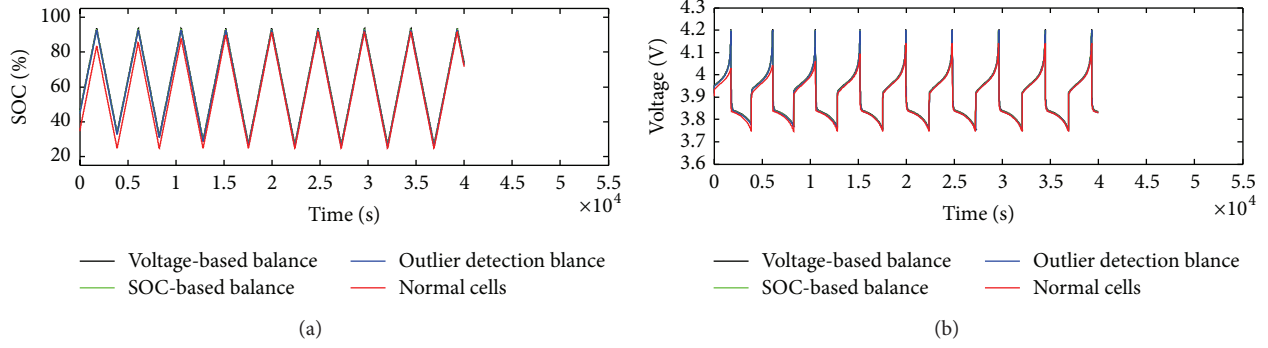
FIGURE 10: SOC (a) and voltage (b) of pack with balancing algorithms during CCCD cycle.

(a)

(b)
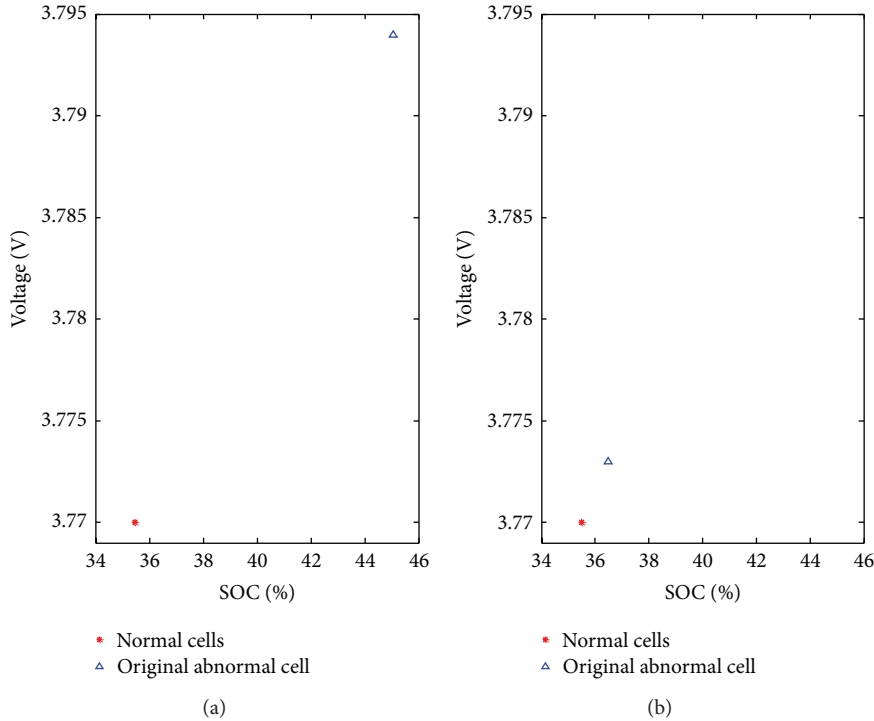
FIGURE 11: The position of the abnormal cell before it was balanced (a) and after it was balanced (b).

and the voltage variance (discharging cut-off) decreased to 0.003, but the frequency of balancing switch on/off reaches as high as 1150. With outlier detection balancing, the frequency of the switching on/off was significantly reduced from 1150 to 2 and reduced the voltage variance (charging cut-off) and the SOC variance to 0 and 0.157, respectively, when compared with voltage-based and SOC-based balancing algorithm. Furthermore, the proposed balancing algorithm increased the total charging cut-off voltage from 161.214 V to 165.738 V when compared with the pack without balancing. And it also reduced the discharging cut-off voltage variance and the SOC variance to 0 and 0.157, respectively.

The process in which the battery pack transferred from unbalanced state to balanced state with different balancing algorithms under ECE + EUDC test cycle is shown as

Figure 16. With outlier detection balancing algorithm, the cells in the battery pack can be completely charged/discharge at the same time and thus increase the available energy stored in the pack.

4. Conclusions

Aiming at the problem that present cell-balancing algorithms cannot identify the unbalanced cells in lithium-ion battery pack accurately in real-time, an algorithm based on outlier detection was proposed in this paper. The unbalanced cells were identified by the proposed balancing algorithms and balanced by shunt method using switches. After validating the efficiency of the balancing algorithms on two simulation models, the advantages of the proposed algorithm have been
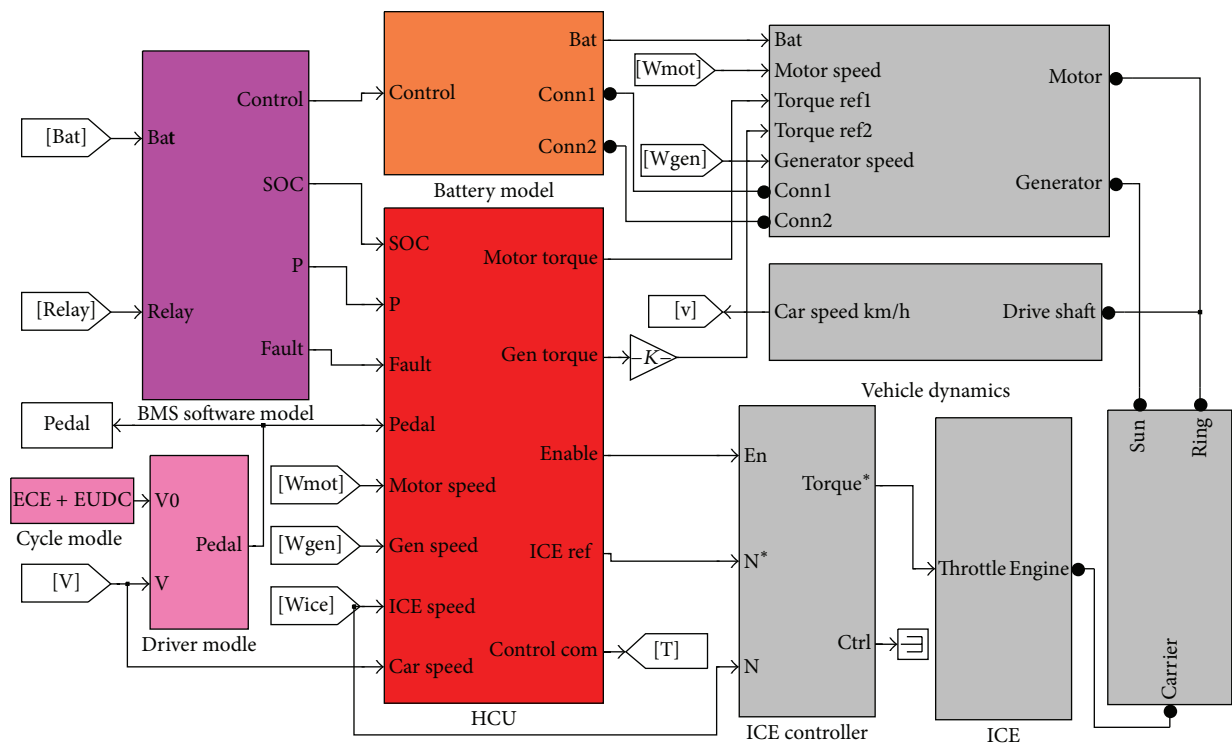
Figure 12: Software-in-the-loop platform model for BMS.

(a)

(b)

(c)
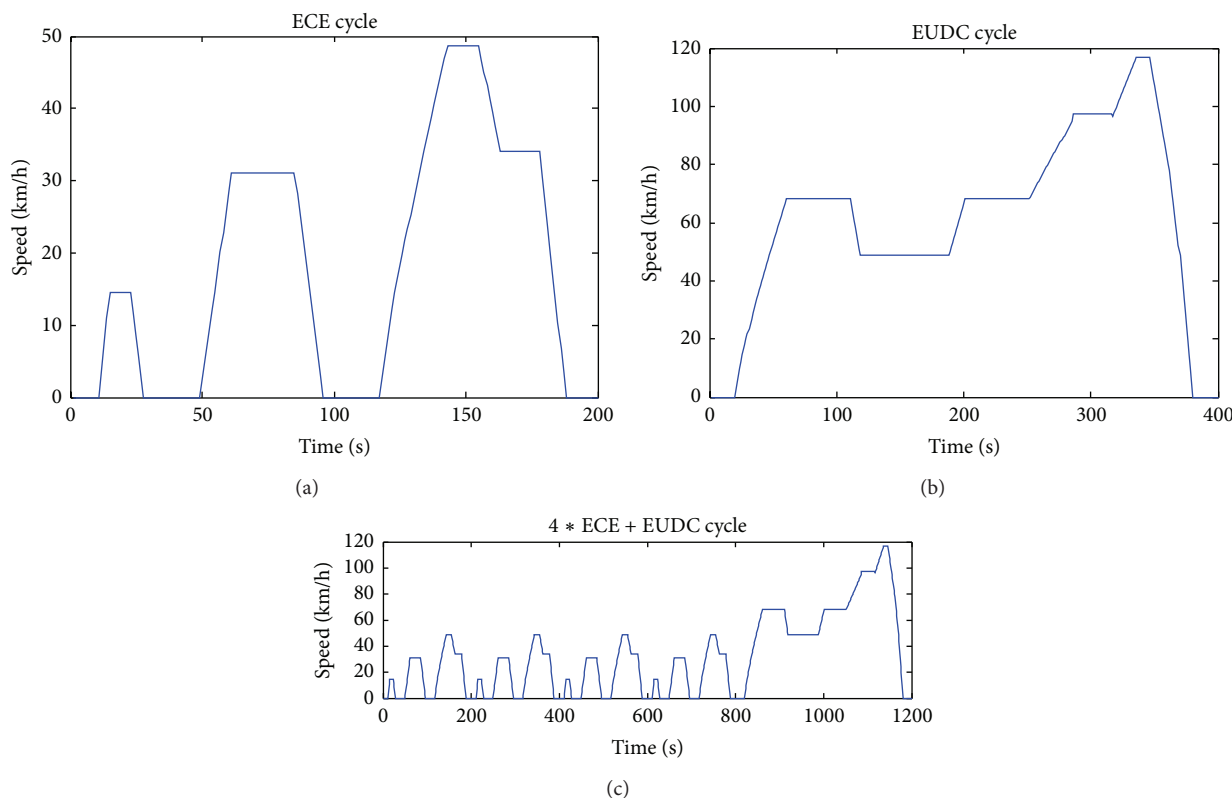
Figure 13: One ECE + EUDC cycle.

TABLE 3: Performance comparisons of 3 algorithms under 4 * ECE + EUDC condition.

Evaluation standard	Balancing algorithm			
	No balance	Voltage-based balancing	SOC-based balancing	Outlier detection balancing
Testing time (S)	—	57264	39074	22052
Balancing time (S)	—	16666	15795	15752
Usable capacity decrease (AH)	—	0.584	0.584	0.584
Frequency of switch on and off	—	1150	2	2
Charging cut-off				
Total voltage (V)	161.214	164.738	164.963	165.738
Voltage range (V)	0.174	0.058	0.003	0.003
Voltage variance	0.028	0.009	0.004	0
Discharging cut-off				
Total voltage (V)	149.994	149.965	150.882	149.965
Voltage range (V)	0.034	0.005	0.002	0.002
Voltage variance	0.005	0.003	0	0
SOC range of cells (%)	10%	1.030%	1.025%	0.999%
SOC variance of cells	1.581	0.162	0.162	0.157
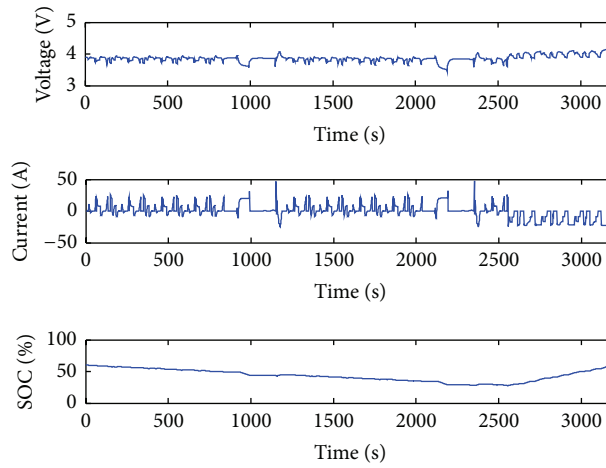
FIGURE 14: Voltage, current, and SOC of the pack during one ECE + EUDC diving cycle.

(a)

(b)
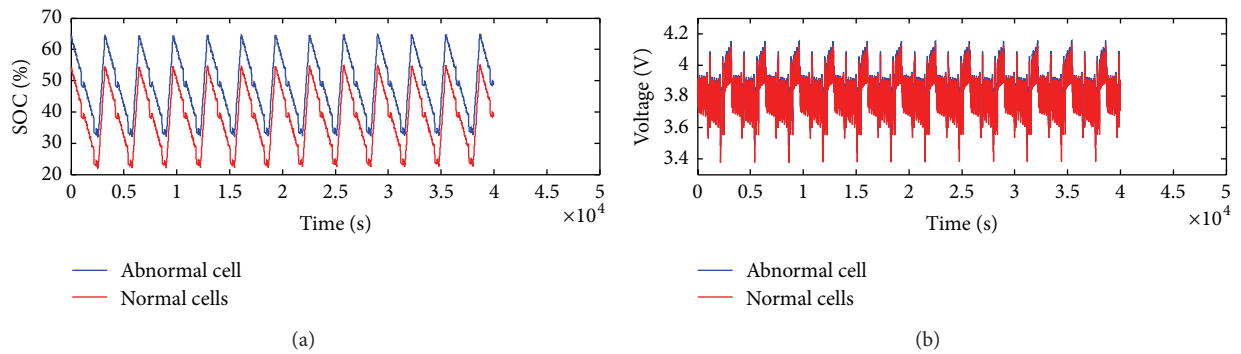
FIGURE 15: SOC (a) and voltage (b) of the unbalanced pack.
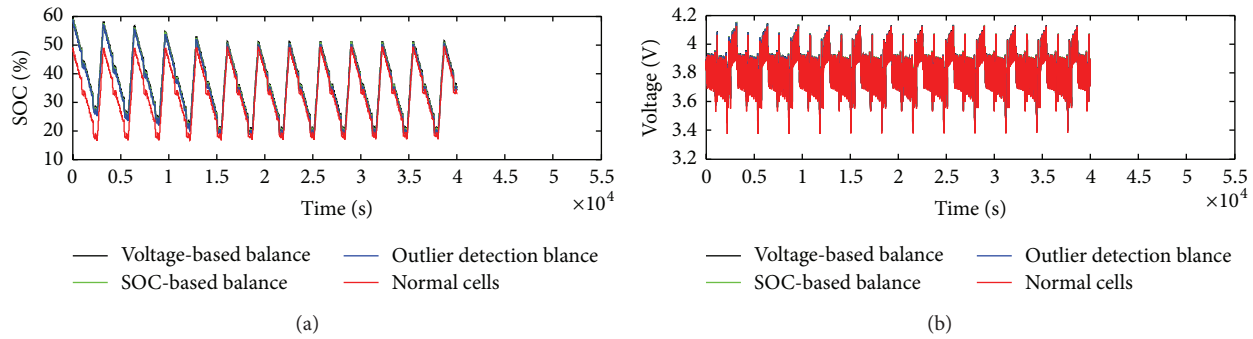
FIGURE 16: SOC (a) and voltage (b) of pack with balancing algorithms during ECE + EUDC test cycle.

pointed out in the context of simulation and analysis. The outlier detection equalization algorithm is able to recognize the abnormal battery cell accurately and to increase the usable energy and extend the lifetime of battery pack, which has extensive application prospect and theory value.

Further work will focus on taking the temperature of cells into account during whole charging and discharging process.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] J. B. Zhang, L. G. Lu, and Z. Li, "Key technologies and fundamental academic issues for traction battery system," *Journal of Automotive Safety and Energy*, vol. 3, no. 2, pp. 87–104, 2012.

[2] M. Einhorn, W. Roessler, and J. Fleig, "Improved performance of serially connected Li-ion batteries with active cell balancing in electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 6, pp. 2448–2457, 2011.

[3] S. G. Xu, Q. S. Zhong, and R. J. Zhu, "Research of equalizing charge control strategy for power battery," *Electric Machines and Control*, vol. 16, pp. 62–65, 2012.

[4] A. G. Xu, S. J. Xie, and X. B. Liu, "Dynamic voltage equalization for series-connected ultracapacitors in EV/HEV applications," *IEEE Transactions Vehicular Technology*, vol. 58, no. 8, pp. 3981–3987, 2009.

[5] L. Maharjan, S. Inoue, H. Akagi, and J. Asakura, "State-of-charge (soc)-balancing control of a battery energy storage system based on a cascade PWM converter," *IEEE Transactions on Power Electronics*, vol. 24, no. 6, pp. 1628–1636, 2009.

[6] T.-H. Kim, N.-J. Park N, R.-Y. Kim, and D.-S. Hyun, "A high efficiency zero voltage-zero current transition converter for battery cell equalization," in *Proceedings of the 27th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '12)*, pp. 2590–2595, IEEE, Orlando, Fla, USA, February 2012.

[7] J. J. Fu, B. J. Qi, and H. J. Wu, "Dynamic bi-direction equalization system research for lithium-ion batteries," *China Measurement Technology*, vol. 31, pp. 10–12, 2005.

[8] M. Einhorn, W. Guertlschmid, T. Blochberger et al., "A current equalization method for serially connected battery cells using a single power converter for each cell," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 9, pp. 4227–4237, 2011.

[9] Y. M. Ye, K. W. E. Cheng, and Y. P. B. Yeung, "Zero-current switching switched-capacitor zero-voltage-gap automatic equalization system for series battery string," *IEEE Transactions on Power Electronics*, vol. 27, no. 7, pp. 3234–3242, 2012.

[10] J. Xu, S. Q. Li, C. Mi, Z. Chen, and B. Cao, "SOC based battery cell balancing with a novel topology and reduced component count," *Energies*, vol. 6, no. 6, pp. 2726–2740, 2013.

[11] S. Yarlagadda, T. T. Hartley, and I. Husain, "A battery management system using an active charge equalization technique based on a DC/DC converter topology," in *Proceedings of the 3rd Annual IEEE Energy Conversion Congress and Exposition (ECCE '11)*, pp. 1188–1195, IEEE, Phoenix, Ariz, USA, September 2011.

[12] Y. Y. Wu and H. Liang, "A study on equalization charging for EV traction battery," *Automotive Engineering*, no. 16, pp. 382–385, 2004.

[13] C. H. Piao, W. L. Fu, G. H. Lei, and C. D. Cho, "Online parameter estimation of the Ni-MH batteries based on statistical methods," *Energies*, vol. 3, no. 2, pp. 206–215, 2010.

[14] Z. Y. Huang and Y. H. Cao, "Estimation for SOC of LiFePO4 Li-ion battery based on GA-RBF neural network," *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, vol. 25, no. 3, pp. 412–417, 2013.

[15] C. H. Piao, Z. Huang, L. Su, and S. Lu, "Research on outlier detection algorithm for evaluation of battery system safety," *Advances in Mechanical Engineering*, vol. 14, no. 1, pp. 65–70, 2014.

[16] K. K. Wang, G. X. Gui, W. Ni, and G. L. Gou, "Fastoutlierdata mining algorithm based on cell in large datasets," *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, no. 5, pp. 673–677, 2010.

[17] H. D. Wang, Y. H. Tong, S. H. Tan, S. E. Tang, and D. Q. Yang, "Research progress on outlier mining," *CAAI Transactions on Intelligent Systems*, no. 5, pp. 67–74, 2006.

[18] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: a survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 5, pp. 823–839, 2012.

[19] F. Angiulli and C. Pizzuti, "Outlier mining in large high-dimensional data sets," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 203–215, 2005.

[20] C. H. Piao, Q. F. Yu, C. X. Duan, L. Su, and Y. Zhang, "Virtual environment modeling for battery management system," *Journal of Electrical Engineering & Technology*, vol. 9, no. 5, pp. 1729–1738, 2014.

[21] O. Tremblay and L.-A. Dessaint, "Experimental validation of a battery dynamic model for EV applications," *World Electric Vehicle Journal*, vol. 2, pp. 930–939, 2009.

[22] F. Angiulli, S. Basta, and C. Pizzuti, "Distance-based detection and prediction of outliers," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 2, pp. 145–160, 2006.

[23] H. Spath and J. Goldschmidt, *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*, Halsted Press, New York, NY, USA, 1985.

[24] M. Daowd, N. Omar, P. van den Bossche, and J. van Mierlo, "Passive and active battery balancing comparison based on MATLAB simulation," in *Proceedings of the 7th IEEE Vehicle Power and Propulsion Conference (VPPC '11)*, pp. 1–7, IEEE, Chicago, Ill, USA, September 2011.

[25] P. Chen, C. H. Piao, C. X. Duan, and S. Lu, "Modeling of battery management system software in virtual simulation environment," *Automotive Safety and Energy*, vol. 4, no. 1, pp. 67–74, 2013.

Research Article

The Study of Operation Modes and Control Strategies of a Multidirectional MC for Battery Based System

Saman Toosi,[1] Norhisam Misron,[1,2] Tsuyoshi Hanamoto,[3] Ishak Aris,[1] Mohd Amran Mohd Radzi,[1] and Hiroaki Yamada[4]

[1] Department of Electrical & Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM),
 43400 Serdang, Selangor, Malaysia
[2] Institute of Advanced Technology (ITMA), Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia
[3] Department of Biological Functions Engineering, Graduate School of Life Science and Systems Engineering,
 Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu 808-0916, Japan
[4] Graduate School of Science and Engineering, Yamaguchi University, 2-16-1 Tokiwadai, Ube-shi, Yamaguchi 755-8611, Japan

Correspondence should be addressed to Norhisam Misron; norhisam@upm.edu.my

To enhance the performance of stand-alone battery based system and to achieve the continuous power transmission, the behavior of multidirectional matrix converter (MDMC) has been analyzed in different operation modes. A systematic method interfacing a renewable source, a storage battery, and a load is proposed for a stand-alone battery based power system (SABBPS) to utilize the MDMC as PWM converter, inverter, or PWM converter and inverter in different operation modes. In this study, the Extended Direct Duty Pulse Width Modulation (EDDPWM) technique has been applied to control the power flow path between the renewable source, load, and the battery. Corresponding to generator voltage, input frequency, and loads demands, several operating states and control strategies are possible. Therefore, the boundaries and distribution of operation modes are discussed and illustrated to improve the system performance. The mathematical equation of the EDDPWM under different operation modes has been derived to achieve the maximum voltage ratio in each mode. The theoretical and modulation concepts presented have been verified in simulation using MATLAB and experimental testing. Moreover, the THD, ripple, and power flow direction have been analyzed for output current to investigate the behavior of system in each operation mode.

1. Introduction

Renewable sources, such as hydro, solar, and wind, have the potential to play an important role in providing energy with sustainability to the vast populations of the world who do not have access to clean energy. Currently the stand-alone power system supplies the local villages or individual users in remote area where the grid extension is difficult or not economical. In stand-alone renewable power system storage elements such as battery or supercapacitor are used to supply power to the loads continuously [1]. The complexity of the control system remarkably rises by increasing the number of input sources as several separate converters are employed to control the power flow direction between the input ports and output ports of system. The type and number of power electronic converters change based on the type and number of energy sources and loads.

Multi-input DC-DC converter has been proposed to combine several types of energy sources and to obtain a desired DC output voltage [2]. This type of converter is commonly employed in the hybrid electric vehicles [3] and renewable power system to stabilize the voltage and frequency of system [4].

Several schemes of converter with two source inputs and single output have been proposed for the stand-alone hybrid renewable energy system in order to increase the power quality and reliability of system [5, 6]. A double-input sources' single-output source power converter was developed to combine the energy sources of wind and solar in one power system [5]. However, it will be difficult to acquire

the regulated voltage output if one of the DC sources is diminished, since the input voltage variation is significant. Therefore, Chen et al. (2001) proposed the high frequency transformer for double source system with isolated electrical circuit to reduce the voltage variation effect. These circuits have no charging path for the backup battery storage and could not control the power flow direction in system [6].

A multidirectional power converter (MDC) has been presented by Mei et al. (2006) for a battery based stand-alone hybrid renewable energy system to supply the battery and to control the power flow direction in a system. The MDC provided a battery storage path to supply the power demand in the days of deficit in solar and wind powers. However, the proposed MDC control was very complicated due to the number of modes and effect of power flow direction in high frequency isolated converter [7]. Later, the directional power converter is changed with bidirectional high-power-density DC-DC converter to interface with multiple energy storage components such as batteries and ultracapacitors. The proposed system needs multiwinding transformer for soft-switching conditions which cannot justify the unique features of low component count and compact structure for the integrated multiport converter [8].

Four-port DC/DC converter with bidirectional capability and isolated output feature has been proposed to reduce the size. Zero-voltage switching is introduced for all four main switches. Three of the four ports were tightly regulated by adjusting their independent duty-cycle values, while the fourth port was left unregulated to maintain the power balance for the system. In addition, a decoupling network is introduced to allow the separate controller design for each power port. This four-port converter is suitable for low-power applications, where the energy storage is required while allowing tight load regulation [9]. However, the major problem is that the input and output port are DC and isolated transformer is essential for this structure.

In order to achieve maximum power tracking (MPT) through rotor speed control under varying wind speeds and control of the magnitude and the frequency of the load voltage, the new multiport system has been introduced based on two back-to-back voltage source converters (VSCs) with a battery energy storage system at their DC link. The proposed hybrid system was able to control the power flow, by which it controls the magnitude and the frequency of the load voltage [10]. The number of converters and passive components is still high and it is not suitable for the integrated multiport converter. Moreover, the size and cost increase and efficiency decreases due to multiple-stage conversion through the converters and transformers.

Most desired feature of multidirectional converter can be fulfilled by using matrix converter (MC) structure. In the MC, several bidirectional switches are used to couple the power sources to load side. With proper switching method, the bidirectional switches in MC can be utilized as inverter or rectifier. The first principle of MC control has been proposed by Venturini and Alesina in 1980, which is known as a "direct transfer function" approach [11]. They also extended the voltage ratio to 0.866 by using the third harmonic injection technique [12]. In 1983, Rodriguez introduced the novel

control method based on "fictitious DC link" to reduce the complexity of direct method [13]. Ziogas et al. expanded Rodriguez's "fictitious DC link" idea to provide a rigorous mathematical explanation [14, 15]. Later, Kastner and Rodriguez (1985) used space vectors modulation in the switching control of matrix converters to increase the voltage ratio and reduce the number of switching states [16, 17]. Several techniques have been reported which may have simplified the modulation [18–21] and solved the commutation problems in MC [22, 23]. Although the SVPWM technique is the proper method for the three-phase matrix converters, the complexity of the implementation remarkably rises by increasing the number of inputs or outputs of MC and there is no attention to the input current.

To synthesize the sinusoidal input current with unity power factor and desired output voltage a new carrier based modulation method has been proposed based on the conventional space vector pulse modulation (SVPWM), with complex calculation [24]. In order to simplify the modulation method, the preliminary concepts of a new carrier based PWM strategy, named direct duty ratio PWM (DDPWM), are presented in [25]. They extended the DDPWM to various topologies of matrix converter and derived the control schemes for alternative structures converters in [26]. This modulation scheme is highly flexible and intuitive and it can be applied to any configuration of the matrix converter.

Toosi et al. in 2014 combined the characteristics of several separate converters in multidirectional matrix converter (MDMC) and proposed a novel modulation method that can control the power flow direction between each of input power supplies and output loads. The proposed modulation method is able to inject power from DC and AC supply to the load simultaneously by using the proper switching pattern. They validated that the MDMC with EDDPWM can work as modular converter, where the frequency and voltage of each output phase are independent of other output phases [27]. However, the system has been tested in simulation and operated in one mode.

Due to the high number of system parameters (i.e., number of inputs/outputs, load parameters and input filter, output frequency, switching frequency, modulation methods, and number of passive components) and the inherent differences between the converter topologies such as the maximum voltage transfer ratio, it is difficult to compare the proposed MDMC with the other multisource converters such as multidirectional power converter [7], ZVS bidirectional DC-DC converter [8], AC-DC-AC converter [28], multiport with several voltage sources converter [10], integrated four-port DC/DC converter [9], multi-input DC/DC converter [2], and double-input single-output power converter [5]. The main advantage of the MDMC for multi-input/output power system compared to other converters is its potential to decrease size of system by combining all characteristics of different rectifiers and inverters in one compact silicon converter and eliminating the passive component such as bulky capacitor and multiwinding transformer. In addition, the EDDPWM technique can provide bidirectional power flow, control input power factor, and synthesize the sinusoidal input current and output voltage waveforms.
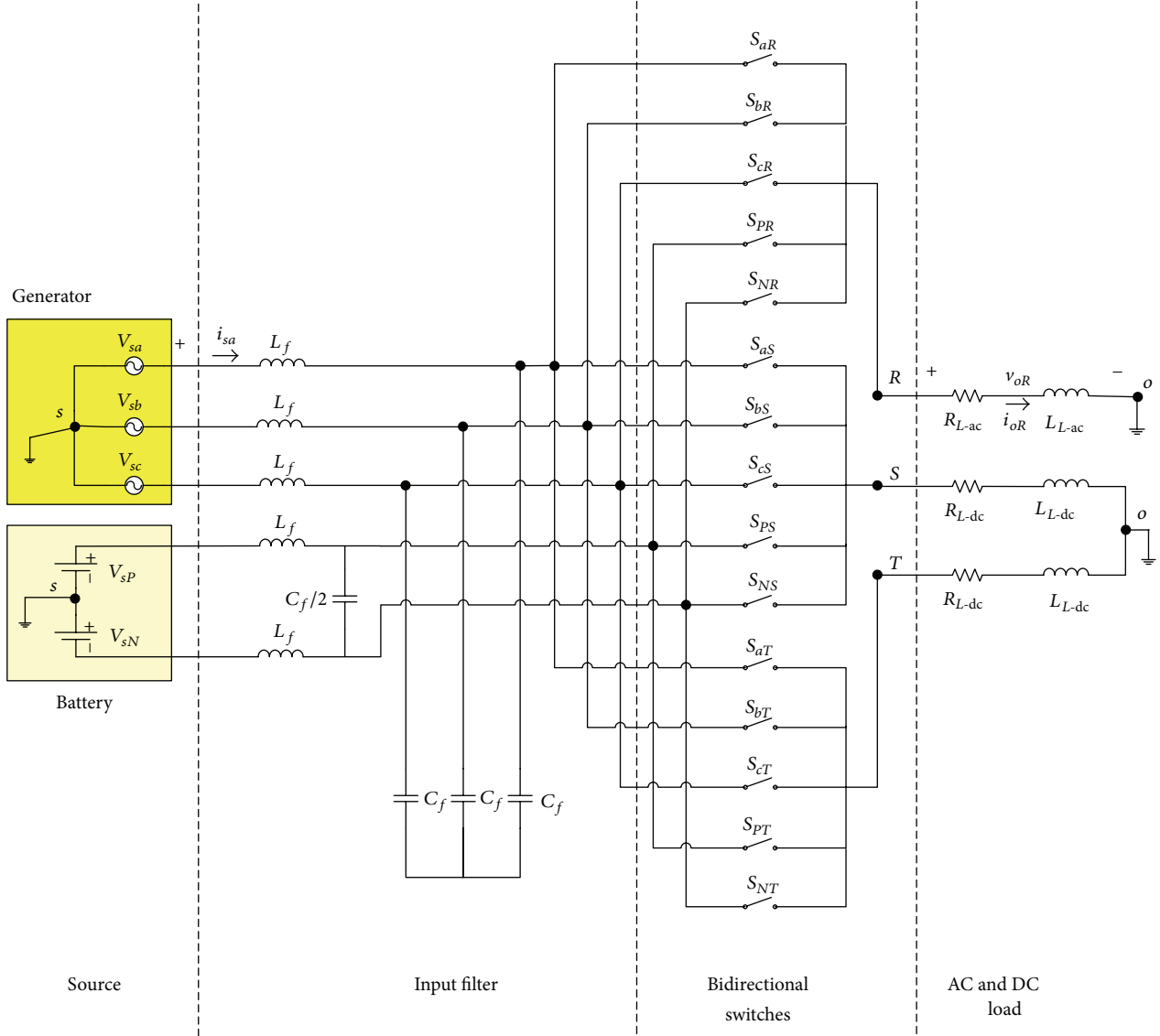
FIGURE 1: Multidirectional matrix converter circuit.

The voltage ratio of MDMC varied when the system switched from one operation mode to another mode. In order to keep the output voltage constant, the output voltage command should be calculated based on the type and number of input sources and output loads which are connected to the system. Accordingly, the behaviour of MDMC has been investigated in different operation modes to achieve the continuous power transmission and increase the efficiency by reducing the number of switches and calculating the maximum voltage ratio in each operation mode. By adjusting the time subinterval in EDDPWM method and analysing the power flow direction among the input and output ports of system, MDMC can work as inverter, PWM converter, or both PWM converter and inverter. In addition, this study is dedicated to analysis of the output current quality and

derives the necessary equation of EDDPWM method in each operation mode.

2. Principle of Extended Direct Duty PWM

The operating principle of the EDDPWM has been described in [27], for MDMC with 15 bidirectional switches. Figure 1 indicates the circuit configuration of the MDMC in stand-alone battery based system (SABBS) when three-phase generator and two batteries are connected to the source side of the MDMC. The $R_{L\text{-dc}}$ and $L_{L\text{-dc}}$ and $R_{L\text{-ac}}$ and $L_{L\text{-ac}}$ indicate the DC and AC load, respectively.

According to Figures 2 and 3, a switching period T_s is divided into two time periods, T_c and T_3. During T_c, the input phases of AC generator are connected to a corresponding
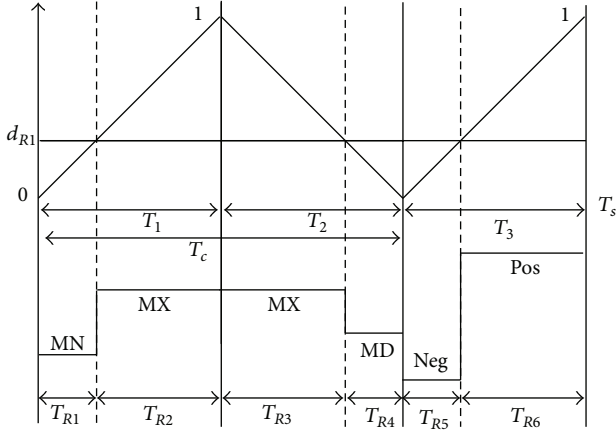
FIGURE 2: Switching pattern I, output R phase switching state.
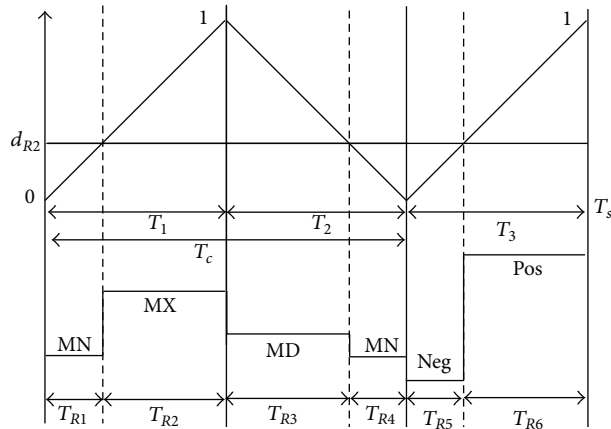
FIGURE 3: Switching pattern II, output R phase switching state.

output terminal, and during T_3 the input phases of DC battery are connected to a corresponding output terminal. In addition, the time interval T_c is divided into two periods, T_1 and T_2. Also, the MX, MD, and MN denote the instantaneous values of maximum, medium, and minimum input voltages of AC generator.

Furthermore, POS and NEG denoted the instantaneous values of positive and negative input voltages of DC battery, respectively. During T_1, the line-to-line voltage between MX and MN is used, which is the maximum line-to-line voltage among three line-to-line input voltages of generator at the sampling instant. During T_2, the second maximum line-to-line voltage is used which is MX to MD for switching pattern

I and MD to MN for switching pattern II. Finally, during T_3 the line-to-line voltage between POS and NEG is employed.

2.1. Switching Pattern I. Figure 2 indicates the switching pattern I, where the R phase duty ratio value (d_{R1}) is compared with triangular carrier waveform to generate the R phase output voltage. The output phase is changed during the switching pattern I from MN $\rightarrow$ MX $\rightarrow$ MX $\rightarrow$ MD $\rightarrow$ NEG $\rightarrow$ POS, consequently. As illustrated in Figure 2, the output phase "R" is connected to the input phase "MN" during T_{R1} and when T_s is the sequence switching period. And it is connected to phases "MX," "MX," "MD," "NEG," and "POS" during time periods T_{R2}, T_{R3}, T_{R4}, T_{R5}, and T_{R6}, respectively. These six time subintervals can be represented as (1), where d_{R1} is the R phase duty ratio value and carrier slops are defined as $m = T_1/T_c$ and $n = T_c/T_s$. Consider

$$
\begin{aligned}
T_{R1} &= d_{R1} \cdot m \cdot n \cdot T_s, \\
T_{R2} &= (1 - d_{R1}) \cdot m \cdot n \cdot T_s, \\
T_{R3} &= (1 - d_{R1}) \cdot (1 - m) \cdot n \cdot T_s, \\
T_{R4} &= d_{R1} \cdot (1 - m) \cdot n \cdot T_s, \\
T_{R5} &= d_{R1} \cdot (1 - n) \cdot T_s, \\
T_{R6} &= (1 - d_{R1}) \cdot (1 - n) \cdot T_s.
\end{aligned}
\tag{1}
$$

The fluctuation of the input voltage is negligible during the switching periods. Thus, the integration of the output voltage v_{oR} over T_s can be expressed in

$$
\int_0^{T_s} v_{oR} dt \cong T_{R1} \cdot \text{MN} + (T_{R2} + T_{R3}) \cdot \text{MX} + T_{R4} \text{MD} \\
+ T_{R5} \cdot \text{NEG} + T_{R6} \cdot \text{POS}.
\tag{2}
$$

Based on (1) and (2), the average output voltage can be expressed in terms of m and n as presented in

$$
\overline{v}_{OR} = \frac{1}{T_s} \int_0^{T_s} v_{oR} dt \cong d_{R1} \left(- (1 - n) \cdot \text{POS} - n \cdot \text{MX} \right. \\
+ (1 - m) \cdot n \cdot \text{MD} + m \cdot n \cdot \text{MN} + (1 - n) \cdot \text{NEG}) \\
+ n \cdot \text{MX} - (1 - n) \cdot \text{POS}.
\tag{3}
$$

Therefore, for present switching cycle, the duty ratio value, d_{R1}, can be written as

$$
d_{R1} = \frac{\left(v_{oR}^* - n \cdot \text{MX} - (1 - n) \cdot \text{POS} \right)}{- (1 - n) \text{POS} - n \cdot \text{MX} + (1 - m) \cdot n \cdot \text{MD} + m \cdot n \cdot \text{MN} + (1 - n) \cdot \text{NEG}},
\tag{4}
$$

where v_{oR}^* is the R phase output voltage command which is equal to $\overline{v}_{OR}$.

2.2. Switching Pattern II. The procedure to drive the equation for switching pattern II is the same as the previous switching

pattern. Figure 3 illustrates the case of switching pattern II where the R phase duty ratio value (d_{R2}) is compared with triangular carrier waveform to generate the R phase output voltage. The output phase is changed during switching pattern II from MN $\rightarrow$ MX $\rightarrow$ MD $\rightarrow$ MN $\rightarrow$ NEG $\rightarrow$ POS, consequently.

Similarly, the integration of the output voltage v_{oR} and the average output voltage $\bar{v}_{OR}$ is presented in

$$\int_0^{T_s} v_{oR} dt \cong (T_{R1} + T_{R4}) \cdot \text{MN} + T_{R2} \cdot \text{MX} + T_{R3} \cdot \text{MD}$$
$$+ T_{R5} \cdot \text{NEG} + T_{R6} \cdot \text{POS},$$
$$\bar{v}_{OR} = \frac{1}{T_s} \int_0^{T_s} v_{oR} dt \cong d_{R2} \left(-(1-n) \cdot \text{POS} - m \cdot n \right. \quad (5)$$
$$\cdot \text{MX} - (1-m) \cdot n \cdot \text{MD} + n \cdot \text{MN} + (1-n) \cdot \text{NEG})$$
$$+ (1-n) \cdot \text{POS} - m \cdot n \cdot \text{MX} + (1-m) \cdot n \cdot \text{MD}.$$

By letting $\bar{v}_{OR}$ be equal to v_{oR}^* the duty ratio value d_{R2} can be written as

$$d_{R2} = \left(v_{oR}^* - n \cdot \text{MX} - (1-n) \cdot \text{POS} \right) \left(-(1-n) \cdot \text{POS} \right.$$
$$- n \cdot \text{MX} + (1-m) \cdot n \cdot \text{MD} + m \cdot n \cdot \text{MN} + (1-n) \quad (6)$$
$$\left. \cdot \text{NEG} \right)^{-1}.$$

When the switching state for output phase "R" is POS, NEG, MX, MD, or MN, the output phase "R" is connected to the input phase voltage POS, NEG, MX, MD, or MN, respectively. In fact, by using logic devices such as FPGA, the circuit for generating the PWM signal can easily be implemented.

It is possible to synthesize the input current and control the power factor in EDDPWM by adjusting the amount of m and n, when m and n are related to the maximum and minimum current as represented:

$$m \equiv \frac{T_1}{T_c} = -\frac{i_{sMX}}{i_{sMN}},$$
$$n \equiv \frac{T_c}{T_s} = \frac{i_{sMN}}{(i_{sMN} - i_{sPOS})}. \quad (7)$$

3. System Operation Modes

According to the type and number of outputs and inputs connected to the system, the operation states of SABBPS system can be classified into five possible modes which are listed in Table 1. Whether energy sources provide power for the load or battery, when the converter is connected to the AC load and when the battery bank absorbs or supplies power. In Table 1 the renewable source, supplying power to the battery, or the AC load is defined by "O" or "X." For the battery bank, "▼" illustrates a discharge, whereas "▲" illustrates a charge. In addition, the AC load connecting to the MDMC is indicated by "O" or by "X."

Power sources of the MDMC change for different operation modes. In operation mode 1, power is transferred from

TABLE 1: The operation states of SABBPS.

Parameters	SABBS operation mode				
	1	2	3	4	5
Three-phase generator	O	O	O	X	O
AC load	O	O	X	O	O
Battery bank	X	▲	▲	▼	▼

O = connected, X = disconnected, ▲ = charging, and ▼ = discharging.

G: generator
L: AC load
BB: battery bank
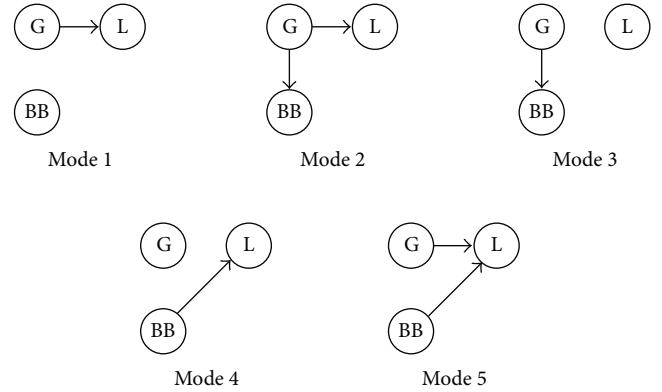
FIGURE 4: Simplified sketch of the SABBS under different operation modes.

the AC generator to the AC output load. Thus, the battery bank is disconnected from the system and the MDMC acts as three-to-single-phase AC/AC converter. In operation mode 2, power is transferred from the AC generator to the battery bank and the AC output load; thus, the MDMC supplies power for the AC load and charges up the battery bank via the charging controller circuit; MDMC acts as inverter and PWM converter. In operation mode 3, the AC load is disconnected from system; the variable speed turbine only exports energy to the battery bank through the bidirectional switches of the MDMC (just charges up the battery bank). Thus, MDMC works as PWM converter. By following the above analytic approach, other operation modes also can be educated.

Figure 4 indicates the simplified sketch of the stand-alone battery based system under five different operation modes, where "G," "BB," and "L" represent the three-phase generator, the battery bank, and the AC load, respectively.

3.1. Battery Charging.
Several methods, such as SOC estimation based on the terminal voltage and internal resistance [29], cell-impedance and impedance variation of cells/batteries [30, 31], error-correction mechanism based on Kalman filter for both state observation and prediction problems [32, 33], SOC estimation based on artificial neural networks [34], and fuzzy logic principles [35], have been reported by researchers to predict SOC. According to [36] the open circuit voltage method is online and cheap and makes it easily to determine SOC in battery based system. Open circuit
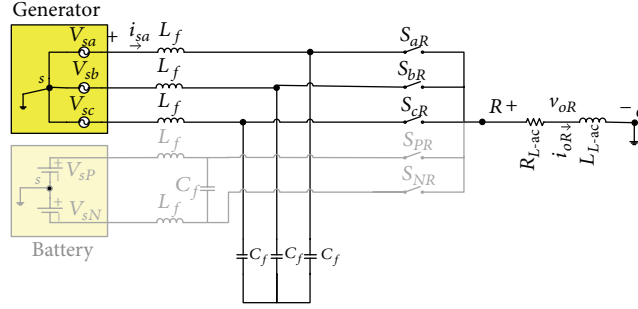
FIGURE 5: Circuit configuration of MDMC in operation mode 1.

voltage technique for SOC estimation monitors the terminal voltage and current under discharging state to determine the voltage of a battery under load.

The MDMC with EDDPWM method introduced by [27] is able to read the voltage at the DC side continually and monitor the amount of current injection to DC load to determine the SOC and charging the battery. Therefore, the output phases S and T indicated in Figure 1 can be used for battery charging or supplying the DC load. However, the extra charge control circuit should be added to the system to increase the safety and reduce the charging time of system.

4. Equivalent Circuit of the Converter

The EDDPWM control method can be applied to the MDMC as a modular structure for each phase where each output phase has the independent reference control signal. This reference control signal can be different in terms of frequency, waveform shape, and amplitude [27]. The MDMC works in operation modes 1, 2, and 3 when the line-to-neutral voltage of AC generator $v_{s\text{-rms}}$ is bigger than battery voltage $v_{s\text{-dc}}$ and works in operation modes 4 and 5 when $v_{s\text{-dc}}$ is bigger than the line-to-neutral $v_{s\text{-rms}}$.

Regarding the semiconductor finite switching times and propagation delays in practical, switches cannot be switched on and off instantaneously in MDMC. In addition, due to the lack of a natural free-wheeling in this structure, reliable current commutation between switches in MDMC is too difficult. Therefore, in this study, current commutation or four-step commutation method introduced by [37, 38] has been improved based on MDMC structure for safe commutation between an outgoing and an incoming switch. Furthermore, in MDMC, any problem in supplying the current to load can be generated over voltages at output phase. Also, overvoltages can appear from the input side caused by line disturbance. Hence, a clamp circuit introduced by [39] has been modified based on MDMC structure to avoid overvoltages coming from the grid and from the load to the system.

4.1. Operation Mode 1. In this mode, the MDMC acts as three-phase to single-phase matrix converter. The power is transferred through the three bidirectional switches which are connected to input phases a, b, and c to output phase R. According to switching patterns I and II when $n = 1$, the input

of MDMC is connected to the AC generator and battery is disconnected from the system. Figure 5 indicates the circuit configuration of MDMC in operation mode 1 with neutral connection.

According to EDDPWM switching method, output terminals can separately be controlled to follow their reference signals. Therefore, the EDDPWM can be used as modular configuration at each output phase. The duty ratio for phase R in operation mode 1 can be represented as (8). Due to the existence of neutral connection in this mode the maximum voltage ratio is limited to $q_{\max} = 0.5$. q is the voltage ratio of the input terminal voltage to the output terminal voltage ($v_{o\text{-rms}}/v_{s\text{-rms}}$). The duty cycle in operation mode 1 for switching pattern I and switching pattern II can be calculated in the same manner of (4) and (6) while $n = 1$ in this operation mode. The duty ratio can be represented as follows for switching patterns I and II, respectively:

$$d_{R1} = \frac{v_{oR}^* - \text{MX}}{-\text{MX} + (1 - m) \cdot \text{MD} + m \cdot \text{MN}}, \tag{8}$$

$$d_{R2} = \frac{v_{oR}^* - n \cdot \text{MX} + (1 - m) \cdot \text{MD}}{-m \cdot \text{MX} + (-1 + m) \cdot \text{MD} + \text{MN}}. \tag{9}$$

The output voltage command v_{oR}^* of the converter can be represented as below:

$$v_{oR}^* = \sqrt{\frac{2}{3}} \cdot q \cdot v_{s\text{-rms}} \cdot \sin\left(2\pi f_o t\right), \tag{10}$$

where $v_{s\text{-rms}}$ is the line-to-line RMS value and f_o is the desired output frequency for the corresponding phase.

4.2. Operation Mode 2. In this mode, the MDMC acts as three-phase to single-phase matrix converter [40] (through the switches S_{aR}, S_{bR}, and S_{cR}) and two three-phase to single-phase PWM converter (through the switches S_{aS}, S_{bS}, and S_{cS} for output phase S and S_{aT}, S_{bT}, and S_{cT} for phase T). The power is transferred through the nine bidirectional switches which are connected to input phases a, b, and c to output phases R, S, and T. As described in operation mode 1 when $n = 1$, the input of MDMC is connected to the AC generator and battery is disconnected from the system. Figure 6 shows the circuit configuration of MDMC in operation mode 2 with three output terminals.
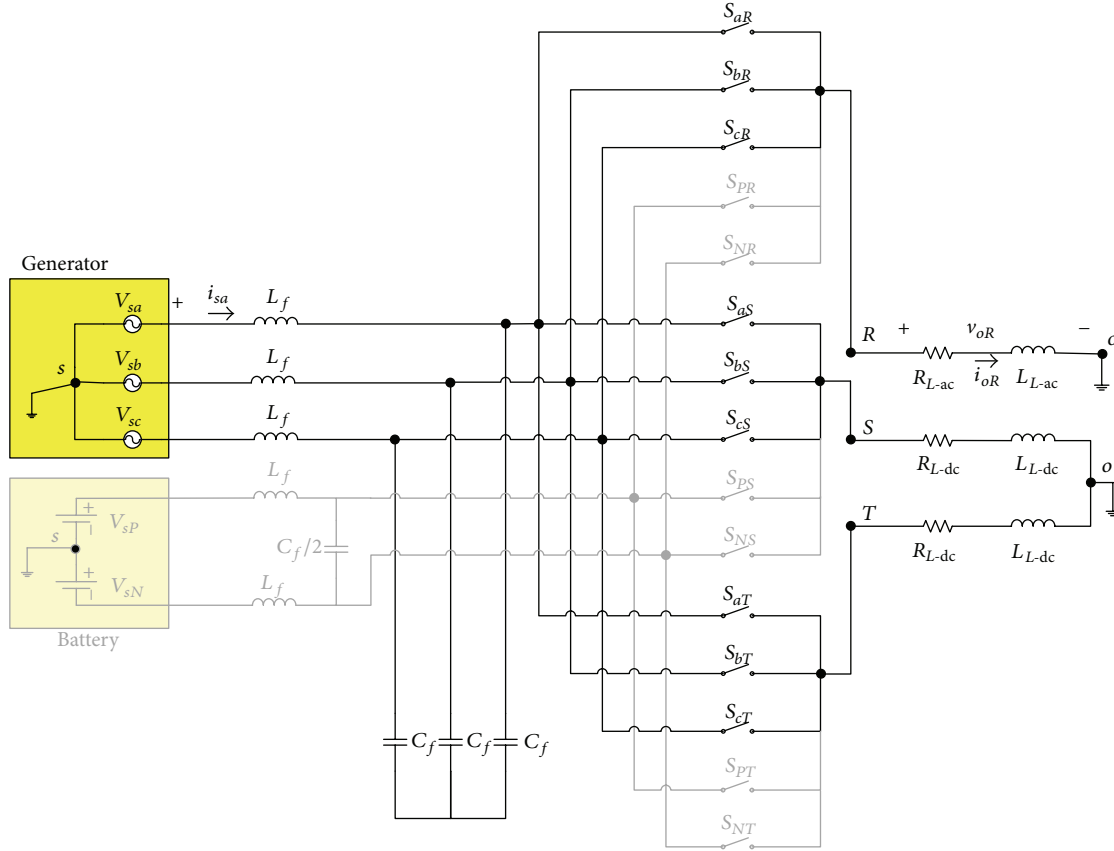
FIGURE 6: Circuit configuration of MDMC in operation mode 2.

Three bidirectional switches are used for each output phase to apply the switching patterns I and II. The POS and NEG input phases are always disconnected while MX, MD, and MN are selected by instantaneous comparison of the AC input phases. When the switching state for output phase "R" is MX, MD, or MN, the output phase "R" is connected to the input phase where the voltage is MX, MD, or MN, respectively. The duty ratio formula is the same as direct duty PWM presented in [25].

The duty ratio of phases S and T is indicated as d_S and d_T and can be derived in the same way of phase R by letting $\bar{v}_{OS}$ and $\bar{v}_{OT}$ be equal to the S and T phase voltage commands v_{oS}^* and v_{oT}^*, respectively. The duty ratio for phases S and T can be presented as

$$d_S = \begin{cases} d_{S1} = \dfrac{v_{oS}^* - \text{MX}}{-\text{MX} + (1-m) \cdot \text{MD} + m \cdot \text{MN}} \\[3mm] d_{S2} = \dfrac{v_{oS}^* - n \cdot \text{MX} + (1-m) \cdot \text{MD}}{-m \cdot \text{MX} + (-1+m) \cdot \text{MD} + \text{MN}} \end{cases}$$

$$d_T = \begin{cases} d_{T1} = \dfrac{v_{oT}^* - \text{MX}}{-\text{MX} + (1-m) \cdot \text{MD} + m \cdot \text{MN}} \\[3mm] d_{T2} = \dfrac{v_{oT}^* - n \cdot \text{MX} + (1-m) \cdot \text{MD}}{-m \cdot \text{MX} + (-1+m) \cdot \text{MD} + \text{MN}} \end{cases}$$

(11)

The voltage command for phase R is the same as (10), and the voltage command for phases S and T can be expressed as follows:

$$v_{oS}^* = -v_{oT}^* = \sqrt{\frac{2}{3}} \cdot q \cdot v_{s\text{-rms}}.$$

(12)

4.3. Operation Mode 3. In this mode, the MDMC acts as double three-phase to single-phase PWM converter. The power is transferred through the six bidirectional switches which are connected to input phases a, b, and c to output phases S and T. The time subinterval for AC switch is equal to switching period ($T_c = T_s$), the input of MDMC is connected to the AC generator, and battery is disconnected from the system. Figure 7 illustrates the circuit configuration of MDMC in operation mode 3 when the bidirectional switches connected to phase R are turned off.

The duty ratio for phases S and T can be calculated in the same way as (11). The voltage command is also the same as (12) for phases S and T. The DC voltage can be used for DC load or charging the battery through the proper charging circuit.

4.4. Operation Mode 4. In this mode, the MDMC acts as single PWM converter. The power is transferred through the two bidirectional switches which are connected to DC input phases P and N to output phase R. The time subinterval for AC switch becomes zero ($n = T_c = 0$), the input
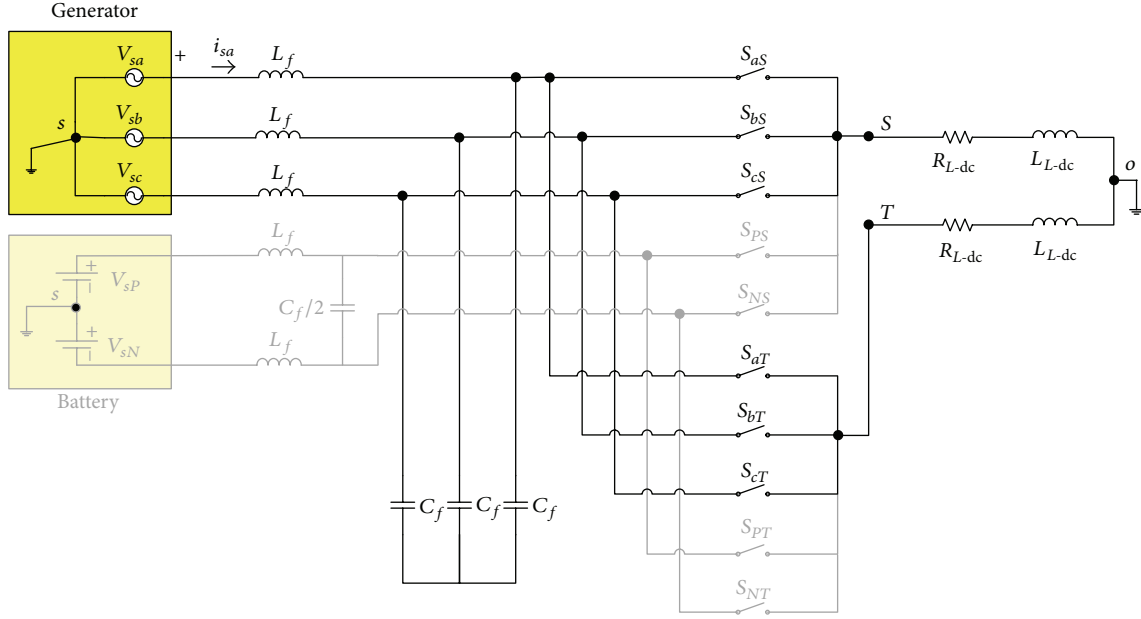
FIGURE 7: Circuit configuration of MDMC in operation mode 3.
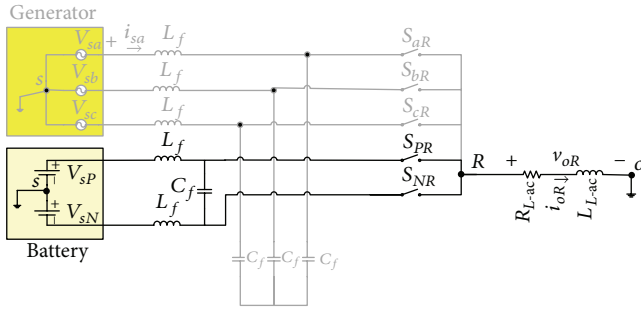
FIGURE 8: Circuit configuration of MDMC in operation mode 4.
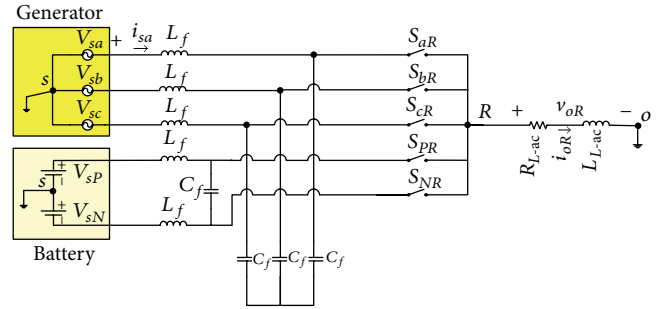
FIGURE 9: Circuit configuration of MDMC in operation mode 5.

of MDMC is connected to the battery, and AC generator is disconnected from the system. Figure 8 demonstrates the circuit configuration of MDMC in operation mode 4 when the bidirectional switches S_{PR} and S_{NR} are connected to phase R.

By putting $n = 0$ in (4) and (6), the duty ratio for operation mode 4 in switching patterns I and II can be expressed as below:

$$d_R = \frac{v_R^* - \text{POS}}{-\text{POS} + \text{NEG}} \qquad (13)$$

when the v_R^* is

$$v_{oR}^* = \frac{1}{2\sqrt{2}} \cdot q \cdot v_{s\text{-}PN} \cdot \sin(2\pi f_o t). \qquad (14)$$

For this operation mode, the voltage ratio can be increased to $1/\sqrt{2} \approx 0.7$ as the input is connected to DC sources.

4.5. Operation Mode 5. In conventional battery based system when the generator power is less than load demands,

the generator is disconnected from system and battery will supply the system. According to the EDDPWM method, the generator and battery can supply the load, simultaneously. Therefore, in this operation mode, the MDMC acts as five-phase to single-phase converter. The power is transferred through the five bidirectional switches which are connected to AC and DC input phases a, b, c, P, and N to output phase R. The duty ratio of phase R is indicated in (4) and (6) for switching patterns I and II, respectively. The voltage command for phase R can be expressed as follows:

$$v_{oR}^* = \frac{1}{2\sqrt{2}} \cdot q \cdot v_{s\text{-}PN} \cdot \sin(2\pi f_o t). \qquad (15)$$

Figure 9 demonstrates the circuit configuration of MDMC in operation mode 5 when the bidirectional switches S_{aR}, S_{bR}, S_{cR}, S_{PR}, and S_{NR} are connected to phase R. For this operation mode, the voltage ratio can be increased to 0.6 as the input is connected to AC and DC sources.

(a) AC output voltage

(b) AC output current
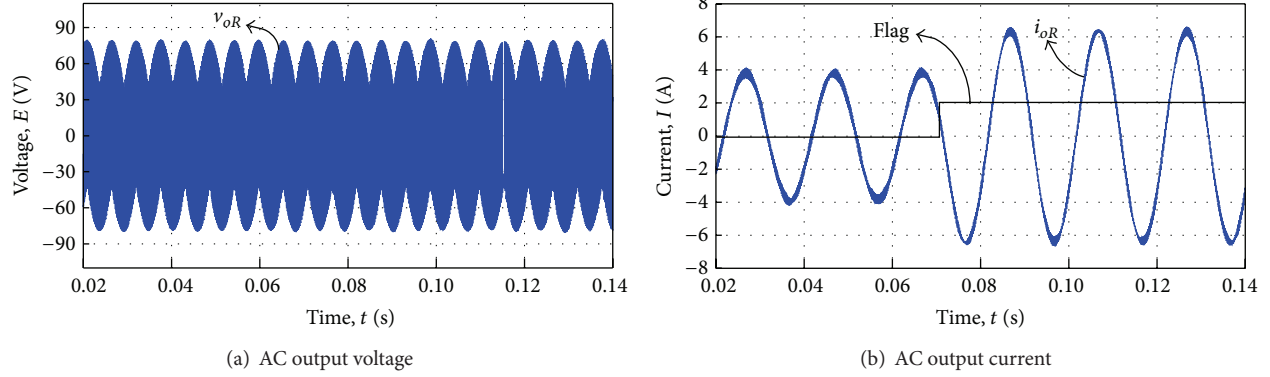
FIGURE 10: Simulation waveforms of operation mode 1.

TABLE 2: Simulation parameter.

Parameter	Value
R-L load	$R = 5\,\Omega$, $L = 10\,\text{mH}$
Input filter inductor L_f	$100\,\mu\text{H}$
Input filter capacitor C_f	$60\,\mu\text{F}$
Input voltage (line-to-neutral) $V_{s\text{-rms}}$	56 V
Battery voltage (line-to-neutral) $V_{s\text{-dc}}$	$\pm 48\,\text{V}$
Input frequency f_s	60 Hz
Output frequency f_o	50 Hz

In this mode, the maximum voltage ratio can change within $0.6 < q < 0.7$ as expressed in (14):

$$q = \begin{cases} 0.7 - \left(\dfrac{k - 0.25}{2.5}\right) & \text{for } k > 0.25, \\[2ex] 0.7 - \left(\dfrac{0.25 - k}{2.5}\right) & \text{for } k < 0.25, \end{cases} \tag{16}$$

where $k = v_{s\text{-rms}}/(v_{s\text{-rms}} + v_{sP})$ indicates the magnitude variation between line-to-neutral RMS value of input AC and DC power supplies and $v_{s\text{-rms}}$ and v_{sP} indicate the RMS value of generator input voltage and DC power supply, respectively.

5. Results and Discussion

5.1. Simulation Result. Simulation of the EDDPWM method for MDMC is performed by using MATLAB software. The voltage ratio has been changed in each operation mode to investigate the stability of system. The switching period T_s is assumed to be $200\,\mu$s in all operation modes. The simulation parameters shown in Table 2 are the same for all operation modes.

The input line-to-neutral input voltage ($V_{s\text{-rms}}$) in operation modes 1, 2, and 3 is bigger than the battery voltage ($V_{s\text{-dc}}$) as shown in Table 2. The AC voltage $V_{s\text{-rms}}$ is less than $V_{s\text{-dc}}$ in operation modes 4 and 5 and it is equal to 35 V.

Figure 10 indicates the AC output voltage and current waveforms of proposed MDMC in operation mode 1, respectively. In this mode the voltage ratio (q) has been increased from 0.3 to 0.5 at time $t = 0.07$ (s). The simulation result indicates that the MDMC is able to reach the maximum voltage ratio ($q = 0.5$) in operation mode 1 without any distortion in output voltage or output current waveforms.

Figures 11(a) and 11(b) illustrate the line-to-line DC output voltage of v_{oST} and line-to-neutral and AC output voltage of v_{oR}. Figure 11(c) shows the simulated responses of MDMC in operation mode 2 when the voltage ratio is changed in AC and DC side at $t = 0.07$ and $t = 0.09$ s, respectively. According to Figure 11(c), the current in i_{oS} and i_{oT} is constant at $t = 0.07$ regardless of the changing in i_{oR} which is increased by 0.4 pu. In addition, when the voltage ratio for DC phase is reduced from 0.5 to 0.3, the current in AC side remains constant at $t = 0.09$ s. The simulation result validates that the proposed EDDPWM is able to track the variation in reference control signal for each phase without disturbing the signal in other output terminals.

Figure 12 indicates the DC output voltage and current waveforms of proposed MDMC in operation mode 3, respectively. In this mode the voltage ratio (q) has been reduced from 0.5 to 0.3 at time $t = 0.09$ (s). The simulation result reveals that the MDMC is able to track the control signal in terms of waveforms, frequency, and amplitude regardless of the type and the number of outputs connected to the system.

Figure 13 illustrates the line-to-neutral AC load voltage and current waveforms in operation mode 3, respectively. According to Figure 13 the voltage ratio can reach 0.7 in operation mode 4, when the AC power supply is disconnected from the system. The simulation result reveals that the output voltage has been synthesized well with maximum voltage ratio of 0.7.

Meanwhile, the maximum voltage ratio is equal to 0.5 for operation modes 1, 2, and 3 when the line-to-neutral voltage of AC generator $v_{s\text{-rms}}$ is bigger than battery voltage $v_{s\text{-dc}}$.

In conventional battery based system when the voltage of generator is less than battery voltage, AC power supply will be disconnected from system and demand power will be supplied by battery bank. By using the EDDPWM for MDMC, it is possible to inject the power from AC and DC

(a) DC output voltage

(b) AC output voltage

(c) Output current waveforms
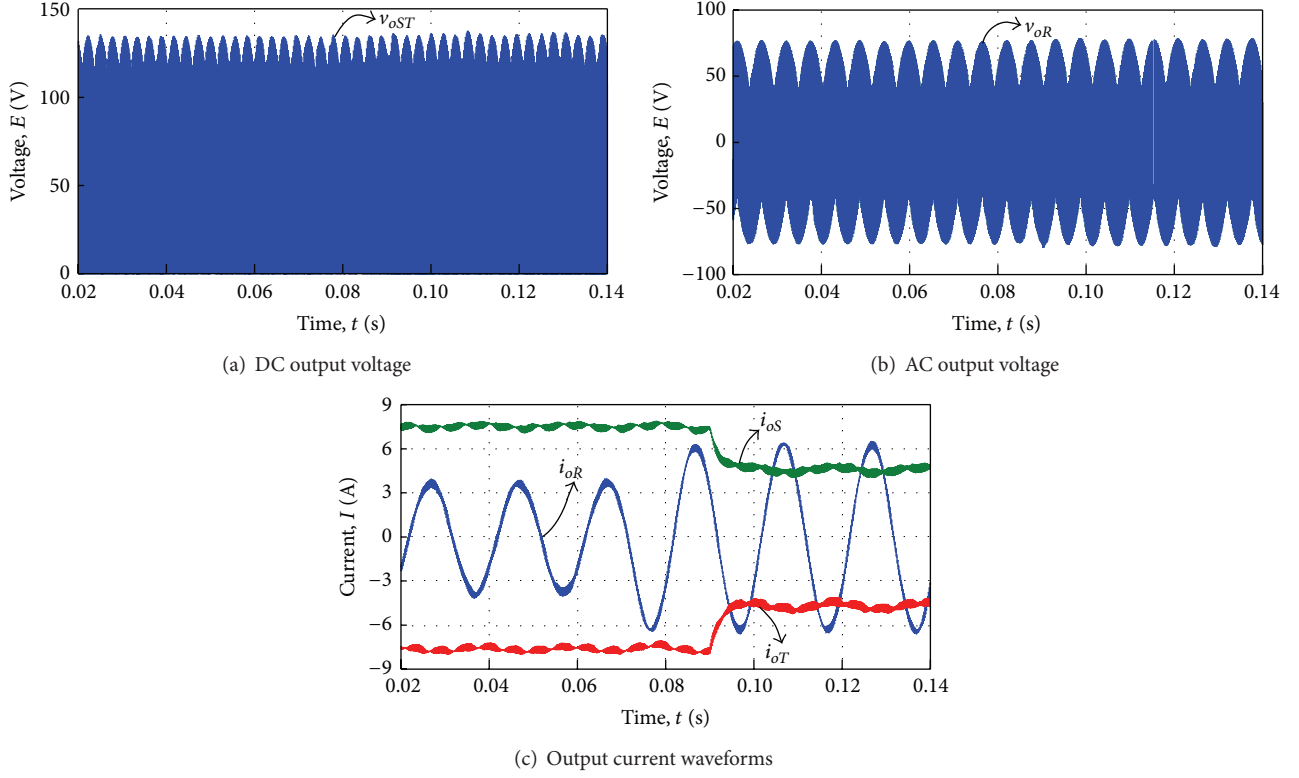
FIGURE 11: Simulation waveforms of operation mode 2.

(a) DC output voltage
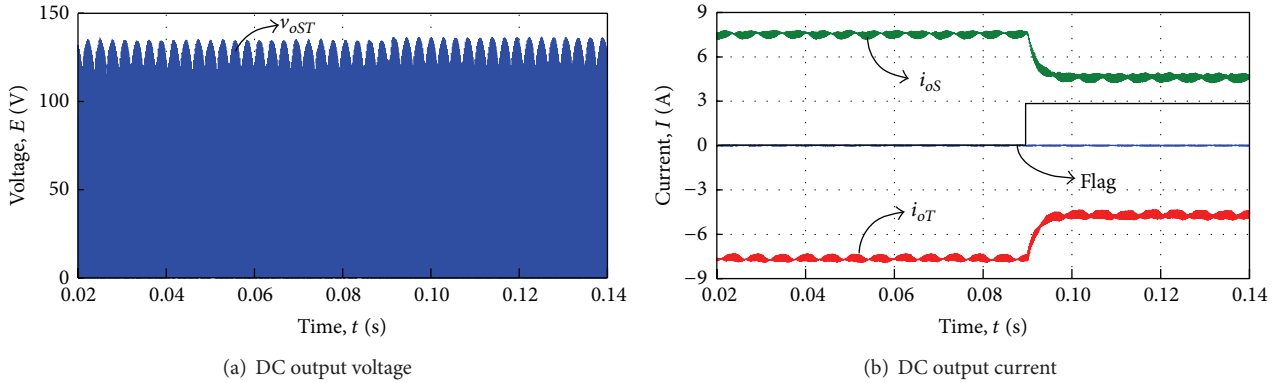
(b) DC output current

FIGURE 12: Simulation waveforms of operation mode 3.

power supply and reach the maximum voltage ratio that is expressed in (12).

Figure 14 indicates the AC output voltage and current waveforms of proposed MDMC in operation mode 5, respectively. In this mode the voltage ratio (q) has been increased from 0.4 to 0.63 at time $t = 0.07$ (s).

It can be clearly seen from Figure 14 that the system is able to track the variation of reference voltage output of terminal R. Moreover, the simulation results exhibited that the undershoot/overshoot and steady-state error for output currents are acceptable in all operation modes.

5.2. Experimental Result.
To verify the feasibility of the novel EDDPWM method for the proposed MDMC, an experimental setup was built and the EDDPWM controller was implemented using Xilinx Virtex-6 FPGA DSP development kit. Figure 15 shows the experimental setup built in laboratory.

In order to test the stability of the system, reference voltage is changed in AC and DC sides while the loads are constant. Table 3 indicates the experimental parameters of the proposed system when reference voltage is changed. The output voltage variation is not visible in voltage waveform as the switching frequency is high in this modulation method.

(a) AC output voltage

(b) AC output current
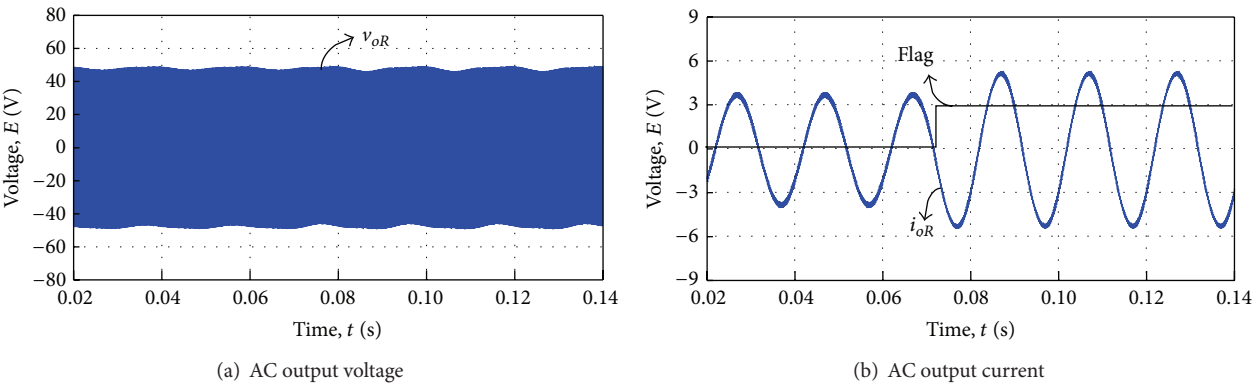
FIGURE 13: Simulation waveforms of operation mode 4.

(a) AC output voltage

(b) AC output current
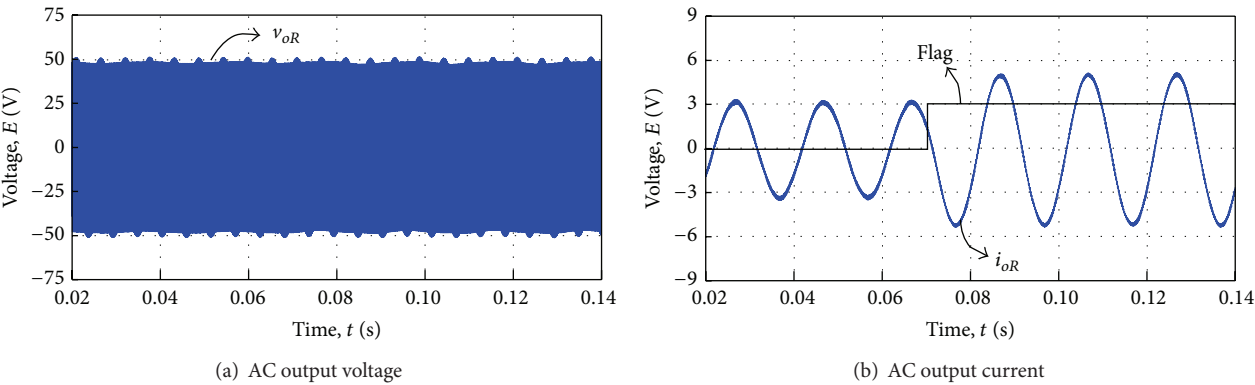
FIGURE 14: Simulation waveforms of operation mode 5.
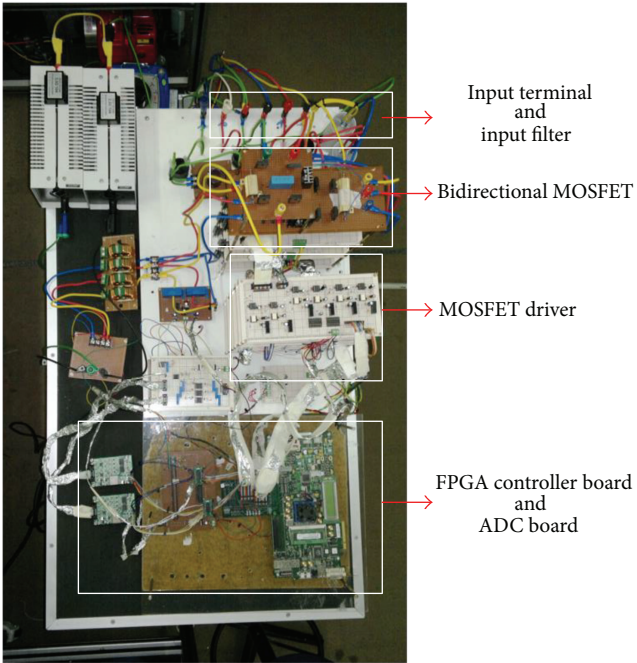
FIGURE 15: Experimental setup of multidirectional matrix converter.

(a) When reference current is reduced by 0.4 pu in phase R

(b) When reference current is reduced by 0.4 pu in phase S
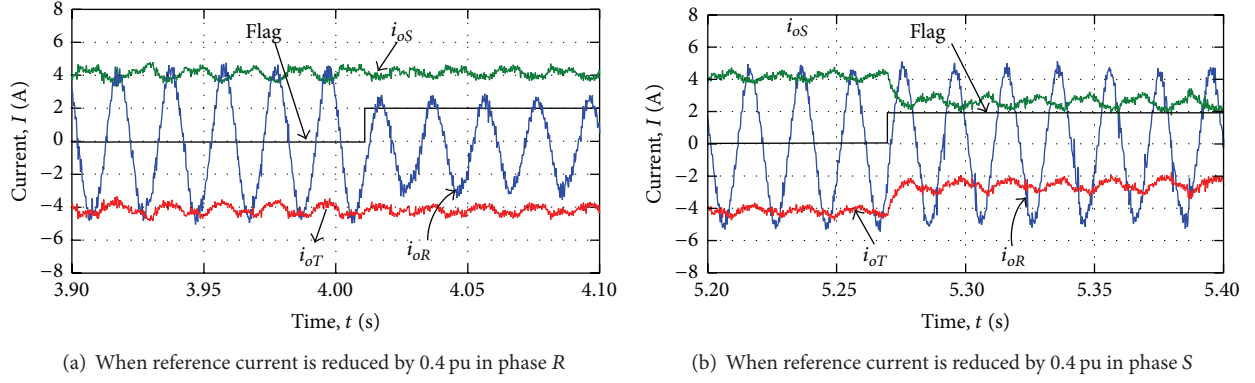
FIGURE 16: The dynamic state of experimental waveform: current reference variation.

TABLE 3: Dynamic state experimental parameter and reference voltage variation.

Parameter	Value
R-L load	$R = 3.1\,\Omega$, $L = 6.4\,$mH
AC input voltage (line-to-neutral RMS) $V_{s\text{-RMS}}$	30 V
DC input voltage $V_{s\text{-dc}}$	±24 V
Input frequency f_s	60 Hz
Output frequency f_o	50 Hz
Switching frequency f_{sw}	5 KHz
Switching period T_s	200 μs

At constant load the current waveform increases or decreases in accordance with sudden change of reference voltage variation. Figure 16 shows the responses of MDMC when the reference voltage is changed in AC and DC side at $t = 4.01$ s and $t = 5.27$ s, respectively, in operation mode 2. According to Figure 16(a), the output current i_{oS} remains constant at $t = 4.01$ s regardless of the changing in i_{oR} which results in voltage reduction of 0.4 pu, where the step change of Flag signal indicates the instant change of the reference voltage. Furthermore, when the DC reference voltage is reduced by 0.4 pu, as shown in Figure 16(b), the current in AC side remains constant at $t = 5.27$ s.

It can clearly be seen from Figure 16 that the system is able to track the variation of reference voltage in each phase separately. The experimental result validates that the reference control signal of each output terminal is independent of other output terminals in terms of frequency, waveform shape, and amplitude. Moreover, the experimental results exhibited that the undershoot/overshoot and dynamic-state error for current waveforms are acceptable for SABBS.

5.3. Total Harmonic Distortion. While there is no national standard dictating total harmonic distortion limits on systems, there are suggested values for acceptable harmonic distortion. IEEE-519 provides recommended harmonic values for power electronic systems. The allowable THD in the input

current of the power converter generally is restricted between 10 and 30% depending on the grid impedance limits [41, 42].

According to IEEE-519 standards, the maximum THD in the grid current cannot exceed 5% for THD and 3% for any single harmonic. It is important to note that the recommendations and values given in this standard are purely voluntary. However, keeping low THD values on a system will further ensure proper operation of equipment and a longer equipment life span.

Table 4 compares the DC output current ripple and AC output current THD of simulation and experimental results for MDMC in different operation modes. The results indicated that the THD is different in each operation mode as the number of switches, type of input/output, and input voltage and frequency have been varied in each operation mode. Furthermore, the nonideal characteristic of component, total propagation delay of system, inherent noise and distortion of input power supply, and finite switching times in practical could be the reason for THD gap between the experimental and simulation results. In addition, both generator and DC power supply in experimental setup have some noise and distortion which have direct effect on system performance and output waveform quality.

According to Table 4 and waveform result, the highest THD and ripple appear in operation mode 2 as the AC and DC appear in output phase together. The THD in operation mode 5 is less than other operation modes since 5 bidirectional switches are used to supply the load. In addition, in operation mode 5 the AC voltage is less than the DC voltage and input current has been synthesized by adjusting two variables, n and m. Therefore, the input current distortion is compensated by proper switching between the two input power supplies.

6. Conclusion

This paper presents a systematic approach for MDMC based on the power flow direction among the input and output ports to enhance the performance of stand-alone battery based system and to achieve the continuous power transmission. According to the proposed power flow strategy, the connected

TABLE 4: Experimental and simulation distortions comparison.

Parameter	Operation mode 1	Operation mode 2	Operation mode 3	Operation mode 4	Operation mode 5
Simulation THD	7.22%	7.29%	—	5.7%	3.59%
Simulation DC current ripple	—	4.1%	3.9%	—	—
Experimental THD	14.32%	14.54%	—	13.95%	13.77%
Experimental DC current ripple	—	11.2%	8.7%	—	—

input and output terminals with corresponding bidirectional switches have been determined to supply the power demand with minimum number of switches. The maximum possible voltage ratio has been calculated in each operation mode. The result validates that the MDMC can work as inverter, PWM converter, or PWM converter and inverter by adjusting the time subinterval in EDDPWM method. In addition, the proposed EDDPWM was able to change the function of the MDMC from inverter to rectifier, rectifier to inverter, rectifier and inverter to inverter, or rectifier based on the operation modes of system.

Particularly, this study exhibited that the maximum voltage ratio is achievable for each operation mode when the undershoot/overshoot and steady-state error for output currents are acceptable in all operation modes. In addition, the quality of output currents has been analysed in terms of THD and ripples.

Based on the literature and result of the present study, as the proposed converter switched from inverter to rectifier in different operation mode, an active power filter can improve the output waveform quality and reduce the noise in input current.

In addition, the modulation control method can be applied to the MDMC as a modular structure for each phase where each output phase has the independent reference control signal. Therefore, the proposed MDMC structure can be used for programmable power supply and extended to the hybrid system with higher number of input and output phases.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] S. Malo and R. Griñó, "Design, construction, and control of a stand-alone energy-conditioning system for PEM-type fuel cells," *IEEE Transactions on Power Electronics*, vol. 25, no. 10, pp. 2496–2506, 2010.

[2] Y.-M. Chen, Y.-C. Liu, and F.-Y. Wu, "Multi-input DC/DC converter based on the multiwinding transformer for renewable energy applications," *IEEE Transactions on Industry Applications*, vol. 38, no. 4, pp. 1096–1104, 2002.

[3] L. Solero, A. Lidozzi, and J. A. Pomilio, "Design of multiple-input power converter for hybrid vehicles," *IEEE Transactions on Power Electronics*, vol. 20, no. 5, pp. 1007–1016, 2005.

[4] A. Iqbal, S. M. Ahmed, and H. Abu-Rub, "Space vector PWM technique for a three-to-five-phase matrix converter," *IEEE*

Transactions on Industry Applications, vol. 48, no. 2, pp. 697–707, 2012.

[5] L. Solero, F. Caricchi, F. Crescimbini, O. Honorati, and F. Mezzetti, "Performance of A 10 kW power electronic interface for combined wind/PV isolated generating systems," in *Proceedings of the 1996 27th Annual IEEE Power Electronics Specialists Conference (PESC '96)*, pp. 1027–1032, January 1996.

[6] Y.-M. Chen, Y.-C. Liu, F.-Y. Wu, and T.-F. Wu, "Multi-input DC/DC converter based on the flux additivity," in *Proceedings of the 36th IAS Annual Meeting. Conference Record of the IEEE Industry Applications Conference*, vol. 3, pp. 1866–1873, IEEE, Chicago, Ill, USA, September-October 2001.

[7] Q. Mei, W.-Y. Wu, and Z.-L. Xu, "A multi-directional power converter for a hybrid renewable energy distributed generation system with battery storage," in *Proceedings of the 5th CES/IEEE International Power Electronics and Motion Control Conference (IPEMC '06)*, pp. 1–5, IEEE, Shanghai, China, August 2006.

[8] D. Liu and H. Li, "A ZVS Bi-directional DC-DC converter for multiple energy storage elements," *IEEE Transactions on Power Electronics*, vol. 21, no. 5, pp. 1513–1517, 2006.

[9] Z. Qian, O. Abdel-Rahman, and I. Batarseh, "An integrated four-port DC/DC converter for renewable energy applications," *IEEE Transactions on Power Electronics*, vol. 25, no. 7, pp. 1877–1887, 2010.

[10] P. K. Goel, B. Singh, S. S. Murthy, and N. Kishore, "Isolated wind-hydro hybrid system using cage generators and battery storage," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 4, pp. 1141–1153, 2011.

[11] M. Venturini and A. Alesina, "The generalised transformer—a new bidirectional sinusoidal waveform frequency converter with continuously adjustable input power factor," in *Proceedings of the Power Electronics Specialists Conference (PESC '80)*, pp. 242–252, 1980.

[12] A. Alesina and M. G. B. Venturini, "Analysis and design of optimum-amplitude nine-switch direct AC-AC converters," *IEEE Transactions on Power Electronics*, vol. 4, no. 1, pp. 101–112, 1989.

[13] J. Rodriguez, "A new control technique for AC-AC converters," in *Proceedings of the IFAC Control in Power Electronics and Electrical Drives Conference*, pp. 203–208, 1983.

[14] P. Ziogas, S. I. Khan, and M. Rashid, "Analysis and design of forced commutated cycloconverter structures with improved transfer characteristics," in *Proceedings of the 16th Annual Power Electronics Specialists Conference (PESC '85)*, Record (A86-40426 19-33), pp. 610–622, Institute of Electrical and Electronics Engineers, Toulouse, France, June 1985.

[15] P. D. Ziogas, S. I. Khan, and M. H. Rashid, "Some improved forced commutated cycloconverter structures," *IEEE Transactions on Industry Applications*, vol. 21, no. 5, pp. 1242–1253, 1985.

[16] G. Kastner and J. Rodriguez, "A forced commutated cyclo-converter with control of the source and load currents," in *Proceedings of the European Conference on Power Electronics and Applications (EPE '85)*, pp. 1141–1146, 1985.

[17] L. Huber and D. Borojevic, "Space vector modulator for forced commutated cycloconverters," in *Proceedings of the Conference Record of the IEEE Industry Applications Society Annual Meeting*, vol. 1, pp. 871–876, San Diego, Calif, USA, October 1989.

[18] L. Huber and D. Borojevic, "Space vector modulated three-phase to three-phase matrix converter with input power factor correction," *IEEE Transactions on Industry Applications*, vol. 31, no. 6, pp. 1234–1246, 1995.

[19] L. Huber, D. Borojevic, and N. Burany, "Digital implementation of the space vector modulator for forced commutated cyclo-converters," in *Proceedings of the 4th International Conference on Power Electronics and Variable-Speed Drives*, pp. 63–68, July 1990.

[20] L. Huber, D. Borojevic, X. F. Zhuang, and F. C. Lee, "Design and implementation of a three-phase to three-phase matrix converter with input power factor correction," in *Proceedings of the 8th Annual Applied Power Electronics Conference and Exposition (APEC '93)*, pp. 860–865, IEEE, San Diego, Calif, USA, March 1993.

[21] L. Huber, N. Burány, and D. Borojević, "Analysis, design and implementation of the space-vector modulator for forced-commutated cycloconvertors," *IEE Proceedings B: Electric Power Applications*, vol. 139, no. 2, pp. 103–113, 1992.

[22] L. Wei and T. A. Lipo, "A novel matrix converter topology with simple commutation," in *Proceedings of the 36th IAS Annual Meeting—Conference Record of the Industry Applications*, pp. 1749–1754, October 2001.

[23] L. Wei, T. A. Lipo, and H. Chan, "Matrix converter topologies with reduced number of switches," in *Proceedings of the IEEE 33rd Annual Power Electronics Specialists Conference (PESC '02)*, pp. 57–63, June 2002.

[24] Y.-D. Yoon and S.-K. Sul, "Carrier-based modulation technique for matrix converter," *IEEE Transactions on Power Electronics*, vol. 21, no. 6, pp. 1691–1703, 2006.

[25] Y. Li, N.-S. Choi, B.-M. Han, K. M. Kim, B. Lee, and J.-H. Park, "Direct duty ratio pulse width modulation method for matrix converters," *International Journal of Control, Automation and Systems*, vol. 6, no. 5, pp. 660–669, 2008.

[26] Y. Li and N.-S. Choi, "Carrier based pulse width modulation for matrix converters," in *Proceedings of the 24th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '09)*, pp. 1709–1715, 2009.

[27] S. Toosi, N. Misron, T. Hanamoto, I. B. Aris, M. A. Mohd Radzi, and H. Yamada, "Novel modulation method for multidirectional matrix converter," *The Scientific World Journal*, vol. 2014, Article ID 645734, 12 pages, 2014.

[28] N. Yamamura, M. Ishida, and T. Hori, "A simple wind power generating system with permanent magnet type synchronous generator," in *Proceedings of the 3rd IEEE International Conference on Power Electronics and Drive Systems (PEDS '99)*, pp. 849–854, July 1999.

[29] S. Sato and A. Kawamura, "A new estimation method of state of charge using terminal voltage and internal resistance for lead acid battery," in *Proceedings of the Power Conversion Conference (PCC-Osaka '02)*, vol. 2, pp. 565–570, IEEE, Osaka, Japan, 2002.

[30] F. Huet, "A review of impedance measurements for determination of the state-of-charge or state-of-health of secondary batteries," *Journal of Power Sources*, vol. 70, no. 1, pp. 59–69, 1998.

[31] S. Rodrigues, N. Munichandraiah, and A. K. Shukla, "Review of state-of-charge indication of batteries by means of a.c. impedance measurements," *Journal of Power Sources*, vol. 87, no. 1, pp. 12–20, 2000.

[32] O. Barbarisi, F. Vasca, and L. Glielmo, "State of charge Kalman filter estimator for automotive batteries," *Control Engineering Practice*, vol. 14, no. 3, pp. 267–275, 2006.

[33] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[34] C. C. Chan, E. W. C. Lo, and S. Weixiang, "The available capacity computation model based on artificial neural network for lead-acid batteries in electric vehicles," *Journal of Power Sources*, vol. 87, no. 1, pp. 201–204, 2000.

[35] A. J. Salkind, C. Fennie, P. Singh, T. Atwater, and D. E. Reisner, "Determination of state-of-charge and state-of-health of batteries by fuzzy logic methodology," *Journal of Power Sources*, vol. 80, no. 1, pp. 293–300, 1999.

[36] S. Piller, M. Perrin, and A. Jossen, "Methods for state-of-charge determination and their applications," *Journal of Power Sources*, vol. 96, no. 1, pp. 113–120, 2001.

[37] L. C. Herrero, S. de Pablo, F. Martín, J. M. Ruiz, J. M. González, and A. B. Rey, "Comparative analysis of the techniques of current commutation in matrix converters," in *Proceedings of the IEEE International Symposium on Industrial Electronics (ISIE '07)*, pp. 521–526, June 2007.

[38] P. W. Wheeler, J. Rodríguez, J. C. Clare, L. Empringham, and A. Weinstein, "Matrix converters: a technology review," *IEEE Transactions on Industrial Electronics*, vol. 49, no. 2, pp. 276–288, 2002.

[39] C. L. Neft, "AC power supplied static switching apparatus having energy recovery capability," Google Patents, 1987.

[40] N. Nguyen-Quang, D. A. Stone, C. M. Bingham, and M. P. Foster, "A three-phase to single-phase matrix converter for high-frequency induction heating," in *Proceedings of the 13th European Conference on Power Electronics and Applications (EPE '09)*, pp. 1–10, September 2009.

[41] A. K. Sahoo, A. Shahani, K. Basu, and N. Mohan, "LCL filter design for grid-connected inverters by analytical estimation of PWM ripple voltage," in *Proceedings of the 29th Annual IEEE Applied Power Electronics Conference and Exposition (APEC '14)*, pp. 1281–1286, March 2014.

[42] A. K. Sahoo, K. Basu, and N. Mohan, "Comparison of filter components of back-to-back and matrix converter by analytical estimation of ripple quantities," in *Proceedings of the 39th Annual Conference of the IEEE Industrial Electronics Society (IECON '13)*, pp. 4831–4837, November 2013.

Research Article

Small-Signal Modeling of Marine Electromagnetic Detection Transmitter Controlled-Source Circuit

Haijun Tao,[1,2] Yiming Zhang,[1] and Xiguo Ren[1]

[1]*College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China*
[2]*School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo, Henan 454003, China*

Correspondence should be addressed to Haijun Tao; taohj99@hpu.edu.cn

Marine electromagnetic transmitter transmits electromagnetic waves with large power frequency conversion to the seabed to obtain the submarine structure and mineral resources. However, the current transmitter presents several problems, such as low efficiency, serious heat, and poor adaptability to the load. Soft-switching controlled-source circuit is used to reduce circuit losses. The mathematical model of controlled-source circuit should be established to realize a closed-loop control for increasing the output transient performance of electromagnetic waves. Given that the soft-switching controlled-source circuit has more status and that direct modeling is difficult, small-signal model of soft-switching controlled-source circuit is established based on that of hard-switching controlled-source circuit by analyzing the effect of output filter inductor current transformer leakage inductance and input voltage soft-switching controlled circuit on change in the duty cycle. Finally, experiments verify the accuracy and validity of the model.

1. Introduction

The exploration and exploitation of marine resources have high degrees of drilling and investment risks. To improve the success rate of drilling, the major oil companies in the world constantly conduct a variety of exploration activities (seismic, gravity, magnetic, etc.) prior to well drilling. Marine controlled-source electromagnetic detection has become an important approach for reducing the risk in deep-water drilling oil and gas resources by distinguishing between oil and water traps nature, which, in turn, exposes "the high impedance body" under the coverage [1]. Marine controlled-source electromagnetic detection system supplies the transmitter on the seabed by tugs. The multicomponent electromagnetic receiver laid on the seabed measures electromagnetic field values. Calculating the apparent resistivity and phase or directly using the electric and magnetic fields detects the distribution characteristics of underground electrical current. Such a method can reveal the distribution of the underlying structure, as well as oil, gas, and other mineral resources [2].

Marine electromagnetic survey techniques and instruments have been developed since the 1970s. The chief organizations in this field include the United States Scripps Institution of Oceanography, the German Leibniz Institute of Marine Sciences, the British Association of Subsea Equipment, and Japanese TIERRA companies [3]. Meanwhile, the major marine electromagnetic detection service companies in the world, such as the Norwegian company Electro Magnetic Geo Services, the British company Offshore Hydrocarbons Mapping, AGO company in the United States, and MTEM company in the UK, have conducted numerous marine electromagnetic detection projects [4].

Marine electromagnetic transmitter has a large size and weight, low efficiency, and poor transient waveform emissions; these properties hinder the fulfillment of the needs of the oil exploration industry practice [2]. Based on conventional hard-switching full-bridge DC/DC controlled-source circuit, a soft-switching controlled-source circuit is designed. The control circuit must be designed accordingly to obtain good dynamic and static output characteristics [5, 6]. The design of the control loop is closely related to the main circuit
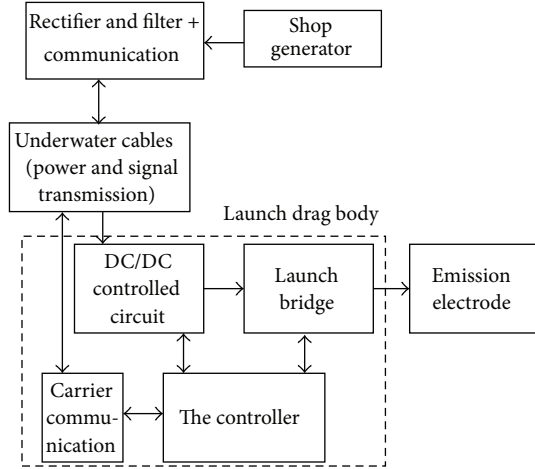
FIGURE 1: Diagram of marine electromagnetic transmitter.

FIGURE 2: Controllable source circuit schematics.

structure and parameters [7]. Therefore, a mathematical model of controlled-source circuit must be established prior to designing the control system. However, the soft-switching controlled-source circuit has 12 types of working conditions in a duty cycle [8–10], thus causing difficulty in directly obtaining the transfer function.

In this paper, the ideal, state-space averaged, and small-signal models of hard-switching controlled-source circuit are established. On this basis, a small-signal model of soft-switching controlled-source circuit is established by analyzing the effect of the output filter inductor current transformer leakage inductance and input voltage source circuit on soft-switching controllable duty cycle. Finally, the accuracy and validity of the model are verified through an experiment.

2. Operation of Marine Electromagnetic Detection Transmitter

The operation of a marine controlled-source electromagnetic detection system is depicted in Figure 1. The system chiefly includes shipboard diesel generators, deck-side rectifier filter circuits, deck-side PC monitor unit, underwater streamers, full-bridge DC/DC converter circuits, launch circuits, emitter electrode, control unit of the transmission system, and carrier communication.

The shipboard generator provides the initial energy for the entire controlled-source electromagnetic detection system. The deck-side rectifier filter circuits rectify the three-phase AC voltage generated by the generator into a direct current, reducing the transmission loss through shipboard towing. After the electromagnetic detection transmitter gets the seabed, the deck-side PC monitor unit can establish remote communications between the underwater streamer and the control unit of the transmitter. Moreover, the deck-side PC monitor unit has a time service function for the entire system. The ship underwater streamers connect the ship and the transmitter for power and signal transmission. Underwater DC/DC converter circuits transform the high-voltage direct current by the underwater streamers to
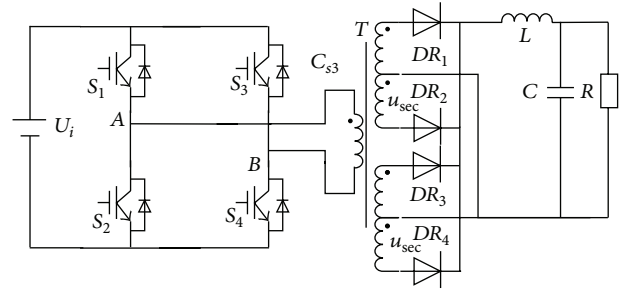
a controllable DC current. The launch circuits subsequently reverse this controllable DC current into a frequency tunable square wave AC current. Ultimately, the final current would be transmitted by the emitter electrode to the seawater. The control unit of the transmission system with dual DSP could control the DC-DC converter circuits and the launch circuits and could detect the transmitter state. The underwater carrier communication (i.e., two-way fiber optic communication) establishes a communication connection with the transmitter and PC monitor unit.

The DC/DC controlled-source transmitter circuit is a key part of an ocean observation system, and its dynamic characteristics and efficiency directly affect the performance of the entire transmitter. Output voltage and current control accuracy and circuit efficiency could be improved using soft-switching controlled-source circuit. However, this circuit requires the use of a closed-loop control system. Therefore, a controlled-source circuit modeling is necessary. However, this circuit requires the use of a closed-loop control system. Thus, controlled-source circuit modeling is necessary, but the soft-switching controlled-source circuit has 12 states for one cycle, and direct modeling is difficult. The processes of hard-switching and soft-switching controlled-source circuits are initially analyzed to identify the differences and similarities between the two circuits. A soft-switching small-signal model of marine transmitter controlled-source circuit is subsequently obtained based on the established hard-switching controlled source of small-signal circuit model.

3. Operation Process Analysis of Controlled-Source Circuit

3.1. Controlled-Source Circuit Structure. As shown in Figure 2, the controlled-source circuit structure includes an inverter bridge consisting of four IGBTs, a high-frequency transformer with a center tap, and a high-frequency rectifier filter. The major difference between hard-switching and soft-switching controlled-source circuits is the value of the transformer leakage inductance L_r. The transformer leakage inductance L_r in hard-switching controlled-source circuit is extremely small, which can be negligible, to reduce the value of impulse voltage of the primary side of the transformer at IGBT turning-off time. For soft-switching circuit, transformer leakage inductance is large to provide bridge arm capacitors with sufficient energy to achieve IGBT ZVS during
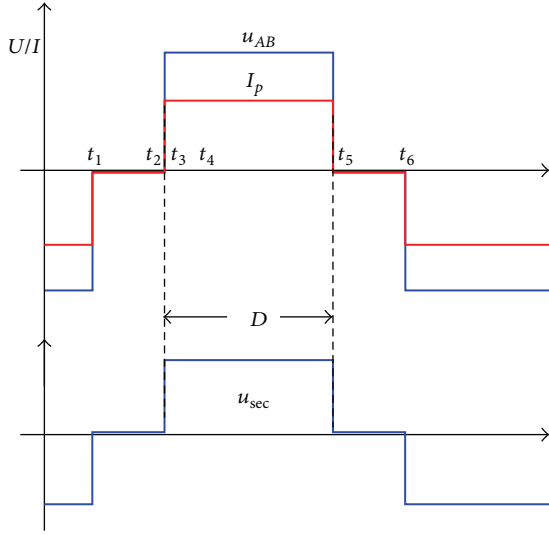
FIGURE 3: Main waveform of hard-switching controlled-source circuit.
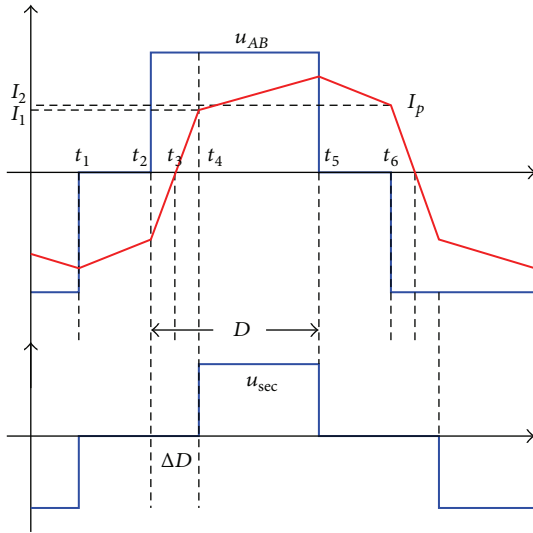
FIGURE 4: Main waveform of soft-switching controlled-source circuit.

the freewheeling period. In terms of control mode, hard-switching circuit uses bipolar control, whereas soft-switching circuit uses phase-shifted control.

3.2. Operation Process Analysis. Primary voltage and current and secondary voltage waveform of high-frequency transformer for hard-switching controlled-source circuit and soft-switching controlled-source circuit are shown in Figures 3 and 4, respectively. These figures show that the voltage waveform type of these forms of circuit in a given duty ratio control is approximation, and the difference is the duty cycle loss because of leakage inductance in the soft-switching controlled-source circuit.

Two equivalent circuits shown in Figure 5 can be obtained through the analysis of controlled-source circuit from

the output. These circuits correspond to operation states when power flows from the primary side of transformer transfer to secondary side and when the secondary side is freewheeling. Thus, a mathematical model of soft-switching controlled-source circuit can be derived by obtaining the mathematical model of hard-switching controlled-source circuit.

4. Hard-Switching Controlled-Source Circuit Modeling

4.1. Ideal Switch Model. According to the equivalent switching states illustrated in Figure 5 and Kirchhoff's voltage and current law, the equation of state of the circuit can be obtained

$$\dot{x} = A_1 x + B_1 u,$$

$$y = C_1 x,$$

$$t \in [t_i, t_i + DT_s],$$

$$(1)$$

$$\dot{x} = A_2 x + B_2 u,$$

$$y = C_2 x,$$

$$t \in [t_i + DT_s, t_{i+1}],$$

where

$$A_1 = \begin{bmatrix} 0 & -\dfrac{1}{L} \\ \dfrac{1}{C} & -\dfrac{1}{RC} \end{bmatrix},$$

$$B_1 = \begin{bmatrix} \dfrac{1}{nL} \\ 0 \end{bmatrix},$$

$$C_1 = \begin{bmatrix} \dfrac{1}{n} & 0 \\ 0 & 1 \end{bmatrix},$$

$$A_2 = \begin{bmatrix} 0 & -\dfrac{1}{L} \\ \dfrac{1}{C} & -\dfrac{1}{RC} \end{bmatrix},$$

$$(2)$$

$$B_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$x = \begin{bmatrix} i_L \\ u_C \end{bmatrix},$$

$$y = \begin{bmatrix} i_i \\ u_o \end{bmatrix},$$

$$u = [u_i].$$
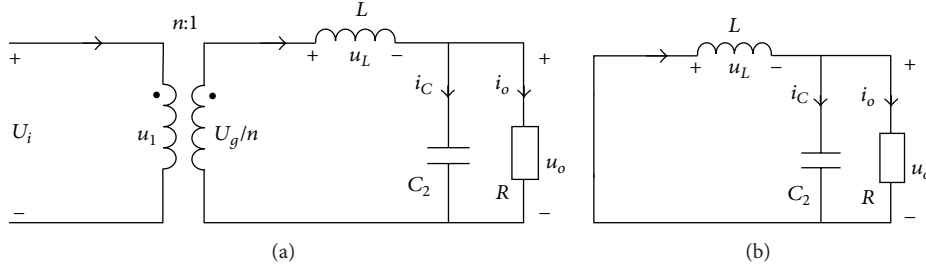
Figure 5: Equivalent switching state.

Ideal switch model is very close to the characteristics of the actual circuit. The results using this model for analysis are in accordance with the actual situation. However, ideal switch model is a typical time-varying system. If the duty cycle is as an input variable, then the product item of the input variable and the input variable u exist. Thus, the system is nonlinear. Obtaining an analytical solution is difficult for nonlinear time-varying systems. Hence, an ideal switch model needs to eliminate the time-varying characteristics to obtain analytical solutions.

4.2. State-Space Averaged Model.
Ideal switch model is time-varying, but its topology and state equation are determined to be time-invariant when the switches are conducting and is turned off. According to the circuit schematic and (1), A_1, B_1, and C_1 apply for the first (on) interval, or during D of the switching time, while A_2, B_2, and C_2 exist during the $1 - D$ (off) switching time interval. The system average approximate state equation in a switching cycle can then be obtained

$$\dot{x} = Ax + Bu,$$
$$y = Cx, \tag{3}$$

where $A = DA_1 + (1 - D)A_2$, $B = DB_1 + (1 - D)B_2$, $C = DC_1 + (1 - D)C_2$.

The model described by (3) is the system state-space average model. The average state-space model is time-invariant and can simplify the task of obtaining the analytical solution. This model is important and effective for control system analysis and design of the controlled-source circuit. The following aspects must be considered:

(1) Compared with solutions derived from ideal switch model, the solutions derived from state-space average model allow greater approximation. In addition, fluctuations produced by state variables, such as the inductor current, and capacitance voltage with the switch turning-on and turning-off did not reflect in the solutions of state-space average model.

(2) State-space average model is only applicable to lower switch frequency range of 1/5–1/20 of the switching frequency, and the result would be meaningless if the frequency involved in the analysis process is close to or greater than the switching frequency.

4.3. Small-Signal Model.
The control circuit controls controlled-source circuit by adjusting the duty cycle D. In this case, the duty cycle D is an input variable quantity of controlled-source circuit but changes over time. Traditionally, d is used and D represents a fixed duty cycle. In the case that a duty cycle is used as an input quantity, the state-space average model is no longer linear. This phenomenon is due to the presence of coupling between the state variables and the control quantity, such as in (3), where the control amount d and the system input amount u multiply. Solving a local linear system, which has been a small-signal model, is usually necessary for conducting system analysis and design.

In state-space averaged model, the state equation of controlled-source circuit can be represented as a unified form as follows [11, 12]:

$$\dot{x} = F(x, u, d). \tag{4}$$

Assuming that the static operating point of the circuit is (x_0, u_0, d_0), the right side of (4) is expanded as a Taylor series in the vicinity of operation point, order $\hat{x} = x - x_0$, $\hat{u} = u - u_0$, $\hat{d} = d - d_0$, and higher-order infinite events are ignored

$$\dot{\hat{x}} = \frac{\partial F(x_0, u_0, d_0)}{\partial x}\hat{x} + \frac{\partial F(x_0, u_0, d_0)}{\partial u}\hat{u}$$
$$+ \frac{\partial F(x_0, u_0, d_0)}{\partial d}\hat{d}. \tag{5}$$

In the above equation, to allow $A = \partial F(x_0, u_0, d_0)/\partial x$, $B = \partial F(x_0, u_0, d_0)/\partial u$, and $C' = \partial F(x_0, u_0, d_0)/\partial d$, small-signal model state equation can be obtained as follows:

$$\dot{\hat{x}} = A\hat{x} + B\hat{u} + C'\hat{d}. \tag{6}$$

Coefficient matrices A and B are the same as those in (3). The static operating point of state variables and output variables of controlled-source circuit can be obtained by (3) [13, 14]

$$X = -A^{-1}BU,$$
$$Y = \left(E - CA^{-1}B\right)U. \tag{7}$$

Then,

$$C' = \left[(A_1 - A_2)X + (B_1 - B_2)U\right]. \tag{8}$$
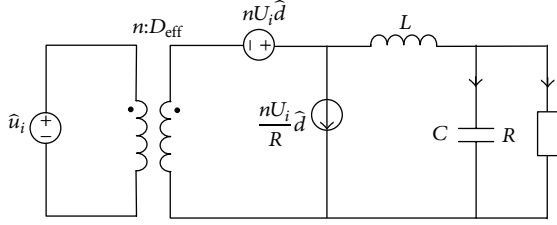
FIGURE 6: Small-signal model of hard-switching full-bridge circuit.

Using Laplace transform on small-signal model state equation (6), we can obtain small model state equation in complex frequency domain

$$s\widehat{x}(s) = A\widehat{x}(s) + B\widehat{u}(s) \\ + \left[(A_1 - A_2)X + (B_1 - B_2)U\right]\widehat{d}(s). \tag{9}$$

Transforming (9), we can derive the solution for small-signal model state equation in the frequency domain:

$$\widehat{x}(s) \\ = (sI - A)^{-1}B\widehat{u}(s) \tag{10} \\ + (sI - A)^{-1}\left[(A_1 - A_2)X + (B_1 - B_2)U\right]\widehat{d}(s).$$

Small-signal model of hard-switching controlled-source circuit is illustrated in Figure 6.

Based on the small-signal model diagram of hard-switching controlled-source circuit shown in Figure 6, we can obtain control-output transfer function

$$\frac{\widehat{u}_o(s)}{\widehat{d}(s)} = \frac{U_i}{n*(LCs^2 + sL/R + 1)}. \tag{11}$$

5. Soft-Switching Controlled-Source Circuit Modeling

From the previous process analysis of hard-switching and soft-switching controlled-source circuits, the major difference between soft-switching and hard-switching circuits is the change of duty cycle, including static duty cycle loss and small-signal duty cycle adjustment.

5.1. Static Duty Cycle Loss. Controlled-source circuit must increase transformer leakage inductance to increase the load range of the zero-voltage switches. However, large leakage inductance induces the decline of primary current rise rate when voltage is applied to the primary side of the transformer. Current ramp reduces the effective duty D_{eff} of transformer secondary voltage and seriously affects the dynamic performance of the converter.

The duty cycle of the primary side of the transformer can be obtained according to Figure 4

$$D = \frac{t_5 - t_2}{T_s/2}. \tag{12}$$

Secondary duty (effective duty) can be derived as follows:

$$D_{\text{eff}} = \frac{t_5 - t_4}{T_s/2}. \tag{13}$$

In the presence of the transformer leakage inductance, effective duty is smaller than the primary duty. The duty cycle loss is given by

$$\Delta D = D - D_{\text{eff}}, \tag{14}$$

where D is the duty cycle of the primary voltage determined by the control circuit and ΔD represents the duty cycle loss.

The primary current at time instant $t = t_4$ can be derived as

$$I_1 = n\left(I_L - \frac{\Delta I}{2}\right), \tag{15}$$

where at time instant $t = t_6$

$$I_2 = n\left(I_L + \frac{\Delta I}{2} - (1 - D)\frac{V_o T_s}{2L}\right). \tag{16}$$

Based on Figure 4, ΔD can be derived as

$$\Delta D = \frac{I_1 + I_2}{(U_i/L_r)(T_s/2)}. \tag{17}$$

By combining (15) and (16) into (17), we obtain

$$\Delta D = \frac{1}{(U_i/L_r)(T_s/2)}\left[2I_L - \frac{u_o}{L}(1 - D)\frac{T_s}{2}\right]. \tag{18}$$

Considering (14), D_{eff} can be obtained as follows:

$$D_{\text{eff}} = D - \frac{1}{(U_i/L_r)(T_s/2)}\left[2I_L - \frac{u_o}{L}(1 - D)\frac{T_s}{2}\right]. \tag{19}$$

5.2. Small-Signal Duty Cycle Adjustment

(1) Duty Cycle Adjustment Caused by the Change of Output Filter Inductor Current. When steady-state operation of controlled-source circuit is perturbed by an increase of the filter inductor current, assuming the filter inductor current is denoted by $\widehat{i}_L$, the primary current will reach the reflected filter inductor current at later time than it would in the steady-state operation. This phenomenon will cause a reduction of the duty cycle, which can be derived as follows:

$$\widehat{d}_i = -\frac{\Delta t}{T_s/2} = -\frac{4nL_r f_s}{U_i}\widehat{i}_L. \tag{20}$$

(2) Duty Cycle Adjustment Caused by the Change of Input Voltage. When steady-state operation of controlled-source circuit is perturbed by an increase of the input voltage by the amount $\widehat{u}_i$, the slope of the primary current will increase to reach the reflected filter inductor current sooner than it would in the unperturbed operation. This phenomenon will cause an increase of the duty cycle, which can be obtained as follows:

$$\widehat{d}_u = \frac{4nf_s L_r I_L}{U_i^2}\widehat{u}_i. \tag{21}$$
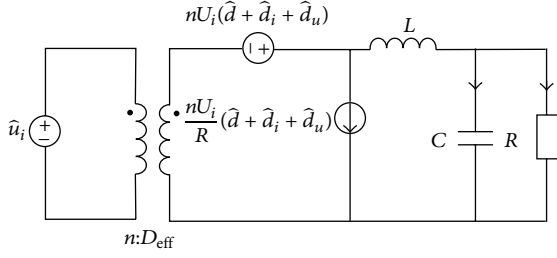
FIGURE 7: Small-signal circuit model of soft-switching controlled-source circuit.

5.3. Small-Signal Model.

The above analysis is introduced into small-signal circuit model of hard-switching full-bridge circuit. In particular, $\widehat{d}_{\text{eff}}$ is used instead of $\widehat{d}$. Equivalent small-signal circuit model of soft-switching phase-shifted full-bridge converter can then be derived

$$\widehat{d}_{\text{eff}} = \widehat{d} + \widehat{d}_i + \widehat{d}_u. \tag{22}$$

Small-signal circuit model of soft-switching controlled-source is illustrated in Figure 7.

Based on the small-signal model of soft-switching controlled-source circuit illustrated in Figure 7, the control-output transfer function of the control system can be obtained

$$\frac{\widehat{u}_o(s)}{\widehat{d}_{\text{eff}}(s)} = \frac{U_i}{n * \left(LCs^2 + s\left(L/R + R_d C \right) + R_d/R + 1 \right)}, \tag{23}$$

where $R_d = 4n^2 L_r f_s$. When $L_r = 0$, the model of soft-switching controlled-source circuit is evolved into the model of hard-switching controlled-source circuit.

As shown in (23), the function of internal current feedback reduces the low frequency gain of the transfer function, which is due to R_d/R. If R_d/R is controlled within reasonable limits, which can be negligible, then

$$\frac{\widehat{u}_o(s)}{\widehat{d}_{\text{eff}}(s)} = \frac{U_i \omega_0^2/n}{s^2 + 2\omega_0 \xi s + \omega_0^2}, \tag{24}$$

where $\omega_0^2 = 1/LC$, $\xi = (1/2R)\sqrt{L/C} + (R_d/2)\sqrt{C/L}$. The first item of ξ is damping item caused by hard-switching controlled-source circuit. The second item is damping item caused by soft-switching controlled-source circuit due to the presence of the leakage inductance of the transformer.

6. Simulation and Experiment

Based on the preceding analysis, the small-signal model is verified by simulation and experiment for a 6 kW marine transmitter circuit shown in Figure 8. The circuit parameters are as follows: input voltage $U_i = 540$ V, switching frequency $f_s = 20$ kHz, output voltage $U_o = 34$ V, output filter inductor $L = 20\,\mu$H, output filter capacitor $C = 1000\,\mu$F, transformer leakage inductance $L_r = 56\,\mu$H, and load $R = 0.17\,\Omega$.

(1) Simulation Analysis. The amplitude-frequency and phase-frequency characteristics of marine transmitter hard-switching controlled-source circuit are depicted by green lines and
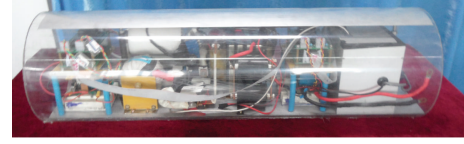
FIGURE 8: Physical picture of marine electromagnetic transmitter circuit.
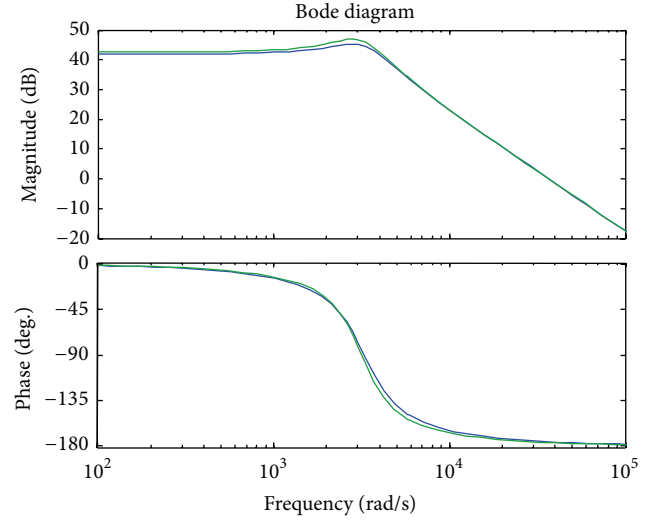
FIGURE 9: Control-to-output voltage transfer functions of hard-switching (green lines) and soft-switching (blue lines) controlled-source circuits.

those of marine transmitter soft-switching controlled-source circuit are depicted by blue lines, as illustrated in Figure 9. This figure shows that the resonance peaks of DC gain between the two circuits are different. An inhibition term is added in the soft-switching circuit because of the function of the leakage inductance, which significantly reduced the resonance peak.

(2) Experiment Verification. The contrast curves of measured and predicted amplitude-frequency characteristics of the control-to-output transfer function are illustrated in Figure 10. This figure shows that the measured curve and prediction curve can be well fitted at low frequency, whereas large error is found in the high frequency because of the effect of high-frequency zero. The error of model prediction at high-frequency range has little effect on the control of the system. Hence, the model accurately reflects the actual circuit.

Transformer primary voltage and current waveforms of soft-switching controlled-source circuit and hard-switching controlled-source circuit are illustrated in Figures 11 and 12, respectively. These figures show that the experimental results are consistent with the theoretical analysis. The voltage waveform is different from the current waveform as a result of phase shifting control mode.

The high-frequency voltage transformer primary and secondary voltage waveforms of soft-switching and hard-switching controlled-source circuits are illustrated in Figures 13 and 14, respectively. These figures show that the experimental results are consistent with the theoretical analysis.
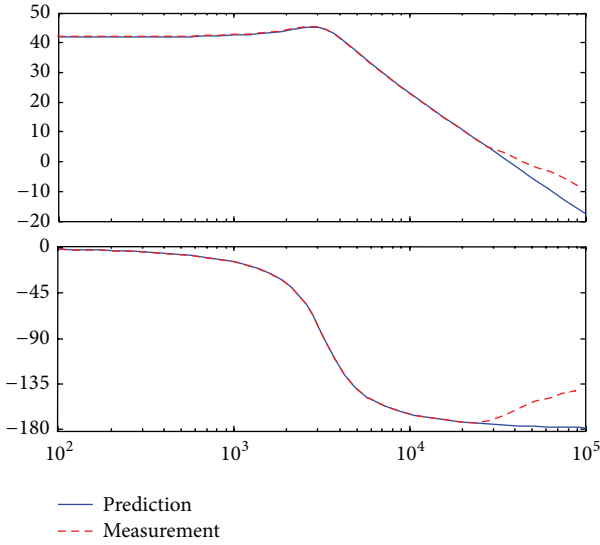
FIGURE 10: Control-to-output voltage transfer function of the controlled-source circuit. Model prediction (solid lines) and experimental measurement (dashed lines).
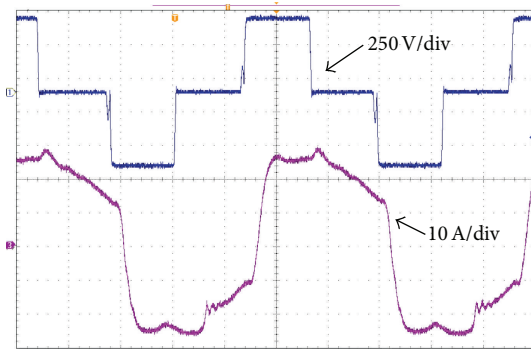
FIGURE 11: Primary voltage and current waveforms in soft-switching mode.
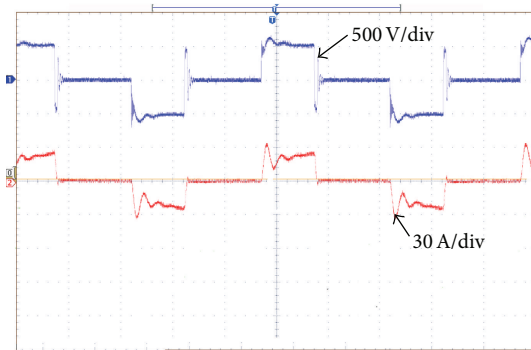
FIGURE 12: Primary voltage and current waveforms in hard-switching mode.

The duty cycle is less than the effective control duty cycle because of the presence of the leakage inductance of the transformer.
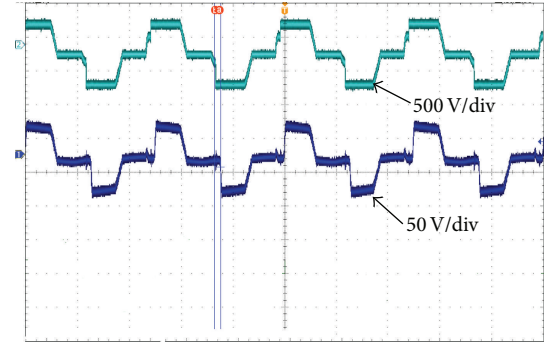
FIGURE 13: Primary and secondary voltage waveforms of transformer in soft-switching mode.
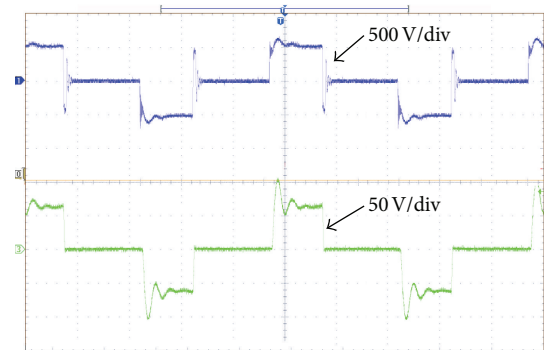
FIGURE 14: Primary and secondary voltage waveforms of transformer in hard-switching mode.

Output voltage and current waveforms of transmitter transmitting electrode in soft-switching and hard-switching modes are depicted in Figures 15 and 16, respectively. The emission frequency is 50 Hz, the transmitting voltage is 34 V, and the emission current is 200 A. The output of electromagnetic transients is significantly improved by using controller designed with the proposed mathematical models.

The measure efficiency curves of marine electromagnetic detection transmitter using soft-switching controlled-source circuit and using hard-switching controlled-source circuit are illustrated in Figure 17. The maximum efficiency of the transmitter circuit using soft-switching controlled-source circuit is at 90% (including self-excitation auxiliary power supply), and the maximum efficiency of the transmitter circuit using hard-switching controlled-source circuit is at 83%.

7. Conclusions

(1) Based on the analysis of the operation process of marine transmitter hard-switching and soft-switching controlled-source circuits, similarities and differences between the two circuits have been identified, and two equivalent circuits of controlled-source operation are obtained. On this basis, a mathematical model of controlled-source circuit can be established.
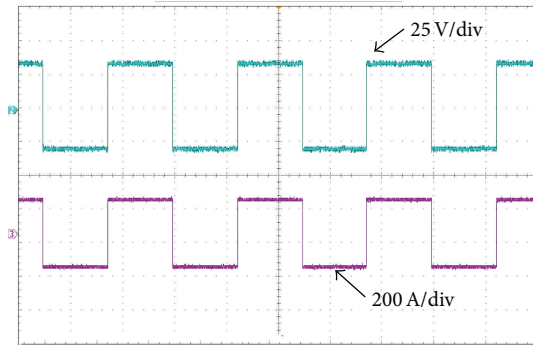
FIGURE 15: Output voltage and current waveforms of transmitter transmitting electrode in soft-switching mode.
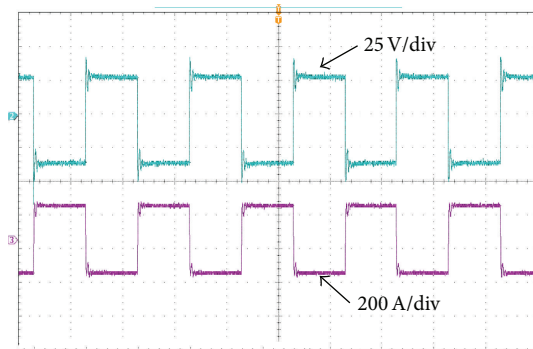
FIGURE 16: Output voltage and current waveforms of transmitter transmitting electrode in hard-switching mode.
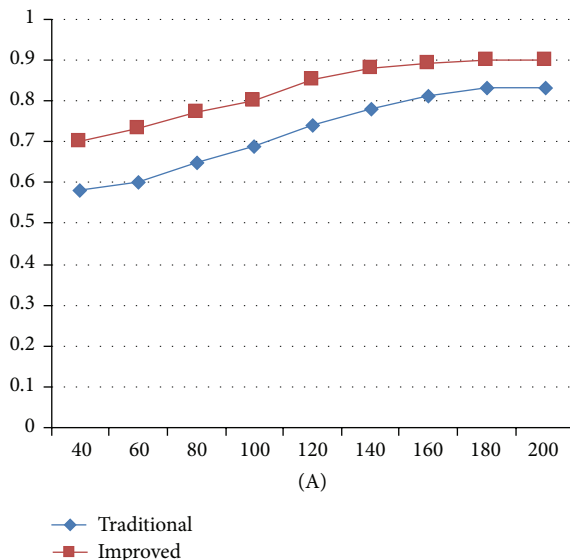
FIGURE 17: Measure efficiency of the proposed transmitter and conventional transmitter.

(2) Ideal switch model, state-space average model, and small-signal model are established based on the idealization method and abstract extent of hard-switching controlled-source circuit.

(3) Small-signal model and transmission function of soft-switching controlled circuit are obtained based on the mathematical modeling of hard-switching controlled-source circuit by analyzing the effect of output filter inductor current transformer leakage inductance and input voltage soft-switching controlled source on change in the duty cycle.

(4) The proposed small-signal model is verified by simulation and experiment. The results are consistent with theoretical analysis. The controller designed with the proposed model has significantly improved the control accuracy and transient performance of the output electromagnetic wave of marine electromagnetic transmitter.

Conflict of Interests

The authors confirm that this paper content has no conflict of interests.

Acknowledgments

References

[1] S. C. Constable, A. S. Orange, G. M. Hoversten, and H. F. Morrison, "Marine magnetotellurics for petroleum exploration part I: a sea–floor equipment system," *Geophysics*, vol. 63, no. 3, pp. 816–825, 1998.

[2] S. Constable and L. J. Srnka, "An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration," *Geophysics*, vol. 72, no. 2, pp. WA3–WA12, 2007.

[3] M. Kurang, N. Misac, and L. Yaoguo, "Controlled source electromagnetic (CSEM) technique for detection and delineation of hydrocarbon reservoirs: an evaluation," in *Proceedings of the 75th Annual Internat SEG Meeting: Expanded Abstracts*, pp. 546–549, 2005.

[4] T. Eidesmo, S. Ellingsrud, L. M. Macgregor et al., "Re-mote detection of hydrocarbon filled layers using marinecontrolled source electromagnetic sounding," in *Proceedings of the EAGE 64thConference & Exhibition*, pp. 27–30, Florence, Italy, May 2002.

[5] R. Erichson and D. Maksimovic, *Fundamentals of Power Electronics*, Kluwer Academic Publishers, New York, NY, USA, 2001.

[6] R. Kollman and J. Betten, "Closing the loop with a popular shunt regulator," *Power Electronics Technology*, vol. 46, no. 2, pp. 30–36, 2003.

[7] R. D. Middlebrook and S. Cuk, "A general unified approach to modelling switching-converter power stages," *International Journal of Electronics*, vol. 42, no. 6, pp. 521–550, 1977.

[8] R. B. Ridley, "A new, continuous-time model for current-mode control," *IEEE Transactions on Power Electronics*, vol. 6, no. 2, pp. 271–280, 1991.

[9] V. Vorperian, "Simplified analysis of PWM converters using the model of the PWM switch: parts I and II," *IEEE Transactions on*

Aerospace and Electronic Systems, vol. 52, no. 26, pp. 490–505, 1990.

[10] S. V. Mollov, M. Theodoridis, and A. J. Forsyth, "High frequency voltage-fed inverter with phase-shift control for induction heating," *IEE Proceedings: Electric Power Applications*, vol. 151, no. 1, pp. 12–18, 2004.

[11] N. S. Bayindir, O. Kukrer, and M. Yakup, "DSP-based PLL-controlled 50~100 kHz 20 kW high-frequency induction heating system for surface hardening and welding applications," *IEE Proceedings—Electric Power Applications*, vol. 150, no. 3, pp. 365–371, 2003.

[12] C. P. Basso, *Switch-Mode Power Supplies, Spice Simulations and Practical Designs*, The McGraw-Hill Companies, Paris, France, 2008.

[13] C. P. Basso, *Designing Control loops for Linear and Switching Power Supplies*, Artech House, Paris, France, 2012.

[14] L. Dixon, "Average current mode control of switching power supplies," in *Product & Application Handbook*, Unitrode, pp. 356–369, 1993-1994.

Research Article

A New Hybrid Model Based on an Intelligent Optimization Algorithm and a Data Denoising Method to Make Wind Speed Predication

Ping Jiang and Qingli Dong

School of Statistics, Dongbei University of Finance and Economics, Dalian, Liaoning 116025, China

Correspondence should be addressed to Qingli Dong; isaacdon525@gmail.com

To mitigate the increase of anxiety resulting from the depletion of fossil fuels and destruction of the ecosystem, wind power, as the most common renewable energy, is a flourishing industry. Thus, accurate wind speed forecasting is critical for the efficient function of wind farms. However, affected by complicated influence factors in meteorology and volatile physical property, wind speed forecasting is difficult and challenging. Based on previous research efforts, an intelligent hybrid model was proposed in this paper in an attempt to tackle this difficult task. First, wavelet transform was utilized to extract the main components of the original wind speed data while eliminating noise. To make better use of the back-propagation artificial neural network, the initial parameters of the network are substituted with optimized ones, which are achieved by using the artificial fish swarm algorithm (AFSA), and the final combination model is employed to conduct wind speed forecasting. A series of data are collected from four different observation sites to test the validity of the proposed model. Through comprehensive comparison with the traditional models, the experiment results clearly indicate that the proposed hybrid model outperforms the traditional single models.

1. Introduction

With fossil fuels gradually drying up and the aftermath of the polluted environment becoming increasingly obvious, the world is sparing no effort to explore and exploit renewable energy for future power generation [1]. However, intermittence is the nature of the most renewable energy sources, which presents a great challenge to maintain and ensure the stability and reliability of power network [2]. Thus, effective energy storage technologies play a paramount role in mitigating the volatility [3]. As key ingredients, electric vehicles (EVs) would be widely regarded as an integral part of such a renewable energy system if they use electricity generated by a renewable energy source [4]. Besides, oil and coal supply uncertainty, growing mobility demand, and increasingly stringent regulations on pollutants and carbon footprints are expediting a paradigm shift towards sustainable power system and transportation [5, 6]. Until now, even though numerous clean energy technologies, such as plug-in hybrid electric vehicles (PHEVs) [7], battery management system (BMS) [8],

and battery electric vehicles (BEVs) [9], are being widely deployed and developed to significantly reduce the fuel consumption and the carbon emission throughout the world, they also increase the demand of power in the meantime. Thus, expanding investment on exploration and exploitation of the supplement of traditional energy is badly needed.

Wind power, the most recognized renewable energy source (RES), definitely possesses the prominent potential to become the resource of electricity, which plays a fundamental and indispensable role in social modernization [10]. Among different types of renewable energy, such as solar, wind, geothermal, and tidal energy, wind power is the fastest-developing and the most widely used one and also meets the demand for large scale commercial exploitation, owing to the maturity of the technology [11].

Currently, the technology of wind power has attracted worldwide attention and occupied a considerable share in the proportion of energy construction. Many countries take the development of wind power as an effective action to improve their domestic energy structures, preserve the ecological

environment, alleviate environmental pollution, and even integrate it into the strategic development planning for the entire power system.

China, with such a great population, will soon replace America as the leading energy consuming country [12]. Depending on a series of positive policies, the wind power industry has stepped into rapid development and the present scale is exciting. The capacity of exploitable wind power is on rise year by year as technology progresses. According to a study started by the National Energy Administration (NEA), in the national wind power long-term plan, it is explicitly proposed that, to 2020, the national wind power installed capacity goal is 20,000,000 kilowatts [13]. In order to achieve such a capacity of power, nearly 708 million tons coal are needed according to the internationally accepted conversion standard, and the accompanying carbon emission even reaches to 480 million tons. Therefore, developing wind power efficiently does not only have economic significance, but also high ecological value.

With such a large capacity of wind power, the integration of wind power becomes an issue. Furthermore, high randomness of wind brings a big problem for a wind farm to provide electric grids to meet its necessary need of power, and this characteristic of wind power is a big challenge to traditional power system which has a relatively mature operation mode. Hence, accurate wind speed forecasting, as a necessary step in wind power integration, is particularly important, and these are the contributions of this paper.

There are three acknowledged methods for forecasting wind speed, including long-term prediction, medium-term prediction, and short-term prediction [14]. Briefly, long-term forecasting is used in the process of site selection for wind farms and in the maintenance and overhaul of the power network. When researching the dynamic economic dispatch scheme in a grid-connected wind power system, the medium-term forecasting of wind that usually spans a few days or weeks comes in handy. As for the latter method, it mostly works to meet the control requirements of wind turbines. Overall, apparently, forecast error will become greater as the forecast horizon increases. Thus, the accuracy of short-term forecasting is particularly important, for it lays the foundation of the renewable energy program, security evaluation, optimal dispatching model, and other critical decisions in power network dispatching.

Until now, many scientists have devoted research efforts to short-term wind speed forecasting, and various achievements have been successfully applied. Numerical simulation, as early results, is usually used to construct mathematical models to simulate the dynamical physical behaviour of wind. These models include many linear, nonlinear, difference, and differential equations related with many conditions of wind, and those numerical solutions are usually achieved by calculation and iterative methods. However, because of the astatic and chaotic characteristics of the factors involved in the numerical simulation, as well as many uncertainties in the forecasting process, the results of these methods usually cannot meet the requirements of precision [15].

To solve these difficulties, many new models have been applied, which can roughly be categorized into physical, such

as the Numerical Weather Prediction systems (NWPs), statistical, containing linear methods such as the Auto Regressive model (AR), Moving Average (MA), Auto Regressive Integrated Moving Average (ARIMA) model, and intelligent methods, which come from artificial intelligence and machine learning fields, such as the artificial neural networks (ANN) [16], and even by hybrid approach methods, which are combination of statistical and physical methods that use the analysis of a time series [17].

ARIMA, as a traditional statistical approach, has been proposed by some researchers to forecast wind speed, though it should be improved when dealing with the volatility and chaotic characteristics of wind speed. Ernst et al. supplied a method using past power measurements and meteorological predictions of the data of wind speed and direction interpolated at a wind farm [18]. The Grey prediction model $GM(1, n)$, as a novel method for wind speed forecasting, as well as a wind power predictor, that reflects a Grey model that has a differential equation with a single variable, was provided by El-Fouly. The structure of the artificial neural network, which was applied for forecasting the mean monthly wind speed in regions of Cyprus, is discussed by Baran, though the data were not sufficient [19]. In addition, a hybrid model that combined the Kalman filter (KF) with the artificial neural network was proposed based on an ARIMA model to further enhance the forecasting accuracy of wind speed. Sideratos and Hatziargyriou [20] employed a hybrid method to forecast wind speed in which three independent algorithms had been adopted. They were the Seasonal Adjustment Algorithm, the Exponential Smoothing Method (ESM), and the Radial Basis Function (RBF) neural network, and the final results indicated that the suggested hybrid method had satisfactory performance.

Back-propagation, as one of the most common learning algorithms among the artificial neural networks, has obvious advantages in nonlinear model fitting and predication because of its unique characteristic in terms of high tolerance with data errors. From the operational principle of the learning algorithm of the BPANN model, it can be observed that the initial weights and thresholds of the back-propagation neural network have a serious impact on the accuracy of the model [21].

In addition to the model, the characteristic of the initial data can also affect the final forecasting of the wind speed. In fact, the original data from the wind farm usually cannot be used in the model directly because there is much chaotic and scrambled information [22]; then, the WT is employed in this work to preprocess the data of ten-minute wind speed to eliminate the high frequency disturbance. For the sake of improving the accuracy of wind speed forecasting further, a new optimization algorithm called the artificial fish swarm algorithm (AFSA) is proposed in this paper. The AFSA adapted here is used to find the proper parameters of the initial back-propagation artificial neural network, whose input is the processed ten-minute wind speed data. The final results of the data showed that the hybrid optimal AFSA and BPANN method can perform better in short-term wind speed forecasting compared with previous models.

The rest of this paper is organized as follows. The analysed data are described in Section 2. A general description of relative methods is provided in Section 3. In Section 4, the hybrid optimal model is discussed in detail; specifically, the process of selection of the optimal parameters of the propagation and the overall structure of the WAFSA-BPANN model will be presented and introduced explicitly. In Section 5, the simulation of the model and the evaluation of its final performance of wind speed forecasting are discussed in detail. The conclusions of this paper are presented in Section 6.

2. The Sources and Characteristics of the Data Set

To meet the demands of the model proposed in this paper, a variety of data were collected from turbines of wind farms from the Shandong Peninsula of China. Owing to its abundant sources of wind from the sea, which are richly endowed by nature, the Shandong Peninsula plays a more and more important role within the entire large sector of the Chinese wind power structure. However, it is difficult to forecast and evaluate wind speed because there are too many factors that can affect the changing of the wind speed. Our effective and reliable method is to find out the valuable and useful information, which the proposed model can then utilize to display the forecasting through the historical data.

The data used in this paper are collected from three different turbines of a wind farm. To ensure the fairness and consider the potential periodicity of seasons, the involved data series of wind speed cover four seasons in 2011. To be specific, data from four representative months in different seasons are chosen to implement the simulation and forecasting, namely, January, May, July, and September. For instance, ten-minute wind speed data series from January 1 to January 31 in 2011 are collected. Each day, the data series span from 00:00 to 23:50, and the wind speed data are recorded every ten minutes. Figure 1 shows the whole profile of the wind speed data series spanning four seasons in one of the three chosen observation sites.

From Figure 1, the obvious feature of fluctuation and undulation of the original wind speed data series can be observed, and there are no rules that may be utilized to conduct forecasting through simple estimation. Besides, the results of parameter estimation and testing indicate that wind speed data series fit the Weibull distribution. Though many models can tolerate the instability of the original data, the accuracy of the forecasting models can still be impaired by the noise and disturbance suffused in them. Thus, there are active imperious demands to preprocess the initial data series to achieve a relatively smooth time series. Herein, wavelet transform was proposed to dispose the wind speed data series firstly, and preprocessed data were applied to execute the hybrid model, which will be discussed in the next section.

3. Description of Relative Models

Several methods are referred to in this paper, such as wavelet transform, back-propagation artificial neural network, and artificial fish swarm algorithm; each of them will be explained briefly below.

3.1. Wavelet Transform (WT). Wavelet transform is a type of denoising method that can decompose signals with different frequencies. The essence of WT is to transform a one-dimensional signal in the time domain into a two-dimensional time-frequency signal. WT can carry out the multiresolution analysis, which indicates that WT has a high time resolution and low frequency resolution in the high frequency part; in the low frequency part, WT has a high frequency resolution and low time resolution [23]. The basic idea of WT is to approach the primitive function by using two of the above functions. Assume that the signal $X(t)$ is the square integrable function; then, the WT of the signal is the inner product with $\psi_{(\alpha,\tau)}(t)$:

$$\begin{aligned} \mathrm{WT}_X(\alpha,\tau) &= \frac{1}{\sqrt{\alpha}} \int X(t)\,\psi^*\left(\frac{t-\tau}{\alpha}\right) dt \\ &= \left\langle X(t), \psi_{(\alpha,\tau)}(t) \right\rangle, \end{aligned} \tag{1}$$

where α and τ are the scaling and shift factor, respectively, and $\psi^*(t)$ is the adjoin function of $\psi(t)$.

The basic steps of WT are as follows.

Step 1 (Decompose the original signal). Choose the wavelet basis and decomposition hierarchy and calculate coefficients of wavelet decomposition in each layer.

Step 2 (Address the threshold value of high frequency coefficients). Choose a threshold value for each decomposition hierarchy and process the high frequency coefficients to eliminate the noise centred on the high frequency part.

Step 3 (Carry out wavelet reconstruction of signal). Carry out the wavelet reconstruction for low frequency coefficients and high frequency coefficients such that the threshold value is processed quantitatively [24].

Through the wavelet transform, the high frequency wind speed data series with significant noise is eliminated, and the residual of the subseries is combined to substitute the original data. From Figure 2, it is obvious that the denoising data series is more smooth and stable than the original data. The simulation and forecasting results upon the preprocessed data employed in this paper also indicate the necessity and superiority of the wavelet transform.

It should be noted here that the entire preprocessed data are only used to train the hybrid model involved hereinafter. To be specific, wind speed data during the whole month are employed to use the wavelet transform, and the final forecasting of the next week is achieved from the original real wind speed data.

3.2. Back-Propagation Artificial Neural Network (BPANN). The schematic of the BP artificial neural network is plotted in Figure 3(a). Applying the classic back-propagation optimal algorithm, the BPANN is widely used in multidomains. This particular BP neural network has three layers: the input layer,
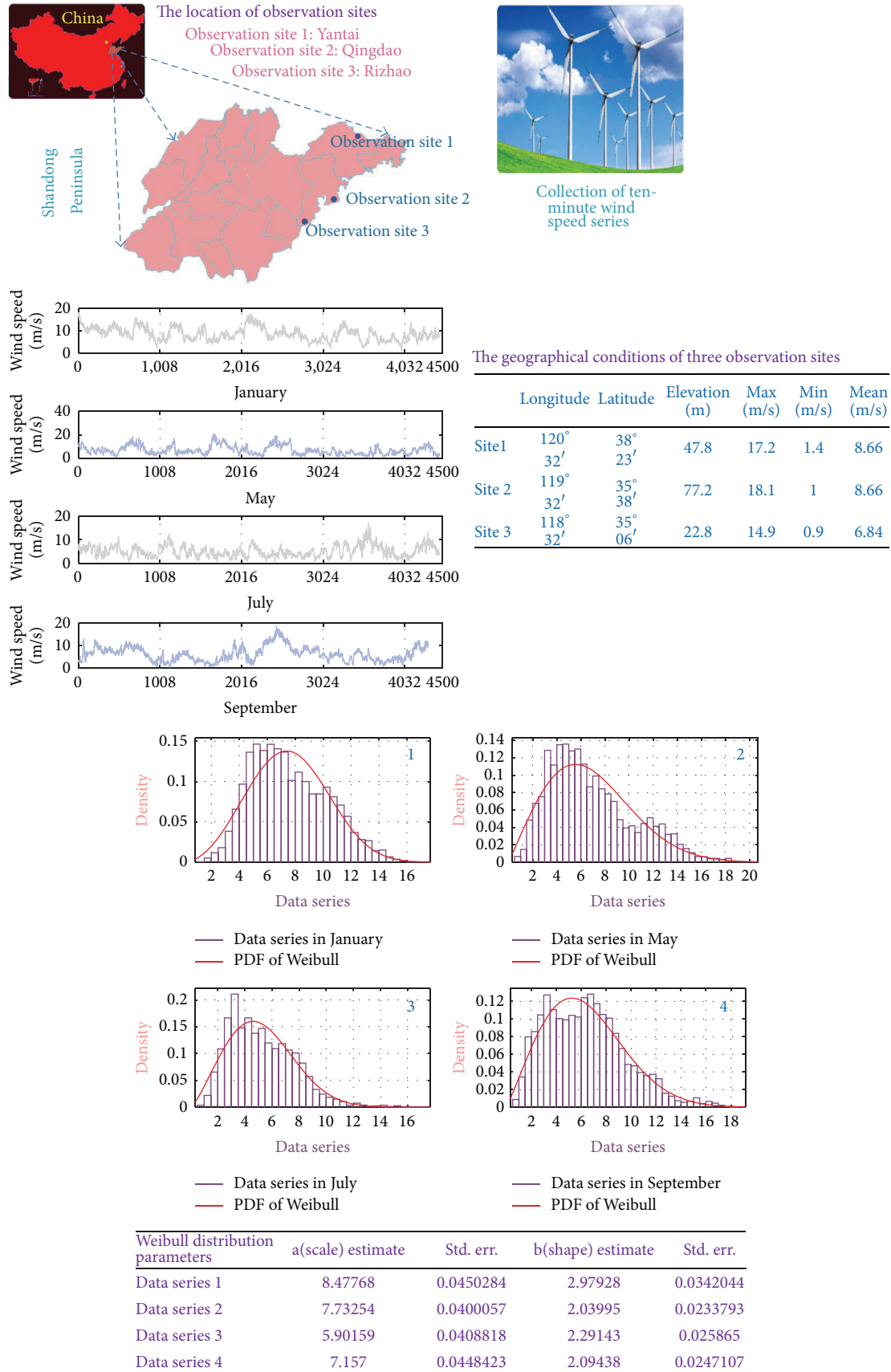
FIGURE 1: Original wind speed data collected every 10 minutes in four seasons in 2011 and basic information of these data series.

The geographical conditions of three observation sites

	Longitude	Latitude	Elevation (m)	Max (m/s)	Min (m/s)	Mean (m/s)
Site1	120° 32′	38° 23′	47.8	17.2	1.4	8.66
Site 2	119° 32′	35° 38′	77.2	18.1	1	8.66
Site 3	118° 32′	35° 06′	22.8	14.9	0.9	6.84

Weibull distribution parameters	a(scale) estimate	Std. err.	b(shape) estimate	Std. err.
Data series 1	8.47768	0.0450284	2.97928	0.0342044
Data series 2	7.73254	0.0400057	2.03995	0.0233793
Data series 3	5.90159	0.0408818	2.29143	0.025865
Data series 4	7.157	0.0448423	2.09438	0.0247107

FIGURE 2: Comparison of the original wind speed data series and the results of WT.

the hidden layer, and the output layer [25]. The number of the nodes of the input layer n is determined by our concrete data of wind speed. Herein, our data series is the ten-minute wind speed time series, and three nodes were selected as the input layer [26, 27]. This choice is based on extensive experiments: the prediction results obtained with three nodes in the input layer are much better than other cases. According to the data series of wind speed, these three input nodes represent the history data at continuous time epochs. Similarly, only one node of the output layer is also used to adapt our data series. As to the hidden layer, it is not difficult to find that it has a great effect on the robustness and stability of the neural network. According to Hecht-Nielsen [28], the proper number of the nodes of the hidden layer, that is, $l = 2n + 1$, was selected.

Before implementing the network, it is necessary to normalize the data sets, with both the input and the output data included. Thereafter, the outputs y_j of all hidden layer nodes are calculated as

$$y_j = f\left(\sum_i w_{ij} x_i + b_j\right),$$

$$(i = 1, \ldots, n; j = 1, \ldots, 2n + 1),$$

$$\text{(2)}$$

where w_{ij} is the weight connected from input nodes i to hidden nodes j, while the bias of neural is b_j, f is the sigmoid

function, and x_i represents the value of each input node. Then, calculate the output W_1 of neural network:

$$W_1 = f_0\left(\sum_j w_j y_j + b_0\right), \quad (j = 1, \ldots, 2n + 1), \quad \text{(3)}$$

where w_j is the connection weight from hidden layer to output layer, b_0 is the corresponding bias of neural cell, and the activation function is f_0. At last, minimize the error via the training process. In this paper, the network was trained firstly; then, the network was validated, and finally we performed model testing. To see more details of the BPANN, [29] can be referred to.

3.3. Artificial Fish Swarm Algorithm (AFSA).
Artificial fish swarm algorithm is an effective optimal algorithm for increasing the exploitation capability during the process of searching. As we all know, fish can easily find the area with more nutrition through individual searching or following after the ambient fishes. Li et al. [30] introduced this efficient algorithm in 2002, which models the behaviours of fishes to get the global optimal solution.

The state of the artificial fish can be described as $X = (x_1, x_2, \ldots, x_n)$, and x_i $(i = 1, 2, \ldots, n)$ represents the initial variable. As the objective function, $Y = f(X)$ is the consistency of the current position where the fish is in. $d_{i,j} = \|x_i - x_j\|$ represents the distance of individual fish. Visual is the cognitive distance, while Step and δ represents the maximum

Wind speed series

X_1 X_2 X_3

Wavelet transform to perform denoising of the original data

W_{ij}

Connection weights

Input layer (nodes)

a

Hidden layer (nodes)

W_{jk}

Thresholds

b Output layer (nodes)

Y

WAFSA-BPANN Forecasts

(a)

AFSA optimization

Parameters set
State and number of fishes
Consistency of position
Distance of fish individuals
Cognitive distance
Maximum step size
Perceptibility

Algorithm's parameters

AFSA is an optimization algorithm developed by Li et al. in 2002

AFSA-algorithm

Initial process

Initial fishes

Number of fishes

Evaluation function

Main function from concrete question

Iteration process

Judgement of fishes

Output the result

Fish "moving"

Select behavior and perform

Local search

Eliminating some fishes

Update the bulletin board

Continue this algorithm?

Reach termination condition?

(b)

FIGURE 3: The structure of the hybrid WAFSA-BPANN model.

step size and factor of congestion degree, respectively. There are four types of behaviours in the AFSA:

(1) Foraging behaviour: the current status of an artificial fish is X_i, and the fish selects a state X_j in its sensing range randomly; if the problem is approximately maximum and $Y_i < Y_j$ (or $Y_i > Y_j$ if the problem is approximately minimum; because problems that are approximately maximum or minimum can be converted to each other; in the following, we only discuss the problem that is approximately maximum), then make a step forward in this direction; otherwise, randomly select state X_j again and determine whether it meets the condition to move forward; after attempting for try_number times, if the condition still cannot be met, then choose a direction randomly to move a step:

$$X_j = X_i + (2\text{rand}() - 1)\,\text{Visual},$$
$$X_i = X_i + (2\text{rand}() - 1)\,\text{Step}.$$
$$(4)$$

(2) Swarming behaviour: X_i represents the current position of an artificial fish; the task is to explore the numbers of partners n_f and the central location X_c in the neighbourhood of the current position. If $Y_c/n_f > \delta y_i$, which indicates that there is more nutrition at the centre of partners and less competitiveness, then make a step forward towards the direction of the central of partner; otherwise, execute foraging behaviour:

$$X_i = X_i \_\text{rand}() \cdot \text{Step} \cdot \frac{X_c - X_i}{\|X_c - X_i\|}. \quad (5)$$

(3) Following behaviour: the current status of the artificial fish is X_i; explore the number of partners Y_j and the biggest partner X_j in the neighbourhood of the current position. If $Y_c/n_f > \delta y_i$, which indicates that the surroundings of X_j have more nutrition and are not so congested, then make a step forward towards the direction of X_j; otherwise, execute foraging behaviour:

$$X_i = X_i + \text{rand}() \cdot \text{Step} \cdot \frac{X_{g\text{best}} - X_i}{\|X_{g\text{best}} - X_i\|}. \quad (6)$$

(4) Random behaviour: implementation of random behaviour is relatively simple; the action is to move a step in a direction that is randomly picked in the field. In fact, this is a default behaviour during foraging behaviour.

There is a bulletin board in the algorithm, which is also defined as an artificial fish, to record the information of the optimal individual [31]. Each of the artificial fish would compare its current state with the state recorded in the bulletin board after every movement, and then substitute it if the current state of the individual is superior to the one recorded in bulletin board before.

According to the nature of the problem to be solved, evaluate the current environment of the fish and then choose a corresponding behaviour. A common method is to choose a behaviour from the alternative behaviours that can make the most progress toward the optimal direction; if there is no choice that can make the next state better than the current state, then the fish has to take a random action.

4. The Structure of Hybrid WAFSA-BPANN Model

Just as in the aforementioned references, the initial data of wind speed usually contain much interference, which can affect the accuracy of forecasting; it is therefore a critical procedure for optimal models to select valuable information while filtering the disturbance. Therefore, a relatively mature WT model was introduced to preprocess the original data series, which is nonstationary wind speed data. The sublayers wind speed is a much more stable and smooth time series. Meanwhile, because the initial weights and thresholds of back-propagation artificial neural networks are initialized randomly, it is difficult to reach the best state in most cases. Therefore, finding the best initial parameters of the initial network is greatly needed [32]. Herein, the artificial fish swarm algorithm was applied to finish this work. Based on the WAFSA-BPANN model, the steps of training are depicted as follows.

Step 1. Given the structure of the back-propagation artificial neural network, there are three layers in the network, and the number of nodes is I, J, K in the input layer, the hidden layer, and the output layer, respectively [33].

Step 2. Set the dimensions of the artificial fish, including the weights and thresholds of the artificial neural network:

$$X = X\left(v_{11}, \ldots, v_{I1}, \mu_1, \ldots, v_{1J}, \ldots, v_{IJ}, \mu_J, \ldots, w_{1K}, \ldots, \right.$$
$$\left. w_{JK}, \theta_K\right). \quad (7)$$

Step 3. Initialize the basic parameters of the artificial fish swarm algorithm, such as the size of the population pop_size, the cognitive distance Visual, the maximum step size Step, and the factor of congestion degree δ, while Max_Gen represents the maximum number of iterations and ε the objective value.

Step 4. Initialize the iteration step Gen = 0. Produce artificial fish individuals with a population of pop_size randomly in the feasible region as the initial fish; meanwhile, the component of each individual fish should strictly be random, with digits between $(-1, 1)$.

Step 5. Calculate the food concentration Y of the current position of each individual fish, choose the best one, and mark it on the bulletin board.

Step 6. Simulate each individual fish to execute the following behaviour and swarming behaviour, and then calculate Y; after comparing and determining the bigger Y, execute

Input:
$x = (x_1, x_2, \ldots, x_d)^T$—A sequence of training data.
Output:
$x_{\text{try\_number|next}}$—The returned value with the best fitness in the search domain.
Parameters:
try_number—The attempted times in the behaviour of prey.
friend_number—The number of friends around AF.
AF_X[n]—The position of AF.
AF_step—The distance that AF can move for each step.
AF_visual—The visual distance of AF.
AF_delta—The condition of jamming.

(1) /*Generate an initial population of n AF $x_i(i = 1, 2, \ldots, n)$*/
(2) /*AF_prey()*/
(3) **FOR EACH** $i : 0 \leq i \leq$ try_number **DO**
(4) $X_i = X_j + \text{Visual} \cdot \text{Rand}()$;
(5) **IF** $Y_i < Y_j$ **THEN**
(6) $X_{i|\text{next}} = X_i + X_j - X_i/\|X_j - X_i\| \cdot \text{Step} \cdot \text{Rand}()$;
(7) **ELSE** $X_{i|\text{next}} = X_i + \text{Step} \cdot \text{Rand}()$ **END END**
(8) /*AF_swarm()*/
(9) **FOR EACH** $i : 0 \leq i \leq$ friend_number
(10) **IF** $(d_{i,j} < \text{Visual})$ **DO**
(11) $n_f = n_f + 1$;
(12) $X_c = X_c + X_j$ **END**
(13) $X_c = X_c/n_f$
(14) **IF** $(Y_c/n_f > \delta y_i)$ **THEN**
(15) $X_{i|\text{next}} = X_i + X_c - X_i/\|X_c - X_i\| \cdot \text{Step} \cdot \text{Rand}()$;
(16) **ELSE** AF_prey() = 0 **END END**
(17) /*AF_follow()*/
(18) **FOR EACH** $i : 0 \leq i \leq$ friend_number
(19) **IF** $(d_{i,j} < \text{Visual}\&\&Y_j > Y_{\max})$ **THEN**
(20) $Y_{\max} > Y_j$;
(21) $X_j > X_{\max}$; **END**
(22) $n_f = 0$;
(23) **IF** $(d_{i,j} < \text{Visual})$ **THEN**
(24) $n_f = n_f + 1$;
(25) **END IF** $(Y_{\max}/n_f > \delta Y_i)$
(26) $X_{i|\text{next}} = X_i + X_{\max} - X_i/\|X_{gbest} - X_c\| \cdot \text{Step} \cdot \text{Rand}()$;
(27) **ELSE** AF_prey()

ALGORITHM 1: AFSA.

the corresponding behaviour (the default action is to execute the foraging behaviour).

Step 7. Update the bulletin board after each iteration, and always substitute the exiting Y with more optimal Y in fish individuals.

Step 8. Judge if the iteration times Gen have already reached Max_Gen or if the error ε has met the requirement, and output the digit Y on the bulletin board once the termination condition is met; otherwise, make Gen = Gen + 1 and go back to Step 6.

The pseudocode of the artificial fish swarm algorithm is listed in Algorithm 1.

Specifics parameters of the AFSA algorithm are presented in Table 1.

TABLE 1: The value of the parameters of AFSA adopted in this paper.

Parameters	Value
Population size	40
Maximum iterations	1000
Visual of each individual	0.5
Step of each fish	0.01
Try_number of each fish	10
Congestion degree	11
Objective value	0.0001
Time complexity	$O(10^6)$
Space complexity	$O(10)$

The hybrid method has better performance than the traditional back-propagation artificial neural network and other single models, and the structure of the WAFSA-BPANN

is briefly shown in Figure 3, from which a visualized understanding of the hybrid optimum model can be got.

5. Simulation and Analysis

By means of the aforementioned hybrid model and the time series wind speed data, a series of simulations and forecasting have been operated using Matlab R2012b, and the results are discussed as follows.

5.1. Related Index. To ensure fairness and real forecasting results, the data volume, training data set, and forecasting data set are kept unchanged. Herein, three common indexes of error, such as the mean square error (MSE), mean absolute percentage error (MAPE), and the mean absolute error (MAE), are adopted to evaluate the performance of related models, and their specific formulations are shown below:

$$
\text{MAPE} = \frac{1}{M} \sum_{k=1}^{M} \left| \frac{\widehat{x}_k - x_k}{x_k} \right|,
$$

$$
\text{MAE} = \frac{1}{M} \sum_{k=1}^{M} \left| \widehat{x}_k - x_k \right|, \tag{8}
$$

$$
\text{MSE} = \frac{1}{M} \sum_{k=1}^{M} \left(\widehat{x}_k - x_k \right)^2,
$$

where $\widehat{x}_k (k = 1, 2, \ldots, M)$ represents the concrete value of forecasting series and M is the length of the series, while the corresponding value of original data series is x_k.

5.2. Discussion of the Results from Different Seasons. Affected by many meteorological factors which are usually erratic, wind speed fluctuates all of the time, making it difficult to forecast. Based on a previous study, a hybrid new model is proposed in this paper. To achieve intuitive results of the hybrid model, the specific prediction data with a certain time interval on the first weeks of February, June, August, and October are collected, while the training data sets span the whole month of January, May, July, and September, respectively. Thus, the length of the prediction series is 1008 because of the time intervals, which makes it improper to list all these data under the limited space. In this case, forecasting data from several comparative models are recorded every 6 hours. The results from our hybrid model are compared with those from other traditional hybrid models. In this experiment, preprocessed data from four months in different seasons are utilized to train the neural network, and upon training the neural network, the final forecasting is made. Herein, data collected from January 1 to January 31 are utilized as the training set to train the network, upon which the prediction of wind speed from February 1 to February 7 is achieved. Next, preprocessed data acquired from May 1 to May 31 are used to train the artificial neural network, and the data from June 1 to June 7 are the relevant testing set. According to the same principle, the remaining training data series are formed. For example, the testing data from August 1

to August 7 are obtained from the neural network, which are trained by wind speed data spanning July 1 to July 31.

Tables 2, 3, 4, and 5 display the prediction results of four seasons, respectively. To be specific, forecasting results in the first week of February obtained from different models are presented in Table 2. Similarly, Tables 3, 4, and 5 represent the forecasting results in other three seasons.

As is depicted in Tables 2, 3, 4, and 5 and Figure 4, the specific forecasting results are enumerated vividly based on four different models: the initial back-propagation artificial neural network, the ARIMA model, the ARIMA with data from wavelet transforming (WT-ARIMA), and the hybrid BP network with optimal parameters as well as the preprocessed data. Through the comparison of data, it can be found that the MAPE of WAFSA-BPANN is relatively more smooth and smaller than the other three models. Here, we can know that the wind speed in February has a remarkable fluctuation, which results in a relatively big error in all four models. Through further analysis, a rough speculation can be formed: compared with other times, the falling of temperature in the winter may lead to the fluctuation of the wind speed.

5.3. Discussion of the Results from Different Observation Sites. To test the stability of the experiments as well as the effectiveness of the proposed hybrid model, the scope of the training data and testing data was enlarged, including four whole months in different seasons. At the same time, horizontal comparison is also employed to strengthen the validity of the proposed hybrid models. Furthermore, to eliminate random error in wind speed data, data fields from three different observation sites are applied in the process of simulation and forecasting. All these efforts aim to reduce the randomness of the experiments as well as testify the effectiveness of the proposed models, and the analysis results of data series gathered from three different observation sites are presented in Tables 6–9.

Obvious improvements in the performance of the hybrid models can be noticed, which can be read from all three error criteria. And the superiority of our WAFSA-BPANN model exists in observation site 1, site 2, and site 3. As is depicted in Table 6, it is not hard to find that there are 1.28%, 1.6%, 1.31%, 1.18%, and 1.13% reductions in MAPE, 0.1419, 0.1586, 0.1447, 0.1113, and 0.1199 reductions in MSE, and 0.0871, 0.1048, 0.0888, 0.073, and 0.0779 reductions in MAE, respectively, of the hybrid model proposed in this paper when compared with BPANN, ARIMA, AFSA-BPANN, WT-BPANN, and WT-ARIMA model.

On other aspects, through the horizontal comparison of the data shown in Tables 6, 7, 8, and 9, WT's superior performance on the three error indexes, compared with the performances of the other traditional models, can be seen. The average index of errors in all tables covering four seasons also indicates this rule: the hybrid model indeed has much more superiority when dealing with the original data set, which contains much noise and fluctuation. In other words, the process of denoising of WT has a critical effect on the final forecasting when analysing the comprehensive comparison, from which we can find an obvious improvement for

Figure 4: The specific results of different models.

the model when the wavelet transform is added. Synthesizing these diagrams, a rough rule can be observed, showing that the wavelet transform has an impact on simulation and forecasting: the tendency of the error index is more flat compared with the original error. The hybrid model (WAFSA-BPANN)

can at least be confirmed to finish forecasting the wind speed with a tolerated error.

Moreover, it is necessary to consider the running time when implementing these models. On average, the BPANN model is the fastest which usually costs less than two seconds,

TABLE 2: The forecasting results of February from observation site 1 on a given time interval.

Date February	Time (h)	Actual value	BPANN		ARIMA		WT-ARIMA		WAFSA-BPANN	
			Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)
Feb. 1	0:00	7.9	7.025	11.071	7.080	10.378	7.759	1.779	7.681	2.778
	6:00	7.4	7.082	4.292	7.085	4.253	7.140	3.513	7.171	3.094
	12:00	3.1	3.123	0.743	3.014	2.782	2.981	3.832	3.158	1.880
	18:00	4.4	4.385	0.339	4.287	2.571	4.268	3.003	4.404	0.102
Feb. 2	0:00	5	4.619	7.614	4.320	13.609	4.500	9.998	4.370	12.597
	6:00	4.7	4.877	3.760	4.789	1.900	4.767	1.429	4.844	3.069
	12:00	2.8	2.825	0.899	2.492	10.995	2.570	8.218	2.823	0.838
	18:00	8.2	7.712	5.946	7.623	7.033	7.859	4.158	7.859	4.153
Feb. 3	0:00	8.3	8.179	1.461	8.180	1.443	8.221	0.947	8.478	2.150
	6:00	8.4	8.388	0.146	8.253	1.746	8.442	0.505	8.520	1.428
	12:00	4.6	5.234	13.778	5.172	12.440	5.239	13.895	5.224	13.568
	18:00	5.4	5.167	4.314	5.139	4.829	5.107	5.431	5.375	0.469
Feb. 4	0:00	3	3.048	1.616	2.948	1.728	2.867	4.435	3.056	1.869
	6:00	4.7	4.553	3.126	4.692	0.165	4.383	6.750	4.831	2.783
	12:00	5.6	5.046	9.896	4.412	21.211	4.785	14.558	4.639	17.166
	18:00	5.6	5.650	0.888	5.596	0.079	5.583	0.304	5.519	1.452
Feb. 5	0:00	6	6.075	1.255	6.073	1.214	5.960	0.669	6.196	3.264
	6:00	3	3.325	10.835	3.256	8.517	3.176	5.874	3.274	9.138
	12:00	6.3	6.925	9.914	6.936	10.096	6.875	9.130	6.701	6.360
	18:00	7.5	8.285	10.467	8.300	10.672	8.118	8.236	8.306	10.742
Feb. 6	0:00	15	13.412	10.590	13.409	10.606	13.222	11.850	14.511	3.263
	6:00	9.4	9.374	0.274	9.309	0.965	9.445	0.475	9.136	2.809
	12:00	8.6	9.010	4.771	8.956	4.136	9.058	5.323	8.938	3.933
	18:00	6.5	6.402	1.502	6.300	3.073	6.421	1.214	6.601	1.556
Feb. 7	0:00	4.6	4.679	1.721	4.659	1.278	4.601	0.023	4.502	2.135
	6:00	3.4	3.716	9.308	3.488	2.591	3.610	6.177	3.556	4.583
	12:00	8.7	10.145	16.614	10.084	15.904	10.176	16.970	9.207	5.830
	18:00	7.9	8.217	4.010	8.219	4.044	7.995	1.196	8.179	3.535

and this can be seen in Tables 6, 7, 8, and 9. It is still very quick even the process of wavelet transform is added. The ARIMA and WT-ARIMA model are relatively fast with the running time about forty seconds. However, the speed of model always slows down when it involves the AFSA algorithm, which usually lasts 150 seconds. Through further analysis, it is not hard to explain the longer period of hybrid model proposed in this paper since there are so many loops and iterations in the AFSA algorithm.

Though much more time is needed when applying the hybrid model, it is still a reasonable choice to make wind speed forecasting for wind farms, for the improvement of prediction accuracy can compensate the costs in time and bring more profits both in technology and economy.

5.4. Comprehensive Comparison between Different Models. Table 10 shows the final comprehensive comparison between the traditional and hybrid models. Based on the metrics tabularized in Table 10, the overall perspective of each model can be seen distinctly. It should be noted here that all of those digits are the means of the whole experiment. For instance, the MAE of the BPANN model is 0.4284, which is the mean of the aforementioned forecasting results from these three

observations sites covering four seasons. The MSE and MAPE are formed in the similar way.

It can be found that, when compared with BPANN, ARIMA, AFSA-BPANN, WT-BPANN, and WT-ARIMA, there are 24.7%, 25.5%, 24.78%, 24.64%, and 25.53% improvements in MAE, 45.25%, 46.2%, 45.01%, 45.58%, and 46.93% improvements in MSE, and 21.28%, 21.17%, 21.87%, 20.1%, and 20.52% improvements in MAPE, respectively, of the hybrid model proposed in this paper. Similarly, reductions in MAPE, MSE, and MAE occurred at all three observation sites in four seasons which are presented in aforementioned tables.

Through the aforementioned table and analysis, a general rule could be summarised now: the proposed WAFSA-BPANN can always perform better than other referred models, which is obviously embodied in MAPE, MAE, and MSE. Above all, a visualized conclusion can be reached: the proposed hybrid model performs better than the others. In summary, the proposed optimal hybrid model can, to some extent, make relatively better forecasts than the traditional prediction models. In other words, the optimal algorithm introduced in this paper is suitable to be applied to optimizing the traditional back-propagation artificial neural network based on the wind speed data.

TABLE 3: The forecasting results of June from observation site 1 on a given time interval.

Date June	Time (h)	Actual value	BPANN		ARIMA		WT-ARIMA		WAFSA-BPANN	
			Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)
June. 1	0:00	2.9	4.423	52.508	4.293	48.044	3.292	13.530	3.529	21.693
	6:00	3.7	2.748	25.728	2.653	28.287	2.540	31.358	3.387	8.449
	12:00	2.6	3.235	24.425	3.144	20.929	3.126	20.230	2.991	15.027
	18:00	1.8	2.309	28.265	2.138	18.759	2.078	15.423	2.151	19.499
June. 2	0:00	1.6	1.690	5.632	1.464	8.529	1.385	13.407	1.730	8.116
	6:00	0.7	1.389	98.435	1.091	55.881	1.105	57.881	1.543	120.427
	12:00	6.1	6.150	0.817	6.133	0.537	6.247	2.403	6.261	2.636
	18:00	9.5	9.949	4.728	9.949	4.726	10.032	5.595	9.857	3.758
June. 3	0:00	10.8	10.788	0.112	10.685	1.063	10.748	0.481	10.982	1.684
	6:00	12	12.724	6.034	12.742	6.183	12.790	6.581	12.276	2.300
	12:00	8.9	9.151	2.815	9.332	4.859	9.199	3.357	9.304	4.542
	18:00	2.4	1.813	24.455	1.785	25.627	1.631	32.040	2.172	9.482
June. 4	0:00	2.4	2.376	0.983	2.200	8.338	2.207	8.022	2.340	2.509
	6:00	1.9	2.143	12.795	1.901	0.049	1.900	0.024	1.922	1.150
	12:00	2.9	3.028	4.404	2.844	1.943	2.915	0.504	2.872	0.970
	18:00	5.4	4.435	17.873	4.395	18.611	4.410	18.337	4.398	18.551
June. 5	0:00	8	7.444	6.956	7.369	7.886	7.566	5.429	7.321	8.491
	6:00	1.3	1.570	20.753	1.400	7.701	1.269	2.365	1.763	35.644
	12:00	2.2	2.652	20.543	2.482	12.801	2.487	13.066	2.665	21.118
	18:00	5.9	5.879	0.351	5.781	2.015	5.895	0.081	5.704	3.321
June. 6	0:00	7.8	7.861	0.776	7.841	0.525	7.884	1.081	8.072	3.482
	6:00	8.3	8.286	0.167	8.312	0.150	8.277	0.283	8.431	1.581
	12:00	9.5	8.627	9.190	8.635	9.106	8.702	8.398	9.268	2.443
	18:00	14.2	12.340	13.098	12.331	13.165	12.412	12.593	12.911	9.078
June. 7	0:00	11.4	10.582	7.179	10.742	5.775	10.578	7.214	11.174	1.982
	6:00	7	6.839	2.302	6.859	2.014	6.914	1.229	6.988	0.175
	12:00	7.3	6.744	7.614	6.845	6.235	6.794	6.938	7.374	1.013
	18:00	10.3	10.616	3.068	10.546	2.388	10.748	4.348	10.445	1.404

6. Conclusions

With the ongoing need for wind power, accurate wind speed forecasting has a strong impact on wind farm management, even on the entire wind power dispatch hybrid system. Therefore, a new hybrid model is proposed in this paper to solve this tough task. Firstly, based on WT, extra noise in the original wind speed data series which are nonstationary is eliminated while the remaining data contain most of the effective information, which is critical for forecasting wind speed. Next, to achieve the most proper parameters of the back-propagation artificial neural network, AFSA, as an effective method, is applied in this paper, and with the optimal parameters, the initial network is formed. Then, the preprocessed wind speed data are utilized to train the BPANN artificial neural network; in the final process, the forecasting of wind speed is obtained from the trained network. A number of sets of wind speed data from four different observation sites are gathered to execute the simulation and forecasting. Meanwhile, three error metrics are built up to evaluate the performance of the models. Through comprehensive comparison, the conclusion elicited is that the proposed intelligent hybrid model (WAFSA-BPANN) outperforms the other traditional models, such as BPANN, ARIMA, AFSA-BPANN, WT-BPANN, and WT-ARIMA. Hence, WAFSA-BPANN is an effective and reliable model for short-term wind speed forecasting in wind farms of China.

Though relatively superior precision of wind speed prediction can be obtained applying the hybrid model proposed in this paper, there is still room for improvements. Investigating these forecasting approaches, such as the combination of numerical simulation and intelligence algorithms, could be a future research topic.

Abbreviations

AFSA: Artificial fish swarm algorithm
ANN: Artificial neural network
AR: Auto Regressive Model
ARIMA: Auto Regressive Integrated Moving Average
BEVs: Battery electric vehicles
BMS: Battery management system
BPANN: Back-propagation artificial neural network
ESM: Exponential Smoothing Method
EVs: Electric vehicles
GM(1, n): Grey prediction model
KF: Kalman filter

TABLE 4: The forecasting results of August from observation site 1 on a given time interval.

Date August	Time (h)	Actual value	BPANN		ARIMA		WT-ARIMA		WAFSA-BPANN	
			Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)
Aug. 1	0:00	4.8	4.539	5.448	4.595	4.263	4.403	8.263	4.639	3.356
	6:00	4.1	4.536	10.642	4.529	10.454	4.442	8.341	3.897	4.951
	12:00	2.4	2.586	7.731	2.434	1.437	2.493	3.869	2.579	7.472
	18:00	1.7	1.751	3.013	1.566	7.872	1.510	11.184	1.838	8.139
Aug. 2	0:00	3.5	3.865	10.423	3.787	8.201	3.775	7.845	3.848	9.951
	6:00	3	2.971	0.959	2.850	4.986	2.816	6.137	2.838	5.413
	12:00	2.4	2.912	21.340	2.796	16.501	2.835	18.126	2.642	10.068
	18:00	3.3	3.421	3.669	3.323	0.702	3.302	0.049	3.352	1.585
Aug. 3	0:00	2.8	2.853	1.900	2.816	0.583	2.682	4.230	2.875	2.676
	6:00	5.1	4.442	12.903	4.376	14.201	4.345	14.800	4.379	14.141
	12:00	4.3	3.970	7.675	3.939	8.400	3.886	9.621	3.775	12.218
	18:00	2	2.168	8.380	2.060	2.978	2.087	4.344	2.383	19.144
Aug. 4	0:00	2.3	1.925	16.306	1.740	24.338	1.743	24.214	2.298	0.088
	6:00	3	3.221	7.362	3.134	4.451	3.099	3.287	3.428	14.257
	12:00	3.1	2.504	19.233	2.418	22.009	2.411	22.242	3.022	2.504
	18:00	2.8	2.912	3.999	2.858	2.064	2.815	0.521	2.661	4.980
Aug. 5	0:00	1.8	1.648	8.435	1.529	15.072	1.479	17.849	1.888	4.861
	6:00	0.8	1.581	97.609	1.474	84.254	1.398	74.703	1.514	89.304
	12:00	1.4	1.718	22.680	1.567	11.911	1.609	14.911	1.897	35.478
	18:00	4.8	4.287	10.696	4.256	11.328	4.387	8.607	4.635	3.429
Aug. 6	0:00	1.3	1.444	11.099	1.284	1.226	1.161	10.729	1.746	34.313
	6:00	3.3	3.136	4.964	3.120	5.445	3.076	6.792	3.039	7.923
	12:00	3.9	3.719	4.651	3.689	5.399	3.600	7.684	3.472	10.980
	18:00	1.9	2.598	36.716	2.455	29.233	2.511	32.145	2.490	31.040
Aug. 7	0:00	2.6	2.796	7.526	2.764	6.291	2.782	7.007	2.787	7.203
	6:00	3.8	3.768	0.837	3.710	2.381	3.772	0.733	3.762	0.989
	12:00	4.5	4.019	10.688	3.996	11.189	3.995	11.233	3.973	11.722
	18:00	12.1	10.417	13.907	10.259	15.216	10.559	12.733	11.377	5.975

TABLE 5: The forecasting results of October from observation site 1 on a given time interval.

Date October	Time (h)	Actual value	BPANN		ARIMA		WT-ARIMA		WAFSA-BPANN	
			Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)	Forecasting	MAPE (%)
Oct. 2	0:00	7.5	7.452	0.645	7.384	1.541	6.322	15.709	7.160	4.527
	6:00	8.5	8.185	3.706	8.129	4.368	8.368	1.548	7.955	6.406
	12:00	9.1	9.412	3.431	9.422	3.543	9.511	4.513	9.228	1.408
	18:00	10.7	11.173	4.423	10.695	0.051	10.442	2.408	10.988	2.690
Oct. 3	0:00	10.5	11.719	11.610	11.723	11.649	11.765	12.052	10.970	4.480
	6:00	6.5	6.817	4.883	6.912	6.337	6.770	4.148	7.017	7.957
	12:00	5.7	5.273	7.492	5.344	6.247	5.082	10.841	5.796	1.683
	18:00	4.4	4.188	4.818	4.166	5.316	4.199	4.571	4.207	4.394
Oct. 4	0:00	3.8	4.424	16.428	4.477	17.817	4.434	16.675	4.179	9.983
	6:00	3.3	3.432	3.986	3.353	1.592	3.392	2.781	3.212	2.677
	12:00	3.1	3.176	2.442	2.955	4.673	3.075	0.795	2.574	16.962
	18:00	2.9	2.579	11.065	2.474	14.683	2.476	14.619	2.667	8.043
Oct. 5	0:00	4.3	4.483	4.267	4.449	3.459	4.499	4.622	4.317	0.392
	6:00	3.8	3.921	3.187	3.918	3.114	3.894	2.481	4.034	6.146
	12:00	6.6	6.225	5.685	6.190	6.214	6.159	6.687	6.106	7.480
	18:00	7.5	6.712	10.508	6.557	12.568	6.865	8.468	7.105	5.272
Oct. 6	0:00	9.4	8.985	4.412	8.982	4.450	9.020	4.038	9.024	4.000
	6:00	8.6	8.160	5.116	8.164	5.067	8.075	6.103	8.192	4.740
	12:00	6.8	7.209	6.008	7.286	7.140	7.216	6.119	7.328	7.765
	18:00	4.2	4.300	2.391	4.402	4.814	4.275	1.794	4.621	10.013
Oct. 7	0:00	3.9	4.456	14.266	4.519	15.864	4.371	12.067	4.354	11.645
	6:00	2.6	2.573	1.038	2.512	3.398	2.484	4.467	2.441	6.111
	12:00	3.1	3.502	12.952	3.448	11.232	3.519	13.511	3.046	1.744
	18:00	4.6	4.641	0.884	4.675	1.624	4.632	0.689	4.670	1.522

Table 6: Errors of different traditional models in February.

	Models	MAE (m/s)	MSE (m²/s²)	MAPE (%)	Running time (s)
	BPANN	0.3911	0.3041	6.68	2.8502
	ARIMA	0.4088	0.3208	7	47.2708
Observation site 1	AFSA-BPANN	0.3928	0.3069	6.71	105.3652
	WT-BPANN	0.3770	0.2735	6.58	5.1081
	WT-ARIMA	0.3819	0.2821	6.53	40.7528
	WAFSA-BPANN	**0.3040**	**0.1622**	**5.4**	156.3590
	BPANN	0.4219	0.3304	9.5	2.0659
	ARIMA	0.4415	0.3526	9.94	35.5574
Observation site 2	AFSA-BPANN	0.4307	0.3389	9.78	106.2827
	WT-BPANN	0.4087	0.3045	9	2.4668
	WT-ARIMA	0.4130	0.3109	9.03	42.0537
	WAFSA-BPANN	**0.3134**	**0.1674**	**7.44**	148.8582
	BPANN	0.3975	0.2845	8.27	0.5424
	ARIMA	0.4098	0.3021	8.43	42.4113
Observation site 3	AFSA-BPANN	0.3949	0.2793	8.24	105.7024
	WT-BPANN	0.3833	0.2654	8.05	0.7321
	WT-ARIMA	0.3849	0.2694	7.92	40.7567
	WAFSA-BPANN	**0.3055**	**0.1601**	**6.87**	146.8292

Table 7: Errors of different traditional models in June.

	Models	MAE (m/s)	MSE (m²/s²)	MAPE (%)	Running time (s)
	BPANN	0.4637	0.4260	10.48	0.8509
	ARIMA	0.4617	0.4215	10.06	37.0557
Observation site 1	AFSA-BPANN	0.4701	0.4267	10.91	104.7545
	WT-BPANN	0.4625	0.4278	10.05	1.4735
	WT-ARIMA	0.4663	0.4315	10.08	37.9540
	WAFSA-BPANN	**0.3381**	**0.2066**	**8.14**	145.9524
	BPANN	0.4446	0.3783	9.99	1.1537
	ARIMA	0.4442	0.3737	9.78	41.4958
Observation site 2	AFSA-BPANN	0.4443	0.3770	10.07	108.9195
	WT-BPANN	0.4450	0.3849	9.76	73.93
	WT-ARIMA	0.4462	0.3859	9.78	42.0491
	WAFSA-BPANN	**0.3441**	**0.2084**	**8.27**	143.1720
	BPANN	0.4716	0.3987	11.78	0.5463
	ARIMA	0.4671	0.3930	11.09	36.7596
Observation site 3	AFSA-BPANN	0.4720	0.3989	11.7	104.6663
	WT-BPANN	0.4700	0.4144	11.17	0.7220
	WT-ARIMA	0.4727	0.4198	11.08	41.9609
	WAFSA-BPANN	**0.3353**	**0.2008**	**8.64**	142.3210

MA: Moving Average Model
MAE: Mean absolute error
MAPE: Mean absolute percentage error
MSE: Mean square error
NWPs: Numerical Weather Prediction Systems
NEA: National energy administration
PHEVs: Plug-in hybrid electric vehicles

RBF: Radial Basis Function
RES: Renewable energy source
WT: Wavelet transform.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

TABLE 8: Errors of different traditional models in August.

	Models	MAE (m/s)	MSE (m²/s²)	MAPE (%)	Running time (s)
Observation site 1	BPANN	0.4595	0.3856	7.95	1.0696
	ARIMA	0.4657	0.3989	8.05	39.0358
	AFSA-BPANN	0.4612	0.3840	7.98	102.2733
	WT-BPANN	0.4668	0.3956	8.05	1.0735
	WT-ARIMA	0.4787	0.4201	8.25	43.0450
	WAFSA-BPANN	**0.3377**	**0.2055**	**5.94**	147.2995
Observation site 2	BPANN	0.4189	0.3351	8.12	0.4442
	ARIMA	0.4189	0.3396	8.13	37.0051
	AFSA-BPANN	0.4139	0.3259	8	105.3182
	WT-BPANN	0.4239	0.3438	8.17	0.9281
	WT-ARIMA	0.4303	0.3563	8.3	37.1158
	WAFSA-BPANN	**0.3043**	**0.1831**	**6.07**	145.4899
Observation site 3	BPANN	0.4601	0.3883	8.26	2.2823
	ARIMA	0.4689	0.3981	8.48	37.7879
	AFSA-BPANN	0.4600	0.3850	8.32	102.6349
	WT-BPANN	0.4665	0.3928	8.39	2.0385
	WT-ARIMA	0.4727	0.4044	8.48	40.7267
	WAFSA-BPANN	**0.3407**	**02140**	**6.35**	142.8692

TABLE 9: Errors of different traditional models in October.

	Models	MAE (m/s)	MSE (m²/s²)	MAPE (%)	Running time (s)
Observation site 1	BPANN	0.3824	0.2768	7.2	0.5651
	ARIMA	0.3827	0.2766	7.18	32.6927
	AFSA-BPANN	0.3827	0.2752	7.22	110.4872
	WT-BPANN	0.3898	0.2954	7.25	0.6419
	WT-ARIMA	0.3930	0.3019	7.27	35.1270
	WAFSA-BPANN	**0.2963**	**0.1503**	**5.79**	151.7133
Observation site 2	BPANN	0.4141	0.3206	8.05	0.8032
	ARIMA	0.4122	0.3257	7.89	33.0404
	AFSA-BPANN	0.4157	0.3238	8.11	109.5214
	WT-BPANN	0.4274	0.3511	8.11	0.8312
	WT-ARIMA	0.4301	0.3526	8.16	35.2028
	WAFSA-BPANN	**0.3164**	**0.1983**	**6.34**	147.8061
Observation site 3	BPANN	0.4156	0.3245	8.75	0.5211
	ARIMA	0.4148	0.3242	8.85	35.9502
	AFSA-BPANN	0.4085	0.3132	8.78	105.3837
	WT-BPANN	0.4167	0.3288	8.9	0.5227
	WT-ARIMA	0.4297	0.3504	9.15	36.1737
	WAFSA-BPANN	**0.3356**	**0.2173**	**7.45**	142.9675

TABLE 10: Comprehensive comparison of the performances of models involved.

Model	MAE (m/s)	Improvement (%)	MSE (m²/s²)	Improvement (%)	MAPE (%)	Improvement (%)	Running time (s)
BPANN	0.4284	**24.70**	0.3461	**45.25**	8.7525	**21.28**	1.14
ARIMA	0.4330	**25.50**	0.3522	**46.20**	8.74	**21.17**	38.0052
AFSA-BPANN	0.4289	**24.78**	0.3446	**45.01**	8.8183	**21.87**	105.9424
WT-BPANN	0.4281	**24.64**	0.3482	**45.58**	8.6233	**20.1**	1.4398
WT-ARIMA	0.4332	**25.53**	0.3571	**46.93**	8.6691	**20.52**	39.4098
WAFSA-BPANN	**0.3226**	—	**0.1895**	—	**6.89**	—	146.8031

Acknowledgments

References

[1] J.-M. Hu, J.-Z. Wang, and G.-W. Zeng, "A hybrid forecasting approach applied to wind speed time series," *Renewable Energy*, vol. 60, pp. 185–194, 2013.

[2] A. J. Lamadrid, "Optimal use of energy storage systems with renewable energy sources," *International Journal of Electrical Power & Energy Systems*, vol. 71, pp. 101–111, 2015.

[3] X. Luo, J. Wang, M. Dooner, and J. Clarke, "Overview of current development in electrical energy storage technologies and the application potential in power system operation," *Applied Energy*, vol. 137, pp. 511–536, 2015.

[4] L. Zhang, Z. Wang, F. Sun, and D. G. Dorrell, "Online parameter identification of ultracapacitor models using the extended Kalman filter," *Energies*, vol. 7, no. 5, pp. 3204–3217, 2014.

[5] N. Juul and P. Meibom, "Road transport and power system scenarios for Northern Europe in 2030," *Applied Energy*, vol. 92, pp. 573–582, 2012.

[6] T. Takeshita, "Assessing the co-benefits of CO_2 mitigation on air pollutants emissions from road vehicles," *Applied Energy*, vol. 97, pp. 225–237, 2012.

[7] D. Kum, H. Peng, and N. K. Bucknor, "Optimal control of plug-in HEVs for fuel economy under various travel distances," in *Proceedings of the 6th IFAC Symposium Advances in Automotive Control (AAC '10)*, pp. 258–263, Munich, Germany, July 2010.

[8] W. Sung and C. B. Shin, "Electrochemical model of a lithium-ion battery implemented into an automotive battery management system," *Computers & Chemical Engineering*, vol. 76, pp. 87–97, 2015.

[9] B. Johansson and A. Mårtensson, "Energy and environmental costs for electric vehicles using CO_2-neutral electricity in Sweden," *Energy*, vol. 25, no. 8, pp. 777–792, 2000.

[10] Y. Kudoh, H. Ishitani, R. Matsuhashi et al., "Environmental evaluation of introducing electric vehicles using a dynamic traffic-flow model," *Applied Energy*, vol. 69, no. 2, pp. 145–159, 2001.

[11] D. MeCimorec and Z. Tomsic, "Portfolio theory application in wind potential assessment," *Renewable Energy*, vol. 76, pp. 494–502, 2015.

[12] X.-C. Wang, P. Guo, and X.-B. Huang, "A review of wind power forecasting models," *Energy Procedia*, vol. 12, pp. 770–778, 2011.

[13] J. Z. Wang, S. S. Qin, Q. P. Zhou, and H. Y. Jiang, "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China," *Renewable Energy*, vol. 76, pp. 91–101, 2015.

[14] C. Croonenbroeck and D. Ambach, "A selection of time series models for short- to medium-term wind power forecasting," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 136, pp. 201–210, 2015.

[15] Y. Zhang, J.-X. Wang, and X.-F. Wang, "Review on probabilistic forecasting of wind power generation," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.

[16] J. Szoplik, "Forecasting of natural gas consumption with artificial neural networks," *Energy*, vol. 85, pp. 208–220, 2015.

[17] Y.-Y. Hong, H.-L. Chang, and C.-S. Chiu, "Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs," *Energy*, vol. 35, no. 9, pp. 3870–3876, 2010.

[18] B. Ernst, B. Oakleaf, M. L. Ahlstrom et al., "Predicting the wind," *IEEE Power and Energy Magazine*, vol. 5, no. 6, pp. 78–89, 2007.

[19] S. Baran, "Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components," *Computational Statistics & Data Analysis*, vol. 75, pp. 227–238, 2014.

[20] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 258–265, 2007.

[21] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Applied Energy*, vol. 67, no. 1-2, pp. 17–35, 2000.

[22] J. C. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332–341, 1992.

[23] W.-Y. Liu and J.-G. Han, "The optimal Mexican hat wavelet filter de-noising method based on cross-validation method," *Neurocomputing*, vol. 108, no. 2, pp. 31–35, 2013.

[24] A. Bakhtazad, A. Palazoglu, and J. A. Romagnoli, "Process data de-noising using wavelet transform," *Intelligent Data Analysis*, vol. 3, no. 4, pp. 267–285, 1999.

[25] H. Liu, H.-Q. Tian, Y.-F. Li, and L. Zhang, "Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions," *Energy Conversion and Management*, vol. 92, pp. 67–81, 2015.

[26] H. D. Tran, N. Muttil, and B. J. C. Perera, "Selection of significant input variables for time series forecasting," *Environmental Modelling & Software*, vol. 64, pp. 156–163, 2015.

[27] M. Ghofrani, M. Ghayekhloo, A. Arabali, and A. Ghayekhloo, "A hybrid short-term load forecasting with a new input selection framework," *Energy*, vol. 81, pp. 777–786, 2015.

[28] R. Hecht-Nielsen, "Kolmogorov's mapping neural network existence theorem," in *Proceedings of the International Conference on Neural Networks*, 1987.

[29] A. Suresh, K. V. Harish, and N. Radhika, "Particle swarm optimization over back propagation neural network for length of stay prediction," *Procedia Computer Science*, vol. 46, pp. 268–275, 2015.

[30] X.-L. Li, Z.-J. Shao, and J.-X. Qian, "An optimizing method based on autonomous animats: fish-swarm Algorithm," *System Engineering Theory and Practice*, vol. 22, no. 11, pp. 32–38, 2002.

[31] N. Fang, J.-Z. Zhou, R. Zhang, Y. Liu, and Y.-C. Zhang, "A hybrid of real coded genetic algorithm and artificial fish swarm algorithm for short-term optimal hydrothermal scheduling," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 617–629, 2014.

[32] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[33] S. S. Qin, F. Liu, J.-Z. Wang, and Y.-L. Song, "Interval forecasts of a novelty hybrid model for wind speeds," *Energy Reports*, vol. 1, pp. 8–16, 2015.

Research Article

A Multiperiod Vehicle Lease Planning for Urban Freight Consolidation Network

Woosuk Yang,[1] Taesu Cheong,[2] and Sang Hwa Song[3]

[1]*Korea Investment & Securities, Seoul 1507450, Republic of Korea*
[2]*School of Industrial Management Engineering, Korea University, Seoul 136713, Republic of Korea*
[3]*Graduate School of Logistics, Incheon National University, Incheon 406130, Republic of Korea*

Correspondence should be addressed to Sang Hwa Song; songsh@incheon.ac.kr

This paper considers a multiperiod vehicle lease planning problem for urban freight consolidation centers (UFCCs) in the urban freight transport network where short-term-leased and long-term-leased vehicles are hired together. The objective is to allocate the two kinds of leased vehicles optimally for direct transportation services from the associated origin node to the associated UFCC or from the associated UFCC to the associated destinations so as to satisfy a given set of period-to-period freight demands over a given planning horizon at total minimum vehicle allocation cost subject to demand-dependent transportation time restriction. The problem is formulated as an integer programming model and proven to be NP-hard in a strong sense. Thus, a Lagrangian heuristic is proposed to find a good solution efficiently. Numerical experiments show that the proposed algorithm finds good lower and upper bounds within reasonable time.

1. Introduction

Recent worldwide economic growth has accelerated urbanization, and people have been moving to cities for better jobs and more fulfilling lifestyle. In the late 2000s, it was reported that more people lived in cities than in rural areas, and it is expected that more people will move to urban areas in the near future [1, 2]. According to Blanco and Fransoo [3], the number of megacities with at least 10 million people is increasing, and in a decade, their contribution to world economy will steadily grow and constitute more than 20% of world GDP. As people start to live in concentrated and sometimes congested areas, traffic jams and air pollution continue to increase. As a result, many researchers have started to investigate the impacts of congestion due to urban transportation systems [2, 4].

Freight transportation is one of the key contributors to traffic congestion and harmful pollutants in cities. Unlike typical passenger cars, freight vehicles are bigger in size and move more frequently due to the nature of their business.

Furthermore, a recent trend toward just-in-time delivery makes the problem even worse [5]. To survive in competition and to meet increased customer expectation about timely delivery, companies have dispatched small packages more frequently to customers inside metropolitan areas. As average vehicle utilization went down and the number of vehicles dispatched increased, traffic congestion and the subsequent problems have been exacerbated to such an extent that they could cause serious social issues. To deal with the problems caused by urban freight transportation, integrated approaches for city logistics systems have been proposed. Crainic et al. [5] proposed an integrated short-term scheduling of operations management of logistics resources. Ehmke [1] and Ehmke et al. [6] proposed efficient routing systems integrating traffic information and logistics. Yang et al. [7] investigated a problem of designing a city logistics network considering green-house gas emissions. Thus, the research goal of urban freight transportation and city logistics should be aligned with congestion and pollution reduction with less degradation of the city center commercial activities. In this

regard, multitier urban freight transportation system with consolidation and in-out synchronization, so-called *urban freight consolidation*, has been proposed as an alternative transportation infrastructure for city logistics.

Urban freight consolidation center (UFCC) is a logistics facility located at the boundary of urban areas to serve freight transportation to city centers including retailers, buildings, and construction sites [8]. Products are transported from their origins such as manufacturers to their destinations in cities through urban consolidation centers. They are first moved to urban consolidation centers, where incoming shipments are unloaded, sorted, and consolidated with other products from different origins. Sorted and consolidated products are then transferred directly to outbound vehicles without storage. In just-in-time delivery environment, it is quite important to send orders to customers when necessary, even though shipment size is not enough to fill a full truck load. By consolidating products in an intermediate point between origins and destinations, urban consolidation centers can contribute to an increase of vehicle utilization, an increase of average size of vehicles involved, and the reduction of delivery frequency. In addition to this, the usage of environmentally friendly vehicles such as electric and clean natural gas power vehicles could often contribute to the decrease of overall harmful emissions from freight transportation vehicles. By improving the overall loading factor of a vehicle destined for congested city centers, urban consolidation centers could effectively reduce the total travel distance and further reduce the impact of freight operations on traffic congestion [9]. In fact, the concept of UFCC has been tested with real business practice in European countries and Toronto, Canada [10–12].

Research on urban freight consolidation centers focuses on the economic analysis of the consolidated freight transportation system. Su and Roorda [12] and Triantafyllou et al. [13] showed that urban consolidation centers can be successfully operated in real urban environments. It was reported that the trial systems were able to reduce harmful emissions and traffic congestion in city centers when properly managed and synchronized. Marcucci and Danielis [8] showed that, in their analysis, urban consolidation centers could attract a considerable amount of freight shipments bound to urban areas. Zhou and Wang [14] studied the issues related to development and construction of consolidation centers and showed that proper strategies based on public-private partnership can increase overall economic benefits of the system to participants in city logistics. While the feasibility of urban freight consolidation centers has been studied in depth, operation and network planning issues have been discussed (e.g., [15–17]) but still not been fully investigated yet. The detailed analysis of operation and planning of urban freight consolidation system should be done to be deployed in real business practice. In the supply chain context, an urban consolidation center is similar to the transportation system with cross-docking terminals. In cross-docking network, goods are moved through cross-docking terminals where shipments are sorted, consolidated, and transferred to outbound vehicles in a synchronized manner.

Research on operation and management of the cross-docking based consolidation strategy has progressed in two directions. One is concerned with problems that are related to the internal operations at the consolidation center. Gue [18] analyzed the effects of scheduling trailers into doors on a layout of a freight consolidation center. Bartholdi III and Gue [19, 20] designed the layout of a freight consolidation center. Li et al. [21] studied a scheduling problem to minimize storage and order picking in a consolidation center. On the other hand, planning problems on a network level have been considered to locate consolidation centers, to allocate vehicles, and to make vehicle consolidation schedules. Ratliff et al. [22] and Chen et al. [23] considered a problem of making vehicle consolidation schedules for a transportation network. Donaldson et al. [24] considered a problem of allocating vehicles and making vehicle consolidation schedules. Sung and Song [25] and Sung and Yang [26] studied an integrated model of locating consolidation centers and allocating vehicles. All the works on a network level about consolidation based transportation have considered static freight demands, not varying with time. However, freight demands in various industries such as food, apparel, electronic goods, and logistics may be dynamic. Especially in urban freight transportation, demands tend to fluctuate over time. As a cost-efficient way of allocating vehicles to satisfy dynamic freight demands, the issue of vehicle supply on lease has received much research attention, in the situation where the unit-period vehicle lease cost depends on the lease term. The unit-period long-term vehicle lease cost is generally cheaper than the unit-period short-term vehicle lease cost [27–29]. In the case of static freight demands, the long-term vehicle lease is obviously better than the short-term vehicle lease, while in the case of dynamic freight demands, either one does not dominate over the other so that the two lease options need to be considered together. Furthermore, in an urban freight consolidation setting, we need to consider consolidation of shipments which makes the lease planning more difficult.

Therefore, this paper considers a multiperiod vehicle lease planning problem in an urban freight consolidation network (MVLPUC). The problem, denoted by $\mathbb{P}_{\text{MVLPUC}}$, is concerned with optimally allocating the two kinds of leased vehicles for inbound and outbound transportation services so as to satisfy a given set of period-to-period freight demands over a given planning horizon at total minimum vehicle allocation cost subject to demand-dependent transportation time restriction. The planning horizon is divided into discrete time periods such as weeks or months. It is assumed that each freight demand is transported through a single path via one urban consolidation center (where the operations of sorting and consolidating are handled) located between origin and destination nodes, and each outbound vehicle at each urban consolidation center departs as soon as all the associated inbound vehicles arrive and the associated freight demands are sorted appropriately. It is also assumed that an unlimited number of homogeneous capacitated vehicles can be acquired through either long-term lease or short-term lease, with the unit-period short-term vehicle lease cost being greater than or equal to the unit-period long-term vehicle lease cost.

This paper is organized as follows: Section 2 introduces the problem formulation for $\mathbb{P}_{\mathrm{MVLPUC}}$. In Section 3, we discuss the procedure for solving $\mathbb{P}_{\mathrm{MVLPUC}}$ based on Lagrangian relaxation and problem decompositions. We present the numerical experiments in Section 4 and then conclude our discussion in Section 5.

2. Model

In this section, we present the mathematical formulation for Problem $\mathbb{P}_{\mathrm{MVLPUC}}$. Before introducing the model formulation, all the parameters and decision variables used in this paper are given as follows.

Sets and Parameters

$\mathcal{I}, \mathcal{J}, \mathcal{K}, \mathcal{V}$: sets of origin nodes representing manufacturers, destination nodes representing retailers, intermediate nodes representing UFCCs, and all nodes (i.e., $\mathcal{V} = \mathcal{I} \cup \mathcal{K} \cup \mathcal{J}$), respectively.

$\mathcal{E}_{\mathrm{IN}}, \mathcal{E}_{\mathrm{OUT}}$: sets of edges representing potential direct services from origin nodes to intermediate nodes and from intermediate nodes to destination nodes, respectively ($\mathcal{E}_{\mathrm{IN}} = \{e = (i,k) \mid i \in \mathcal{I}, k \in \mathcal{K}\}$ and $\mathcal{E}_{\mathrm{OUT}} = \{e = (k,j) \mid k \in \mathcal{K}, j \in \mathcal{J}\}$).

$\mathcal{E}$: set of all edges (i.e., $\mathcal{E} = \mathcal{E}_{\mathrm{IN}} \cup \mathcal{E}_{\mathrm{OUT}}$).

$\mathcal{Q}$: set of freight demands, defined by an ordered pair of two nodes (i, j) for all $i \in \mathcal{I}$ and $j \in \mathcal{J}$.

$\mathcal{Q}_k$: subset of $\mathcal{Q}$, each demand of which can be transported through the intermediated node $k \in \mathcal{K}$ within the associated transportation time restriction (to be explained later).

$\mathcal{T}$: set of time periods.

d_{ijp}: quantity of freight demand $(i, j) \in \mathcal{Q}$ at time period $p \in \mathcal{T}$.

h_{ik}^L, h_{kj}^L: unit-period long-term vehicle lease cost for $(i, k) \in \mathcal{E}_{\mathrm{IN}}$ and $(k, j) \in \mathcal{E}_{\mathrm{OUT}}$, respectively.

h_{ikp}^S, h_{kjp}^S: unit-period short-term vehicle lease cost for $(i, k) \in \mathcal{E}_{\mathrm{IN}}$ and $(k, j) \in \mathcal{E}_{\mathrm{OUT}}$, respectively, at time period $p \in \mathcal{T}$ ($h_{ik}^L \leq h_{ikp}^S$ and $h_{kj}^L \leq h_{kjp}^S$).

Γ: vehicle capacity.

t_{ik}, t_{kj}: transportation time elapsed for each $(i, k) \in \mathcal{E}_{\mathrm{IN}}$ and $(k, j) \in \mathcal{E}_{\mathrm{OUT}}$, respectively.

o_k: handling (sorting and consolidating) time at the intermediate node $k \in \mathcal{K}$.

TL_{ij}: transportation time restriction required for freight demand $(i, j) \in \mathcal{Q}$.

Decision Variables

X_{ijkp}: 1 if freight demand $(i, j) \in \mathcal{Q}$ is transported in period $p \in \mathcal{T}$ through the intermediate node $k \in \mathcal{K}$ and 0 otherwise.

Y_{ik}, Y_{kj}: numbers of long-term-leased vehicles allocated for $(i, k) \in \mathcal{E}_{\mathrm{IN}}$ and $(k, j) \in \mathcal{E}_{\mathrm{OUT}}$, respectively.

Z_{ikp}, Z_{kjp}: numbers of short-term-leased vehicles allocated for $(i, k) \in \mathcal{E}_{\mathrm{IN}}$ and $(k, j) \in \mathcal{E}_{\mathrm{OUT}}$, respectively, in period $p \in \mathcal{T}$.

We note that, for notational simplicity, notations (or subscripts) e and (i, k) (or (k, j)) are used alternately to refer to freight demand (i, k) (or (k, j)) on the corresponding edge e. Furthermore, the two terms "edge" and "direct service" are used interchangeably in this paper. Then, we now present the problem formulation for Problem $\mathbb{P}_{\mathrm{MVLPUC}}$ as follows.

Problem $\mathbb{P}_{MVLPUC}$

$$\min \quad \sum_{(i,k)\in\mathcal{E}_{\mathrm{IN}}} \left(|\mathcal{T}| h_{ik}^L Y_{ik} + \sum_{p\in\mathcal{T}} h_{ikp}^S Z_{ikp} \right)$$

(1a)

$$+ \sum_{(k,j)\in\mathcal{E}_{\mathrm{OUT}}} \left(|\mathcal{T}| h_{kj}^L Y_{kj} + \sum_{p\in\mathcal{T}} h_{kjp}^S Z_{kjp} \right)$$

$$\text{subject to} \quad \sum_{\{k\in\mathcal{K}|(i,j)\in\mathcal{Q}_k\}} X_{ijkp} = 1,$$

(1b)

$$\forall (i,j) \in \mathcal{Q}, \ \forall p \in \mathcal{T}$$

$$\sum_{\{j\in\mathcal{J}|(i,j)\in\mathcal{Q}_k\}} d_{ijp} X_{ijkp} \leq \Gamma \left(Y_{ik} + Z_{ikp} \right),$$

(1c)

$$\forall (i,k) \in \mathcal{E}_{\mathrm{IN}}, \ \forall p \in \mathcal{T}$$

$$\sum_{\{i\in\mathcal{I}|(i,j)\in\mathcal{Q}_k\}} d_{ijp} X_{ijkp} \leq \Gamma \left(Y_{kj} + Z_{kjp} \right),$$

(1d)

$$\forall (k,j) \in \mathcal{E}_{\mathrm{OUT}}, \ \forall p \in \mathcal{T}$$

$$Y_{ik}, Z_{ikp}, Y_{kj}, Z_{kjp} \in \mathbb{Z}^+,$$

$$X_{ijkp} \in \{0, 1\}.$$

(1e)

The objective function (1a) represents the cost of allocating any long-term-leased and short-term-leased vehicles for edges. Constraints (1b) imply that all the freight demands have to be serviced for each period. Constraints (1c) and (1d) require that the total amount of demands transported through any edge should not exceed the total capacity of any allocated vehicles for each period. In regard to the set $\mathcal{Q}_k$ in constraints (1b), (1c), and (1d), we assume that there exists the transportation time limit TL_{ij} for each freight demand (i, j) such that the sum of the transportation times between nodes and the handling time at an intermediate node k does not exceed TL_{ij} (i.e., $t_{ik} + o_k + t_{kj} \leq \mathrm{TL}_{ij}$). Thus, for each $k \in \mathcal{K}$, the elements of set $\mathcal{Q}_k$ can be identified as freight demands (i, j) in $\mathcal{Q}$ that satisfy the aforementioned condition with TL_{ij} at a preprocessing stage.

3. Solution Approach

In this section, we propose a heuristic approach based on Lagrangian relaxation. Problem $\mathbb{P}_{\mathrm{MVLPUC}}$ can be proven as NP-hard in a strong sense in the same manner as

shown in Sung and Song [25]. Moreover, $\mathbb{P}_{\text{MVLPUC}}$ considers time-varying demands, so it may be too complex to derive an exact algorithm to find optimal solutions even for small-sized problem instances. Therefore, we here propose an efficient heuristic method based on the Lagrangian relaxation.

3.1. Lagrangian Relaxation.
Before relaxing constraints in $\mathbb{P}_{\text{MVLPUC}}$ for Lagrangian relaxation heuristic, the following problem $\mathbb{P}_{\text{MVLPUC}_+}$ is derived by introducing another decision variable $\overline{X}_{ijkp}$ (equivalent to variables X_{ijkp}) and adding a set of constraints $X_{ijkp} \leq \overline{X}_{ijkp}$ to $\mathbb{P}_{\text{MVLPUC}}$.

Problem $\mathbb{P}_{\text{MVLPUC}_+}$

$$v_{\text{OPT}} = \min \sum_{(i,k)\in\mathscr{E}_{\text{IN}}} \left(|\mathscr{T}| h_{ik}^L Y_{ik} + \sum_{p\in\mathscr{T}} h_{ikp}^S Z_{ikp} \right)$$

$$+ \sum_{(k,j)\in\mathscr{E}_{\text{OUT}}} \left(|\mathscr{T}| h_{kj}^L Y_{kj} + \sum_{p\in\mathscr{T}} h_{kjp}^S Z_{kjp} \right) \tag{2a}$$

$$\text{subject to} \sum_{\{k\in\mathscr{K}|(i,j)\in\mathcal{Q}_k\}} X_{ijkp} = 1, \tag{2b}$$

$$\forall (i,j) \in \mathcal{Q}, \ \forall p \in \mathscr{T}$$

$$\sum_{\{j\in\mathscr{J}|(i,j)\in\mathcal{Q}_k\}} d_{ijp} X_{ijkp} \leq \Gamma \left(Y_{ik} + Z_{ikp} \right), \tag{2c}$$

$$\forall (i,k) \in \mathscr{E}_{\text{IN}}, \ \forall p \in \mathscr{T}$$

$$X_{ijkp} \leq \overline{X}_{ijkp}, \tag{2d}$$

$$\forall k \in \mathscr{K}, \ \forall (i,j) \in \mathcal{Q}_k, \ \forall p \in \mathscr{T}$$

$$\sum_{\{i\in\mathscr{J}|(i,j)\in\mathcal{Q}_k\}} d_{ijp} \overline{X}_{ijkp} \leq \Gamma \left(Y_{kj} + Z_{kjp} \right), \tag{2e}$$

$$\forall (k,j) \in \mathscr{E}_{\text{OUT}}, \ \forall p \in \mathscr{T}$$

$$Y_{ik}, Z_{ikp}, Y_{kj}, Z_{kjp} \in \mathbb{Z}^+,$$

$$X_{ijkp}, \overline{X}_{ijkp} \in \{0, 1\}. \tag{2f}$$

Problem $\mathbb{P}_{\text{MVLPUC}_+}$ has some obviously redundant variables $\overline{X}_{ijkp}$ and the associated constraints (2d) while it yields the interesting problem structure which is good to apply the Lagrangian relaxation method. That is, if constraints (2b) and (2d) in $\mathbb{P}_{\text{MVLPUC}_+}$ are Lagrangian relaxed, then the resulting problem can be decomposed into single-edge problems. Thus, constraints (2b) and (2d) are Lagrangian relaxed with Lagrange multipliers μ'_{ijp} (unrestricted) and μ''_{ijkp} (≥ 0), respectively. For given $\mu \equiv (\mu'_{ijp}, \mu''_{ijkp})_{\forall i,j,k,p}$, the resulting problem $\mathbb{P}_{\text{LR}_\mu}$ can be derived as follows.

Problem $\mathbb{P}_{\text{LR}_\mu}$

$$v(\mu) = \min \sum_{(i,j)\in\mathcal{Q}} \sum_{p\in\mathscr{T}} \mu'_{ijp} + \sum_{(i,k)\in\mathscr{E}_{\text{IN}}} \left[|\mathscr{T}| h_{ik}^L Y_{ik} + \sum_{p\in\mathscr{T}} \left(h_{ikp}^S Z_{ikp} - \sum_{\{j\in\mathscr{J}|(i,j)\in\mathcal{Q}_k\}} \theta_{ijkp} X_{ijkp} \right) \right]$$

$$+ \sum_{(k,j)\in\mathscr{E}_{\text{OUT}}} \left[|\mathscr{T}| h_{kj}^L Y_{kj} + \sum_{p\in\mathscr{T}} \left(h_{kjp}^S Z_{kjp} - \sum_{\{i\in\mathscr{J}|(i,j)\in\mathcal{Q}_k\}} \delta_{ijkp} \overline{X}_{ijkp} \right) \right] \tag{3}$$

$$\text{subject to} \quad (2c), (2e), (2f),$$

$$\text{where} \quad \theta_{ijkp} = \mu'_{ijp} - \mu''_{ijkp}, \ \delta_{ijkp} = \mu''_{ijkp}.$$

It is evident that $v(\mu)$ is a lower bound on v_{OPT} for any given μ. Let $v^*(\mu) = \max_\mu v(\mu)$, and let v_{LP} be a lower bound (on v_{OPT}) obtained by solving the linear programming (LP) relaxation problem of $\mathbb{P}_{\text{MVLPUC}_+}$. Then, the proposed Lagrangian relaxation $\mathbb{P}_{\text{LR}_\mu}$ provides a good lower bound as indicated in Proposition 1. Since the proposition below is obvious to show, we omit the proof.

Proposition 1. *Consider $v_{\text{OPT}} \geq v^*(\mu) \geq v_{\text{LP}}$.*

Furthermore, as mentioned earlier, Problem $\mathbb{P}_{\text{LR}_\mu}$ can be decomposed into single-edge problems of selecting demand and allocating vehicles (SEPDV), resulting in $\mathbb{P}_{\text{SEPDV}_{ik,\mu}}$ and $\mathbb{P}_{\text{SEPDV}_{kj,\mu}}$ for each $(i,k) \in \mathscr{E}_{\text{IN}}$ and $(k,j) \in \mathscr{E}_{\text{OUT}}$, respectively, as follows.

Problem $\mathbb{P}_{\text{SEPDV}_{ik,\mu}}$

$$v_{\text{SEPDV}}^{ik}(\mu) = \min \quad |\mathscr{T}| h_{ik}^L Y_{ik} + \sum_{p\in\mathscr{T}} \left(h_{ikp}^S Z_{ikp} - \sum_{\{j\in\mathscr{J}|(i,j)\in\mathcal{Q}_k\}} \theta_{ijkp} X_{ijkp} \right) \tag{4a}$$

$$\text{subject to} \quad \sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_k\}} d_{ijp} X_{ijkp} \leq \Gamma \left(Y_{ik} + Z_{ikp} \right), \quad \forall p \in \mathcal{T} \tag{4b}$$

$$Y_{ik}, Z_{ikp} \in \mathbb{Z}^+, \; X_{ijkp} \in \{0,1\}, \quad \forall p \in \mathcal{T}. \tag{4c}$$

Problem $\mathbb{P}_{SEPDV_{kj,\mu}}$

$$v_{SEPDV}^{kj}(\mu) = \min \quad |\mathcal{T}| h_{kj}^L Y_{kj}$$

$$+ \sum_{p \in \mathcal{T}} \left(h_{kjp}^S Z_{kjp} - \sum_{\{i \in \mathcal{J} | (i,j) \in \mathcal{Q}_k\}} \delta_{ijkp} \overline{X}_{ijkp} \right)$$

$$\text{subject to} \quad \sum_{\{i \in \mathcal{J} | (i,j) \in \mathcal{Q}_k\}} d_{ijp} \overline{X}_{ijkp} \leq \Gamma \left(Y_{kj} + Z_{kjp} \right), \tag{5}$$

$$\forall p \in \mathcal{T}$$

$$Y_{kj}, Z_{kjp} \in \mathbb{Z}^+, \; \overline{X}_{ijkp} \in \{0,1\}, \quad \forall p \in \mathcal{T}.$$

It is straightforward to show that, for given μ,

$$v(\mu) = \sum_{(i,j) \in \mathcal{Q}} \sum_{p \in \mathcal{T}} \mu'_{ijp} + \sum_{(i,k) \in \mathcal{E}_{IN}} v_{SEPDV}^{ik}(\mu)$$

$$+ \sum_{(k,j) \in \mathcal{E}_{OUT}} v_{SEPDV}^{kj}(\mu) \tag{6}$$

and hence, $v(\mu)$ can be evaluated through $v_{SEPDV}^{ik}(\mu)$ and $v_{SEPDV}^{kj}(\mu)$. Since $\mathbb{P}_{SEPDV_{ik,\mu}}$ and $\mathbb{P}_{SEPDV_{kj,\mu}}$ have the same problem structure, we only discuss the procedure for solving $\mathbb{P}_{SEPDV_{ik,\mu}}$ in the next section.

3.2. Solving $\mathbb{P}_{SEPDV_{ik,\mu}}$. In this section, we present the solution procedure for $\mathbb{P}_{SEPDV_{ik,\mu}}$ so that we can eventually evaluate the value $v(\mu)$ of $\mathbb{P}_{LR_\mu}$ for each μ. First, for given μ, Proposition 2 and Corollary 3 are characterized so as to reduce the solution space of X and (Y, Z) variables, respectively, in $\mathbb{P}_{SEPDV_{ik,\mu}}$.

Proposition 2. *For given μ, the following statements hold in $\mathbb{P}_{SEPDV_{ik,\mu}}$:*

(a) *If $\theta_{ijkp} \leq 0$ for some $j \in \mathcal{J}$ and $p \in \mathcal{T}$, then there is an optimal solution with $X_{ijkp} = 0$.*

(b) *If $\theta_{ijkp} \geq h_{ikp}^S \lceil d_{ijp}/\Gamma \rceil$ for some $j \in \mathcal{J}$ and $p \in \mathcal{T}$, then there is an optimal solution with $X_{ijkp} = 1$.*

Proof. (a) Let us compare the situations of $X_{ijkp} = 1$ and $X_{ijkp} = 0$ for some $j \in \mathcal{J}$ and $p \in \mathcal{T}$ such that $\theta_{ijkp} \leq 0$. The value of the objective function (4a) for the former case cannot be smaller than that for the latter case because the coefficient of the X variable in (4a) is nonnegative. Also, the former case makes the constraint (4b) tighter than the latter case. Therefore, the former case (i.e., $X_{ijkp} = 1$) cannot give a better solution than the latter case (i.e., $X_{ijkp} = 0$).

(b) This can be shown in the same manner as in the proof of (a), and hence we omit its proof. □

Let $\mathcal{Q}_{ik}$ be a set of freight demands that can be transported through (i,k) where $i \in \mathcal{J}$ and $k \in \mathcal{K}$. Note that $\mathcal{Q}_k = \bigcup_{i \in \mathcal{J}} \mathcal{Q}_{ik}$. For $p \in \mathcal{T}$, let $\widetilde{\mathcal{Q}}_{ikp}^0$ and $\widetilde{\mathcal{Q}}_{ikp}^1$ be subsets of $\mathcal{Q}_{ik}$ satisfying Propositions 2(a) and 2(b) in $\mathbb{P}_{SEPDV_{ik,\mu}}$, respectively, and let $\widetilde{\mathcal{Q}}_{ikp} = \widetilde{\mathcal{Q}}_{ikp}^0 \cup \widetilde{\mathcal{Q}}_{ikp}^1$.

Corollary 3. *For given μ and $\{X_{ijkp}\}$ fixed at the associated values according to Proposition 2, there is an optimal solution in $\mathbb{P}_{SEPDV_{ik,\mu}}$ with $(Y_{ik}, Z_{ikp}) \in \{(y, z_p) \mid \widetilde{Y}_{ik}^{\min} \leq y \leq \widetilde{Y}_{ik}^{\max}, \; 0 \leq z_p \leq \widetilde{Z}_{ikp,y}^{\max}, \; \forall p \in \mathcal{T}\}$, where*

$$\widetilde{Y}_{ik}^{\min} = \left\lceil \frac{\min_{p \in \mathcal{T}} \sum_{\{j \in \mathcal{J} | (i,j) \in \widetilde{\mathcal{Q}}_{ikp}^1\}} d_{ijp}}{\Gamma} \right\rceil,$$

$$\widetilde{Y}_{ik}^{\max} = \left\lceil \frac{\max_{p \in \mathcal{T}} \sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \widetilde{\mathcal{Q}}_{ikp}^0\}} d_{ijp}}{\Gamma} \right\rceil, \tag{7}$$

$$\widetilde{Z}_{ikp,y}^{\max} = \left\lceil \frac{\sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \widetilde{\mathcal{Q}}_{ikp}^0\}} d_{ijp}}{\Gamma} \right\rceil - y.$$

Proof. The constraint (4b) can be rewritten as

$$\sum_{\{j \in \mathcal{J} | (i,j) \in \widetilde{\mathcal{Q}}_{ikp}^1\}} d_{ijp} + \sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \widetilde{\mathcal{Q}}_{ikp}\}} d_{ijp} X_{ijkp}$$

$$\leq \Gamma \left(Y_{ik} + Z_{ikp} \right) \tag{8}$$

for all $p \in \mathcal{T}$. Since all the variables are binary variables, $\widetilde{Y}_{ik}^{\min} \leq \lceil \sum_{\{j \in \mathcal{J} | (i,j) \in \widetilde{\mathcal{Q}}_{ikp}^1\}} d_{ijp}/\Gamma \rceil \leq (Y_{ik} + Z_{ikp})$ holds for all $p \in \mathcal{T}$. Since it is assumed that $0 < h_{ik}^L \leq h_{ikp}^S$ for all $p \in \mathcal{T}$, any feasible solution with $(X_{ijkp}, Y_{ik}, Z_{ikp}) = (x_{jp}, y, z_p)$, where $y < \widetilde{Y}_{ik}^{\min}$ cannot yield a better solution value than the feasible solution with $(X_{ijkp}, Y_{ik}, Z_{ikp}) = (x_{jp}, \widetilde{Y}_{ik}^{\min}, \max\{0, z_p - \widetilde{Y}_{ik}^{\min} + y\})$. Moreover, any feasible solution with $(X_{ijkp}, Y_{ik}, Z_{ikp}) = (x_{jp}, y, z_p)$, where $y > \widetilde{Y}_{ik}^{\max}$ cannot yield a better solution value than the feasible solution with $(X_{ijkp}, Y_{ik}, Z_{ikp}) = (x_{jp}, \widetilde{Y}_{ik}^{\max}, 0)$. Therefore, there is an optimal solution with $\widetilde{Y}_{ik}^{\min} \leq Y_{ik} \leq \widetilde{Y}_{ik}^{\max}$. Also, with Y_{ik} fixed at $y \in \{y' \in \mathbb{Z}^+ \mid \widetilde{Y}_{ik}^{\min} \leq y' \leq \widetilde{Y}_{ik}^{\max}\}$, it is obvious that there is an optimal solution with $0 \leq Z_{ikp} \leq \widetilde{Z}_{ikp,y}^{\max}$. □

Then, the restricted problem with Y_{ik} fixed at $y \in \{y' \in \mathbb{Z}^+ \mid \widetilde{Y}_{ik}^{\min} \leq y' \leq \widetilde{Y}_{ik}^{\max}\}$ in $\mathbb{P}_{SEPDV_{ik,\mu}}$, denoted by $\mathbb{P}_{SEPDV_{ik,\mu,y}}$, can be decomposed into single-period bounded knapsack problems (BKP), $\mathbb{P}_{BKP_{ikp,\mu,y}}$, as follows.

Problem $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$

$$v_{\text{BKP}}^{ikp}(\mu, y) = \max \quad h_{ikp}^S \overline{Z}_{ikp} + \sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \overline{\mathcal{Q}}_{ikp}^0\}} \theta_{ijkp} X_{ijkp}$$

$$\text{subject to} \quad \sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \overline{\mathcal{Q}}_{ikp}\}} d_{ijp} X_{ijkp} + \Gamma \overline{Z}_{ikp} \leq \Gamma \left\lceil \frac{\sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \overline{\mathcal{Q}}_{ikp}^0\}} d_{ijp}}{\Gamma} \right\rceil - \sum_{\{j \in \mathcal{J} | (i,j) \in \overline{\mathcal{Q}}_{ikp}^1\}} d_{ijp} \tag{9}$$

$$\overline{Z}_{ikp} \in \left\{ 0, 1, \ldots, \left(\left\lceil \frac{\sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \overline{\mathcal{Q}}_{ikp}^0\}} d_{ijp}}{\Gamma} \right\rceil - y \right) \right\}, \quad X_{ijkp} \in \{0, 1\}.$$

Note that $\overline{Z}_{ikp} = \widetilde{Z}_{ikp,y}^{\max} - Z_{ikp}$. We next discuss how to solve $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$. Let $v_{\text{SEPDV}}^{ik}(\mu, y)$ denote the optimal value of $\mathbb{P}_{\text{SEPDV}_{ik,\mu,y}}$. We first transform $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$ into the associated 0-1 knapsack problem, and the knapsack problem can be solved by the dynamic programming algorithm proposed by Toth [30]. We remark that $\overline{Z}_{ikp}$ takes the finite number of integer values in the set $\Psi \equiv \{0, 1, \ldots, (\lceil \sum_{\{j \in \mathcal{J} | (i,j) \in \mathcal{Q}_{ik} \setminus \overline{\mathcal{Q}}_{ikp}^0\}} d_{ijp}/\Gamma \rceil - y)\}$. Furthermore, for each value of $\overline{Z}_{ikp}$ in Ψ, it is straightforward to show that the problem $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$ reduces to a 0-1 knapsack problem with decision variables X_{ijkp} for all j. Therefore, we iteratively solve a 0-1 knapsack problem for each value of $\overline{Z}_{ikp}$ in Ψ, compare the objective function values for each case, and then identify the optimal solution for $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$. Once all the problems of $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$ are solved for each value of Y_{ik} as mentioned above, the optimal value of the problem $\mathbb{P}_{\text{SEPDV}_{ik,\mu}}$, $v_{\text{SEPDV}}^{ik}(\mu)$, is computed as

$$v_{\text{SEPDV}}^{ik}(\mu) = \min_{\widetilde{Y}_{ik}^{\min} \leq y \leq \widetilde{Y}_{ik}^{\max}} v_{\text{SEPDV}}^{ik}(\mu, y) = \min_{\widetilde{Y}_{ik}^{\min} \leq y \leq \widetilde{Y}_{ik}^{\max}} \left[|\mathcal{T}| \right.$$

$$\cdot h_{ik}^L y \tag{10}$$

$$\left. + \sum_{p \in \mathcal{T}} \left(h_{ikp}^S \widetilde{Z}_{ikp,y}^{\max} - v_{\text{BKP}}^{ikp}(\mu, y) - \sum_{\{j \in \mathcal{J} | (i,j) \in \overline{\mathcal{Q}}_{ikp}^1\}} \theta_{ijkp} \right) \right].$$

The number of problems $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$ and $\mathbb{P}_{\text{SEPDV}_{ik,\mu,y}}$ to be solved can be further reduced by Propositions 4(a) and 4(b), respectively. Let $(X_{ijkp,y}^*, \overline{Z}_{ikp,y}^*)$ be the optimal solution of $\mathbb{P}_{\text{BKP}_{ikp,\mu,y}}$ with explicit dependence on y, and furthermore, let us denote $\mathcal{T}_{ry}$ as $\mathcal{T}_{ry} = \{p \in \mathcal{T} \mid \overline{Z}_{ikp,y}^* \leq \widetilde{Z}_{ikp,y}^{\max} - r = \widetilde{Z}_{ikp,(y+r)}^{\max}\}$ for $\widetilde{Y}_{ik}^{\min} \leq y \leq \widetilde{Y}_{ik}^{\max}$ and $0 < r \leq \widetilde{Z}_{ikp,y}^{\max}$. Note that $\mathcal{T}_{ry} \subset \mathcal{T}_{ly} \subset \mathcal{T}$ for $0 \leq l < r \leq \widetilde{Z}_{ikp,y}^{\max}$.

Proposition 4. *(a) Given $(X_{ijkp,y}^*, \overline{Z}_{ikp,y}^*)$ for some $y \in \{y' \in \mathbb{Z}^+ \mid \widetilde{Y}_{ik}^{\min} \leq y' \leq \widetilde{Y}_{ik}^{\max}\}$ and some $p \in \mathcal{T}$, if $p \in \mathcal{T}_{ry}$ for some $r \in \{r' \in \mathbb{Z}^+ \mid 0 < r' \leq \widetilde{Z}_{ikp,y}^{\max}\}$, then $(X_{ijkp,(y+r)}^*, \overline{Z}_{ikp,(y+r)}^*) = (X_{ijkp,y}^*, \overline{Z}_{ikp,y}^*)$.*

(b) Given $(X_{ijkp,y}^*, \overline{Z}_{ikp,y}^*)$ for some $y \in \{y' \in \mathbb{Z}^+ \mid \widetilde{Y}_{ik}^{\min} \leq y' \leq \widetilde{Y}_{ik}^{\max}\}$ and all $p \in \mathcal{T}$, if $\sum_{p \in \mathcal{T}_{ry}} h_{ikp}^S \geq |\mathcal{T}| h_{ik}^L$ for some $r \in \{r' \in \mathbb{Z}^+ \mid 0 < r' \leq \widetilde{Z}_{ikp,y}^{\max}\}$, then $v_{\text{SEPDV}}^{ik}(\mu, y+l) \geq v_{\text{SEPDV}}^{ik}(\mu, y+r)$ for $0 \leq l < r$.

In summary, we now present the solution procedure for $\mathbb{P}_{\text{SEPDV}_{ik,\mu}}$, PROC($\text{SEPDV}_{ik,\mu}$), in Algorithm 1 based on the whole discussion above. For reference, Figure 1 illustrates the relationship among all the problems discussed in this section.

3.3. Finding the Lagrange Multipliers. So far, we examined how to address the problem $\mathbb{P}_{\text{SEPDV}_{ik,\mu}}$ for each $(i, k) \in \mathcal{E}_{\text{IN}}$ when the Lagrangian multiplier μ is given. We now discuss the procedure for finding the Lagrangian multiplier μ. For the purpose, we adapt the subgradient optimization procedure which is one of the most popular methods to find a good set of Lagrange multipliers [31]. Specifically, the Lagrange multipliers $\mu \equiv (\mu_{ijp}', \mu_{ijkp}'')_{\forall i,j,k,p}$ are generated iteratively as

(i) $\mu_{ijp,s+1}' = \mu_{ijp,s}' + \gamma_s G_{ijp,s}'$ for all $(i, j) \in \mathcal{Q}_k$, $p \in \mathcal{T}$,

(ii) $\mu_{ijkp,s+1}'' = \max\{0, (\mu_{ijkp,s}'' + \gamma_s)\} + \gamma_s G_{ijp,s}'$ for all $k \in \mathcal{T}$, $(i, j) \in \mathcal{Q}_k$, $p \in \mathcal{T}$,

where $G_{ijp,s}'$ and $G_{ijkp,s}''$ are the subgradients for constraints (2b) and (2d), respectively, and γ_s is the stepsize at iteration s. When it comes to the subgradients $G_{ijp,s}'$ and $G_{ijkp,s}''$ in the formulas above, they are determined by the solution of $\mathbb{P}_{\text{LR}_\mu}$ given μ as follows: $G_{ijp,s}' = 1 - \sum_{\{k \in \mathcal{K} | (i,j) \in \mathcal{Q}_k\}} X_{ijkp}$ for all $(i, j) \in \mathcal{Q}_k$, $p \in \mathcal{T}$, and $G_{ijkp,s}'' = X_{ijkp} - \overline{X}_{ijkp}$ for all $k \in \mathcal{K}$, $(i, j) \in \mathcal{Q}_k$, $p \in \mathcal{T}$. Moreover, the stepsize γ_s at iteration s is determined by

$$\gamma_s = \frac{\eta \left[\text{UB} - v(\mu_s) \right]}{\sum_{(i,j) \in \mathcal{Q}} \sum_{p \in \mathcal{T}} \left[\left(G_{ijp,s}' \right)^2 + \sum_{\{k \in \mathcal{K} | (i,j) \in \mathcal{Q}_k\}} \left(G_{ijkp,s}'' \right)^2 \right]}, \tag{11}$$

where UB represents the best upper bound of v_{OPT} found up to iteration s and η is a control parameter for γ_s. In this paper, η is set to 2 at the beginning and then halved if the lower bound is not improved in a predetermined number of consecutive iterations.

```
PROC(SEPDV_{ik,μ})
begin
(1) Let Y*_{ik}, Z*_{ikp} for all p ∈ 𝒯 and X*_{ijkp} for all j ∈ {j ∈ 𝒥 | (i, j) ∈ 𝒬_k} be Y, Z and X-variables of
    the optimal solution of ℙ_{SEPDV_{ik,μ}} respectively;
(2) Set v^{ik}_{SEPDV}(μ) = ∞;
(3) Fix some X_{ijkp}s at the associated values according to Proposition 2 and determine Ỹ^{min}_{ik},
    Ỹ^{max}_{ik} and Z̃^{max}_{ikp,y} for all y ∈ {y' ∈ ℤ^+ | Ỹ^{min}_{ik} ≤ y' ≤ Ỹ^{max}_{ik}} by Corollary 3;
(4) Set 𝒯_{ry} = φ for all r ∈ {r' ∈ ℤ^+ | 0 < r' ≤ Z̃^{max}_{ikp,y}} and y ∈ {y' ∈ ℤ^+ | Ỹ^{min}_{ik} ≤ y' ≤ Ỹ^{max}_{ik}};
(5) Set α_y = 0 and β_{py} = 0 for all y ∈ {y' ∈ ℤ^+ | Ỹ^{min}_{ik} ≤ y' ≤ Ỹ^{max}_{ik}} and p ∈ 𝒯 where α_y and
    β_{py} are binary variables having 1 only if ℙ_{SEPDV_{ik,μ}} and ℙ_{BKP_{ikp,μ,y}} are not necessary
    to be solved respectively, and 0 otherwise;
(6) Set y = Ỹ^{min}_{ik};
(7) while y ≤ Ỹ^{max}_{ik} do
(8)     if α_y = 0 then
(9)         Set p = 1;
(10)        while p ≤ |𝒯| do
(11)            if β_{py} = 0 then
(12)                Solve ℙ_{BKP_{ikp,μ,y}} and set the resulting solution as (X*_{ijkp,y}, Z̄*_{ikp,y});
(13)                Set 𝒯_{r-l,y+l} = 𝒯_{r-l,y+1} ∪ {p} for all r ∈ {r' ∈ ℤ^+ \ {0} | Z̄*_{ikp,y} ≤ Z̃^{max}_{ikp,y} − r'} and l ∈ {l' ∈ ℤ^+ | 0 ≤ l' < r};
                   end
(14)            Set (X*_{ijkp,y+r}, Z̄*_{ikp,y+r}) = (X*_{ijkp,y}, Z̄*_{ikp,y}) and β_{p,y+r} = 1 for all
                   r ∈ {r' ∈ ℤ^+ \ {0} | Z̄*_{ikp,y} ≤ Z̃^{max}_{ikp,y} − r'} by Proposition 4(a);
(15)            Set p = p + 1;
               end
(16)        Set α_{y+l} = 1 for all l ∈ {l' ∈ ℤ^+ | 0 < l' < r*} by Proposition 4(b) where
               r* = max_{r∈ℛ}r such that ℛ = {r ∈ ℤ^+ | 0 < r ≤ Z̃^{max}_{ikp,y}, Σ_{p∈𝒯_{ry}} h^S_{ikp} ≥ |𝒯|h^L_{ik}};
(17)        if v^{ik}_{SEPDV}(μ, y) < v^{ik}_{SEPDV}(μ) then
(18)            Set v^{ik}_{SEPDV}(μ) = v^{ik}_{SEPDV}(μ, y) (refer to (10));
(19)            Set Y*_{ik} = y and (X*_{ijkp}, Z̄*_{ikp}) = (X*_{ijkp,y}, Z̃^{max}_{ikp,y} − Z̄*_{ikp,y}) for all p ∈ 𝒯;
               end
           end
(20) Set y = y + 1;
       end
   end
end
```

ALGORITHM 1: Solution procedure for ℙ_{SEPDV_{ik,μ}}.

3.4. Finding a Feasible Solution.

A solution obtained by solving ℙ_{LR_μ} based on the discussion in Section 3.2 may be infeasible to ℙ_{MVLPUC_+}, and hence we propose a two-phase heuristic to deal with infeasibility issues as follows: In the first phase, a construction heuristic (CH) is derived to modify any infeasible solutions to ℙ_{MVLPUC_+} into feasible ones, and then, in the second phase, a tabu search-based heuristic (TSH) is utilized to improve the solution of (CH). The details are presented in the subsequent sections.

3.4.1. Construction Heuristic (CH).

As mentioned above, a construction heuristic (CH) intends to deal with possible infeasibility issues of solutions obtained by solving ℙ_{LR_μ}, and (CH) is performed at each subgradient iteration. Let X^*_{ijkp} and $\overline{X}^*_{ijkp}$ be the X- and $\overline{X}$-variable values of the optimal solution of ℙ_{SEPDV_{ik,μ}} and ℙ_{SEPDV_{kj,μ}}, respectively. First, the corresponding demand for all $(i, j) ∈ 𝒬$ and $p ∈ 𝒯$ is

reassigned to the UFCC with the lexicographical maximum value of the three-tuple relation $((X^*_{ijkp} + \overline{X}^*_{ijkp}), (θ_{ijkp}X^*_{ijkp} + δ_{ijkp}\overline{X}^*_{ijkp}), -(h^L_{ik} + h^L_{kj}))$. Then, from the reassigned demands, demand quantity D_{ep} for all $e ∈ 𝒠$ and $p ∈ 𝒯$ is computed as $D_{ep} = \sum_{\{j∈𝒥|(i,j)∈𝒬_k\}} d_{ijp}X_{ijkp}$ or $D_{ep} = \sum_{\{i∈𝒥|(i,j)∈𝒬_k\}} d_{ijp}\overline{X}_{ijkp}$. Finally, to allocate vehicles for each $e ∈ 𝒠$ at the minimum vehicle cost, we solve the following vehicle allocation problem (VAP), ℙ_{VAP_e}.

Problem ℙ_{VAP_e}

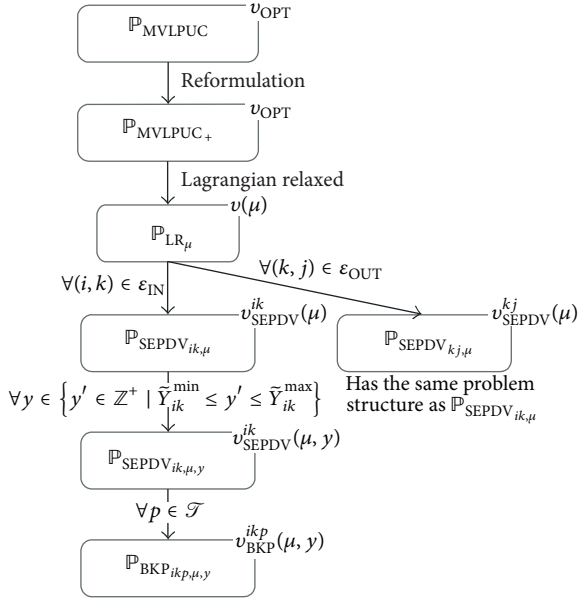$$v^e_{VAP} = \min \quad |𝒯| h^L_e Y_e + \sum_{p∈𝒯} h^S_{ep} Z_{ep} \tag{12a}$$

$$\text{subject to} \quad D_{ep} ≤ Γ (Y_e + Z_{ep}), \quad ∀p ∈ 𝒯 \tag{12b}$$

$$Y_e, Z_{ep} ∈ ℤ^+, \quad ∀p ∈ 𝒯. \tag{12c}$$

Proposition 5 characterizes the optimal solutions of ℙ_{VAP_e}.

Figure 1: Problem decomposition of $\mathbb{P}_{\text{MVLPUC}}$.

Proposition 5. *There is an optimal solution of $\mathbb{P}_{VAP_e}$ with $(Y_e, Z_{ep}) \in \{(y, Z_{ep,y}) \mid y = y_{[1]}, \ldots, y_{[|\mathcal{T}|]}\}$, where $Z_{ep,y} = \max\{0, (\lceil D_{ep}/\Gamma \rceil - y)\}$ and $y_{[i]}$ is the ith largest y.*

Proof. As in Corollary 3, it is straightforward to show that there is an optimal solution of $\mathbb{P}_{VAP_e}$ with $(Y_e, Z_{ep}) \in \{(y, Z_{ep,y}) \mid y_{[1]} \leq y \leq y_{[|\mathcal{T}|]}\}$. For an integer m such that $y_{[n]} \leq m < y_{[n+1]}$ and for $n \in \{1, \ldots, |\mathcal{T}| - 1\}$,

$$Z_{e[p],m} = \begin{cases} 0 & \text{if } p \leq n \\ (y_{[p]} - m) & \text{otherwise,} \end{cases}$$

$$Z_{e[p],m+1} = \begin{cases} 0 & \text{if } p \leq n \\ (y_{[p]} - (m+1)) & \text{otherwise.} \end{cases} \tag{13}$$

Let $v_{\text{VAP}}^e(y)$ be the value of the objective function (12a) associated with Y_e fixed at $y \in [y_1, y_{[|\mathcal{T}|]}]$ in $\mathbb{P}_{VAP_e}$. Then, $v_{\text{VAP}}^e(m+1) - v_{\text{VAP}}^e(m) = [(m+1)|\mathcal{T}|h_e^L - \sum_{p=1}^{|\mathcal{T}|} Z_{e[p],m+1}] - [m|\mathcal{T}|h_e^L - \sum_{p=1}^{|\mathcal{T}|} Z_{e[p],m}] = |\mathcal{T}|h_e^L - \sum_{p=n+1}^{|\mathcal{T}|} h_{e[p]}^S$. It implies that $v_{\text{VAP}}^e(m)$ is a linear function in m over the range $y_{[n]} \leq m < y_{[n+1]}$. Therefore, there is an optimal solution of $\mathbb{P}_{VAP_e}$ with $(Y_e, Z_{ep}) \in \{(y, Z_{ep,y}) \mid y = y_{[1]}, \ldots, y_{[|\mathcal{T}|]}\}$. □

By Proposition 5, $\mathbb{P}_{VAP_e}$ can be solved by finding a solution of having the minimum cost among $|\mathcal{T}|$ number of solutions which are associated with Y_e fixed at $y_{[p]}$ for all $p \in \mathcal{T}$.

3.4.2. Tabu Search-Based Heuristic (TSH).

After the first phase (CH), a tabu search-based heuristic (TSH) is performed to improve the solution obtained from (CH). We note that (TSH) is expected to require much more computational time than (CH), and hence, to save the elapsed time of the overall algorithm, (TSH) is executed only at the subgradient iterations satisfying the following two conditions ((a) and (b)) together: (a) the best lower bound is improved and (b) the solution obtained by (CH) at the current subgradient iteration is different from the solutions obtained by (CH) so far. In this paper, we call the conditions (a) and (b) as "(TSH)-activating conditions."

Tabu search is one of the popular metaheuristics designed to escape the trap of local optimality [32], and it has been used frequently in various network-related problems as in Sung and Song [25], Crevier et al. [33], and Üster and Maheshwari [34]. We now explain the main features of the tabu search used in (TSH), including *move operation*, *tabu restriction* and *aspiration criterion*, *move evaluation*, and *termination condition*.

The *move operation* used in this paper is to modify assignment of some freight demand, say $q = (i, j) \in \mathcal{Q}$, from its current UFCC, say $l \in \mathcal{K}$, to the alternative UFCC, say $k \in \mathcal{K}$, for some period, say $p \in \mathcal{T}$. It corresponds to modifying the value of X_{ijlp} from 1 to 0 and the value of X_{ijkp} from 0 to 1 in Problem $\mathbb{P}_{\text{MVLPUC}_+}$. The *tabu restriction* is employed such that the UFCC assignment for any freight demand should not be changed during *tabu_size* number of iterations except for the case that satisfies the aspiration criterion (to be explained later). To implement the tabu restriction, as a new solution is selected, the information of the freight demand associated with the move operation is recorded in a list (called *tabu list*) and kept in the list over the *tabu_size* number of iterations. In this paper, the *tabu_size* is chosen randomly from the discrete uniform distribution in the closed interval [10, 20]. The *aspiration criterion* adapted here is as follows: a solution from the move operation for the freight demands recorded in the *tabu list* is eligible to be chosen as a new solution only if the move improves the best upper bound. Moreover, when we choose a new solution among neighbor solutions at each iteration, the *move evaluation* function used in this paper is to maximize $\{mv - A \cdot n\}$ where mv, n, and A represent the decrease amount of the objective function value, the frequency that the corresponding move is used to produce a new solution so far, and the associated penalty with the frequency, respectively. When it comes to mv, it is computed simply by subtracting the objective value of the new solution from the objective value of the current solution, and if mv is positive, A is set to 0 and 50 otherwise. We note that this move evaluation function is known to have an advantage in diversifying the search direction efficiently [32]. Lastly, if the total number of performed iterations amounts to *max_iter* or the best upper bound has not been improved during *max_no_imp* consecutive iterations, (TSH) is *terminated*. In this paper, *max_iter* and *max_no_imp* are set at 500 and 30, respectively.

3.5. Overall Solution Algorithm.

We now summarize all the discussion and present the algorithm PROC(MVLPUC) for solving $\mathbb{P}_{\text{MVLPUC}}$ in Algorithm 2. It can be terminated either after a predetermined number (1000 in this paper) of iterations or when η gets too small (i.e., $\eta \leq 0.005$ in this paper). It can also be terminated when $(((UB-LB)/LB) \times 100)$

```
PROC(MVLPUC)
begin
(1) Initialize the necessary parameters and Lagrange multipliers (i.e., s = 1, γ_s = 2, η = 2,
    LB = −∞, UB = ∞, μ'_{ijp,s} = 1 and μ''_{ijkp,s} = 1);
(2) Initialize NI_LB = 0 where NI_LB represents the number of consecutive iterations for
    which the best lower bound has not been improved;
    repeat
(3)     Solve the Lagrangian problem ℙ_{L(μ_s)}, and let υ(μ_s) be the resulting objective value;
        if LB < υ(μ_s) then
(4)         Set LB = υ(μ_s) and NI_LB = 0;
        else
(5)         Set NI_LB = NI_LB + 1
        end
(6)     Find a feasible solution (X, X̄, Y, Z)_{UB,s} by (CH) where (X, X̄, Y, Z)_{UB,s} represents the
        feasible solution to ℙ_{MVLPUC₊} at iteration s;
        if (TSH)-activating conditions are satisfied then
(7)         Update (X, X̄, Y, Z)_{UB,s} by (TSH);
        end
(8)     Let υ_{UB,s} be the objective value of the solution (X, X̄, Y, Z)_{UB,s} for ℙ_{MVLPUC₊};
(9)     if UB > υ_{UB,s} then
(10)        Set UB = υ_{UB,s}, and set (X, X̄, Y, Z)* = (X, X̄, Y, Z)_{UB,s} where (X, X̄, Y, Z)* represents
             the best feasible solution searched until iteration s;
        end
        if NI_LB = 30 then
(11)        Set η = η/2 and NI_LB = 0;
        end
(12)    Update μ by the subgradient optimization procedure (refer to Section 3.3);
(13)    Set s = s + 1;
        until the aforementioned termination conditions (i.e., s > 1000, η ≤ 0.005 or ((UB − LB/LB) × 100) < 0.1) are met;
end
```

ALGORITHM 2: Overall solution procedure for $\mathbb{P}_{\text{MVLPUC}}$.

is less than 0.1%, where UB and LB represent the best upper and lower bounds searched so far, respectively.

4. Numerical Experiments

This section presents the numerical results on the performance of the proposed Lagrangian heuristic algorithm PROC(MVLPUC) for $\mathbb{P}_{\text{MVLPUC}}$. The algorithm PROC(MVLPUC) was implemented in C/C++ language, and the performance of PROC(MVLPUC) is compared with those of the following three algorithms as benchmarks: (a) the Lagrangian heuristic where only (CH) is used in PROC(MVLPUC) while (TSH) is excluded, (b) a commercial optimization software *IBM ILOG CPLEX Optimization Studio*, and (c) the algorithm where (CH) is applied at every node of the branch-and-bound tree for *IBM ILOG CPLEX*. The three algorithms are denoted by (LCH), (CX), and (CXCH), respectively.

For the numerical experiments, a variety of problem instances were randomly generated according to the underlying network structure, the number of nodes (i.e., $|\mathcal{V}|$), and the number of time periods (i.e., $|\mathcal{T}|$). Three types of underlying networks were considered in the experiments including origin-centric, equal, and destination-centric networks according to the number of origin and destination nodes. Origin-centric networks (destination-centric networks) have

larger numbers of origin (destination) nodes than those of destination (origin) nodes, and equal networks have almost equivalent numbers of origin nodes as those of destination nodes. Moreover, three $|\mathcal{V}|$ values including 30, 40, and 50 and three $|\mathcal{T}|$ including 3, 6, and 12 were considered in the experiment.

To generate problem instances, each node was randomly selected on a 100×100 grid. Some of the nodes were connected by arcs, which were also chosen randomly. Unit-period long-term vehicle lease cost h_e^L was calculated based on Euclidean distance between nodes, and unit-period short-term vehicle lease cost h_{ep}^S was randomly chosen in the range from 1.5 to 2 times of h_e^L. Such a structure of vehicle lease costs reflects the actual practice in Korean logistics industry. Moreover, transportation time t_e was randomly chosen in the range from 0.8 to 1.2 times of h_e^L, and handling time at $k \in \mathcal{K}$, o_k, was set as $o_k = \min\{\min_{i \in \mathcal{J}} t_{ik}, \min_{j \in \mathcal{J}} t_{kj}\}$. Transportation time restriction TL_{ij} was set as $\text{TL}_{ij} = 0.1(\min_{k \in \mathcal{K}}\{t_{ik} + o_k + t_{kj}\}) + 0.9(\max_{k \in \mathcal{K}}\{t_{ik} + o_k + t_{kj}\})$. Demand quantity d_{ijp} was randomly selected in the range from 1 to 90, and vehicle capacity Γ was set at 100. For each combination of underlying network structure, number of nodes, and number of periods, four problem instances were generated, and hence, in total, 108 problem instances were generated.

Tables 1–3 show the experimental results for the problem instances with $|\mathcal{V}| = 30, 40$ and 50 nodes, respectively. The tables present the best lower bound (LB), the best upper bound (UB), the percentage deviation between the two (%Gap), and the elapsed time in seconds (Time [s]). The percentage deviation, %Gap, is computed as $((UB-LB)/LB) \times 100$.

For the problem instances with $|\mathcal{V}| = 30$, PROC(MVLPUC) gave solutions with the following performance: (a) the duality gap (shown in the "%Gap" column) ranges from 5.26% to 12.99% with the average of 7.98%, and (b) the elapsed time ranges from 33.77 s to 691.83 s with the average of 175.12 s. The duality gap does not seem to be dependent on both $|\mathcal{T}|$ and the network structure, as observed that the duality gap does not change consistently according to them. On the other hand, the elapsed time seems to increase not exponentially but linearly as $|\mathcal{T}|$ increases while it does not seem to be dependent on the network structure. For example, the average elapsed times for the problem instances with $|\mathcal{T}| = 3$, 6, and 12 are 80.06 s, 158.05 s, and 287.26 s, respectively. The aforementioned trends can be observed in the problem instances with $|\mathcal{V}| = 40$ and 50 as well.

From all the experiments, the performance changes of the proposed algorithm according to the number of nodes (i.e., $|\mathcal{V}|$) can be summarized as follows. As $|\mathcal{V}|$ increases, the duality gap seems to marginally increase. For example, the average duality gaps for the problem instances with $|\mathcal{V}| = 30$, 40, and 50 are 7.98%, 8.58%, and 9.44%, respectively, and this may be due to the fact that, as $|\mathcal{V}|$ increases, the number of the associated Lagrange multipliers greatly increases as shown in μ''_{ijkp}, and hence, the associated lower bound may become looser accordingly. We also observe that, as $|\mathcal{V}|$ increases, the elapsed time increases rapidly, but not exponentially. For example, the average elapsed times for the problem instances with $|\mathcal{V}| = 30$, 40, and 50 are 175.12 s, 654.59 s, and 2591.26 s, respectively.

We now compare the performance of PROC(MVLPUC) with those of the aforementioned three algorithms including (LCH), (CX), and (CXCH), respectively. They were tested on the small-sized problem instances (or instances with $|\mathcal{V}| = 30$). For (LCH), the same termination conditions as in PROC(MVLPUC) were used, while for (CX) and (CXCH), an execution time limit of five hours was applied as a termination condition.

Table 4 shows the numerical results of the compared algorithms for the problem instances with $|\mathcal{V}| = 30$. The table gives the percentage deviations between the best lower bounds obtained from each of the compared algorithms and the best lower bound obtained by the proposed algorithm (LB in Table 1) and the percentage deviations between the best upper bounds obtained by each of the compared algorithms and the best upper bound obtained by the proposed algorithm (UB in Table 1). For (LCH) and both of (CX) and (CXCH), the percentage deviations of the lower bounds are denoted by ILB_{LCH} and ILB_{CX}, respectively. Note that (CX) and (CXCH) give the same lower bound obtained from the LP relaxation of Problem $\mathbb{P}_{MVLPUC}$. For (LCH), (CX), and

(CXCH), the percentage deviations of the upper bounds are denoted by IUB_{LCH}, IUB_{CX}, and IUB_{CXCH}, respectively. In detail, the percentage deviations of the lower bounds and the percentage deviations of the upper bounds were computed by $((LB - LB_C)/LB_C) \times 100$ and $((UB_C - UB)/UB_C) \times 100$, respectively, where LB_C and UB_C are the best lower bound and the best upper bound obtained by the corresponding compared algorithm, respectively. For example, if IUB_{LCH} is positive, it means that the best upper bound of the proposed algorithm is better than that of (LCH) at the associated rate. From the table, the proposed algorithm may give the almost same lower bounds as (LCH) and better upper bounds than (LCH) from the fact that ILB_{LCH} is almost zero and ULB_{LCH} ranges from 0.36% to 11.88%. This may imply that (TSH) embedded in PROC(MVLPUC) improves the upper bound significantly. Moreover, PROC(MVLPUC) may give better lower bounds than (CX) and (CXCH). For the long planning horizon, PROC(MVLPUC) may give better upper bounds than (CX) and (CXCH), while for the short planning horizon, it may not. For example, the associated average values of IUB_{CX} with $|\mathcal{T}| = 3$, 6, and 12 are −2.35%, 2.56%, and 6.44%, respectively. This may imply that the proposed algorithm is suitable for the large-sized problem instances. Also, it can be observed that (CXCH) outperforms (CX) marginally.

In summary, the proposed algorithm PROC(MVLPUC) may give good lower and upper bounds within reasonable time, and the duality gap seems not to be dependent on the network structure and the number of time periods, while it seems to increase a little, as $|\mathcal{V}|$ increases. Furthermore, as presented in Table 3, the average duality gap for the problem instances with $|\mathcal{V}| = 50$ is still less than 10%, and the elapsed time seems not to be dependent on the network structure, while it increases as $|\mathcal{V}|$ or $|\mathcal{T}|$ increases. However, the increase in the elapsed time is not exponential in either case.

5. Concluding Remarks

In this paper, a multiperiod vehicle lease planning problem in an urban freight distribution network with UFCCs is considered where vehicles for each direct service are leased in short term or long term. The problem under consideration is motivated by the realistic situation where the freight demands fluctuate dynamically in time, and urban freight distribution can be made through UFCCs. The problem is formulated as an integer programming model and can be proven to be NP-hard in a strong sense. Thus, we propose a Lagrangian-based heuristic to address the concomitant problem. Various numerical experiments are performed to evaluate the performance of the proposed algorithm, and they indicate that the proposed algorithm works well. The proposed algorithm may be applied to vehicle lease planning for industrial logistic services for handling various products including food, apparel, electronic goods, and logistics companies.

As extended research, several models derived from the proposed model in this paper can be considered. For example, a problem where each freight demand can be delivered through more than a single UFCC may be interesting.

TABLE 1: Test results of the proposed algorithm for the problem instances with $|\mathcal{V}| = 30$.

$|\mathcal{J}|$	$|\mathcal{K}|$	$|\mathcal{F}|$	$|\mathcal{T}| = 3$				$|\mathcal{T}| = 6$				$|\mathcal{T}| = 12$			
			LB	UB	%Gap	Time [s]	LB	UB	%Gap	Time [s]	LB	UB	%Gap	Time [s]
20	5	5	212996.81	224714	5.50	53.16	461155.84	489788	6.21	98.66	929519.50	1002248	7.82	230.75
20	5	5	216811.19	241662	11.46	65.11	460583.37	504220	9.47	107.86	948275.80	1025174	8.11	255.70
22	3	5	279876.66	298848	6.78	240.95	573183.77	611568	6.70	461.84	1192304.31	1289542	8.16	691.83
22	3	5	272613.72	294054	7.86	212.24	554945.75	600348	8.18	432.34	1170739.75	1265118	8.06	419.22
13	5	12	332774.01	355556	6.85	57.92	683250.91	728424	6.61	127.08	1434397.96	1536928	7.15	272.66
13	5	12	304624.33	340794	11.87	48.02	650095.80	714146	9.85	112.13	1343684.15	1470888	9.47	296.05
13	3	14	413042.22	440862	6.74	50.27	872661.27	943590	8.13	109.41	1819085.70	1956382	7.55	325.81
13	3	14	408811.15	439939	7.61	58.99	850207.47	916393	7.78	132.24	1768298.40	1911291	8.09	302.02
5	5	20	243255.92	257018	5.66	42.53	532275.53	565122	6.17	75.63	1088310.70	1168786	7.39	191.48
5	5	20	211330.59	238778	12.99	41.20	510637.65	562726	10.20	96.13	1066249.72	1139144	6.84	211.16
5	3	22	268067.28	282174	5.26	33.77	547072.94	578438	5.73	75.20	1129752.50	1206642	6.81	120.70
5	3	22	287200.52	320211	11.49	56.61	563613.65	613835	8.91	68.11	1188680.85	1282241	7.87	129.75
Average					8.34	80.06			7.83	158.05			7.78	287.26

TABLE 2: Test results of the proposed algorithm for the problem instances with $|\mathcal{V}| = 40$.

$|\mathcal{J}|$	$|\mathcal{K}|$	$|\mathcal{F}|$	$|\mathcal{T}| = 3$				$|\mathcal{T}| = 6$				$|\mathcal{T}| = 12$			
			LB	UB	%Gap	Time [s]	LB	UB	%Gap	Time [s]	LB	UB	%Gap	Time [s]
26	7	7	424766.97	471054	10.90	183.97	1411851.73	1522332	7.60	409.52	2868060.92	3075664	7.24	1393.03
26	7	7	360749.62	399556	10.76	215.84	751912.45	836474	11.25	387.16	1539449.29	1722836	11.91	952.08
29	4	7	418927.58	441254	5.33	105.27	865273.19	934006	7.94	209.19	1771774.46	1933194	9.11	485.31
29	4	7	485306.04	515006	6.12	170.86	1016416.93	1075553	5.82	263.59	2071972.70	2216043	6.95	625.59
17	7	16	550512.84	608016	10.45	200.25	1180080.13	1298552	10.04	655.23	2426278.35	2681688	10.53	2074.97
17	7	16	528692.03	585014	10.65	190.52	1099526.11	1211485	10.18	502.03	2236024.28	2462131	10.11	1370.17
18	4	18	734759.20	783204	6.59	412.23	1587132.99	1749738	10.25	1085.08	3176697.85	3560520	12.08	3414.08
18	4	18	647147.85	700289	8.21	312.63	1383698.67	1485373	7.35	665.97	2769306.69	2975719	7.45	1468.78
7	7	26	376082.83	409574	8.91	190.22	825644.77	894690	8.36	477.67	1678824.58	1874644	11.66	1249.97
7	7	26	387078.61	419447	8.36	146.84	784096.19	854760	9.01	495.31	1594309.42	1748548	9.67	1147.27
7	4	29	493481.06	515596	4.48	129.53	1026893.95	1077740	4.95	281.89	2091129.61	2201314	5.27	644.69
7	4	29	527232.35	568513	7.83	121.63	1103782.36	1189023	7.72	259.09	2240109.02	2419115	7.99	667.61
Average					8.22	198.32			8.37	474.31			9.17	1291.13

TABLE 3: Test results of the proposed algorithm for the problem instances with $|\mathcal{V}| = 50$.

$|\mathcal{J}|$	$|\mathcal{K}|$	$|\mathcal{F}|$	$|\mathcal{T}| = 3$				$|\mathcal{T}| = 6$				$|\mathcal{T}| = 12$			
			LB	UB	%Gap	Time [s]	LB	UB	%Gap	Time [s]	LB	UB	%Gap	Time [s]
30	10	10	504429.05	547776	8.59	603.09	1044268.14	1151346	10.25	2322.48	2138696.38	2348434	9.81	3356.77
30	10	10	601784.87	679393	12.90	588.67	1083093.64	1227322	13.32	1893.45	2069257.08	2307938	11.53	3370.17
34	6	10	619464.52	669098	8.01	531.14	1308349.43	1417378	8.33	1819.80	2636765.90	2854142	8.24	4683.91
34	6	10	603913.87	664708	10.07	485.77	1267603.33	1396176	10.14	1293.69	2626944.97	2887162	9.91	5836.92
20	10	20	810314.00	875940	8.10	1041.22	1671447.84	1809086	8.23	2197.13	3393505.82	3669690	8.14	5413.17
20	10	20	822869.60	892081	8.41	933.94	1669648.98	1843641	10.42	2719.97	3427822.66	3727153	8.73	5259.75
22	6	22	1056175.36	1146126	8.52	722.27	2185505.69	2372490	8.56	2221.61	4404884.25	4825116	9.54	4340.27
22	6	22	937230.54	1019646	8.79	938.88	2215333.07	2404507	8.54	1831.72	3958931.06	4289352	8.35	6382.42
10	10	30	651651.93	701858	7.70	839.84	1322310.58	1428274	8.01	1842.22	2753031.46	2960746	7.54	4175.56
10	10	30	669202.65	737403	10.19	932.56	1381848.96	1505385	8.94	2470.44	2829709.60	3071734	8.55	4343.27
10	6	34	685536.17	755326	10.18	592.84	1447676.89	1591850	9.96	1656.83	2938744.04	3228866	9.87	6059.58
10	6	34	685081.79	764624	11.61	917.94	1431087.35	1595460	11.49	3033.92	2912678.41	3215532	10.40	5632.03
Average					9.42	760.68			9.68	2108.60			9.22	4904.48

TABLE 4: Test results of the compared algorithms for the problem instances with $|\mathcal{V}| = 30$.

$|\mathcal{J}||\mathcal{K}|$	$|\mathcal{J}|$	$|\mathcal{T}| = 3$					$|\mathcal{T}| = 6$					$|\mathcal{T}| = 12$				
		ILB_{LCH}	ILB_{CX}	IUB_{LCH}	IUB_{CX}	IUB_{CXCH}	ILB_{LCH}	ILB_{CX}	IUB_{LCH}	IUB_{CX}	IUB_{CXCH}	ILB_{LCH}	ILB_{CX}	IUB_{LCH}	IUB_{CX}	IUB_{CXCH}
20 5	5	0.01	10.58	5.52	−0.17	−1.17	−0.19	11.45	3.73	1.38	1.16	−0.09	9.37	3.53	4.14	0.26
20 5	5	0.00	11.14	3.90	−6.46	−6.35	−0.14	10.38	4.68	−2.77	−2.27	−0.08	10.21	4.66	3.61	−0.32
22 3	5	0.45	9.55	11.88	2.30	−0.43	0.00	9.18	8.11	10.27	2.05	0.09	9.22	4.08	7.54	4.11
22 3	5	−0.01	8.48	4.00	−3.38	−0.76	0.02	8.88	6.29	−0.47	−1.72	0.04	8.65	3.97	9.44	4.98
13 5	12	−0.14	10.05	2.80	1.12	−1.10	−0.13	9.25	7.05	3.18	1.54	−0.71	8.34	3.92	6.77	3.88
13 5	12	−0.31	9.50	5.78	−2.84	−2.72	−0.56	8.76	6.76	1.07	1.87	0.11	9.65	6.44	7.89	6.85
13 3	14	−0.26	5.17	5.96	1.13	0.01	−0.17	5.52	4.07	8.08	3.13	−0.09	5.60	5.75	11.74	7.45
13 3	14	−0.20	6.19	5.09	−1.86	−1.45	−0.07	6.33	5.61	8.67	5.70	−0.08	6.29	6.29	9.87	8.04
5 5	20	−0.05	15.65	4.44	−1.48	−1.26	−0.56	16.84	5.09	1.00	0.12	−0.26	16.59	3.19	0.36	1.32
5 5	20	−0.02	12.99	0.36	−8.46	−8.34	0.00	16.94	3.87	−5.58	−5.31	−0.16	16.57	4.13	1.42	0.17
5 3	22	−0.06	13.93	2.91	−1.57	−1.87	0.02	13.92	3.82	0.99	−0.39	−0.58	12.42	3.86	4.14	6.70
5 3	22	−0.01	13.45	3.48	−6.51	−6.97	−0.16	12.66	3.17	−2.62	−2.75	−0.68	11.69	2.96	5.66	3.84
Average		−0.05	10.56	4.68	−2.35	−2.70	−0.18	10.39	5.13	2.56	0.69	−0.16	9.97	4.47	6.44	4.32

Moreover, it may be necessary to deal with vehicle specific lease periods. Lastly, proposing the efficient heuristic-based approach for solving large-scale problems would be the topic for the future research.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] J. F. Ehmke, "City logistics," in *Integration of Information and Optimization Models for Routing in City Logistics*, chapter 2, Springer New York, 2012.

[2] E. Taniguchia, R. G. Thompsonb, and T. Yamada, "Recent trends and innovations in modelling city logistics," *Procedia—Social and Behavioral Sciences*, vol. 125, pp. 4–14, 2014.

[3] E. E. Blanco and J. C. Fransoo, "Reaching 50 million nanostores: retail distribution in emerging megacities," TUE Working Paper 404, 2013.

[4] N. Anand, H. Quak, R. van Duin, and L. Tavasszy, "City logistics modeling efforts: trends and gaps—a review," *Procedia—Social and Behavioral Sciences*, vol. 39, pp. 101–115, 2012.

[5] T. G. Crainic, N. Ricciardi, and G. Storchi, "Models for evaluating and planning city logistics systems," *Transportation Science*, vol. 43, no. 4, pp. 432–454, 2009.

[6] J. F. Ehmke, A. Steinert, and D. C. Mattfeld, "Advanced routing for city logistics service providers based on time-dependent travel times," *Journal of Computational Science*, vol. 3, no. 4, pp. 193–205, 2012.

[7] J. Yang, J. Guo, and S. Ma, "Low-carbon city logistics distribution network design with resource deployment," *Journal of Cleaner Production*, 2013.

[8] E. Marcucci and R. Danielis, "The potential demand for a urban freight consolidation centre," *Transportation*, vol. 35, no. 2, pp. 269–284, 2008.

[9] M. Browne, J. Allen, and J. Leonardi, "Evaluating the use of an urban consolidation centre and electric vehicles in central London," *IATSS Research*, vol. 35, no. 1, pp. 1–6, 2011.

[10] L. Dablanc, G. Giuliano, K. Holliday, and T. O'Brien, "Best practices in urban freight management: lessons from an international survey," *Transportation Research Record*, no. 2379, pp. 29–38, 2013.

[11] C. Macharis and S. Melo, *City Distribution and Urban Freight Transport: Multiple Perspectives*, Edward Elgar Publishing, Cheltenham, UK, 2011.

[12] F. Su and M. J. Roorda, "The potential of urban freight consolidation for the Toronto central business district," in *Proceedings of the 93rd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2014.

[13] M. K. Triantafyllou, T. J. Cherrett, and M. Browne, "Urban freight consolidation centers: a case study in the UK retail sector," in *Proceedings of the Transportation Research Board 93rd Annual Meeting*, 2014.

[14] Y. Zhou and X. Wang, "Decision-making process for developing urban freight consolidation centers: analysis with experimental economics," *Journal of Transportation Engineering*, vol. 140, no. 2, 2014.

[15] J. F. Ehmke, S. Meisel, and D. C. Mattfeld, "Floating car based travel times for city logistics," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 338–352, 2012.

[16] M. A. Figliozzi, "Analysis of the efficiency of urban commercial vehicle tours: data collection, methodology, and policy implications," *Transportation Research Part B: Methodological*, vol. 41, no. 9, pp. 1014–1032, 2007.

[17] J.-B. Sheu, "A novel dynamic resource allocation model for demand-responsive city logistics distribution operations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 42, no. 6, pp. 445–472, 2006.

[18] K. R. Gue, "Effects of trailer scheduling on the layout of freight terminals," *Transportation Science*, vol. 33, no. 4, pp. 419–428, 1999.

[19] J. J. Bartholdi III and K. R. Gue, "Reducing labor costs in an LTL crossdocking terminal," *Operations Research*, vol. 48, no. 6, pp. 823–832, 2000.

[20] J. J. Bartholdi III and K. R. Gue, "The best shape for a crossdock," *Transportation Science*, vol. 38, no. 2, pp. 235–244, 2004.

[21] Y. Li, A. Lim, and B. Rodrigues, "Crossdocking—JIT scheduling with time windows," *Journal of the Operational Research Society*, vol. 55, no. 12, pp. 1342–1351, 2004.

[22] H. D. Ratliff, J. V. Vate, and M. Zhang, "Network design for load-driven cross-docking systems," GIT Technical Report, 1999.

[23] P. Chen, Y. Guo, A. Lim, and B. Rodrigues, "Multiple crossdocks with inventory and time windows," *Computers and Operations Research*, vol. 33, no. 1, pp. 43–63, 2006.

[24] H. Donaldson, E. L. Johnson, H. D. Ratliff, and M. Zhang, "Schedule-driven cross-docking networks," Tech. Rep., GIT, 1999.

[25] C. S. Sung and S. H. Song, "Integrated service network design for a cross-docking supply chain network," *Journal of the Operational Research Society*, vol. 54, no. 12, pp. 1283–1295, 2003.

[26] C. S. Sung and W. Yang, "An exact algorithm for a cross-docking supply chain network design problem," *Journal of the Operational Research Society*, vol. 59, no. 1, pp. 119–136, 2008.

[27] D. A. Serel, M. Dada, and H. Moskowitz, "Sourcing decisions with capacity reservation contracts," *European Journal of Operational Research*, vol. 131, no. 3, pp. 635–648, 2001.

[28] R. W. Seifert, U. W. Thonemann, and W. H. Hausman, "Optimal procurement strategies for online spot markets," *European Journal of Operational Research*, vol. 152, no. 3, pp. 781–799, 2004.

[29] L. van Norden and S. van de Velde, "Multi-product lot-sizing with a transportation capacity reservation contract," *European Journal of Operational Research*, vol. 165, no. 1, pp. 127–138, 2005.

[30] P. Toth, "Dynamic programming algorithms for the Zero-One Knapsack problem," *Computing*, vol. 25, no. 1, pp. 29–45, 1980.

[31] C. R. Reeves, *Modern Heuristic Techniques for Combinatorial Problems*, John Wiley & Sons, 1993.

[32] F. Glover and M. Laguna, *Tabu Search*, Kluwer Academic Publishers, Boston, Mass, USA, 1997.

[33] B. Crevier, J.-F. Cordeau, and G. Laporte, "The multi-depot vehicle routing problem with inter-depot routes," *European Journal of Operational Research*, vol. 176, no. 2, pp. 756–773, 2007.

[34] H. Üster and N. Maheshwari, "Strategic network design for multi-zone truckload shipments," *IIE Transactions*, vol. 39, no. 2, pp. 177–189, 2007.

Research Article

Comparison of Nonlinear Filtering Methods for Estimating the State of Charge of $Li_4Ti_5O_{12}$ Lithium-Ion Battery

Jianping Gao[1] and Hongwen He[2]

[1]*College of Vehicle & Transportation Engineering, Henan University of Science and Technology, Luoyang 471023, China*
[2]*National Engineering Laboratory for Electric Vehicles, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Hongwen He; hwhebit@bit.edu.cn

Accurate state of charge (SoC) estimation is of great significance for the lithium-ion battery to ensure its safety operation and to prevent it from overcharging or overdischarging. To achieve reliable SoC estimation for $Li_4Ti_5O_{12}$ lithium-ion battery cell, three filtering methods have been compared and evaluated. A main contribution of this study is that a general three-step model-based battery SoC estimation scheme has been proposed. It includes the processes of battery data measurement, parametric modeling, and model-based SoC estimation. With the proposed general scheme, multiple types of model-based SoC estimators have been developed and evaluated for battery management system application. The detailed comparisons on three advanced adaptive filter techniques, which include extend Kalman filter, unscented Kalman filter, and adaptive extend Kalman filter (AEKF), have been implemented with a $Li_4Ti_5O_{12}$ lithium-ion battery. The experimental results indicate that the proposed model-based SoC estimation approach with AEKF algorithm, which uses the covariance matching technique, performs well with good accuracy and robustness; the mean absolute error of the SoC estimation is within 1% especially with big SoC initial error.

1. Introduction

To address the two urgent tasks nowadays of protecting the environment and achieving energy sustainability, it is of a strategic significance on a global scale to replace the oil-dependent vehicles with electric vehicles. Lithium-ion batteries are currently considered to be the development trends of traction batteries and have been widely used in plug-in hybrid electric vehicles (PHEVs) due to its high power and energy density, its high voltage, being pollution-free, having no memory effect, its long cycle life, and its low self-discharge [1–3]. Battery management system (BMS) is essential for the lithium-ion battery to maximize its performance, ensure its safety, and extend its life. Estimation for battery state of charge (SoC) is one of the most key techniques in the BMS. Nevertheless, it is difficult to accurately estimate SoC, because SoC is an inner state of each battery cell which cannot be directly measured and is greatly influenced by many factors, including ambient temperature, discharging current, and battery aging [4, 5]. Therefore, the battery SoC has to be estimated with specific estimation techniques according to measured battery parameters, such as voltage, current, and temperature.

A wide variety of SoC estimation methods have been put forward to improve battery SoC determination, each one having its own advantage, most of which can be divided into four categories: looking-up table based methods, ampere-hour integral method, data-driven estimation methods, and model-based estimation methods [6–21]. In terms of the looking-up table based methods, the measurements of battery impedance, open circuit voltage, and so forth are commonly used to infer the SoC of batteries [6–9]. References [10, 11] put forward methods to determine battery SoC based on impedance measurements. Unfortunately, due to the uncertainty of driving cycles and the variable application environment, it is hard to measure these characteristic parameters accurately in real-time. The second type is the ampere-hour integral method, which has been widely applied to BMS and battery simulations of PHEVs [12]. The ampere-hour integral method acquires the SoC

by integrating the current over the time. This method is simple and can be easily implemented on-board, so that it has been widely used in practice. However, as an open-loop estimation algorithm, it cannot deal with problems caused by measurement noise and inaccurate initial SoC [13]. The third type is the data-driven estimation methods. Data-driven control methods merely use input-output data of the system to develop a controller. Since these methods do not require an accurate plant model, the estimations and assumptions introduced in the plant modeling step are omitted. For instance, the battery state estimations based on artificial neural network and fuzzy logic models are demonstrated with good accuracy [14, 15]. However, these algorithms are very sensitive to their parameters and they may even be not convergent with bad parameters selection when the train data cannot completely cover the present operating conditions. The last type is model-based estimation method [16–20]. Plett used an extended Kalman filter (EKF) to identify unknown parameter and adaptively estimate the battery's SoC [16–18]. An inappropriate matrix of the system noise may lead to remarkable errors and even divergence [19]. As an alternative approach, SoC estimation via SoC-OCV look-up table is put forward and the battery open circuit voltage (OCV) is estimated in real-time with a recursive least squares algorithm [21]. Additionally, the unscented Kalman filter (UKF) is also investigated to estimate the lithium-ion battery (LiB) SoC [20]. However, the OCV-based SoC estimation method can hardly safeguard the estimation accuracy dynamically for its open-loop characteristics.

In view of battery behavior and performance being relatively vulnerable to operating conditions and aging levels, what is important for us to do is to achieve the accurate SoC estimation in the long-term. In this point, there are three difficulties that should be considered seriously for achieving efficient and reliable battery SoC estimation for BMS: (1) accurate parametric modeling approach, (2) reliable and robust state estimation algorithm, and (3) systematic modeling and estimation scheme.

In solving the first problem, different kinds of battery models have been proposed and applied to their application field. However, if one applied them to BMS, they can hardly achieve desired performance. It is because of that the adaptive parameter update technique has been neglected; as a result, the model error will be larger as the battery aged or operated conditions changed. In dealing with the second problem, several advanced Kalman filters [13, 16–18] have been widely used and the estimation accuracies are appropriate for BMS application in a limited operating condition. For overcoming the last difficulties, the offline model and adaptive filters are widely used. However, it is limited by the unupdated battery model; the SoC estimation will be influenced by the unavoidable uncertainties from battery which aged and operating conditions varied.

The $Li_4Ti_5O_{12}$ which can release lithium ions repeatedly for recharging and quickly for high current has been accepted as a novel anode material in $Li_4Ti_5O_{12}$ LiB. Nevertheless, its dynamic behavior is very different from other LiBs. Traditional battery model fails to ensure high prediction precision in its voltage prediction. A key contribution of this study is that a general three-step model-based battery SoC estimation scheme, which includes adaptive model parameter updating technique for improving the parametric modeling performance, an open interface for employing adaptive filters to solve the hidden states from strong time-varying dynamic system and series structure based systematic modeling and estimation approach. With the proposed scheme, we have compared the performance of three commonly used filters, which include EKF algorithm-based SoC observer, UKF algorithm-based SoC observer, and AEKF algorithm-based SoC observer. In addition, the Gaussian model describing the open circuit voltage behavior has been developed to improve the performance of the battery model, and then the improved lumped-parameter battery model was applied to the three filters based SoC estimator to improve their prediction performance for $Li_4Ti_5O_{12}$ LiB (its voltage bounds are 2.7 V and 1.5 V, resp.).

This paper is organized as follows. A description of the general model-based battery SoC estimation scheme is given in Section 2. The verification experiments and results are described in Section 3. A detailed comparison among the EKF-based, UKF-based, and AEKF-based SoC estimation approaches is illustrated in Section 4. Finally, conclusions are drawn in Section 5.

2. Model-Based SoC Estimator

2.1. General Model-Based Battery SoC Estimation Scheme. Figure 1 presents the model-based battery SoC estimation scheme. It contains three steps.

Step 1 (real-time measurement). After the driving cycles loaded on the battery, the data-sampling module can measure battery current and voltage in real-time. The measured current and voltage are served as input data for the next modeling and state estimation steps.

Step 2 (parametric modeling). Based on the previous research experience on battery model selection [21, 22], one RC network based lumped-parameter battery model (shown in Figure 2) has been chosen for describing the dynamic behavior of the LiB cell. In parametric modeling step, the battery model has been divided into two parts: OCV and dynamic voltage. Considering that OCV varies with battery SoC, it is common to use variation of OCV to achieve an accurate feedback of SoC. Therefore, we use a nonlinear battery OCV model to describe battery voltage source. With the estimated SoC, the OCV will be updated accordingly. For dynamic voltage part, we use recursive least squares with optimal forgetting factor to identify the parameter in real-time. The description of recursive least squares can be found in [22]. Through the process of Step 2, we can get online model parameter with real-time measurements of battery current and voltage.

Step 3 (model-based approach). With the real-time battery model from Step 2, we can build the state-space function of the battery accordingly. With the state-space function, the adaptive filters can be used to estimate the hidden state (SoC)
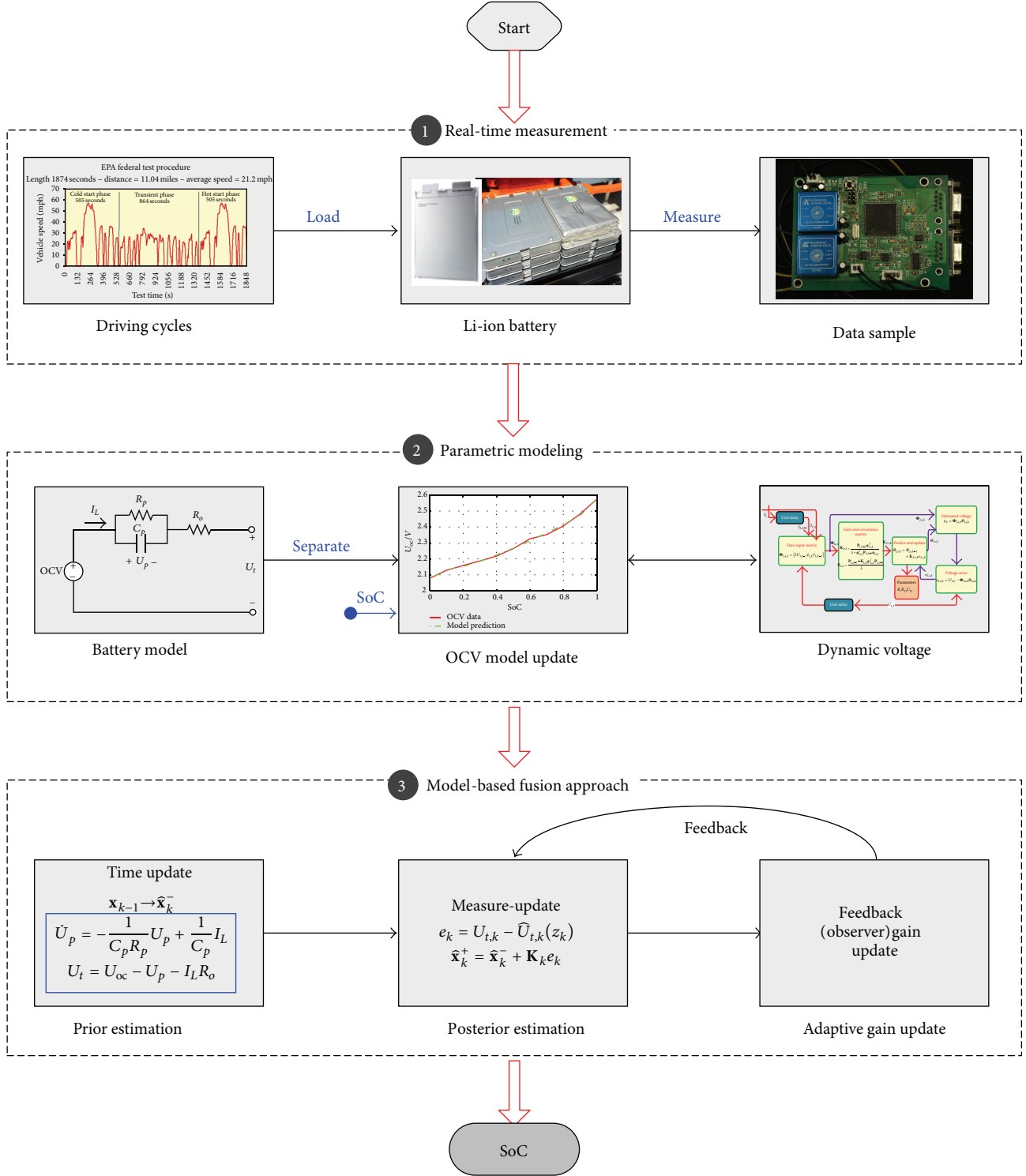
FIGURE 1: General model-based battery SoC estimation scheme.

through operation processes of prior estimation, posterior estimation, and feedback control. In this study, we use three kinds of Kalman filters to verify this approach [21, 23, 24]. The three kinds of advanced Kalman filters have been employed to develop the SoC estimator and their performance has been analyzed and evaluated in Section 4.

It is noted that in essence the Kalman filter based SoC estimation approach is a fusion method. It fuses the estimation results of OCV-based look-up table method and ampere-hour counting method through the state-space function of battery. SoC serves the communication link between OCV-based look-up table method and ampere-hour counting
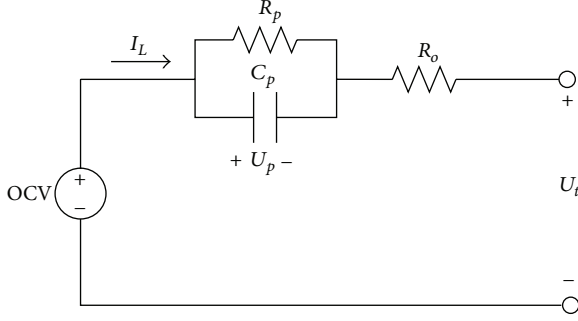
FIGURE 2: The schematic diagram of the lumped-parameter model.

method; an inaccurate ampere-hour counting method will lead to inaccurate SoC estimate and results in a larger erroneous OCV, which will bring bigger voltage error in turn. Thus, the OCV can be served as a feedback control link to regulate the SoC estimation error in ampere-hour counting method. Therefore, the fusion method has optimal SoC estimation from its capacities of correction, weighting, and filtering. Lastly, together with the real-time measurements of battery current and voltage for online identification of the parameter of battery model, the proposed scheme in fact is model-based fusion method.

2.2. Battery Model. In order to execute the state estimation with state-space-based filter algorithm, we need a model to describe relationship between SoC with battery voltage. According to our previous experience in battery modeling [21], one RC network based lumped-parameter model was chosen for this study. However, to improve the model's accuracy and enhance the relationship between the SoC with battery voltage, a Gaussian model [17] is developed to improve its prediction precision; the model is illustrated in Figure 2. The main dynamic voltage behavior is described by

$$\dot{U}_p = -\frac{1}{C_p R_p} U_p + \frac{1}{C_p} I_L,$$
$$U_t = U_{oc} - U_p - I_L R_o,$$
(1)

where U_{oc} and I_L are the OCV and load current, respectively. U_t and R_o denote the terminal voltage and ohmic resistance of battery, respectively. The RC network is employed to describe the dynamic voltage behavior including polarization resistance R_p and polarization capacitance C_p, and U_p represents the polarization voltage across C_p. It is noted that the positive value of battery current is assumed for discharging operation while the negative value is assumed for charging operation.

To describe the nonlinear characteristic of battery open circuit voltage, the Gaussian model has been selected:

$$U_{oc} = a_1 \exp\left(-\left(\frac{(z - b_1)}{c_1}\right)^2\right)$$
$$+ a_2 \exp\left(-\left(\frac{(z - b_2)}{c_2}\right)^2\right)$$
(2)
$$+ a_3 \exp\left(-\left(\frac{(z - b_3)}{c_3}\right)^2\right),$$

where a_i, b_i, c_i ($i = 1, 2, 3$) are the constants chosen to make the U_{oc} model fit the SoC-OCV well and z is the abbreviation of battery SoC.

2.3. Parameters Identification. For identifying the parameters of the lumped-parameter battery model, a regression equation in discretization form for (1) is needed which has been rewritten as follows [3]:

$$U_{p,k+1} = e^{-\Delta t/\tau} U_{p,k} + \left(1 - e^{-\Delta t/\tau}\right) R_p I_{L,k},$$
$$U_{t,k} = U_{oc,k} - U_{p,k} - I_{L,k} R_o,$$
(3)

where Δt denotes the sampling interval and $U_{p,k+1}$ can be equivalent to $U_p((k + 1)\Delta t)$ at the $(k + 1)$th sampling times. τ denotes the time constant and equals $\tau = R_p \times C_p$. $I_{L,k}$ is the current at the time index k, and $U_{t,k}$ and $U_{oc,k}$ are terminal voltage and OCV at the time index k, respectively.

With (3),

$$U_{t,k+1} = U_{oc,k+1} - \left(e^{-\Delta t/\tau} U_{p,k} + \left(1 - e^{-\Delta t/\tau}\right) R_p I_{L,k}\right)$$
$$- I_{L,k+1} R_o.$$
(4)

Defining $E_t = U_t - U_{oc}$, then (4) can be rewritten by

$$E_{t,k+1} = -\left(e^{-\Delta t/\tau} U_{p,k} + \left(1 - e^{-\Delta t/\tau}\right) R_p I_{L,k}\right)$$
$$- I_{L,k+1} R_o.$$
(5)

The difference equation of (5) can be used to eliminate $U_{p,k}$; then,

$$E_{t,k+1} = e^{-\Delta t/\tau} E_{t,k} + (-R_o) I_{L,k+1}$$
$$+ \left(e^{-\Delta t/\tau} R_o - \left(1 - e^{-\Delta t/\tau}\right) R_p\right) I_{L,k}.$$
(6)

Thus,

$$E_{t,k+1} = \alpha_1 E_{t,k} + \alpha_2 I_{L,k+1} + \alpha_3 I_{L,k},$$
(7)

where

$$\alpha_1 = e^{-\Delta t/\tau},$$
$$\alpha_2 = -R_o,$$
(8)
$$\alpha_3 = e^{-\Delta t/\tau} R_i - \left(1 - e^{-\Delta t/\tau}\right) R_p.$$

From (8), we can obtain

$$R_i = -\alpha_2,$$
$$R_p = \frac{\alpha_1 \alpha_2 + \alpha_3}{\alpha_1 - 1},$$
(9)
$$C_p = \frac{(1 - \alpha_1) \Delta t}{(\alpha_1 \alpha_2 + \alpha_3) \log(\alpha_1)}.$$

We assume $y_k = \Phi_{Ls,k} \theta_{Ls,k}$, where

$$y_k = E_{t,k},$$
$$\Phi_{Ls,k} = \begin{bmatrix} U_{t,k-1} & I_{L,k} & I_{L,k-1} \end{bmatrix},$$
(10)
$$\theta_{Ls,k} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{bmatrix}^T,$$

where $\boldsymbol{\Phi}_{Ls,k}$ and $\boldsymbol{\theta}_{Ls,k}$ denote matrixes of input data and parameters at time index k, respectively.

We use the recursive least squares to implement the parameters estimating process; the parameter estimates are updated at each sampling intervals. The forgetting factor λ is optimized by the genetic algorithm and the optimization objective is to achieve the minimum sum of squares of the estimated voltage error. After presetting initial vector $\boldsymbol{\theta}_{Ls,0}$ and its error covariance matrix $\mathbf{P}_{Ls,0}$, the parameter vector $\boldsymbol{\theta}_{Ls,k}$ can be determined.

2.4. SoC Estimation with the Kalman Filter Algorithm. As mentioned in the introduction section, SoC is a necessary index for ensuring the safety operation of the batteries. SoC can be calculated by the following equation:

$$z_k = z_{k-1} - \frac{\eta_i I_{L,k} \Delta t}{C_a}, \tag{11}$$

where z_k denotes the SoC at the kth sampling interval and η_i denotes the columbic efficiency of battery. Δt is the sampling interval. C_a denotes the maximum available capacity of battery.

Based on the dynamic voltage model of battery and the SoC computational equation, we can build the state equation for recursive prediction and the state equation is described by the following equation:

$$\begin{aligned} \mathbf{X}_{k+1} &= f\left(\mathbf{X}_k, \mathbf{u}_k\right) + \omega_k, \\ \mathbf{Y}_k &= h\left(\mathbf{X}_k, \mathbf{u}_k\right) + v_k, \end{aligned} \tag{12}$$

where the function $f(X, u)$ is used to describe the state transition of the nonlinear system, the parameter $\mathbf{X}$ denotes the system state, and the $\mathbf{u}$ denotes the input of this system. The function $g(X, u)$ is used to describe the measurement process of this nonlinear system, Y denotes the measurement of the system. To describe the unavoidable noise information, a Gaussian white noise was assumed to describe the process noise ($\mathbf{w}$) and measurement noise ($\mathbf{v}$). These two types of noise possess zero of mean values, and their covariance is described by vectors $\mathbf{Q}$ and $\mathbf{R}$, respectively.

A few of nonlinear filtering methods have been applied to determining the SoC for electric vehicles batteries, especially of Kalman filtering methods. They are extensively used not only for parameter identification and stats estimation, but also for other typical engineering problems such as global positioning system [16–18, 21]. Based on [16–18], we can conclude the detailed computational process for EKF algorithm and AEKF algorithm, which have been listed below. And the algorithm of UKF algorithm can be found in [13].

Summary of the EKF algorithm is as follows.

Step 1 (initialization). For $k = 0$, set

$$\begin{aligned} \widehat{\mathbf{X}}_0^+ &= E\left[X_0\right], \\ \mathbf{P}_0^+ &= E\left[\left(X_0 - \widehat{\mathbf{X}}_0^+\right)\left(X_0 - \widehat{\mathbf{X}}_0^+\right)^T\right]. \end{aligned} \tag{13}$$

Step 2 (computation). For $k = 1, 2, \ldots$, compute the following.

State estimate time update is as follows:

$$\widehat{\mathbf{X}}_k^- = f\left(\widehat{\mathbf{X}}_{k-1}^+, \mathbf{u}_k\right). \tag{14}$$

Error innovation is as follows:

$$\mathbf{e}_k = \mathbf{Y}_k - g\left(\widehat{\mathbf{X}}_k^-, \mathbf{u}_k\right). \tag{15}$$

Error covariance time update is as follows:

$$\mathbf{P}_k^- = \mathbf{A}_k \mathbf{P}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_k. \tag{16}$$

Kalman gain matrix is as follows:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}_k^T \left(\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}_k\right)^{-1}. \tag{17}$$

State estimate measurement update is as follows:

$$\widehat{\mathbf{X}}_k^+ = \widehat{\mathbf{X}}_k^- + \mathbf{K}_k \mathbf{e}_k. \tag{18}$$

Noise and error covariance measurement update is as follows:

$$\mathbf{P}_k^+ = \left(\mathbf{I} - \mathbf{K}_k \mathbf{C}_k\right) \mathbf{P}_k^-, \tag{19}$$

where

$$\begin{aligned} \mathbf{A}_k &= \left.\frac{\partial f\left(\mathbf{X}_k, \mathbf{u}_k\right)}{\partial \mathbf{X}}\right|_{\mathbf{X} = \widehat{\mathbf{X}}_k^-}, \\ \mathbf{C}_k &= \left.\frac{\partial h\left(\mathbf{X}_k, \mathbf{u}_k\right)}{\partial \mathbf{X}}\right|_{\mathbf{X} = \widehat{\mathbf{X}}_k^-}, \end{aligned} \tag{20}$$

where $\mathbf{K}_k$ is Kalman gain matrix at the kth sampling time; $\widehat{\mathbf{X}}_k^-$ is the *priori* estimate of $\mathbf{X}_k$ *before* the measurement $\mathbf{Y}_k$ is taken into account, $\widehat{\mathbf{X}}_k^+$ is the estimate of $\mathbf{X}_k$ after the measurement $\mathbf{Y}_k$ is taken into account, which is called *posteriori* estimate; $U_{t,k}$ and $\widehat{U}_{t,k}$ are the terminal voltage measured and estimated by the battery model at the kth sampling time, respectively.

Summary of the AEKF algorithm is as follows.

Step 1 (initialization). For $k = 0$, set:

$$\begin{aligned} \widehat{\mathbf{X}}_0^+ &= E\left[X_0\right], \\ \mathbf{P}_0^+ &= E\left[\left(X_0 - \widehat{\mathbf{X}}_0^+\right)\left(X_0 - \widehat{\mathbf{X}}_0^+\right)^T\right]. \end{aligned} \tag{21}$$

Step 2 (computation). For $k = 1, 2, \ldots$, compute the following.

State estimate time update is as follows:

$$\widehat{\mathbf{X}}_k^- = f\left(\widehat{\mathbf{X}}_{k-1}^+, \mathbf{u}_k\right). \tag{22}$$

Error innovation is as follows:

$$\mathbf{e}_k = \mathbf{Y}_k - g\left(\widehat{\mathbf{X}}_k^-, \mathbf{u}_k\right). \tag{23}$$

TABLE 1: Results of the coulomb efficiency.

Current/A	3	10	20	40	60	80
Coulombic efficiency in discharging process	1	1	0.99	0.98	0.95	0.91
Coulombic efficiency in charging process	1	1	0.99	0.97	—	—

Adaptive law-covariance matching is as follows:

$$\mathbf{H}_k = \frac{1}{M}\sum_{i=k-M+1}^{k}\mathbf{e}_k\mathbf{e}_k^T,$$

$$\mathbf{R}_k = \mathbf{H}_k - \mathbf{C}_k\mathbf{P}_k^-\mathbf{C}_k^T. \tag{24}$$

Error covariance time update is as follows:

$$\mathbf{P}_k^- = \mathbf{A}_k\mathbf{P}_{k-1}\mathbf{A}_k^T + \mathbf{Q}_k. \tag{25}$$

Kalman gain matrix is as follows:

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{C}_k^T\left(\mathbf{C}_k\mathbf{P}_k^-\mathbf{C}_k^T + \mathbf{R}_k\right)^{-1}. \tag{26}$$

State estimate measurement update is as follows:

$$\widehat{\mathbf{X}}_k^+ = \widehat{\mathbf{X}}_k^- + \mathbf{K}_k\mathbf{e}_k. \tag{27}$$

Noise and error covariance measurement update is as follows:

$$\mathbf{Q}_k = \mathbf{K}_k\mathbf{H}_k\mathbf{K}_k^T,$$

$$\mathbf{P}_k^+ = \left(\mathbf{I} - \mathbf{K}_k\mathbf{C}_k\right)\mathbf{P}_k^-, \tag{28}$$

where $\mathbf{H}_k$ is the innovation covariance matrix based on the innovation sequence $\mathbf{e}_k$ inside moving estimation window of size M at the kth sampling time.

3. Experimental Configurations and Battery Test

A test platform introduced in our previous work [21] was employed to implement the systematic test for evaluating the performance of the proposed model-based SoC estimation algorithm. Battery used in the test is the $Li_4Ti_5O_{12}$ lithium-ion battery. The battery was placed in a thermal chamber and the test schedule can be found in [21]. The maximum available capacity of the battery is 9.8 Ah [25]. After the static capacity test, an efficiency test has been implemented for determining the coulomb efficiency of the $Li_4Ti_5O_{12}$; the results are listed in Table 1 [21].

When we know the exact capacity and coulomb efficiency, we can carry out the OCV test to calibrate the relationship between battery SoC and OCV, and we use 3 A to charge and discharge the cell. In this paper, the hysteresis is ignored. The OCV curves are plotted in Figure 3.

Figure 3 gives the OCV comparison between the model estimates and the test data. It suggests that the proposed
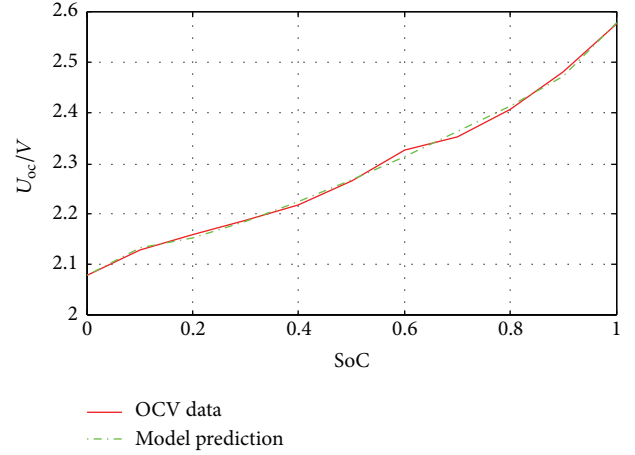
FIGURE 3: The OCV curves of the test data and the model prediction.

model can track the battery OCV behavior closely. The OCV function can be described by the following equation:

$$U_{oc} = 2.85\exp\left(-\left(\frac{(z-1.614)}{1.78}\right)^2\right)$$
$$+ 0.69\exp\left(-\left(\frac{(z-0.027)}{0.56}\right)^2\right)$$
$$+ 0.28\exp\left(-\left(\frac{(z-0.071)}{0.025}\right)^2\right). \tag{29}$$

In addition to the above three types of test, the hybrid pulse power characteristic (HPPC) [26] and the dynamic stress test (DST) have been carried out for verifying the proposed general model-based SoC estimation method. The measurements of HPPC test and DST are plotted in Figures 4 and 5, respectively.

To achieve an exact SoC, we first charged the battery with CCCV charge mode at the nominal current. Then we discharged some capacity of the battery with the nominal current to achieve an accurate initial SoC. Afterwards, the DST test was loaded and executed. Lastly, a further nominal current discharge experiment was conduct to gain an accurate terminal SoC. Based on the known exact SoC and accurate coulomb efficiency, we can determine the reference SoC profiles with the SoC definition-based ampere-hour counting method. The SoC profiles of DST test shown in Figure 5(c) will serve as a reference SoC for the evaluation the robust SoC estimation, and its actual initial SoC is 0.8075.
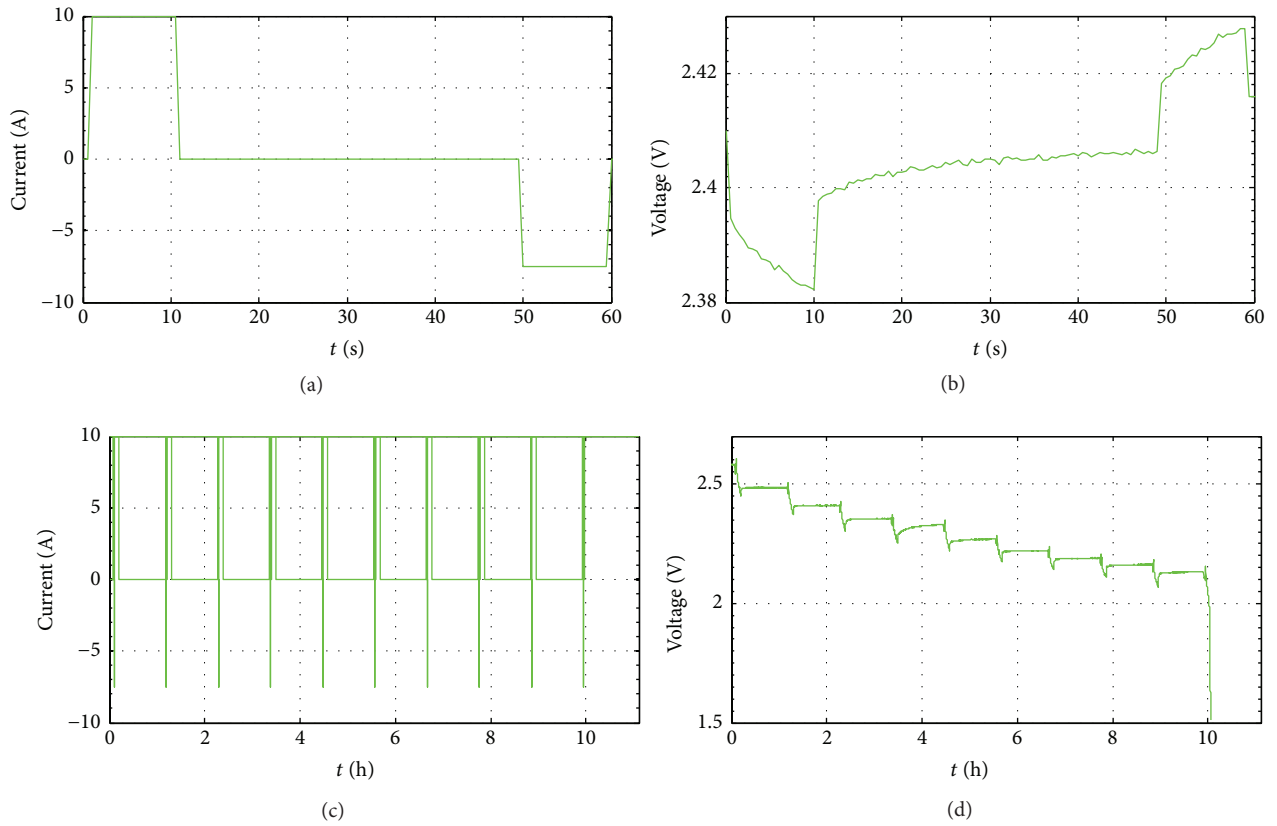
FIGURE 4: Profiles of the HPPC test: (a) battery current of one cycle; (b) battery terminal voltage (SoC = 0.8); (c) battery current for ten cycles; and (d) complete voltage profile.
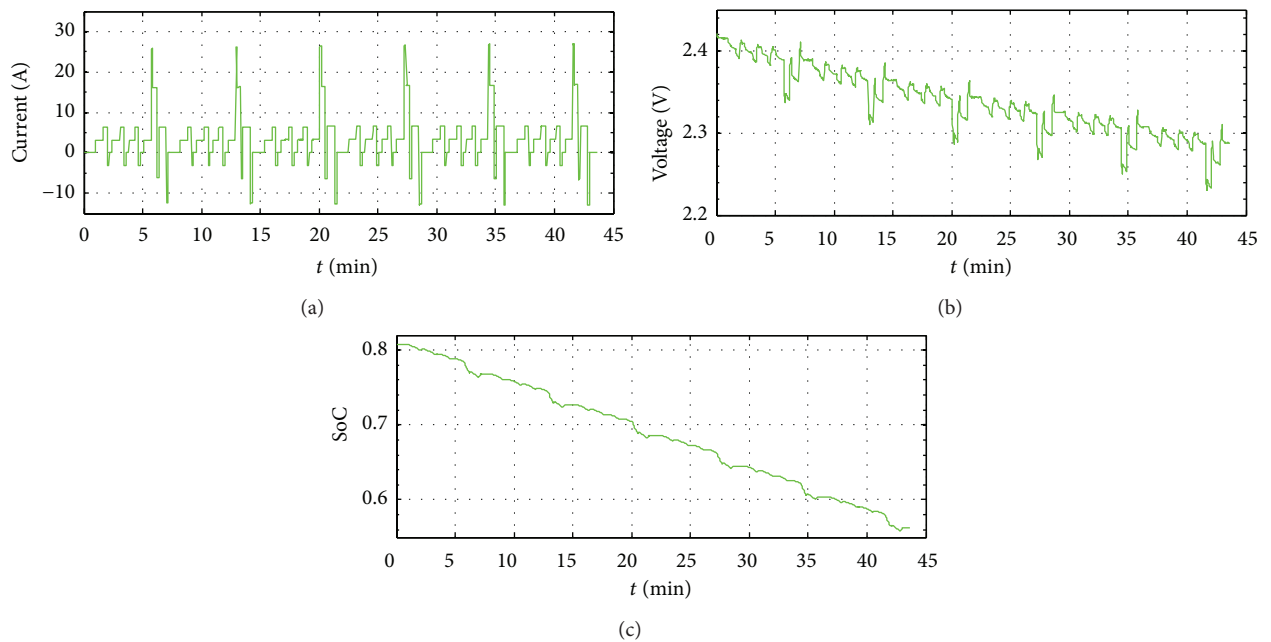

FIGURE 5: Profiles of the DST cycles: current and voltage are plotted in (a) and (b); SoC is plotted in (c).
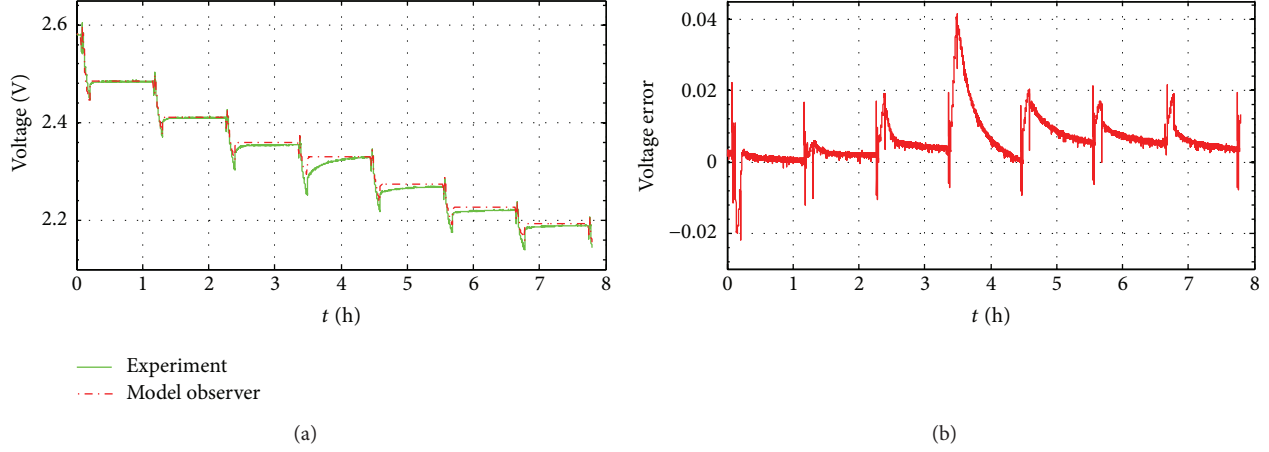
(a)

(b)

FIGURE 6: The comparisons profiles of the observer and experiment: (a) voltage; (b) voltage error.

4. Verification Analysis and Discussion

4.1. Parametric Modeling. Through the online parameter identification operation, we can get the real-time battery model. The voltage profiles of the experimental data and observer are presented in Figure 6(a), and the voltage error is shown in Figure 6(b). A statistical analysis on the terminal voltage errors has been conducted and the result is listed in Table 2. It shows that the maximum error of the terminal voltage is less than 2% of its nominal voltage. Additionally, its mean error is 6 mV. It can be concluded that the proposed improved model has good dynamic voltage prediction performance.

4.2. Analysis on the SoC Estimation Results. The following verification and analysis are based on the AEKF algorithm, and the other two Kalman filters will be discussed in the Section 4.3. The SoC estimation is conducted and the results are shown in Figure 8. Figure 7(a) presents the comparison profiles of the voltages between the estimations and the experimental result. Figure 7(b) presents the voltage error curve for Figure 7(a). Figure 7(c) presents the SoC comparison profiles between the estimations and the reference SoC. Figure 7(d) presents the SoC estimation error.

We can observe that the prediction inaccuracy of the battery terminal voltage is below 1%, which is lower than the prediction result plotted in Figure 6(b). The prediction precision depends adaptively on correction performance by adjusting the Kalman filter gain matrix based on the voltage error bound. Figure 7(d) indicates that the SoC estimation error arises with the terminal voltage estimation error, while the AEKF approach can correct the voltage error adaptively and quickly. It also shows that the model precision is important during the estimation for battery SoC.

4.3. Evaluation on the Robust Performance. With an accurate initial SoC, most of the SoC estimators can achieve desired estimation performance in a period of time, such as ampere-hour integral method. However, the estimation accuracy against different unknown initial SoC makes lots of methods

TABLE 2: The statistics list of the terminal voltage error.

Error	Maximum/V	Minimum/V	Mean/V	Covariance/V^2
Value	0.0416	−0.0218	0.0060	$3.88e-005$

unacceptable for electric vehicles application. In this section, we will discuss whether the AEKF-based SoC estimation can achieve accurate SoC estimation with the erroneous initial SoC. Two types of erroneous initial SoC, 0.95 and 0.50, are applied to implement the evaluation. The estimation results are plotted in Figure 8.

From Figures 8(a) and 8(c), we can observe that the performance of the model-based SoC estimator cannot be affected by the erroneous initial SoC. In contrast, the convergence trajectories of the SoC estimations are very similar, especially when the initial error in battery SoC has been corrected. We also can observe that the SoC estimation errors are less than 1%. In order to evaluate the estimation performance more intuitively, we select the index of the mean absolute error (MAE) to describe the convergence trajectory. The MAE can be calculated by the following equation:

$$z_{\text{mae},k} = \frac{\sum_{j=0}^{k} \left| \text{SoC}_k - \widehat{\text{SoC}_k} \right|}{k+1}, \tag{30}$$

where $z_{\text{mae},k}$ represents the MAE index of the battery SoC estimations up to and involving the kth sample time; $\widehat{\text{SoC}_k}$ denotes the SoC estimation value under the time index k.

Figure 9 shows the MAE results. It indicates that the MAE of the SoC estimations which started with incorrect initial SoC values can converge to within 1% in 1 minute (60 sample intervals).

4.4. Discussion. Based on the above analysis, we can find that the AEKF algorithm is suitable for applying to proposed general model-based battery SoC estimation scheme, and it can reach high estimation accuracy. To discuss the suitability of the proposed general SoC estimation scheme and compare the AEKF algorithm with other widely used methods, UKF
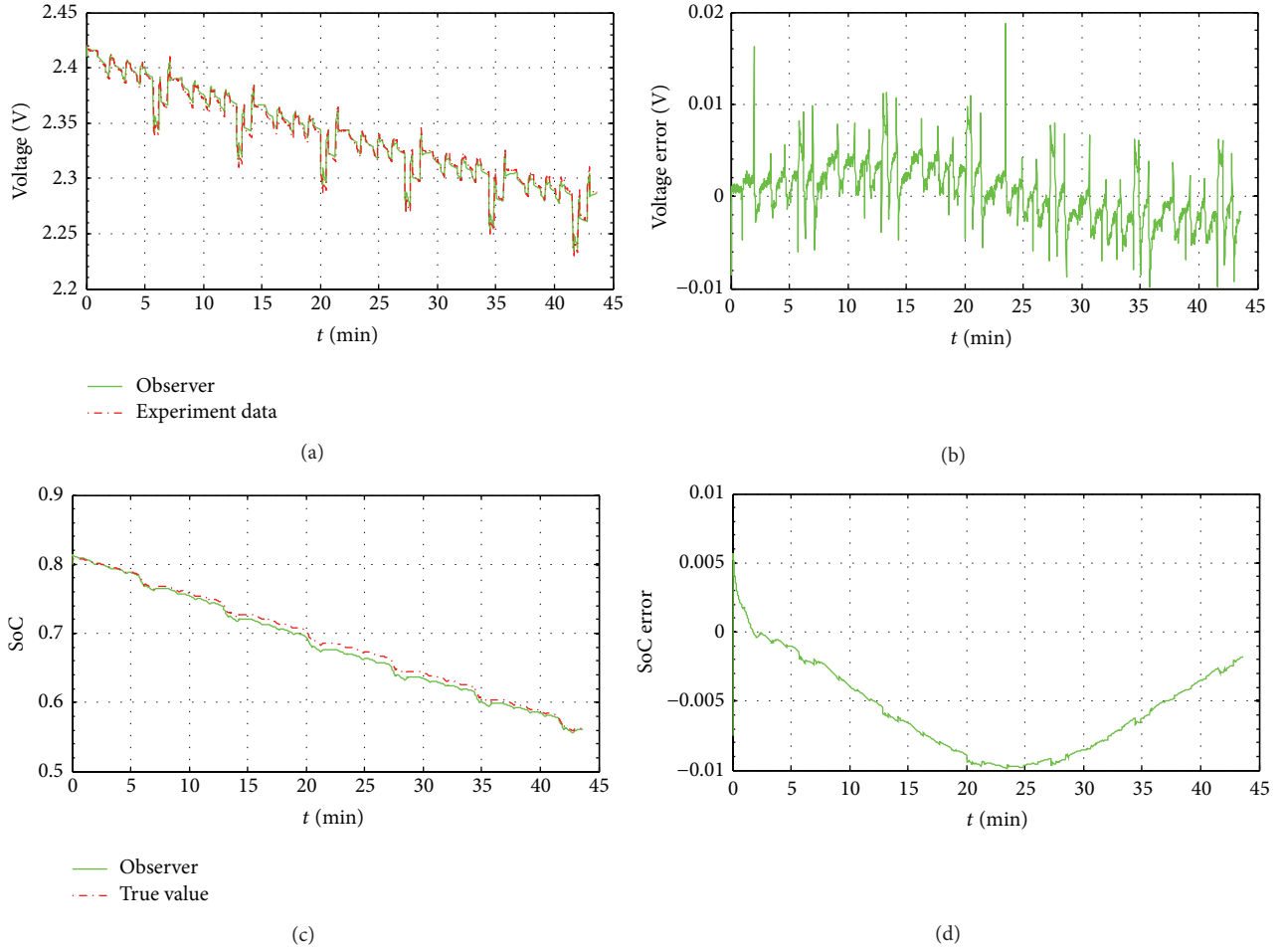
Figure 7: SoC estimation with AEKF approach: (a) voltage; (b) voltage error; (c) SoC; (d) SoC error.

algorithm and EKF algorithm, we have made systematical analysis. The real-time model-based SoC estimation results using the EKF and UKF are plotted in Figure 10. The estimation errors between the EKF-based SoC estimations and UKF-based estimations are plotted in Figure 11 and the MAE index is plotted in Figure 12.

From Figures 8–12, it can be observed that the AEKF-based approach achieves the best accuracy in these three approaches. Additionally, the AEKF-based SoC estimation is more precise than the EKF-based SoC estimation. The better estimation performance of the AEKF algorithm based method is due to the fact that the adaptively updating for the error covariance greatly improves the estimation performance. The SoC estimation error between the EKF-based method and the UKF-based method is virtually indiscernible.

The comparisons between the estimation inaccuracies and MAE index of the AEKF-based, EKF-based and UKF-based methods, show that the maximum MAE of the EKF-based and UKF-based approaches are around 3%, which is higher than the AEKF-based approach. In conclusion, the proposed general model-based battery SoC estimation scheme can be applied to estimate the SoC of batteries accurately with good robust performence. More importantly,

the performance of the proposed scheme is not sensitive with the operated nonlinear filtering methods. For the algorithms of AEKF, EKF, and UKF, all of their estimation errors are less than 5%. It is acceptable for the current requirements of the battery management system. Furthermore, the AEKF algorithm, which can update the error covariance matrix adaptively, has the best estimation accuracy when applied to the proposed scheme and $Li_4Ti_5O_{12}$ lithium-ion battery cell.

5. Conclusions

This paper presents a comparison of nonlinear filtering methods for estimating the SoC of $Li_4Ti_5O_{12}$ lithium-ion battery. The Gaussian model has been selected to improve the prediction precision of the dynamic battery model. With the new battery model, general model-based battery SoC estimation has been proposed. It contains the adaptive model parameter updating technique for improving the parametric modeling performance, an open interface for employing adaptive filters to solve the hidden states from strong time-varying dynamic system, and series structure based systematic modeling and estimation approach. Three Kalman filters are employed to build model-based SoC estimator. With
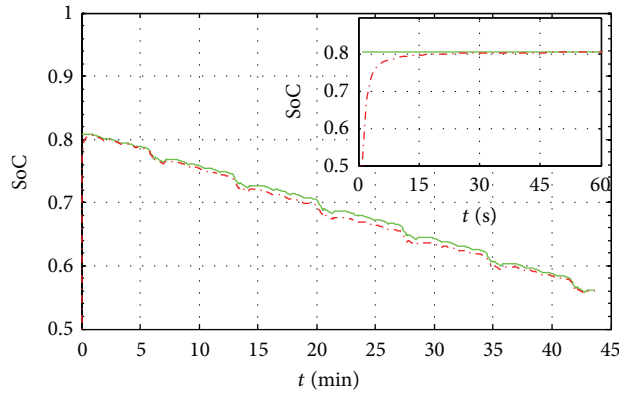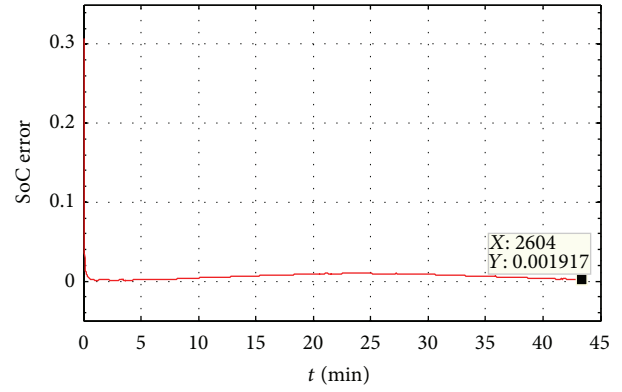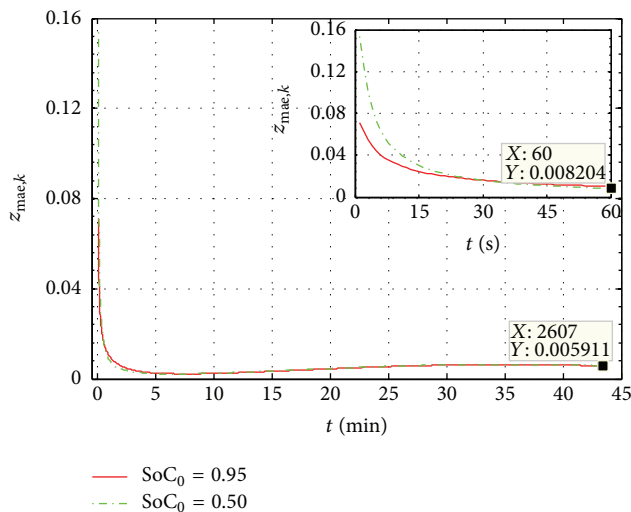
(a)

(b)
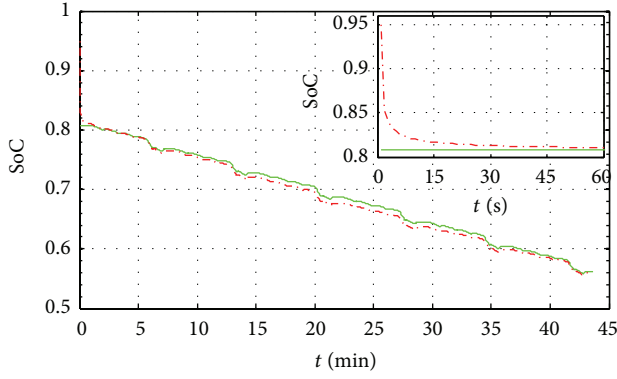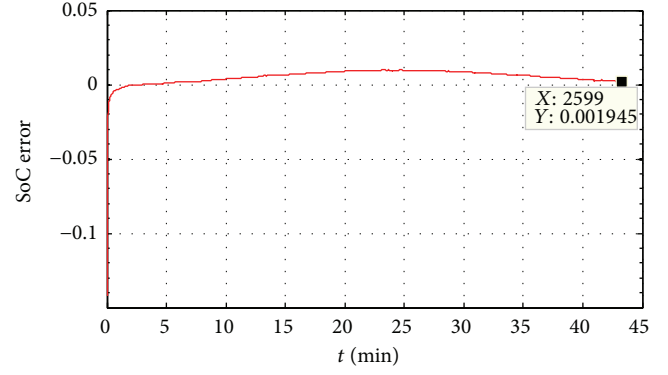
(c)

(d)

Figure 8: Self-correcting capability for erroneous initial SoC: (a) estimation with $SoC_0 = 0.95$; (b) SoC estimation error for (a); (c) estimation with $SoC_0 = 0.50$; and (d) SoC estimation error for (c).

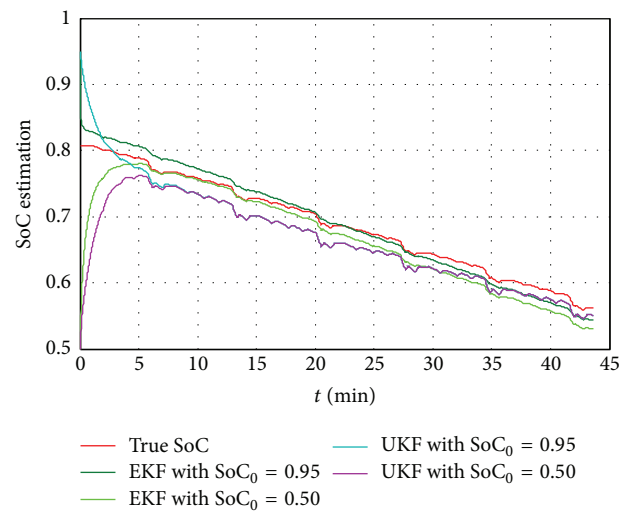Figure 9: MAE results of SoC estimation with AEKF approach.

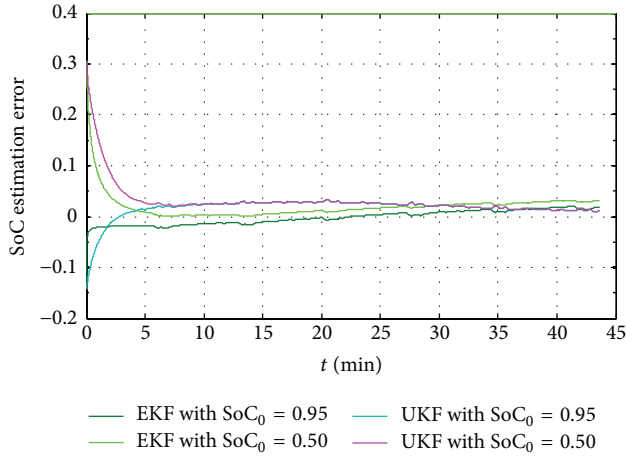Figure 10: SoC results of EKF-based estimation and UKF-based estimation with two different types of SoC_0.

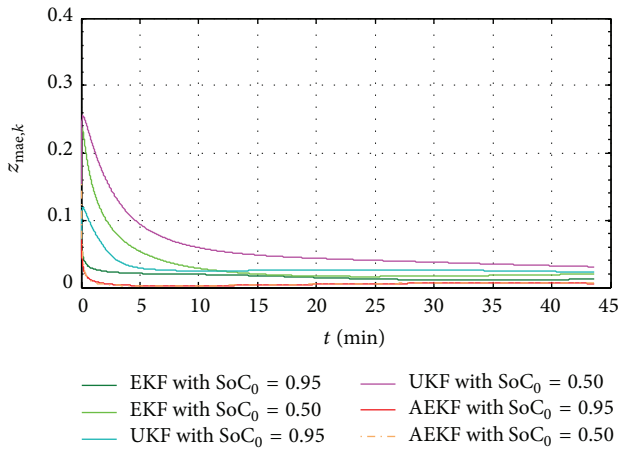FIGURE 11: Errors of EKF-based estimation and UKF-based estimation.

FIGURE 12: MAE results of EKF-based estimation and UKF-based estimation.

the proposed model-based scheme, three advanced Kalman filters, which include extended Kalman filter, unscented Kalman filter, and adaptive extended Kalman filter, have been employed to develop the SoC estimator.

The detailed evaluation and comparison are made for model-based SoC estimator. A comparison for the SoC estimation approach among the AEKF-based, EKF-based, and UKF-based algorithms with the $Li_4Ti_5O_{12}$ lithium-ion battery shows that the proposed method has superior performance, which indicates that the covariance matching approach for EKF is a useful way to improve its filter performance. Adaptive extended Kalman filter is an optimal choice for battery SoC estimation. Experimental results show that the AEKF-based approach can estimate the battery SoC accurately. Further, for different SoC initial values with big error, the mean absolute errors of the SoC estimation are all within 1%; more importantly, the AEKF-based approach can ensure the estimates converge to true values quickly, less than 60 sample intervals.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] J. C. Álvarez Antón, P. J. García Nieto, F. J. de Cos Juez, F. S. Lasheras, C. B. Viejo, and N. Roqueńí Gutiérrez, "Battery state-of-charge estimator using the MARS technique," *IEEE Transactions on Power Electronics*, vol. 28, no. 8, pp. 3798–3805, 2013.

[2] C. Blanco, L. Sanchez, M. Gonzalez, J. C. Anton, V. Garcia, and J. C. Viera, "An equivalent circuit model with variable effective capacity for LiFePO$_4$ batteries," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3592–3599, 2014.

[3] R. Xiong, F. Sun, H. He, and T. D. Nguyen, "A data-driven adaptive state of charge and power capability joint estimator of lithium-ion polymer battery used in electric vehicles," *Energy*, vol. 63, pp. 295–308, 2013.

[4] Y. Xing, E. W. M. Ma, K. L. Tsui, and M. Pecht, "Battery management systems in electric and hybrid vehicles," *Energies*, vol. 4, no. 11, pp. 1840–1857, 2011.

[5] R. Xiong, F. Sun, Z. Chen, and H. He, "A data-driven multi-scale extended Kalman filtering based parameter and state estimation approach of lithium-ion olymer battery in electric vehicles," *Applied Energy*, vol. 113, pp. 463–476, 2014.

[6] C. C. Chan and K. T. Chau, *Modern Electric Vehicle Technology*, Oxford University Press, 2001.

[7] F. Huet, "A review of impedance measurements for determination of the state-of-charge or state-of-health of secondary batteries," *Journal of Power Sources*, vol. 70, no. 1, pp. 59–69, 1998.

[8] J. H. Aylor, A. Thieme, and B. W. Johnson, "A battery state-of-charge indicator for electric wheelchairs," *IEEE Transactions on Industrial Electronics*, vol. 39, no. 5, pp. 398–409, 1992.

[9] C. C. Christianson and R. F. Bourke, "Battery state of charge gauge," US Patent 4595880, 1975.

[10] E. J. Dowgiallo Jr., "Method for determining battery state of charge by measuring A.C. electrical phase angle change," US Patent 3984762, 1975.

[11] E. Zaugg, "Process and apparatus for determining the state of charge of a battery," US Patent 4433295, 1982.

[12] V. Pop, H. J. Bergveld, D. Danilov, P. P. L. Regtien, and P. H. L. Notten, *Battery Management Systems: Accurate State-of-Charge Indication for Battery-Powered Applications*, Springer, London, UK, 2008.

[13] F. Sun, X. Hu, Y. Zou, and S. Li, "Adaptive unscented Kalman filtering for state of charge estimation of a lithium-ion battery for electric vehicles," *Energy*, vol. 36, no. 5, pp. 3531–3540, 2011.

[14] S. Grewal and D. A. Grant, "A novel technique for modelling the state of charge of lithium ion batteries using artificial neural networks," in *Proceedings of the 23rd International Telecommunications Energy Conference*, pp. 174–179, October 2001.

[15] A. J. Salkind, C. Fennie, P. Singh, T. Atwater, and D. E. Reisner, "Determination of state-of-charge and state-of-health of batteries by fuzzy logic methodology," *Journal of Power Sources*, vol. 80, no. 1-2, pp. 293–300, 1999.

[16] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: part 1. Background," *Journal of Power Sources*, vol. 134, no. 2, pp. 252–261, 2004.

[17] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: part 2. Modeling and identification," *Journal of Power Sources*, vol. 134, no. 2, pp. 262–276, 2004.

[18] G. L. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: part 3. State and parameter estimation," *Journal of Power Sources*, vol. 134, no. 2, pp. 277–292, 2004.

[19] A. H. Mohamed and K. P. Schwarz, "Adaptive Kalman filtering for INS/GPS," *Journal of Geodesy*, vol. 73, no. 4, pp. 193–203, 1999.

[20] H. Dai, X. Wei, and Z. Sun, "Design and implementation of a UKF-based SOC estimator for $LiMnO_2$ batteries used on electric vehicles," *Przeglad Elektrotechniczny*, vol. 88, no. 1, pp. 57–63, 2012.

[21] R. Xiong, F. Sun, X. Gong, and C. Gao, "A data-driven based adaptive state of charge estimator of lithium-ion polymer battery used in electric vehicles," *Applied Energy*, vol. 113, pp. 1421–1433, 2014.

[22] F. Sun and R. Xiong, "A novel dual-scale cell state-of-charge estimation approach for series-connected battery pack used in electric vehicles," *Journal of Power Sources*, vol. 274, pp. 582–594, 2015.

[23] R. K. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on Automatic Control*, vol. 17, no. 5, pp. 693–698, 1972.

[24] J. Han, D. Kim, and M. Sunwoo, "State-of-charge estimation of lead-acid batteries using an adaptive extended Kalman filter," *Journal of Power Sources*, vol. 188, no. 2, pp. 606–612, 2009.

[25] Technical specification of Battery Management System for Electric vehicles, http://www.catarc.org.cn/Upload/file/bzyj/PDF/zhengqiuyijian-sc27-19.pdf.

[26] Idaho National Engineering & Environmental Laboratory, *Battery Test Manual for Plug-In Hybrid Electric Vehicles*, Assistant Secretary for Energy Efficiency and Renewable Energy (EE) Idaho Operations Office, Idaho Falls, Idaho, USA, 2010.

Research Article

Fault Characteristics and Control Strategies of Multiterminal High Voltage Direct Current Transmission Based on Modular Multilevel Converter

Fei Chang,[1] Zhongping Yang,[1] Yi Wang,[2] Fei Lin,[1] and Shihui Liu[1]

[1]*School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China*
[2]*School of Electrical Engineering, Tsinghua University, Beijing 100084, China*

Correspondence should be addressed to Fei Chang; 14117397@bjtu.edu.cn

The modular multilevel converter (MMC) is an emerging voltage source converter topology suitable for multiterminal high voltage direct current transmission based on modular multilevel converter (MMC-MTDC). This paper presents fault characteristics of MMC-MTDC including submodule fault, DC line fault, and fault ride-through of wind farm integration. Meanwhile, the corresponding protection strategies are proposed. The correctness and effectiveness of the control strategies are verified by establishing a three-terminal MMC-MTDC system under the PSCAD/EMTDC electromagnetic transient simulation environment.

1. Introduction

The rapid development of power electronic technology has promoted the development of sustainable transportation and power systems [1–5]. The modular multilevel converter (MMC) was first introduced in 2001 [6] and has drawn great attention due to its excellent output waveform and high efficiency [7, 8]. As a new topology of voltage sourced converter based high voltage direct current transmission (VSC-HVDC), MMC-HVDC has prodigious potential in transmission and distribution applications, such as wind farm connection [9–13], multiterminal operation [14], and a passive network power supply [15].

Multiterminal HVDC transmission based on MMC (MMC-MTDC) is defined as the flexible HVDC transmission system which has three or more voltage source converters (VSCs) under the same DC grid [16]. Its prominent feature lies in providing multiple power supplies, power receiving in multiple places. As a more flexible and efficient power transmission mode, MMC-MTDC shows great potential in renewable energy connection, urban DC distribution network, and so on. In the world, there are only two MMC-MTDC projects and they are all in China [17]. One of which is Nanao three-terminal MMC-MTDC project constructed in Dec. 2013 which is the world's first MMC-MTDC project; the other one is Zhoushan five-terminal MMC-MTDC project constructed in Jul. 2014 which is the world's largest number of terminals in MMC-MTDC projects.

At present, the research of MMC-MTDC is focused on DC voltage stability [17], which can be divided into two categories, including control with communication or no communication. The control with no communication is basically adopted in the actual project which includes DC voltage slope control and DC voltage deviation control. However, the related research on fault protection is also rarely reported [18], in which, a multipoint DC voltage control strategy based on DC voltage margin method is proposed. Furthermore, the impact of different DC faults of the system is analyzed and the corresponding control and protection strategies are given. This paper has been further research on fault characteristics and control strategies of MMC-MTDC, including submodule fault, DC line fault, and fault ride-through of wind farm integration.

2. MMC-MTDC System

MMC-MTDC system is composed of three or more MMC converter stations and DC power transmission interconnection lines, as shown in Figure 1. Wherein, the structure of
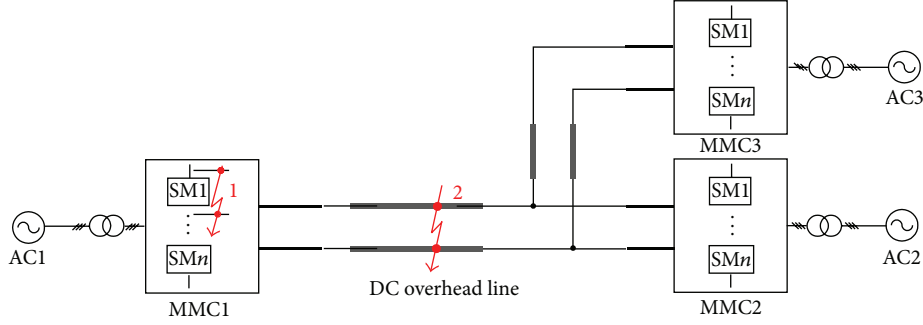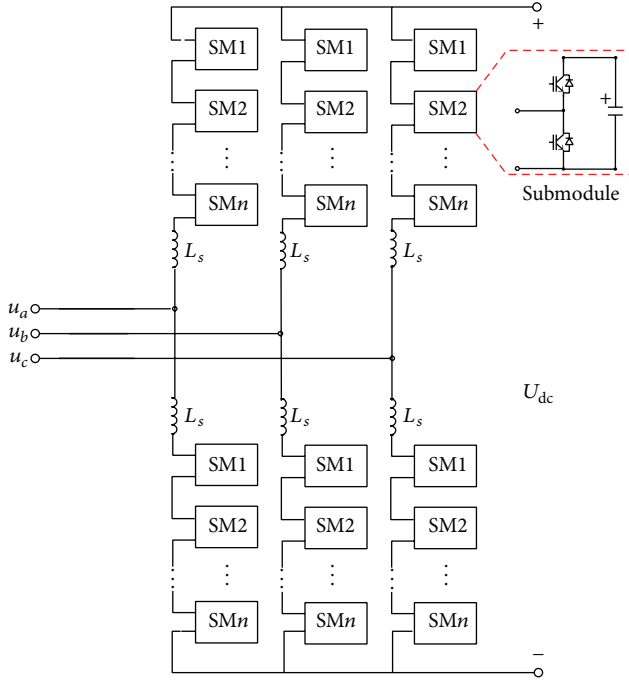
FIGURE 1: Structure of MMC-MTDC system.

FIGURE 2: Structure of MMC converter station.

MMC converter station is shown in **Figure 2**. The system has the advantages of providing multiple power supplies, power receiving in multiple places, and linking several AC systems or separating one AC system into several independent grids.

2.1. Topology of MMC. The main circuit topology of a three-phase MMC is shown in **Figure 2**; the basic circuit unit of MMC is known as submodule (SM). Each bridge arm is constructed by a certain number of submodules and an arm reactance L in series. The MMC topology can change the output voltage and power level of converter in a flexible way, only by changing the number of submodules. As a consequence, the MMC topology has less switching losses and harmonic distortion. In addition, the MMC topology has positive and negative DC bus, which is especially suitable for HVDC applications.

2.2. Mathematical Model of MMC. Considering the circumstances of bridge reactance, the simplified equivalent circuit of MMC is illustrated in (1), where u_{sa}, u_{sb}, and u_{sc} are the fundamental components of the three-phase voltage in AC side, respectively. i_{sa}, i_{sb}, and i_{sc} are the fundamental components of the three-phase current in AC side, separately. L is the sum of bridges' inductance which is in single-phase as well as leakage inductance of the converter transformer. R is the equivalent resistance which consists of bridge reactor and converter transformer. u_a, u_b, and u_c are the fundamental components of the three-phase voltage in converter side, respectively [19]:

$$
\begin{aligned}
L\frac{di_{sa}}{dt} + i_{sa}R &= u_{sa} - u_a, \\
L\frac{di_{sb}}{dt} + i_{sb}R &= u_{sb} - u_b, \\
L\frac{di_{sc}}{dt} + i_{sc}R &= u_{sc} - u_c.
\end{aligned}
\tag{1}
$$

3. Submodule Fault

Normally, the submodule fault occurs mainly due to overvoltage, overcurrent or excessive dv/dt, di/dt, or the control fault due to false triggering pulses. The system operation should not be influenced by one or several fault submodules, so the submodule needs fault redundancy protection to make the converter have the ability of fault tolerance and improve the reliability of the system.

3.1. Fault Characteristics. Taking phase a, for example, the upper and lower arms energy of MMC W_{pa} and W_{na} can be expressed as [20]

$$
\begin{aligned}
W_{pa} &= \frac{1}{2}CNu_{cpa}^2 \\
&= \int_0^T \frac{U_{dc}}{2}\left(1 - m\cos\omega t\right)i_{pa}dt + \frac{1}{2}CNU_C^2, \\
W_{na} &= \frac{1}{2}CNu_{cna}^2 \\
&= \int_0^T \frac{U_{dc}}{2}\left(1 + m\cos\omega t\right)i_{na}dt + \frac{1}{2}CNU_C^2,
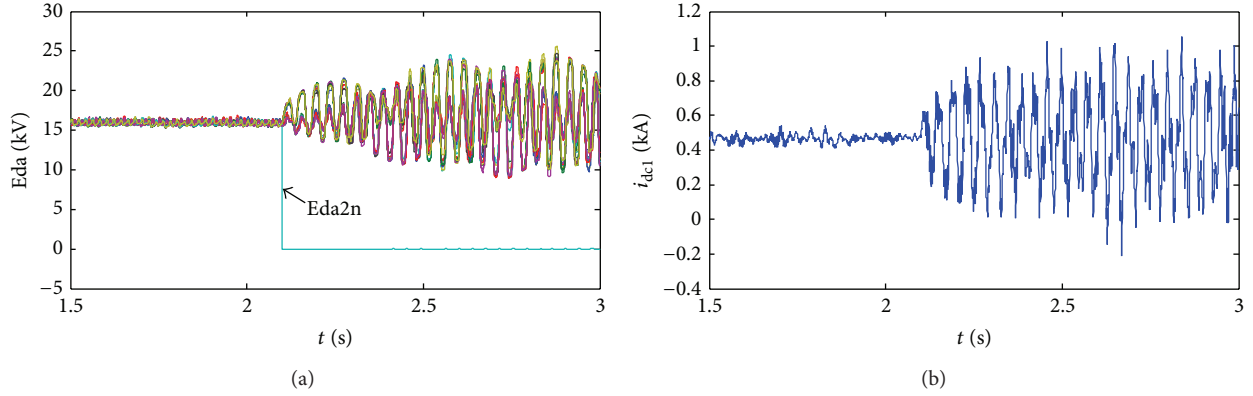\end{aligned}
\tag{2}
$$

FIGURE 3: Fault characteristics of submodule. (a) Capacitor voltage of submodules. (b) DC current.

wherein C is capacitance value; N is the number of submodules of each bridge arm; u_{cpa}, u_{cna} are, respectively, any submodule voltage of upper and lower arms of phase a; T is frequency cycle; U_{dc} is DC voltage; m is voltage modulation ratio, which ranges within $(0, 1]$; i_{pa}, i_{na} are, respectively, the upper and lower arms' currents of phase a; and U_C is the rated voltage of submodule during normal operation.

By formula (2), it can be seen, when the submodule of upper bridge arm of phase a was fault and bypass, the number of submodules of bridge arm will be less than N. In this case, the energy of upper bridge arm of phase a will be less than the other bridge arm, causing the fluctuation of submodule capacitor voltage increase, eventually leading to fluctuations in DC current.

This part describes short-circuit fault of submodule caused by false triggering pulses which correspond to a fault at point 1 in Figure 1 and its simulation parameters are shown in Table 1. Moreover, the simulation waveforms of fault characteristics of submodule are shown in Figure 3. When $t = 2.1$ s, fault occurs in number 2 submodule of lower bridge arm of MMC1 phase a because the upper and lower IGBT simultaneously turned on. Figure 3 shows that the capacitor voltage of fault submodule rapidly drops to 0 that means this submodule stops working and the output voltage of the fault phase will decrease. In addition, because of the parallel connection of three phases, DC current oscillates between the fault phase and the other two phases and may flow into bridge arm and pass through IGBT to cause the fluctuation of capacitor voltage of submodule.

3.2. Fault Redundancy Protection.
Redundancy protection in cascaded H-bridge converter obtains lots of research and can be classified into two methods [21, 22].

Method One. In normal working state, the minority of redundancy submodules are in hot standby mode and the majority are in cold standby mode. Once the submodule fails, the hot standby submodules will replace the cold standby ones and the cold standby submodules will become hot state. The shortcoming is that it takes some time for the action of redundancy submodules and capacitor recharging.

TABLE 1: Simulation parameters of MMC-MTDC system.

Parameters	Values
Rated capacity of MMC1	100 MVA
Rated capacity of MMC2	50 MVA
Rated capacity of MMC3	25 MVA
Transformer ratio of MMC1 (Y/Δ)	110 kV/86 kV
Transformer ratio of MMC2 (Y/Δ)	110 kV/86 kV
Transformer ratio of MMC3 (Y/Δ)	110 kV/86 kV
DC voltage	160 kV
Number of submodules of bridge arm	10
Modulation strategy	Nearest level modulation
Capacitor voltage balancing strategy	Capacitor voltage sort
Control mode of MMC1	U_{dc}, Q
Control mode of MMC2	P, Q
Control mode of MMC3	P, Q

Method Two. The redundant submodules will not be in hot standby state or in cold standby state but will be directly involved in the normal operation. And once fault occurs in the submodules which are being bypassed, DC voltage will be shared by the remaining submodules in the bridge arm. In order to maintain symmetric operation, the remaining normal operation phases can be bypassed by the same number of submodules in fault phase.

By analyzing the redundancy protection method of cascaded H-bridge converter, this paper proposes a redundancy protection method of MMC. This method will bypass the monitored submodules when fault occurs and then bypass the same number of submodules in the other bridge arm of the same phase to keep the upper and lower bridge arms symmetric. Finally by adjusting the control strategies of MMC a transition is achieved from a full submodules operation mode to $(N - x)$ submodules operation mode, where X means the number of fault submodules. Generally, $x < 4$. If $x \geqslant 4$, the system should stop.

Taking that fault occurring in one submodule, for example, the specific processes of fault redundancy protection are shown in Figure 4.
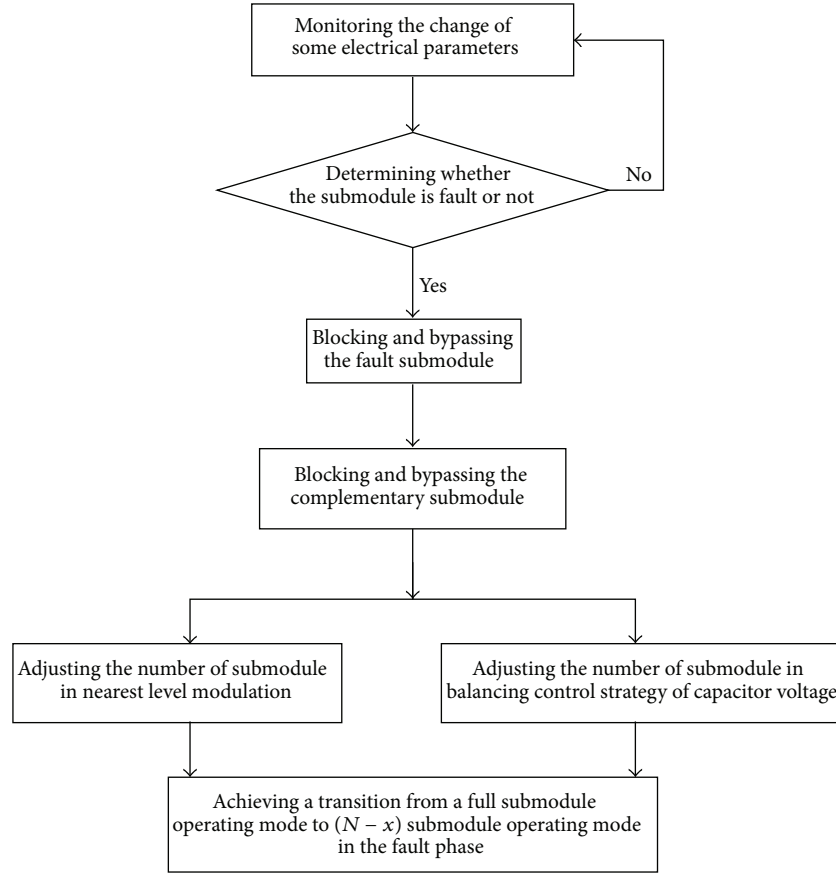
FIGURE 4: Protection strategies flow chart of submodules fault.

(1) Monitor some electrical parameters including capacitor voltage, capacitor current, and PWM pulses. Once the submodule is at fault, block the fault submodule and bypass it. At this time, the number of submodules in the fault bridge arm changes to $(N - x)$, while the number of submodules in nonfault arm bridge of the same phase is still N.

(2) To maintain a constant DC voltage and the same number of submodules in the upper and lower bridge arms of the same phase, it is needed to bypass x submodules in another arm bridge of the same phase. At this time, the total number of submodules in the fault phase becomes $2 * (N - x)$, the number of conduction submodules becomes $(N - x)$, and the capacitor voltage of each submodule rises to $U_{dc}/(N - x)$.

(3) The number of submodules will affect the control strategies of pulses and DC balance; therefore the number of submodules needs to be adjusted correspondingly in the two control strategies. For example, the total number of levels in NLM (Nearest Level Modulation) should be reduced by one, from $(N + 1)$ to N; sorting control should only work in the remaining $2 * (N - 1)$ submodules.

But it is important to note that the redundancy protection should cooperate with other protections. After the fault submodule and its complementary submodule being bypassed, because of the three-phase is in parallel, the normal phase will charge the capacitor of the remaining submodules of the fault phase and the capacitor voltage of the fault phase will gradually rise to $U_{dc}/(N - x)$, so the fault phase inevitably undergoes transient process of DC current rising. The transient process may cause the bridge arm short-time overcurrent and the overcurrent will disappear after one cycle. Because the overcurrent time is too short to accumulate enough heating power to burn the device, and the submodule fault should not cause overcurrent protection of the entire bridge arm to act, this lastly causes the whole converter to block or even shut down. Therefore it is reasonable to set bridge arm overcurrent protection threshold and submodules protection threshold or to extend the action time of bridge arm overcurrent protection to prevent the protection malfunction.

The simulation after adding redundancy protection is shown in Figure 5. Because of the rapid blocking of the 2nd fault submodule of lower bridge arm in a phase (Eda2n), Eda2n retains a certain amount of capacitor voltage, the capacitor voltage of the complementary submodule remains near the rating, the capacitor voltage of the rest of the normal submodules in the fault phase rises to rating $N/(N - 1)$ times of rating value and will be stable after a short transient process
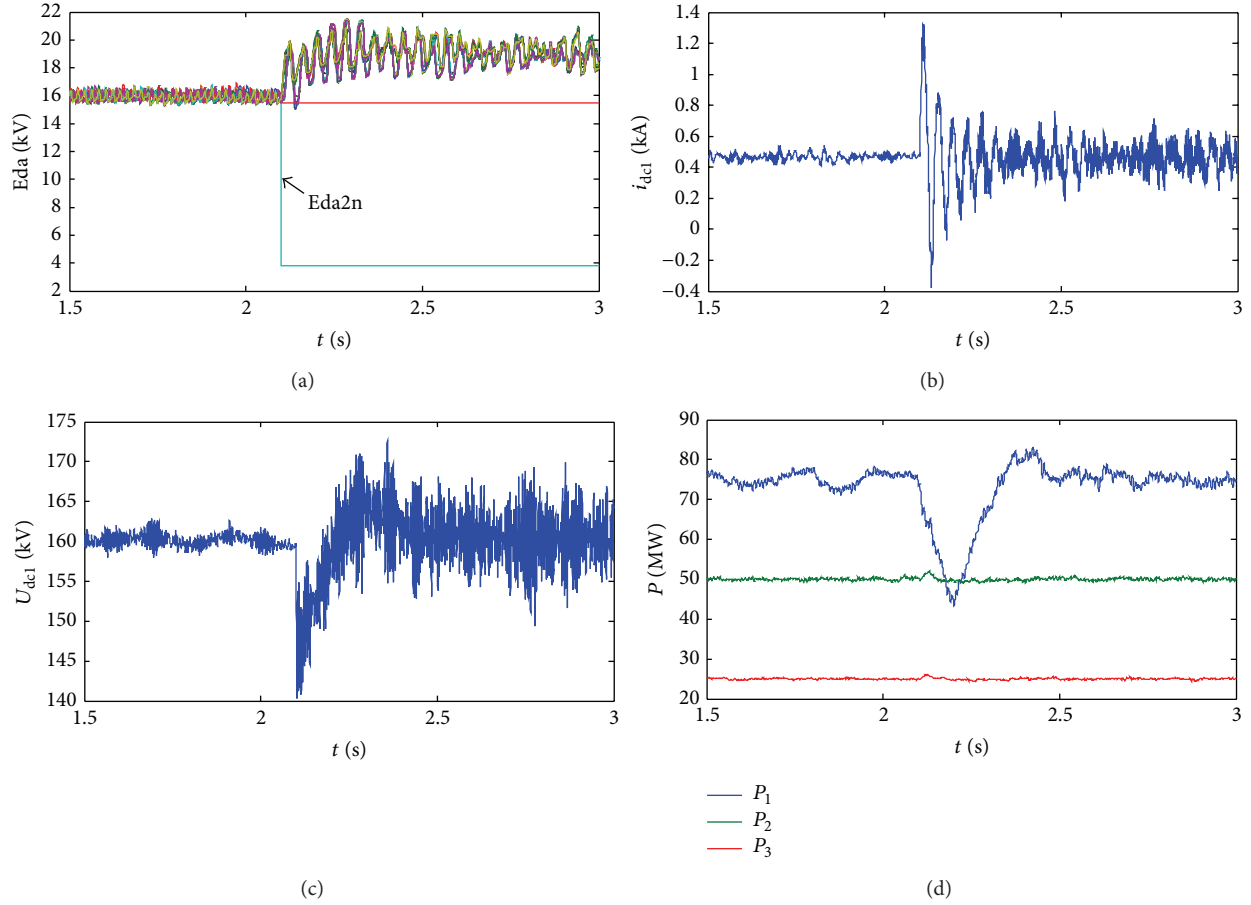
FIGURE 5: Simulation effect after adding fault redundancy protection. (a) Capacitor voltage of submodule of phase *a*. (b) DC current. (c) DC voltage. (d) Active power transmission.

(as shown in Figure 5(a)); the oscillation component of DC current gradually decays to zero (as shown in Figure 5(b)); and DC voltage stability and power transmission normal are shown in Figures 5(c) and 5(d). In summary, the redundancy protection can make the entire system stable when the fault occurs in submodule.

4. DC Line Fault

MMC-MTDC is a potential candidate for renewable energy integration over long distances. DC fault is an issue that MMC-MTDC must deal with, especially for the nonpermanent faults when using overhead lines. This section proposed a protection scheme to implement fast fault clearance and automatic recovery for nonpermanent faults on DC lines.

DC overhead line may cause bipolar short-circuit fault due to tree branches. Compared with unipolar ground short-circuit fault, the probability of bipolar short-circuit fault is smaller, but the fault consequences are much more serious. So it is necessary to research fault feature and design fault protection specially.

Most of the overhead line faults are nonpermanent faults and should not result in the system outage, so the system should automatically restart after fault source disappeared

and quickly restore power supply. Therefore, the protection is designed with the following objectives. Firstly, IGBT and freewheeling diode should be protected. Secondly, the protection should eliminate DC arc of the fault point under the premise of still working. Thirdly, the system can automatically restart and quickly restore power supply after fault source disappear and DC arc extinguish for nonpermanent fault, but the system needs outage and overhaul for permanent fault. Specific protection methods are as follows.

4.1. Double Thyristor Switches. A single thyristor is usually enough if the aim is just to protect the diode from overcurrent. In this paper, in order to make MMC able to quickly clear the fault current and restart power transmission after nonpermanent faults on DC overhead line, double thyristor switches are alternatively employed as shown in Figure 6. The two thyristors are controlled by the same gate signal. During normal operation, the thyristor switches are kept in off-state condition. During DC fault, the thyristor switches are switched on. Since bidirectional thyristor switches are employed, not only is the fault current transferred from diodes to thyristors, but also the aforementioned diode freewheeling effect can be eliminated, which makes it possible to extinguish the DC fault current [23].
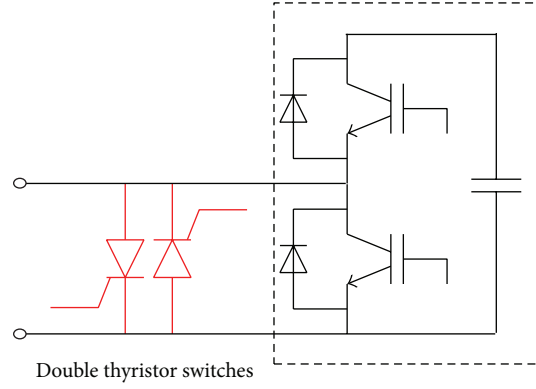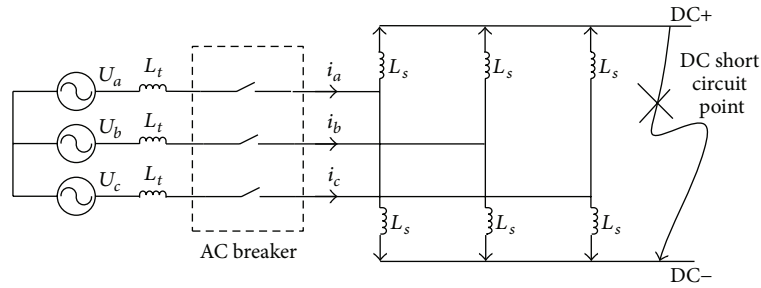
FIGURE 6: Double thyristor switches.

FIGURE 7: Equivalent circuit after adding protection under the overhead line fault.

In the proposed protection scheme, all IGBTs should be blocked as soon as DC fault is monitored. Simultaneously, all thyristors should be switched on to eliminate the rectifier mode of MMC. The fault equivalent circuit of MMC using the proposed protection scheme is shown in Figure 7. Different from the rectifier bridges, six MMC arms become six R-L branches after all thyristors are switched on. Because of the three-phase upper and lower arm symmetrical, DC line positive and negative electric potential is basically the same; DC side of MMC can be equivalent to withstand a relatively small voltage. DC short-circuit current gradually attenuates and disappears; the short point is naturally cut off. The role of the bypass switch is equivalent to convert DC fault into AC fault, thereby enabling the fault natural arcing.

4.2. The Specific Protection Process as Shown in Figure 8. Setting the protection action threshold of DC current is I_{act} and the rated DC line current is I_{dc} in normal operation mode. Generally, I_{act} is set to be two or three times the size of I_{dc}. If $I_{dc} < I_{act}$, MMC converter works in the normal operation mode. If $I_{dc} > I_{act}$, it indicates that DC current increases because short circuit occurs in DC line and protection acts to make MMC converter work in fault protection mode.

Set the protection returning threshold of DC line current as I_{ret} in fault protection mode. Generally, I_{ret} is set to be a little bigger than zero. If $I_{dc} > I_{ret}$, it indicates that DC current was not completely interrupted and still makes MMC converter maintained in fault protection mode. If $I_{dc} < I_{ret}$, it

indicates that short-circuit current has disappeared, so MMC converter goes into automatic recovery mode.

In automatic recovery mode, compare I_{dc} with I_{act} once again. If $I_{dc} < I_{act}$, it indicates that DC fault is nonpermanent fault, making MMC converter transfer to normal operation mode and restore power. If $I_{dc} > I_{act}$, it indicates that DC fault is permanent fault, so making protection act open the breaker of AC side and conduct outage maintenance.

When $t = 0.8$ s, bipolar short-circuit fault occurs in DC overhead lines and the simulation waveforms after adding protection are shown in Figures 9 and 10. System structure, fault point, and simulation parameters are shown in Figure 1, point 2 in Figure 1, and Table 1, respectively. Figure 9 shows the simulation waveform when nonpermanent fault occurs in overhead lines. Set I_{act} as 3 kA and I_{ret} as 0. After monitoring DC lines fault, fast blocking pulse acts to protect the switching devices in case of overcurrent and keeps the capacitor voltage of the submodule. The turning-on of bypass switch changes the structure of fault circuit and DC fault current attenuates. After the fault current arc extinguishing automatically, applied the zero level signal to bypass thyristors, until after 20 ms, all thyristors reliable shutdown, and then deblocking IGBT, MMC converter can achieve the automatic restart. Due to the fact that the fault is nonpermanent, the fault source has disappeared; the DC line will not cause the second-time DC line overcurrent and MMC converter works in the normal mode.

The simulation waveform of permanent fault is shown in Figure 10. Second-time overcurrent occurs when autorestart
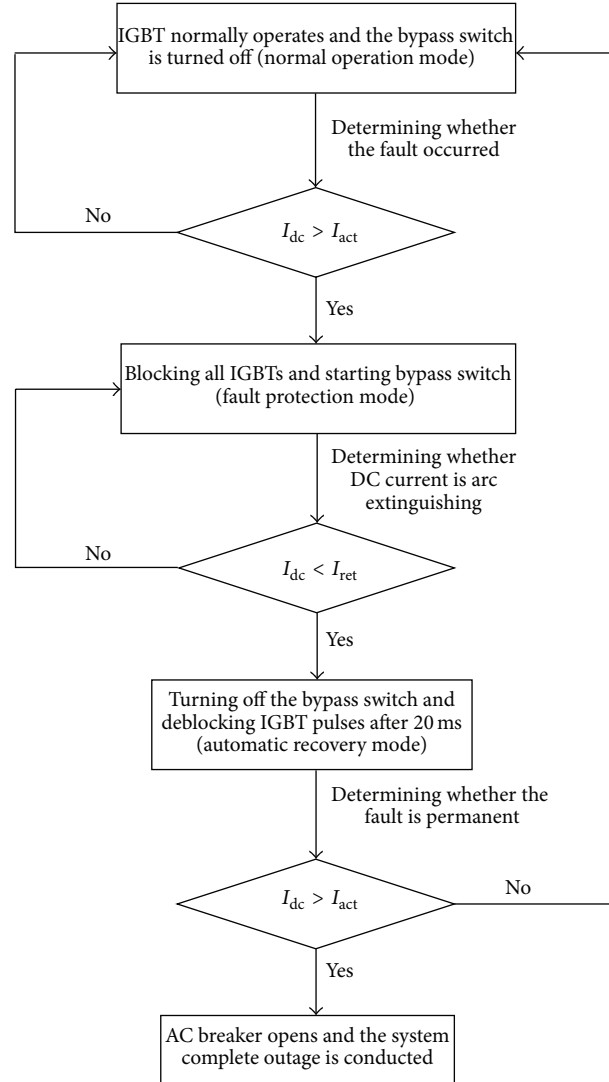
Figure 8: Specific flowchart of fault protection methods.

takes place. Protection AC breaker turns off, system stops totally, and it is time-consuming to recover power supply.

5. Fault Ride-through of Wind Farm Integration

When the wind farm connects AC grid through MMC-MTDC system, if AC grid of the receiving end fails, then the output power capacity of the receiving end will decrease, while power transmission of the wind farm will be not affected, so that active power transmission between the sending and receiving end becomes unbalanced and that will result in DC line voltage being too high. Therefore, control strategy must be taken to make MMC-MTDC system pass through AC grid fault of the receiving end, that is, the issue of fault ride-through (FRT). By installing unloading load in parallel in DC side to eliminate power imbalance in order to maintain a constant DC voltage. This paper further proposes the small and distributed unloading load which adopts

a unified control, which can not only reduce the design and construction difficulty of the unloading load, but also can improve the reliability of the unloading load.

The unloading load can be installed in parallel in DC side which is a resistor controlled by IGBT, as shown in Figure 11. When triggering IGBT to conduct, the unloading load begins to consume energy; if IGBT works in PWM mode, the energy consumption of the unloading load can be quantitatively controlled. The unloading load can consume power difference that MMC-MTDC system cannot eliminate so as to maintain DC line voltage constant. The control strategies of suppressing DC over voltage are as follows:

(1) Monitoring DC line voltage to determine whether DC line is overvoltage.

(2) Setting overvoltage allowable value that is typically 1.01 to 1.05 times of rated voltage.

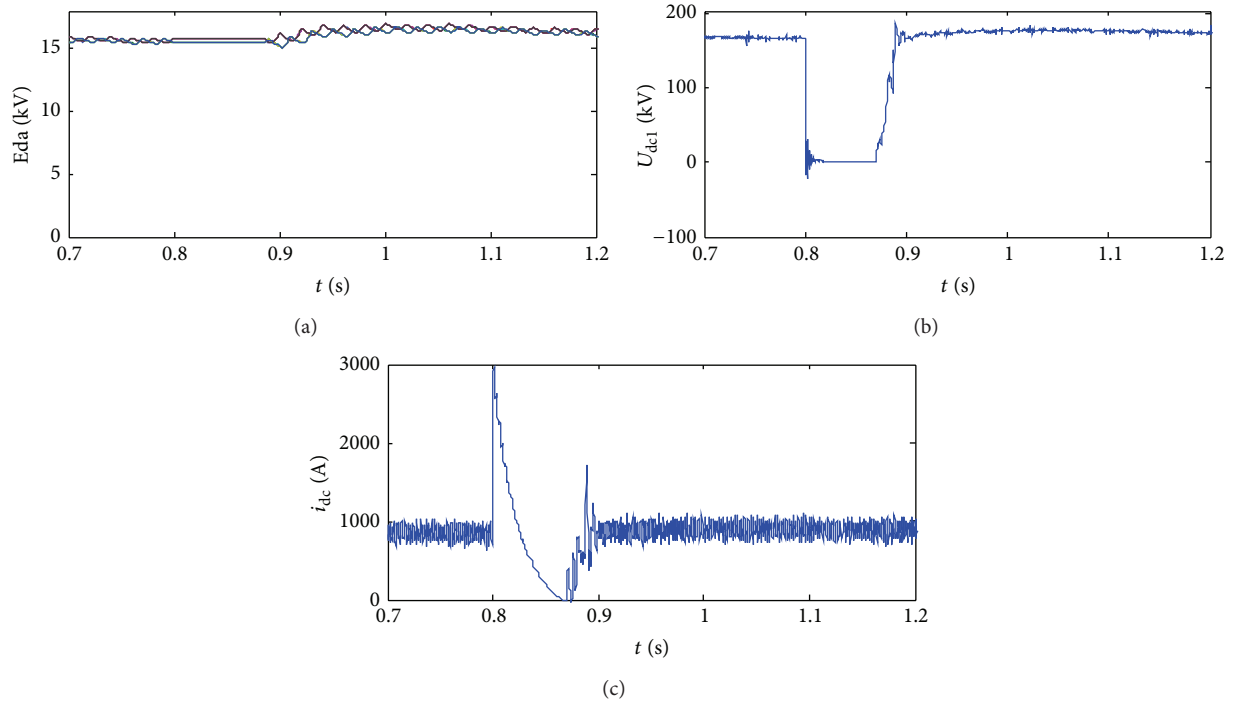(3) When the monitored value of DC voltage rises more than the allowable value, measuring the input and

FIGURE 9: Simulation waveforms in nonpermanent fault after adding protection. (a) Submodule capacitor voltage of phase *a*. (b) DC line voltage. (c) DC line current.
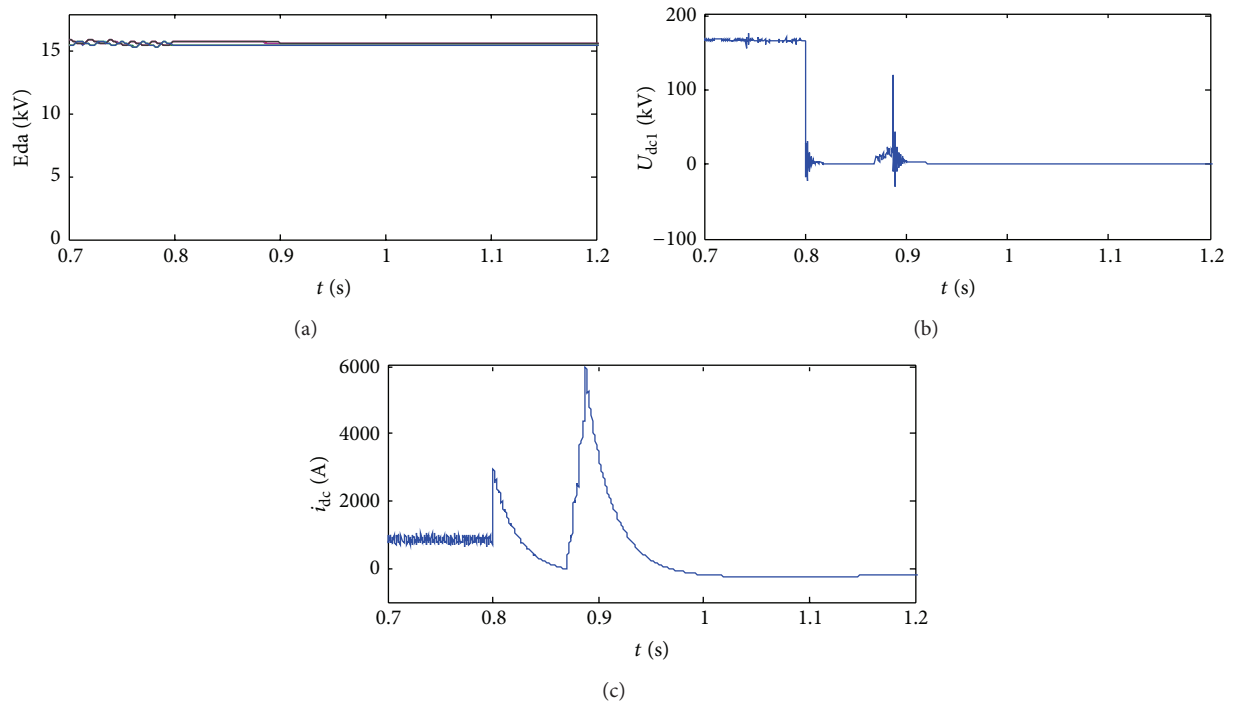

FIGURE 10: Simulation waveforms in permanent fault after adding protection. (a) Submodule capacitor voltage of phase *a*. (b) DC line voltage. (c) DC line current.
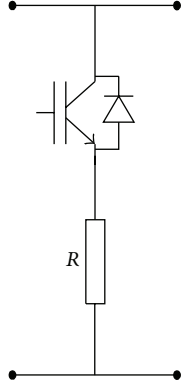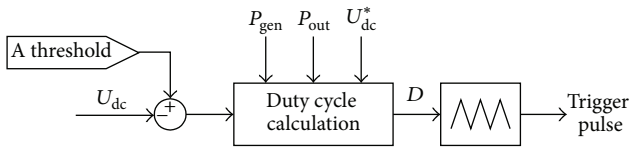
FIGURE 11: Structure of the unloading load.

FIGURE 12: Control block diagram of the unloading load.

FIGURE 13: New control block diagram of the unloading load.

TABLE 2: Simulation parameters of three-terminal MMC-HVDC system.

Parameters	Value
Rated capacity of AC grid	300 MVA
Rated capacity of wind farm 1	100 MVA
Rated capacity of wind farm 2	200 MVA
Transformer ratio of MMC1 (Y/Δ)	220 kV/150 kV
Transformer ratio of MMC1 (Y/Δ)	110 kV/150 kV
Transformer ratio of MMC1 (Y/Δ)	110 kV/150 kV
Voltage of DC bus	±150 kV
Number of submodules of bridge arm	4
Control mode of MMC1	U_{dc}, Q
Control mode of MMC2	V, f
Control mode of MMC3	V, f
Modulation strategy	Carrier phase-shifted modulation
Capacitor voltage balancing strategy	Capacitor voltage sort

output power of transmission system P_{gen}, P_{out}. In order to eliminate the second harmonic fluctuations brought by negative sequence component, P_{gen} and P_{out} need to filter by low pass filter (LPF).

The duty cycle of IGBT can be calculated according to power difference, as in the following formula; the control block diagram is shown in Figure 12:

$$D = \frac{\sqrt{\left(P_{\mathrm{gen}} - P_{\mathrm{out}}\right) \cdot R}}{U_{\mathrm{dc}}^*}. \tag{3}$$

This paper puts forward small and distributed unloading load by a unified control. That is,

(1) Installed locations dispersion: it should be set at DC outlet of MMC2 and MMC3 in the wind farm side rather than only at DC outlet of MMC1.

(2) Installed capacity dispersion: multiple smaller capacity unloading load should be chosen, whose capacity is proportional to the capacity of MMC converter station in the wind farm side, respectively, rather than only a large unloading load matching with MMC1. This not only reduces the design and construction difficulty of the unloading load but also improves the reliability of the unloading load.

(3) Unified control by using a set of controller: set the input power as P_{gen_i} ($i = 1, 2, \ldots, n$) and the output power as P_{out}. According to the proportion of rated capacity P_{N_i} to allocate the balance of power, and respectively calculate the turn-on duty cycle of IGBT of each unloading load D_i ($i = 1, 2, \ldots, n$), as in
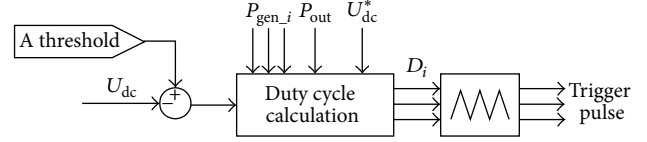
the following formula; the control block diagram is shown in Figure 13:

$$D_i = \frac{\sqrt{\left(\sum_{i=1}^{n} P_{\mathrm{gen}\_i} - P_{\mathrm{out}}\right) \cdot \left(P_{N\_i} / \sum_{i=1}^{n} P_{N\_i}\right) \cdot R_i}}{U_{\mathrm{dc}}^*}. \tag{4}$$

Taking grounding short-circuit fault of phase a, for example, which corresponds to point 1 in Figure 14, the simulation parameters are shown in Table 2. The simulation waveforms according to the above control strategy are shown in Figure 15. The out power of wind farm 1 is 160 MW/10 Mvar and the out power of wind farm 2 is 80 MW/20 Mvar in steady-state operation. Unloading load 1 is placed in MMC2 and the resistance value is 900 ohm; the maximum 100 MW power can be consumed by unloading load 1. Unloading load 2 is placed in MMC3 and the resistance value is 1800 ohm; the maximum 50 MW power can be consumed by unloading load 2. The trigger threshold of both unloading loads is set to 1.05 times of DC voltage reference value. When $t = 1.5$ s, the grounding short-circuit fault of phase a occurs, the asymmetric component of AC line voltage emerges (Figure 15(a)). Due to the short-circuit fault, the active power transmission of MMC1 drops to about 120 MW (Figure 15(b)). But due to the inertia effect, active power transmission sent by two wind farms remains 160 MW and 80 MW (Figure 15(c)). The imbalance of active power transmission is reflected to DC voltage and then the unloading load is triggered to consume excess energy, so that making DC voltage in the vicinity of
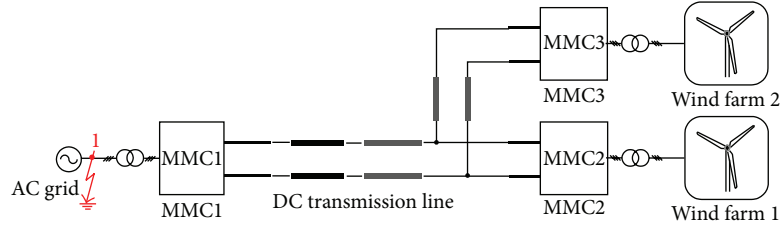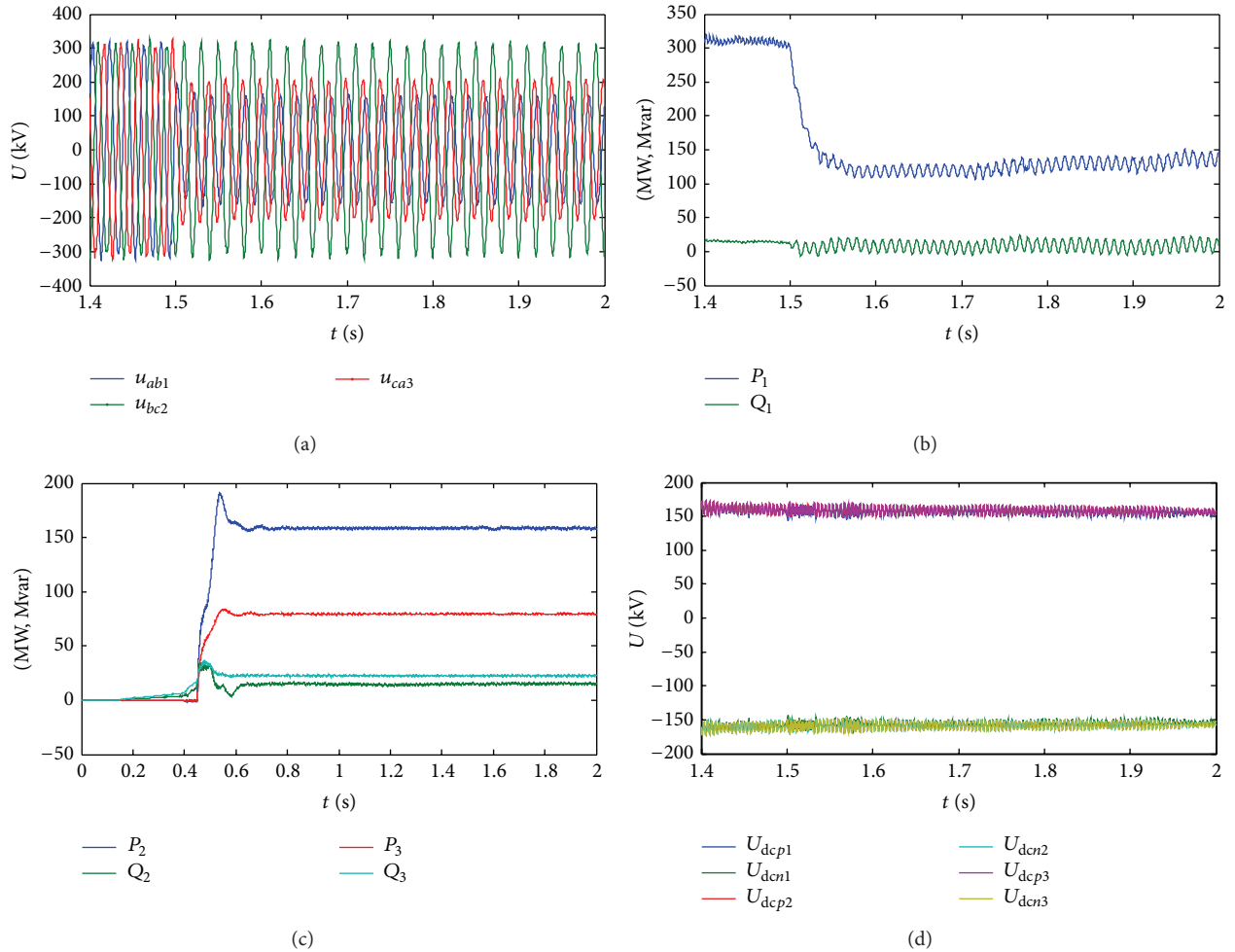
FIGURE 14: Structure of three-terminal MMC-HVDC system.

FIGURE 15: Simulation waveforms of grounding short-circuit fault of phase *a* in AC grid. (a) Line voltage of AC grid. (b) Power transmission of MMC1. (c) Power transmission of MMC2 and MMC3. (d) DC bus voltage.

the reference value (Figure 15(d)) and ensuring that MMC-MTDC system maintains the maximum power transmission during fault.

6. Conclusions

This paper firstly described submodule fault characteristics and proposed submodule redundancy protection for MMC-MTDC system. Secondly, we proposed a protection scheme to implement fast fault clearance and automatic recovery for nonpermanent faults on DC lines. Lastly, a new fault ride-through method for wind farm connection was proposed.

Our future work would focus on the experiment using RT-LAB.

Conflict of Interests

The authors declare no conflict of interests.

Acknowledgment

References

[1] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[2] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Optimal dimensioning and power management of a fuel cell/battery hybrid bus via convex programming," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 1, pp. 457–468, 2015.

[3] C. Sun, S. J. Moura, X. Hu, J. K. Hedrick, and F. Sun, "Dynamic traffic feedback data enabled energy management in plug-in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 3, pp. 1075–1086, 2015.

[4] S. Chu and A. Majumdar, "Opportunities and challenges for a sustainable energy future," *Nature*, vol. 488, no. 7411, pp. 294–303, 2012.

[5] D. MacKay, *Sustainable Energy: Without the Hot Air*, UIT Cambridge, Cambridge, UK, 2009.

[6] R. Marquardt, "Stromrichterschaltungen mit Verteilten Energiespeichern," German Patent DE10103031A1, 2001.

[7] S. Allebrod, R. Hamerski, and R. Marquardt, "New transformerless, scalable modular multilevel converters for HVDC-transmission," in *Proceedings of the IEEE Power Electronics Specialists Conference (PESC '08)*, pp. 174–179, IEEE, Rhodes, Greece, June 2008.

[8] J. Peralta, H. Saad, S. Dennetière, J. Mahseredjian, and S. Nguefeu, "Detailed and averaged models for a 401-level MMC-HVDC system," *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1501–1508, 2012.

[9] S. Xu, H. Rao, Q. Song, W. Liu, and X. Zhao, "Experimental research of MMC based VSC-HVDC system for wind farm integration," in *Proceedings of the IEEE 22nd International Symposium on Industrial Electronics (ISIE' 13)*, pp. 1–5, May 2013.

[10] G. Ramtharan, A. Arulampalam, J. B. Ekanayake, F. M. Hughes, and N. Jenkins, "Fault ride through of fully rated converter wind turbines with AC and DC transmission systems," *IET Renewable Power Generation*, vol. 3, no. 4, pp. 426–438, 2009.

[11] H. G. Jeong, U. M. Choi, and K. B. Lee, "Control strategies for wind power systems to meet grid code requirements," in *Proceedings of the 37th Annual Conference on IEEE Industrial Electronics Society (IECON '11)*, pp. 1250–1255, IEEE, Melbourne, Australia, November 2011.

[12] C. Feltes, H. Wrede, F. W. Koch, and I. Erlich, "Enhanced fault ride-through method for wind farms connected to the grid through VSC-based HVDC transmission," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1537–1546, 2009.

[13] S. K. Chaudhary, R. Teodorescu, P. Rodriguez, and P. C. Kjar, "Chopper controlled resistors in VSC-HVDC transmission for WPP with full-scale converters," in *Proceedings of the 1st IEEE-PES/IAS Conference on Sustainable Alternative Energy (SAE '09)*, pp. 1–8, IEEE, Valencia, Spain, September 2009.

[14] S. Liu, Z. Xu, W. Hua, G. Tang, and Y. Xue, "Electromechanical transient modeling of modular multilevel converter based multi-terminal hvdc systems," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 72–83, 2014.

[15] G. Zhang, Z. Xu, and H. Liu, "Supply passive networks with VSC-HVDC," in *Proceedings of the IEEE Power Engineering Society Summer Meeting*, vol. 1, pp. 332–336, July 2001.

[16] G. Tang, *HVDC Based on Voltage Source Converter*, China Electric Power Press, Beijing, China, 2010.

[17] G. Tang, X. Luo, and X. Wei, "Multi-terminal HVDC and DC-grid technology," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 33, no. 10, pp. 8–17, 2013.

[18] X. Zhang, C. Zhao, H. Pang, and C. Lin, "A control and protection scheme of multi-terminal DC transmission system based on MMC for DC line fault," *Automation of Electric Power Systems*, vol. 37, no. 15, pp. 140–145, 2013 (Chinese).

[19] J. Xu, C. Zhao, and B. Zhang, "Control design and operational characteristics comparison for VSC-HVDC supplying active/passive networks," in *Proceedings of the 6th IEEE Conference on Industrial Electronics and Applications (ICIEA '11)*, pp. 1381–1386, June 2011.

[20] P. Hu, D. Jiang, Y. Zhou, Y. Liang, J. Guo, and Z. Lin, "Energy-balancing control strategy for modular multilevel converters under submodule fault conditions," *IEEE Transactions on Power Electronics*, vol. 29, no. 9, pp. 5021–5030, 2014.

[21] M. Guan and Z. Xu, "Redundancy protection for sub-model faults in modular multilevel converter," *Automation of Electric Power Systems*, vol. 35, no. 16, pp. 94–104, 2011 (Chinese).

[22] W. Song and A. Q. Huang, "Fault-tolerant design and control strategy for cascaded H-bridge multilevel converter-based STATCOM," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 8, pp. 2700–2708, 2010.

[23] X. Li, Q. Song, W. Liu, H. Rao, S. Xu, and L. Li, "Protection of nonpermanent faults on DC overhead lines in MMC-based HVDC systems," *IEEE Transactions on Power Delivery*, vol. 28, no. 1, pp. 483–490, 2013.

Research Article

Modified Quasi-Steady State Model of DC System for Transient Stability Simulation under Asymmetric Faults

Jun Liu,[1] Zhanhong Wei,[1] Wanliang Fang,[1] Chao Duan,[1] Junxian Hou,[2] and Zutao Xiang[2]

[1] *State Key Laboratory of Electrical Insulation and Power Equipment, School of Electrical Engineering, Xi'an Jiaotong University,*
 Xi'an, Shaanxi 710049, China
[2] *China Electric Power Research Institute, Haidian District, Beijing 100085, China*

Correspondence should be addressed to Wanliang Fang; eewlfang@mail.xjtu.edu.cn

As using the classical quasi-steady state (QSS) model could not be able to accurately simulate the dynamic characteristics of DC transmission and its controlling systems in electromechanical transient stability simulation, when asymmetric fault occurs in AC system, a modified quasi-steady state model (MQSS) is proposed. The model firstly analyzes the calculation error induced by classical QSS model under asymmetric commutation voltage, which is mainly caused by the commutation voltage zero offset thus making inaccurate calculation of the average DC voltage and the inverter extinction advance angle. The new MQSS model calculates the average DC voltage according to the actual half-cycle voltage waveform on the DC terminal after fault occurrence, and the extinction advance angle is also derived accordingly, so as to avoid the negative effect of the asymmetric commutation voltage. Simulation experiments show that the new MQSS model proposed in this paper has higher simulation precision than the classical QSS model when asymmetric fault occurs in the AC system, by comparing both of them with the results of detailed electromagnetic transient (EMT) model of the DC transmission and its controlling system.

1. Introduction

By the end of 2013, China has built over 15 EHV and UHV DC transmission lines; both the total length and transmission capacity are the largest in the world [1]. These DC transmission systems are suitable to transmit large-scale renewable electric power generation [2–9] to the remote load center or link the energy storage [10] and electric vehicle-to-grid [11] devices. Due to the intervention of large-scale DC transmission systems, especially the popular multiterminal VSC-based DC transmission [12–15], large amount of power electronic devices and other nonlinear elements have been introduced into the traditional AC power systems; the fast dynamic process of these components might increase the difficulty in performing the electromechanical transient simulation of hybrid AC/DC systems.

In order to improve the precision and speed of electromechanical transient simulation in hybrid AC/DC system, researchers have developed a variety of models for the DC system, including equivalent circuit model, dynamic phasor model, small signal linearized model, and classic quasi-steady state (QSS) model, for transient stability simulation.

The equivalent circuit DC system models mainly adopted the variable topology of converters, to establish the "center-process" method [16], Kron's method of tensor analysis by studying the state matrix of inverters [17], and the cut set matrix analysis method based on graph theoretical framework [18]. All the equivalent circuit based DC system models need a relatively large amount of computation for complicated operation states of thyristors. In order to save calculation time, other models such as the piecewise linearized model are used for DC system dynamic simulation [19]. However, these methods cannot be used in complex DC systems, such as bipolar or multiterminal DC systems, as it would be difficult to segment the complex converter topology.

Dynamic phasor method was firstly presented in [20]; its application for modeling of DC system, through Fourier series expansion, has been mainly used for harmonic analysis of hybrid AC/DC system [21–26].

Small signal linearized model was established by applying the actual sampling data to model the DC transmission system in the dq reference frame [19, 27]. It is able to take into account the nonlinear dynamic characteristics of the converters but only suitable for the small signal dynamic analysis, which cannot directly be transplanted to the electromechanical transient stability simulation problem.

Classical QSS model for the DC transmission and its controlling system has the advantage of fast computation; thus it has been widely used in hybrid AC/DC system simulation [28]. Although several techniques have been presented to improve the accuracy, such as adopting different step lengths during simulation [29], classical QSS model would still give wrong results under asymmetric faulted condition. To deal with the DC system modeling problem for transient stability simulation, a modified quasi-steady state (MQSS) model for DC system is established in this study.

The structure of the paper is as follows. The next section introduces the error causes for the classical QSS model. The modified QSS model is presented in Section 3; it uses the integral of the actual half-cycle voltage on the DC terminals and then calculates the average DC voltage, commutation angle, and the extinction advance angle accordingly. Effectiveness of the proposed model is tested on CIGRE HVDC benchmark system in Section 4, by comparing with the simulation results of EMT simulation software PSCAD/EMTDC. Finally, the conclusions are given in Section 5, which show that the proposed MQSS model has higher accuracy than classical QSS model, while increasing very limited computational complexity.

2. Classic QSS Model of DC System

The basic function of DC transmission system is to complete the AC to DC (rectifier) and DC to AC (inverter) conversion and transmission of electrical energy [30]. Taking the single bridge rectifier as an example, the connection topology of the converter system is shown in Figure 1. The single bridge converter has six bridge arms; each bridge arm is composed of one thyristor valve together with its triggering pulse control circuit. The most important advantage of DC system lies in its capability of controlling the converters' firing angles to adjust the operation mode of power systems very quickly.

2.1. Converter Bridge Model. This section describes the classical quasi-steady state model succinctly [31]. First of all, it is necessary to introduce the assumptions for classical QSS model as follows:

(1) The AC system is assumed to be three-phase symmetric sinusoidal system, with a frequency of 50 Hz (60 Hz in other countries or regions), regardless of the harmonics and the influence of the neutral shift.

(2) The inductance value of the series smoothing reactor on DC side is large enough, and the performance of DC filters is ideal, so that the influence of the ripples can be neglected in the direct current.
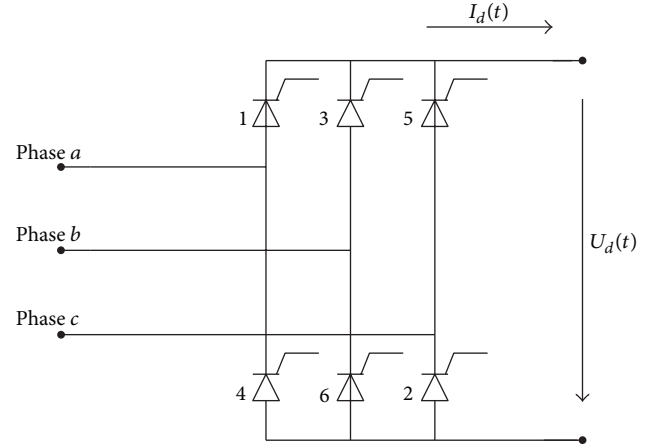

FIGURE 1: The principal wiring diagram of single bridge rectifier.

(3) The converter transformer is thought of as ideal, regardless of the saturation effect, excitation impedance and copper loss, and so forth.

(4) The characteristics of thyristor valves are ideal, namely, the voltage drop during conducting state and the leakage current during blocking state can be ignored, and the six valves are triggered to enter the conducting state in turn with an equal time interval of 1/6 cycle.

Define the firing delay angle as α_R and commutation angle as μ_R; then the DC voltage of both ends of the rectifier to the neutral point can be shown in Figure 2, where e_a, e_b, and e_c denote the three-phase symmetric voltage on the AC side of rectifier and the symbol of $V_{i,j}$ means that valve i commutates to valve j.

The instantaneous three-phase voltages can be expressed in (1), in which U is the RMS value of phase voltage:

$$
\begin{aligned}
e_a(\omega t) &= \sqrt{2}U \sin(\omega t), \\
e_b(\omega t) &= \sqrt{2}U \sin\left(\omega t - \frac{2\pi}{3}\right), \\
e_c(\omega t) &= \sqrt{2}U \sin\left(\omega t + \frac{2\pi}{3}\right).
\end{aligned}
\tag{1}
$$

In classical QSS model, the average DC voltage can be directly calculated according to the symmetric three-phase commutation voltage waveform. Taking the commutation period of valve 3 to valve 4 (the shaded area in Figure 2), for example, the area of the shaded part can be seen as the average DC voltage of the rectifier bridge:

$$
U_{dR} = \int_{\alpha_R + 5\pi/6}^{\alpha_R + 7\pi/6} (e_b - e_c)\, d(\omega t) - \Delta V,
\tag{2}
$$

where ΔV denotes the voltage drop caused by the commutation process, which can be generally expressed as the product of the equivalent commutation resistance R_R and the DC current on rectifier side I_{dR} (the subscript R indicates
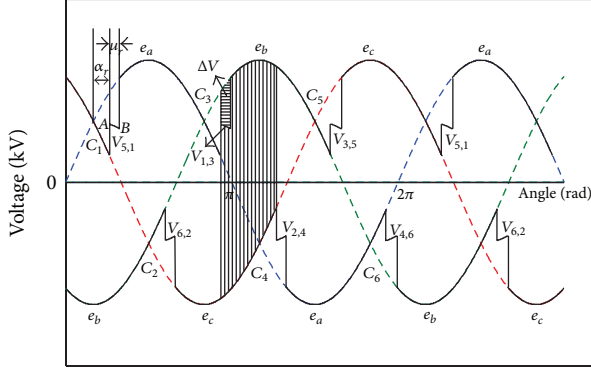
FIGURE 2: DC voltage on rectifier side under symmetric AC voltage waveform.

FIGURE 3: The equivalent circuit of the DC transmission lines.

the variables are on the rectifier side; the subscript I shows those belong to the inverter side).

Substitute (1) into (2):

$$U_{dR} = \frac{3\sqrt{2}}{\pi} E_R \cos \alpha_R - R_R I_{dR},$$

$$R_R = \frac{3\omega L_R}{\pi},$$ (3)

where E_R is the RMS value of the commutation line voltage on rectifier side/kV, I_{dR} is the DC current on rectifier side, α_R is the firing delay angle on rectifier side/rad, and L_R is the equivalent commutation inductance of the rectifier side/H.

The commutation angle on rectifier side can be given as

$$\mu_R = \cos^{-1}\left(\cos \alpha_R - \frac{\sqrt{2}\omega L_R}{E_R}\right) - \alpha_R.$$ (4)

The RMS value of AC current on rectifier side will be

$$I_R = \frac{\sqrt{6}}{\pi} I_{dR}.$$ (5)

The active power consumption by the converters and corresponding power factor on rectifier side can be written as

$$P_R = P_{dR} = U_{dR} I_{dR},$$

$$\cos \varphi_R \approx \frac{1}{2}\left[\cos \alpha_R + \cos(\alpha_R + \mu_R)\right].$$ (6)

The formulas on inverter side of the DC transmission system are similar to the rectifier side in the classical QSS model; we only need to replace the variable of firing delay angle α_R with the extinction advance angle δ_I.

2.2. DC Transmission Line Model. DC line model can be generally classified as lumped parameter circuit model, segmented π-type equivalent circuit model [32], Bergeron model based on distributed parameter [33], and frequency-dependent circuit model [34], and so forth. The researchers can choose among these different DC line models, according
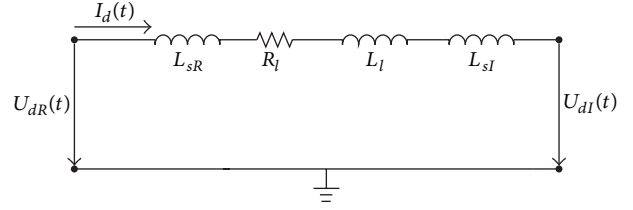
to different accuracy requirements. In this study, it is mainly focused on the influence of asymmetric voltage on the firing angles of converters among different quasi-steady state models during electromechanical transient simulations; therefore, the DC transmission line model is selected as simple lumped parameter circuit model, such as the *R-L* circuit shown in Figure 3.

According to Figure 3, it is easy to write the differential equation of DC transmission line during the electromechanical transient simulation as

$$\left(L_l + L_{sR} + L_{sI}\right)\frac{dI_d(t)}{dt} + R_l I_d(t) = U_{dR}(t) - U_{dI}(t),$$ (7)

where R_l is the equivalent resistance of the DC transmission line/Ω, L_l is the equivalent inductance of the DC transmission line/H, and L_{sR} and L_{sI} are the equivalent inductances of smoothing reactors on each side of the DC system/H.

2.3. DC Control System Model. The DC control system model is adopted as the CIGRE HVDC control system, the block diagram is shown in Figure 4, and it is easy to get the corresponding differential equation according to the transfer function of each block. The initial value of all state variables can be obtained from steady state power flow results. Combining the differential equations of both the control system and the DC transmission line equations, it is sufficient to solve the key parameters in DC systems, such as the firing delay angle $\alpha_R(t + \Delta t)$ and the extinction advance angle $\delta_I(t + \Delta t)$, in which Δt denotes the step length for time domain simulation.

2.4. Error Analysis of the Classical Quasi-Steady State Model. From the modeling of the three parts of DC systems from Sections 2.1 to 2.3, it can be seen that the classical QSS model only considers the situation when the commutation voltages are symmetric. In fact, during electromechanical transient simulation, the last three assumptions in Section 2.1 are easily satisfied, but the first assumption may not always be obeyed, because the AC bus voltage will no longer be symmetric during single phase or double phase short-circuit faults in the AC system. If we still use the symmetric waveform related formulas to fire the thyristors, it may bring serious deviation to the simulation results.

The asymmetric commutation voltage in the AC system may cause potential calculation errors for the classical QSS model in the following ways:

(1) If the average voltage of DC side is still computed according to formulas under symmetric voltage
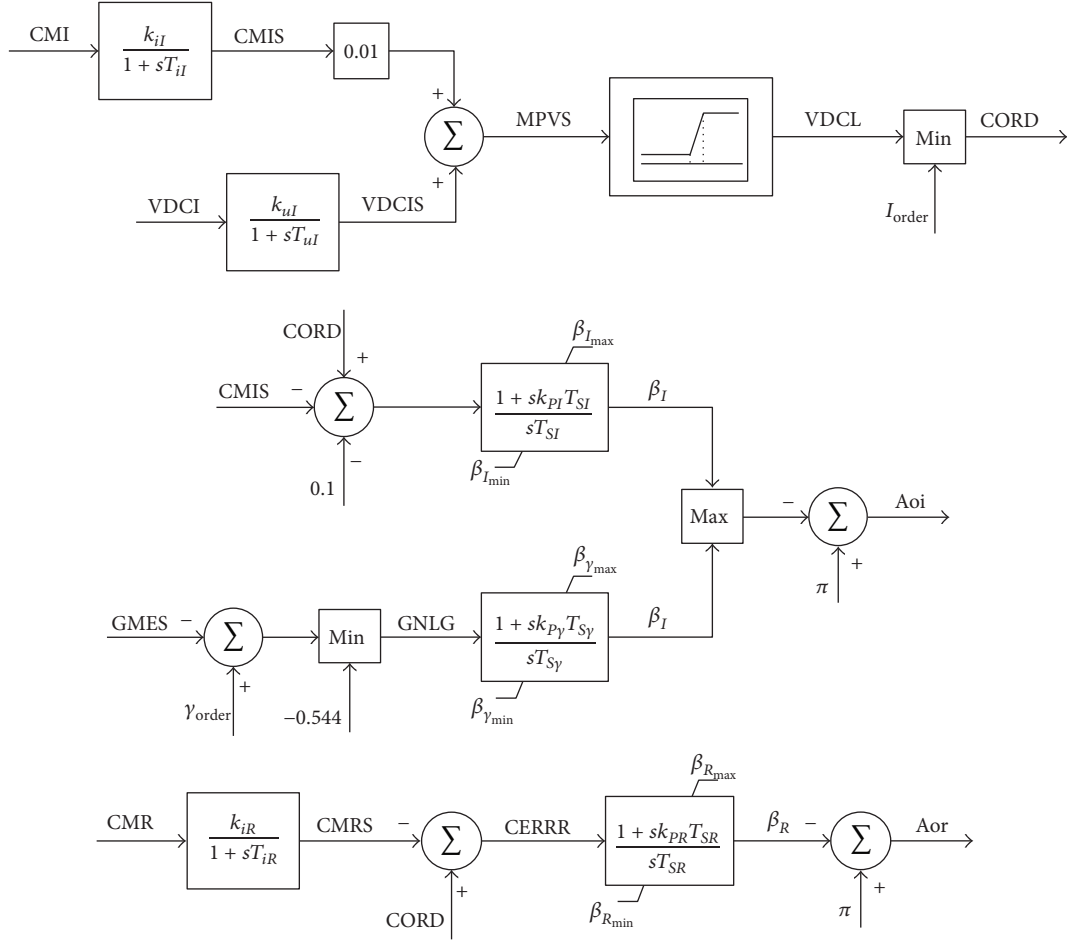
FIGURE 4: The control system of CIGRE HVDC test system.

assumption, all other parameters in the DC side will have inevitable deviation and thus affect the accuracy of calculation. Therefore, more precise formulas of the average DC voltage should be derived, according to the actual AC system operation status.

(2) If the rectifier firing delay angle and the inverter extinction advance angle are calculated using the symmetric waveform, the triggering pulse cannot consider the influence of the commutation voltage zero offset, which might lead to commutation failure or pole blocking for the DC system.

3. The Modified QSS Model for the DC System

It has been shown that the classical QSS model could not provide reliable simulation results under the condition of asymmetric commutation voltage. To address this main defect, a modified quasi-steady state model of DC system is proposed in this paper. It would be better to use the actual voltage waveform on the DC terminals to calculate the average DC voltage, so as to avoid the error caused by symmetric assumption. Before this, it should be better to find the exact commutation voltage zero point for the triggering

pulse of each valve so as to compute the DC voltage and extinction advance angle. Taking the inverter side of six-pulse converter as an example, it is easy to illustrate the situation.

3.1. Exact Zero Point Prediction of Commutation Line Voltage. Since we only care about the fundamental components during electromechanical simulation, the influence of harmonics and interharmonics is not considered in this study. For a given operation status of three-phase voltage amplitude and phase angle, formulas can be derived for predicting the six zero points within one cycle; the detailed process is as follows.

Suppose the instantaneous three-phase asymmetric commutation voltages are expressed as

$$
\begin{aligned}
e_a(\omega t) &= \sqrt{2}U_a \sin(\omega t + \theta_a), \\
e_b(\omega t) &= \sqrt{2}U_b \sin(\omega t + \theta_b), \\
e_c(\omega t) &= \sqrt{2}U_c \sin(\omega t + \theta_c),
\end{aligned}
\tag{8}
$$

using phase a and phase c to derive the prediction formula of the first line voltage zero point C_1. From (8), the line voltage $e_{ac}(\omega t)$ can be written as

$$
e_{ac}(\omega t) = \sqrt{2}U_a \sin(\omega t + \theta_a) - \sqrt{2}U_c \sin(\omega t + \theta_c). \tag{9}
$$

At the line voltage zero point C_1, it should satisfy that $e_{ac}(C_1) = 0$, according to (9):

$$U_a \sin (C_1 + \theta_a) = U_c \sin (C_1 + \theta_c). \tag{10}$$

Applying the trigonometric transformation, the zero point C_1 can be calculated by (12). Consider

$$\sin (C_1) [U_a \cos (\theta_a) - U_c \cos (\theta_c)] \\ = \cos (C_1) [U_c \sin (\theta_c) - U_a \sin (\theta_a)], \tag{11}$$

$$C_1 = \arctan \left(\frac{U_c \sin \theta_c - U_a \sin \theta_a}{U_a \cos \theta_a - U_c \cos \theta_c} \right) \pm \pi. \tag{12}$$

The calculation formulas for the rest five line voltage zero points within one cycle can be acquired similarly, which are listed in

$$C_2 = \arctan \left(\frac{U_b \sin \theta_b - U_c \sin \theta_c}{U_c \cos \theta_c - U_b \cos \theta_b} \right) \pm \pi,$$

$$C_3 = \arctan \left(\frac{U_a \sin \theta_a - U_b \sin \theta_b}{U_b \cos \theta_b - U_a \cos \theta_a} \right) \pm \pi,$$

$$C_4 = C_1 + \pi, \tag{13}$$

$$C_5 = C_2 + \pi,$$

$$C_6 = C_3 + \pi.$$

3.2. Determination of Triggering Pulse. During the simulation process of asymmetric faults, due to the influence of the line voltage zero offset, phase locking device, and the DC control system, the actual firing delay angles for thyristor valves will not be equal to the initial angles given by the triggering pulse control system.

In order to address the influence of zero offset on the DC control system, the firing angle α by the control system is calculated by average value of the three adjacent zeros to determine a more accurate triggering pulse for each valve. This method is able to avoid the inaccurate triggering effect of classical QSS model during asymmetric faults. According to the exact commutation line voltage zero instants calculated by (12) and (13), triggering pulse of valve 3 can be acquired, and then the other five pulses can be obtained through the equidistant firing control as follows:

$$p_3 = \frac{(C_1 + \alpha_I + 2\pi/3 + C_2 + \alpha_I + \pi/3 + C_3 + \alpha_I)}{3},$$

$$p_2 = p_3 - \frac{\pi}{3},$$

$$p_1 = p_2 - \frac{\pi}{3},$$

$$p_6 = p_1 - \frac{\pi}{3}, \tag{14}$$

$$p_5 = p_6 - \frac{\pi}{3},$$

$$p_4 = p_5 - \frac{\pi}{3},$$

TABLE 1: The relationship between p, q, and r and three-phase a, b, and c for each triggering pulse.

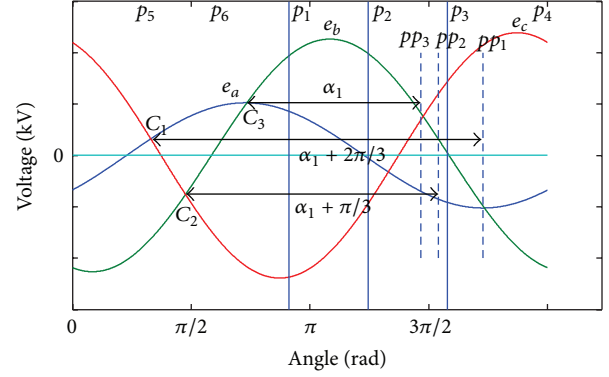| Pulses	p_1	p_2	p_3	p_4	p_5	p_6
p	a	c	b	a	c	b
q	c	b	a	c	b	a
r	b	a	c	b	a	c

FIGURE 5: The determination of triggering pulse for the valves on the inverter side.

where α_I is the inherent firing delay angle on the inverter side that is initially set by the DC control system/rad.

Figure 5 shows the schematic diagram of triggering pulse of each valve on the inverter side, according to fault fundamental component of asymmetric commutation voltage by (8), where the parameters are chosen as $\alpha_I = 2.3$ rad, $U_a = 51.249$ kV, $U_b = 113.268$ kV, $U_c = 119.422$ kV, $\theta_a = -0.7187$ rad, $\theta_b = -1.8384$ rad, and $\theta_c = -4.3136$ rad.

It can be seen in Figure 5 that the three dashed lines of pulse pp_1, pp_2, and pp_3 are firing time instants for valve 3 by applying the three zeros C_1, C_2, and C_3, respectively. The solid line pulse p_3 is the average of these three firing angles, according to (14), and then the firing angles for other valves can be obtained by the equidistant firing control.

During the commutation process, three valves are participating; thus the valves can be divided into three classes. In this study, the valve that is entering into the commutation status is defined as p phase, and exiting phase is defined as the q phase, and the other half bridge that remains conducting is called the r phase. For example, the relationship between p, q, and r and original three-phase a, b, and c for each triggering pulse plotted in Figure 5 can be shown in Table 1.

3.3. Average DC Voltage Calculation. Similar to the classical QSS model in Section 2.1, the average DC voltage under asymmetric faults can be calculated by the integral of actual commutation line voltage waveform on the DC terminals. Also, the voltage drop ΔV caused by DC current during the commutation process can still be replaced by the voltage drop on the equivalent commutation resistance.

According to Table 1, it is easy to obtain the entering, exiting, or conducting state of all the valves on the inverter side during the commutation period, and it is sufficient to

calculate the exact DC voltage waveform and thus the DC voltage calculation formulas, which are shown in Table 2. In Table 2, p_i denotes the firing angle for valve i ($i = 1, 2, \ldots, 6$), and μ_i is the commutation angle of the valve i after it has been triggered.

Using the last column of Table 2 to succinctly express the exact DC voltage waveform, then the average DC voltage can be calculated by the integral of six segments within one cycle:

$$U_{dI} = -\frac{1}{2\pi} \sum_{i=1}^{6} (-1)^{i-1} \left(\int_{p_i}^{p_{i+1}} e_{pri} (\omega t) \, d\omega t - \Delta V \right), \quad (15)$$

where $e_{pri}(\omega t)$ is the general DC voltage formula in the last column of Table 2 within each time period, ΔV is the voltage drop caused by commutation/kV, and p_i is the electric angle of the triggering pulse for valve i/rad.

Similar to classical QSS model, the voltage drop ΔV caused by commutation can also be calculated approximately through the voltage drop on equivalent resistance of the inverter side. Accordingly, the DC average voltage on the inverter side of the new MQSS model can be derived as follows:

$$U_{dI} = -\frac{1}{2\pi} \sum_{i=1}^{6} (-1)^{i-1} \left(\int_{p_i}^{p_{i+1}} e_{pri} (\omega t) \, d\omega t \right) - R_I I_{dI}. \quad (16)$$

Considering that the commutation voltage contains only the fundamental component, it is sufficient to integrate the average DC voltage by half-cycle voltage waveform. Taking the half-cycle voltage waveform from the triggering point of valve 6 to valve 3 of the inverter side as an example, the actual DC terminal voltage waveform on the inverter side can be shown as the shaded part in Figure 6, according to the expressions of DC voltages of Table 2.

Then the average DC voltage from p_6 to p_3 can be calculated as

$$U_{dI} = - \left\{ \frac{1}{\pi} \left(\int_{p_6}^{p_1} e_{cb} (\omega t) \, d\omega t + \int_{p_1}^{p_2} e_{ab} (\omega t) \, d\omega t \right. \right.$$
$$\left. \left. + \int_{p_2}^{p_3} e_{ac} (\omega t) \, d\omega t \right) - R_I I_{dI} \right\}. \quad (17)$$

Solving the integral formula above, we can get the average DC voltage on the inverter side as shown in (18), under asymmetric commutation voltage:

$$U_{dI} = - \left\{ \frac{\sqrt{2}}{\pi} \left[U_b \left(\cos \left(p_6 + \theta_b \right) - \cos \left(p_5 + \theta_b \right) \right) \right. \right.$$
$$- U_c \left(\cos \left(p_6 + \theta_c \right) - \cos \left(p_5 + \theta_c \right) \right)$$
$$+ U_b \left(\cos \left(p_5 + \theta_b \right) - \cos \left(p_4 + \theta_b \right) \right)$$
$$- U_a \left(\cos \left(p_5 + \theta_a \right) - \cos \left(p_4 + \theta_a \right) \right) \quad (18)$$
$$+ U_c \left(\cos \left(p_4 + \theta_c \right) - \cos \left(p_3 + \theta_c \right) \right)$$
$$\left. \left. - U_a \left(\cos \left(p_5 + \theta_a \right) - \cos \left(p_3 + \theta_a \right) \right) \right] - d_I I_{dI} \right\}.$$
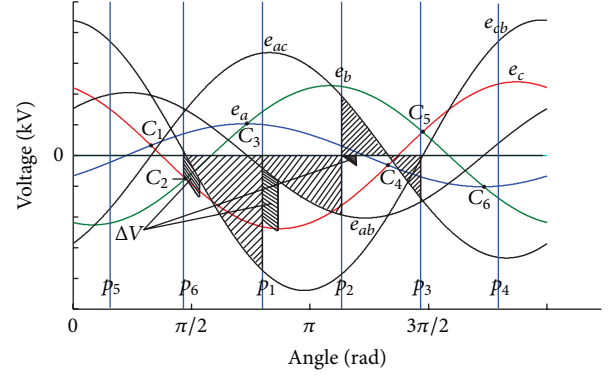

FIGURE 6: Actual DC voltage waveform analysis on the inverter side during asymmetric commutation line voltage.

The calculation process of average DC voltage on the rectifier side is similar. After obtaining the DC voltages on both sides of each DC system, other variables of DC system can then be calculated.

4. Model Validation

In order to test the validity of the proposed MQSS model, simulations are performed on CIGRE HVDC benchmark test system; according to the new model, the classical QSS model, and the power system EMT model from PSCAD software, results from PSCAD are chosen as the comparison reference, because it contains the full electromagnetic transient models that can take into consideration all dynamic performance of the DC system.

4.1. Three-Phase Symmetric Short-Circuit Fault on Inverter Side. At 1.0 s, a three-phase symmetric short-circuit grounding fault occurs on the AC bus of the inverter side, the grounding resistance is set as $0 \, \Omega$, the fault lasts for 0.1 s, and the total simulation time is 1.5 s. With the same initial steady state and faulted operation conditions, the calculation results of the three models for the DC system, namely, the new modified quasi-steady state model (marked with "new"), the electromagnetic transient model (marked with "EMT"), and classical quasi-steady state model (marked with "QSS"), are shown in Figures 7–9. Among them, the ordinates are DC current of the inverter side, DC voltage on inverter side, and DC voltage on rectifier side in per unit system, and the abscissas are simulation time in seconds.

It can be seen from the curves of Figures 7–9 that the simulation results of both the new MQSS model (the solid blue curve) and the classical QSS model (the dotted green curve) match well with results of the EMT simulation (the dashed magenta curve), which demonstrates that both models are suitable to model three-phase faults since the commutation voltages are symmetric. The percent overshoots of the new MQSS model when the fault occurs and is being cleared are a bit more than the classical QSS model, because we use the measured half-cycle AC voltage waveform to calculate the DC variables, but the AC voltage waveform

TABLE 2: The commutation process and the corresponding DC voltage instantaneous value for the single bridge inverter within one cycle.

Trigger pulse	Time period	Upper bridge valve state	Lower bridge valve state	Commutation state	DC voltage waveform	Corresponding relationship	DC voltage calculation
p_5	$p_5 \sim p_5 + \mu_5$	Valve 3 commutes to 5	Valve 4 conducting	b to c	$(e_b + e_c)/2 - e_a$	p-c, q-b, r-a	$e_{pr} + (e_q - e_p)/2$
	$p_5 + \mu_5 \sim p_6$	Valve 5 conducting	Valve 4 conducting	—	e_{ca}	p-c, r-a	e_{pr}
p_6	$p_6 \sim p_6 + \mu_6$	Valve 5 conducting	Valve 4 commutes to 6	a to b	$e_c - (e_a + e_b)/2$	p-b, q-a, r-c	$e_{rp} - (e_q - e_p)/2$
	$p_6 + \mu_6 \sim p_1$	Valve 5 conducting	Valve 6 conducting	—	e_{cb}	p-b, r-c	e_{rp}
p_1	$p_1 \sim p_1 + \mu_1$	Valve 5 commutes to 1	Valve 6 conducting	c to a	$(e_c + e_a)/2 - e_b$	p-a, q-c, r-b	$e_{pr} + (e_q - e_p)/2$
	$p_1 + \mu_1 \sim p_2$	Valve 1 conducting	Valve 6 conducting	—	e_{ab}	p-a, r-b	e_{pr}
p_2	$p_2 \sim p_2 + \mu_2$	Valve 1 conducting	Valve 6 commutes to 2	b to c	$e_a - (e_b + e_c)/2$	p-c, q-b, r-a	$e_{rp} - (e_q - e_p)/2$
	$p_2 + \mu_2 \sim p_3$	Valve 1 conducting	Valve 2 conducting	—	e_{ac}	p-c, r-a	e_{rp}
p_3	$p_3 \sim p_3 + \mu_3$	Valve 1 commutes to 3	Valve 2 conducting	a to b	$(e_a + e_b)/2 - e_c$	p-b, q-a, r-c	$e_{pr} + (e_q - e_p)/2$
	$p_3 + \mu_3 \sim p_4$	Valve 3 conducting	Valve 2 conducting	—	e_{bc}	p-b, r-c	e_{pr}
p_4	$p_4 \sim p_4 + \mu_4$	Valve 3 conducting	Valve 2 commutes to 4	c to a	$e_b - (e_c + e_a)/2$	p-a, q-c, r-b	$e_{rp} - (e_q - e_p)/2$
	$p_4 + \mu_4 \sim p_5$	Valve 3 conducting	Valve 4 conducting	—	e_{ba}	p-a, r-b	e_{rp}

The expression of "$p_5 \sim p_5 + \mu_5$" means a time period from time instant "p_5" to the time instant "$p_5 + \mu_5$".

TABLE 3: Steady state DC values and errors during the 100 ms three-phase short-circuit fault for both QSS and MQSS models.

	EMT model (pu)	QSS model (pu)	Modified model (pu)	Absolute error of QSS model	Absolute error of modified model
I_d	0.5500	0.5500	0.5500	**0**	**0**
U_{DI}	0	0.0561	0.056	**0.0561**	**0.0560**
U_{DR}	0.0050	0.0670	0.0670	**0.0170**	**0.0170**
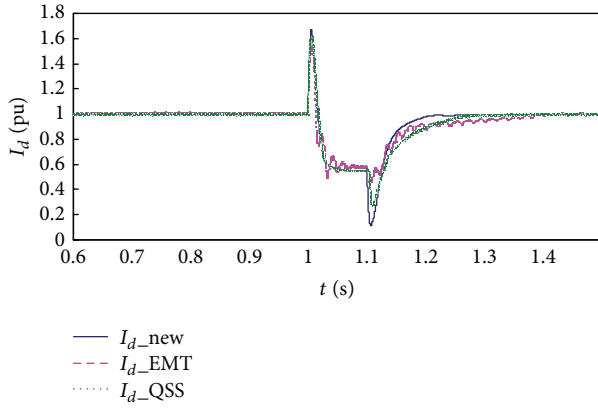
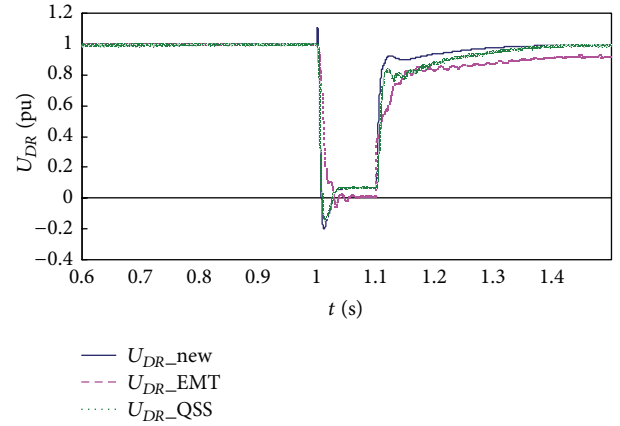FIGURE 7: DC current transient waveform on the inverter side during three-phase grounding fault.

FIGURE 9: DC voltage transient waveform on the rectifier side during three-phase grounding fault.
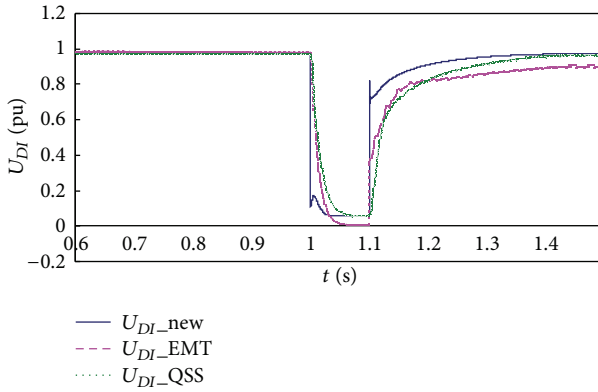

FIGURE 8: DC voltage transient waveform on the inverter side during three-phase grounding fault.

immediately after fault's occurrence and clearance is highly distorted. In electromechanical transient stability study, we care more about the AC fundamental components and the steady state DC values during and after faults; therefore, the comparison of steady state DC values during faults (from time 1.0 s to 1.1 s) and corresponding errors of the two QSS models to EMT model are given in Table 3.

It can be seen from Table 3 that there is only slight differences between the calculation results of all three models; the steady state DC values during the 100 ms fault period for classical QSS model and new MQSS model are similar. This demonstrates that new MQSS model is reasonable and credible under symmetric commutation voltage condition,

although it utilizes the half-cycle measured faulted voltage waveform to predict the commutation voltage zero points and triggering pulse even during symmetric short-circuit faults.

4.2. Single Phase Grounding Fault on Inverter Side. Under asymmetric fault occurring on the AC bus of the inverter side, the voltage asymmetry is mainly decided by fault resistance. If the fault resistances are different, the asymmetric degree of the AC bus voltage will also be different. In this study, we use different values of the AC fault resistance under asymmetric faults to investigate the validity of the proposed MQSS model.

At 1.0 s, a single phase grounding fault occurs on the AC bus of the inverter side, grounding resistance is 10 Ω, and the fault lasts for 0.1 s, and the simulation results for the new model, EMT model, and classical QSS model are shown in Figure 10 to Figure 12.

It can be seen from Figures 10–12 that the calculation results of the new MQSS model are closer to EMT model than to the classical QSS model with same asymmetric fault conditions, although the percent overshoots of our MQSS model when the fault occurs and is being cleared are a bit more than the classical QSS model, which is caused by the fast controlling effect of the converters in the DC systems. The steady state values of the DC variables have higher accuracy using the new model, because it is able to address the zero point offsets and triggering pulse shifts during electromechanical transient stability simulation.

The comparison of steady state DC values during the single phase grounding fault (from time 1.0 s to 1.1 s) and corresponding errors of the two QSS models to EMT model

TABLE 4: Steady state DC values and errors during the 100 ms single phase grounding fault for QSS and MQSS models.

| Single phase grounding resistance (Ω)	Parameters	EMT model (pu)	QSS model (pu)	Modified model (pu)	Relative error of QSS model (%)	Relative error of modified model (%)
0	I_d	0.5500	0.6000	0.5481	**9.09**	**0.34**
	U_{DI}	0	0.4267	0.1954	—	—
	U_{DR}	0.0507	0.4600	0.1843	**807.30**	**263.51**
10	I_d	0.5820	0.7338	0.5512	**44.81**	**5.29**
	U_{DI}	0.2929	0.5925	0.3331	**102.29**	**13.72**
	U_{DR}	0.3000	0.6139	0.3440	**104.63**	**14.67**
20	I_d	0.6823	0.8426	0.7164	**23.49**	**5.00**
	U_{DI}	0.5400	0.6929	0.5846	**28.31**	**8.26**
	U_{DR}	0.5600	0.7360	0.6001	**34.73**	**7.16**
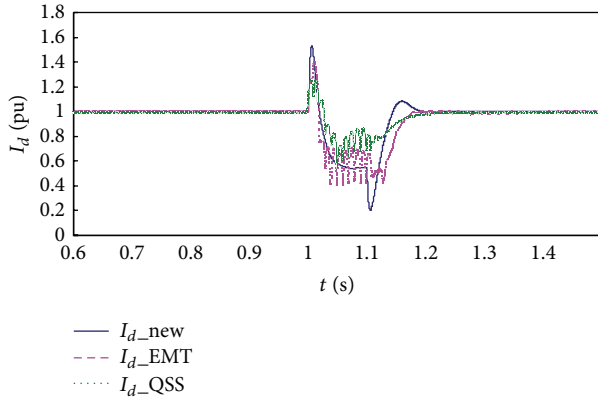
FIGURE 10: DC current transient waveform on the inverter side under single phase grounding fault.

FIGURE 11: DC voltage transient waveform on the inverter side under single phase grounding fault.

TABLE 5: Simulation results for QSS and MQSS models under double phase-grounded faults with small grounding resistances.

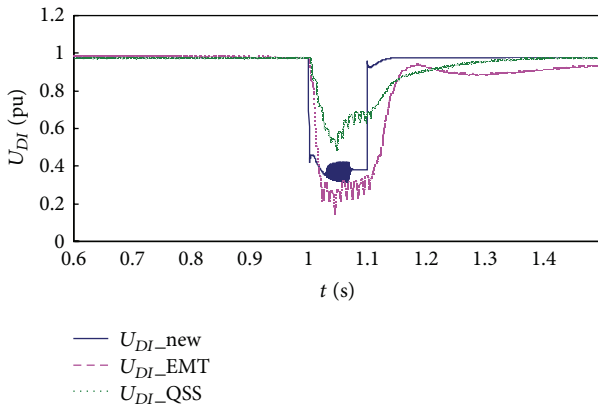| Double phase-grounded resistance (Ω)	EMT Model	QSS model α_I (rad)	Modified model
0	Blocking	Blocking	Blocking
20	Blocking	2.1817	Blocking
40	Blocking	2.2689	Blocking
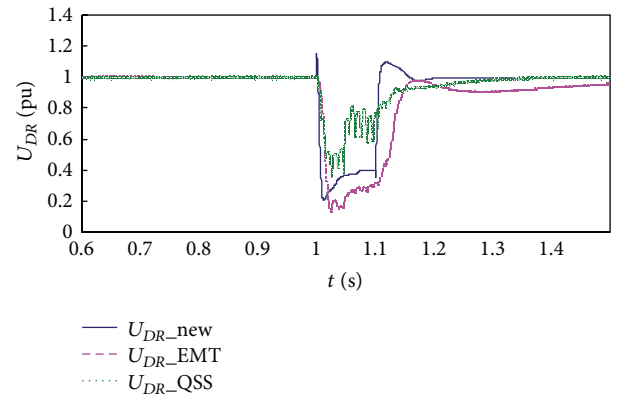
FIGURE 12: DC voltage transient waveform on the rectifier side under single phase grounding fault.

are given in Table 4. It can be seen from the data in Table 4 that the simulation errors of proposed new MQSS model are comparatively smaller than the classical QSS model, because the new model is able to consider the asymmetric commutation voltage under single phase grounding fault, and the deviations of classical QSS model to the EMT model are relatively larger. For example, when the grounding resistance is 10 Ω, the error rate of the calculated DC average voltage on the inverter side U_{DI} during the 100 ms fault period by the new model has reduced by 88.57%, comparing to the calculated value by the classical QSS model; the error rate of the DC average voltage on the rectifier side U_{DR} is reduced by 89.96%, and the error rate of the DC average current I_d is reduced by 39.52%. It demonstrates that the new modified quasi-steady state model has excellent ability to simulate the DC system under asymmetric faulted situation for electromechanical transient stability simulations.

TABLE 6: Steady state DC values and errors during the 100 ms double phase-grounded fault for QSS and MQSS models.

Double phase-grounded resistance (Ω)	Parameters	EMT model (pu)	QSS model (pu)	Modified model (pu)	Relative error of QSS model (%)	Relative error of modified model (%)
60	I_d	0.6500	0.8500	0.7470	**30.77**	**14.92**
	U_{DI}	0.4800	0.7400	0.6217	**54.17**	**29.31**
	U_{DR}	0.4750	0.7250	0.6380	**52.63**	**34.32**
80	I_d	0.7050	0.8840	0.7069	**25.39**	**0.27**
	U_{DI}	0.5700	0.7790	0.5733	**36.67**	**5.79**
	U_{DR}	0.5750	0.7800	0.6424	**35.65**	**11.72**
100	I_d	0.7900	0.9200	0.8104	**16.46**	**2.58**
	U_{DI}	0.6500	0.8000	0.6486	**23.08**	**0.22**
	U_{DR}	0.6520	0.8150	0.6623	**25.00**	**1.58**

4.3. Double Phase-Grounded Fault on Inverter Side. Since double phase-grounded fault usually induces more severe asymmetry than phase-to-phase short-circuit, the phase-to-phase short-circuit asymmetric fault type is not included in this study. At 1.0 s, a double phase-grounded fault (assuming to be phase a and phase b) occurs on the AC bus of the inverter side, and the fault lasts for 0.1 s. Since the waveforms for the DC variables of the new model, EMT model, and classical QSS model are similar to the figures shown in Section 4.2, the figures are omitted, and the simulation results of the new MQSS model are closer to EMT model than to the classical QSS model under identical fault conditions.

Typically, the smaller the short-circuit grounding resistance, the lower the AC voltage; thus the commutation voltage asymmetry degree is larger. It is essential to simulate the more serious condition, such as commutation failure and pole blocking for the DC system.

In the accurate EMT simulation model, continuous commutation failure and HVDC pole blocking for the DC system can be encountered when the grounding resistance is from 0 to 40 Ω, with the initial firing delay angle of the inverter side α_I equal to 1.57 rad. However, no pole blocking phenomenon appears when the grounding resistance is 20 to 40 Ω for the classical QSS model, and the pole blocking phenomenon can be reflected accurately by our new MQSS model, as shown in Table 5. It can be indicated from Table 5 that the modified QSS model is able to simulate the continuous commutation failure and pole blocking of the DC system during severely asymmetric AC faults; the simulation results are consistent with the accurate EMT model.

For larger grounding resistance from 60 to 100 Ω under double phase-grounded fault, the simulation results and corresponding errors of the two QSS models to referenced EMT model are given in Table 6. It can be seen from the data in Table 6 that the simulation errors of new MQSS model proposed in this paper are comparatively smaller than the classical QSS model. During the 100 ms (1.0 s–1.1 s) fault period, the relative errors of the average DC current and voltage variables by the new MQSS model are smaller than classical QSS model; the error can be reduced by 15% at least.

To conclude, the QSS-type models are developed for electromechanical transient simulation, which has much larger time steps and simulation duration than electromagnetic transient simulation. Therefore, the requirements are to reduce the amount of calculation time of electromagnetic computation, while maintaining high accuracy. It is well accepted that the detailed electromagnetic transient simulation programs, such as PSCAD/EMTDC, can provide the reference values; thus we compare both the proposed MQSS and the conventional QSS with the results of PSCAD. The accuracy of our MQSS model has shown to be substantially improved comparing with the conventional QSS model during asymmetric faults. And the computation time does not increase much, as indicated by the additional multiplications and additions in (12), (13), and (14) during the processes of zero point prediction and triggering pulse determination.

5. Conclusion

In power system electromechanical transient stability studies, the classical quasi-steady state (QSS) model is not able to accurately simulate the dynamic characteristics of DC transmission and its controlling system when asymmetric fault occurs in AC system; therefore, a new modified quasi-steady state model (MQSS) is proposed in this paper. The new MQSS model utilizes the actual half-cycle voltage waveform on the DC terminals to predict the exact zero points of commutation voltages and then calculate the average DC voltages and the extinction advance angles, so as to avoid the negative effect of the asymmetric commutation voltage distortion. Simulation experiments show that the new MQSS model proposed in this paper can reduce the simulation error by 15% at least compared to the classical QSS model, under single phase grounding and double phase-grounded asymmetric faults in the AC system, by comparing both of the two models with the results of the detailed EMT model. Because the new MQSS model is capable of reflecting the dynamic characteristics of DC systems without considering the complicated electromagnetic transient processes in typical EMT models, it is very suitable for transient stability simulation in hybrid AC/DC power systems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] J. Cao and J. Cai, "HVDC in China," in *Proceedings of the C-EPRI HVDC & FACTS Conference*, pp. 1–23, Beijing, China, August 2013, http://dsius.com/cet/HVDCinChina_EPRI2013_HVDC.pdf.

[2] D. Van Hertem and M. Ghandhari, "Multi-terminal VSC HVDC for the European supergrid: obstacles," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 3156–3163, 2010.

[3] F. Chang, Z. Yang, Y. Wang, and S. Liu, "Fault characteristics and control strategies of multiterminal high voltage direct current transmission based on modular multilevel converter," *Mathematical Problems in Engineering*. In press.

[4] D. Niu, L. Ji, Q. Ma, and W. Li, "Knowledge mining based on environmental simulation applied to wind farm power forecasting," *Mathematical Problems in Engineering*, vol. 2013, Article ID 597562, 8 pages, 2013.

[5] A. Marucci, D. Monarca, M. Cecchini, A. Colantoni, A. Manzo, and A. Cappuccini, "The semitransparent photovoltaic films for Mediterranean greenhouse: a new sustainable technology," *Mathematical Problems in Engineering*, vol. 2012, Article ID 451934, 14 pages, 2012.

[6] J. Liu, W. Fang, X. Zhang, and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 2, pp. 434–442, 2015.

[7] G. Bergna, E. Berne, P. Egrot et al., "An energy-based controller for HVDC modular multilevel converter in decoupled double synchronous reference frame for voltage oscillation reduction," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 6, pp. 2360–2371, 2013.

[8] D. Zhao, S. Meliopoulos, R. Fan, Z. Tan, and Y. Cho, "Reliability evaluation with cost analysis of alternate wind energy farms and interconnections," in *Proceedings of the 44th IEEE North American Power Symposium (NAPS '12)*, pp. 1–6, IEEE, Urbana, Ill, USA, September 2012.

[9] J. Liu, G. M. Huang, Z. Ma, and Y. Geng, "A novel smart high-voltage circuit breaker for smart grid applications," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 254–264, 2011.

[10] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[11] X. Hu, N. Murgovski, L. Johannesson, and B. Egardt, "Energy efficiency analysis of a series plug-in hybrid electric bus with different energy management strategies and battery sizes," *Applied Energy*, vol. 111, pp. 1001–1009, 2013.

[12] S. Liu, Z. Xu, W. Hua, G. Tang, and Y. Xue, "Electromechanical transient modeling of modular multilevel converter based multi-terminal hvdc systems," *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 72–83, 2014.

[13] G. O. Kalcon, G. P. Adam, O. Anaya-Lara, S. Lo, and K. Uhlen, "Small-signal stability analysis of multi-terminal VSC-based DC transmission systems," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1818–1830, 2012.

[14] J.-H. Li, W.-L. Fang, Z.-C. Du, and D.-Z. Xia, "Calculation method of power flow in hybrid power system containing HVDC and FACTS," *Power System Technology*, vol. 29, no. 5, pp. 31–36, 2005.

[15] J. Beerten and R. Belmans, "Modeling and control of multi-terminal VSC HVDC systems," *Energy Procedia*, vol. 24, pp. 123–130, 2012.

[16] N. G. Hingorani, J. L. Hay, and R. E. Crosbie, "Dynamic simulation of HVDC transmission systems on digital computers," *Proceedings of the Institution of Electrical Engineers. IET Digital Library*, vol. 113, no. 5, pp. 793–802, 1966.

[17] S. Williams and I. R. Smith, "Fast digital computation of 3-phase thyristor bridge circuits," *Proceedings of the Institution of Electrical Engineers*, vol. 120, no. 7, pp. 791–795, 1973.

[18] K. R. Padiyar, "Digital simulation of multiterminal HVDC systems using a novel converter model," *IEEE Transactions on Power Apparatus and Systems*, vol. 102, no. 6, pp. 1624–1632, 1983.

[19] C. M. Osauskas, D. J. Hume, and A. R. Wood, "Small signal frequency domain model of an HVDC converter," *IEE Proceedings—Generation, Transmission and Distribution*, vol. 148, no. 6, pp. 573–578, 2001.

[20] C. Demarco L and C. Verghese G, "Bringing phasor dynamics into the power system load flow," in *Proceedings of the 25th North American Power Symposium*, pp. 31–38, Washington, DC, USA, October 1993.

[21] S. Cole, J. Beerten, and R. Belmans, "Generalized dynamic VSC MTDC model for power system stability studies," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1655–1662, 2010.

[22] S. Cole and R. Belmans, "A proposal for standard VSC HVDC dynamic models in power system stability studies," *Electric Power Systems Research*, vol. 81, no. 4, pp. 967–973, 2011.

[23] C. Hahn, A. Semerow, M. Luther, and O. Ruhle, "Generic modeling of a line commutated HVDC system for power system stability studies," in *Proceedings of the IEEE PES on T&D Conference and Exposition*, pp. 1–6, IEEE, Chicago, Ill, USA, 2014.

[24] H. Zhu, Z. Cai, H. Liu, Q. Qi, and Y. Ni, "Hybrid-model transient stability simulation using dynamic phasors based HVDC system model," *Electric Power Systems Research*, vol. 76, no. 6-7, pp. 582–591, 2006.

[25] D. A. Woodford, A. M. Gole, and R. W. Menzies, "Digital simulation of dc links and ac machines," *IEEE Transactions on Power Apparatus and Systems*, vol. 102, no. 6, pp. 1616–1623, 1983.

[26] J. Reeve and S. P. Chen, "Digital simulation of a multiterminal HVDC transmission system," *IEEE Transactions on Power Apparatus and Systems*, vol. 103, no. 12, pp. 3634–3642, 1984.

[27] C. Osauskas and A. Wood, "Small-signal dynamic modeling of HVDC systems," *IEEE Transactions on Power Delivery*, vol. 18, no. 1, pp. 220–225, 2003.

[28] Y.-X. Ni, V. Vittal, W. Kliemann, and A. A. Found, "Nonlinear modal interaction in hvdc/ac power systems with dc power modulation," *IEEE Transactions on Power Systems*, vol. 11, no. 4, pp. 2011–2017, 1996.

[29] R. M. Brandt, U. D. Annakkage, D. P. Brandt, and N. Kshatriya, "Validation of a two-time step HVDC transient stability simulation model including detailed HVDC controls and DC line L/R dynamics," in *Proceedings of the IEEE Power Engineering Society General Meeting*, pp. 1–6, IEEE, Montreal, Canada, July 2006.

[30] V. K. Sood, *HVDC and FACTS Controllers: Applications of Static Converters in Power Systems*, Springer Science & Business Media, New York, NY, USA, 2004.

[31] N. Rostamkolai, A. G. Phadke, W. F. Long, and J. S. Thorp, "An adaptive optimal control strategy for dynamic stability enhancement of AC/DC power systems," *IEEE Transactions on Power Systems*, vol. 3, no. 3, pp. 1139–1145, 1988.

[32] M. Khatir, S. A. Zidi, M. K. Fellah, S. Hadjeri, and O. Dahou, "HVDC transmission line models for steady-state and transients analysis in SIMULINK environment," in *Proceedings of the 32nd Annual Conference on IEEE Industrial Electronics (IECON '06)*, pp. 436–441, IEEE, Paris, France, November 2006.

[33] F. E. Menter and L. Grcev, "EMTP-based model for grounding system analysis," *IEEE Transactions on Power Delivery*, vol. 9, no. 4, pp. 1838–1847, 1994.

[34] J. R. Marti, "Accurate modelling of frequency-dependent transmission lines in electromagnetic transient simulations," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 1, pp. 147–157, 1982.

Research Article

The Effect of the Integrated Service Mode and Travel Time Uncertainty on Taxis Network Equilibrium

Jian Wang, Zhu Bai, and Xiaowei Hu

School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73, Huanghe River Road, Nangang District, Harbin, Heilongjiang Province 150090, China

Correspondence should be addressed to Zhu Bai; baizhu1979@aliyun.com

This paper aims to discuss the trip mode choice problem by using cumulative prospect theory (CPT) rather than utility maximization from the network uncertainty perspective and evaluates the effect of the integrated service mode on taxi network equilibrium. The integrated service mode means taxis either are actively moving through traffic zones to pick up customers (cruising mode) or are queued at the center of a zone waiting for customers (dispatch mode). Based on this, CPT models are adopted to analyze the choice of customers' trip mode. The travel time uncertainty of the network and the applicability of CPT are considered first, and the Nested Logit model was used to complete the trip mode split problem. Further, several relevant relationships including supply-demand equilibrium, network conditions, taxi behavior, and customer behavior perspectives were analyzed with respect to the integrated mode. Moreover, a network equilibrium model was established and its algorithm was designed. Finally, this paper presented a numerical example and discussed the taxi network equilibrium's characteristic after introducing the integrated service mode.

1. Introduction

Public transport via taxi has increasingly become an important service to modernity because of its personalized and flexible service, comfort, and available speed among other things. Despite its growing popularity, problems have emerged in the practical operation of the taxi market. For instance, it is well known that automobiles used in the taxi service are responsible for consuming a considerable amount of energy. This energy consumption is magnified under the conditions where there is a higher vacancy ratio in the taxi market. The pollution produced as a by-product of this industry is therefore much worse and seriously threatens energy sustainability as well as posing a serious risk to the environment. Based on this, two types of measures are needed in order to alleviate the negative impacts of such an important transportation service. To address the environmental impact of taxi trip, both the alternatively fueled vehicles and cleaner fuels have been proposed to provide taxi service [1]. Management policies which effectively advance the operational efficiency

needed to be explored have also been proposed as a means of curbing the concerns associated with the industry. This investigation seeks to elucidate ways to improve the level of service and the operational situation through improvement to the management policies.

Different taxi policies have been heavily researched since the 1970s with current policies in the taxi field primarily concentrated on market regulations (i.e., entry restrictions and price control). Douglas, who first studied the aggregated model as a precursor in the realm of taxi networks, introduced the realm to economists that have proceeded to investigate the characteristics of different taxi markets, enabling them to understand the equilibrium between supply and demand in regulated markets [2–7]. Despite all the research, it is known and accepted that the taxi market is not the idealized market for conventional economic analyses since the spatial features of the road network are not considered [8]. From this standpoint, Yang and Wong et al. devised a series of models considering fluctuations in supply and demand over variations in time and space of Hong Kong for 1998–2011.

In a previous attempt, Yang et al. sought to characterize taxi movements in a road network for a given origin-destination demand pattern, discussing the relationship between customer demand and taxi utilization by analyzing the impact of congestion, various markets, multiple user classes, stochastic travel time, and price formulation among other things [9–14]. Yang et al. [15] analyzed bilateral searching and meeting relationships between taxis and customers and studied fare control in competitive taxi markets. Further, by considering Pareto efficiency, Yang et al. [16] were able to evaluate quality of service and taxi utilization.

Despite all of the research over the years, models discussed typically account for the cruising service mode when determining regulations and the equilibrium problem; limited attention has been paid to other modes. In one of the first studies discussing multiple modes, Arnott [17] suggested that the dispatch mode was more effective in small cities and the cruising mode was more effective in large cities. von Massow and Canbolat [18] developed a model used to simulate the dispatch mode. One of the first to suggest this, Maria Salanora et al. [19], asserted that the integrated taxi service mode (e.g., cruising and dispatch modes) would be better in big cities.

Recently, in terms of prospect theory which assumes that lotteries are evaluated in a two-step process: an initial phase of editing and a subsequent phase of evaluation, a lot of achievements have been also published [20]. van de Kaa [21, 22] investigated the application of prospect theory and expected utility theory (EUT) where it is assumed that decision maker's attitude toward risk can be rationalized by an expected utility function in trip choice behavior and concluded that, by extending the prospect theory further, it would be suitable to describe traveler behavior. Supporting this assumption, Li and Hensher [23] demonstrated that prospect theory was more suitable for researching travelers decision by analyzing the application of prospect theory in physics, behavior economics, and transportation fields and finally summarized the application constraints of prospect theory in transportation field.

By constructing and analyzing a supply-demand model, the impact factor on the market is evaluated as well as the taxis. The aim was to explore the essential characteristics of the taxi market and provide effective, valuable management policies based on the findings.

In this model, the expected utility theory was replaced with cumulative prospect theory to describe the split in customer trip mode. Cumulative prospect theory is regarded as a bounded rationality model reflective of the uncertainty and complication inherent to real road networks, expressing a decision maker's attitude toward risk. Based on this, we consider the coexistence of cruising and dispatch modes to generate the model used for this paper. The equilibrium model describes the customer-search behavior of vacant-taxi drivers and the interactions between cruising and dispatch modes. Finally, parameters, such as customer-waiting time and taxi-searching time, were identified and evaluated based on variation in traffic zones and fleet sizes.

The rest of this paper is structured as follows: Section 2 is an analysis of the applicable characteristics of cumulative prospect theory and establishes the trip mode model for travelers based on CPT. Section 3 describes the relationship between cruising mode and dispatch modes presenting a supply-demand equilibrium relationship with customer demand variation in taxi market. Section 4 promotes a joint taxi network model of trip distribution and assignment. Section 5 explains the detailed steps of the algorithm developed to establish the supply-demand equilibrium mode. Section 6 describes the network conditions and parameters utilizing a numerical example and discusses the characteristic of equilibrium state with respect to the results. Finally, the last section makes a conclusion and proposes future extensions for the current taxi model.

2. The Trip Mode Choice Model Based on CPT

It was assumed that there are two modes of taxi service available (i.e., cruising and dispatch) as well as other transit modes (bus, rail system, etc.) for customers on a given network. In terms of cruising mode, it means that taxi cruises on roadway to pick up next customers. In terms of dispatch mode, its implementation effect is greatly dependent on the installation rate of in-vehicle navigation system. However, there is a smaller popularizing rate of intelligent system in most current taxi markets. Thus, this paper adopts a simple dispatch policy which means that taxis were required to queue at the center of zone or the gravity point of zone to search for next customers by means of the achievements of von Massow and Canbolat [18]. Based on this assumption, the trip mode choice process may be broken down as follows.

From Figure 1, it can be seen that subsets of alternatives whose properties are similar are grouped in hierarchies or nests. The multinominal logit model (MNL) is used to estimate the probability of trip mode choice for the alternatives of the lower nest, including both cruising and dispatching service modes. And furthermore, at the higher nest, the MNL model consisting of composite taxi mode and other transit modes can be also estimated. Thus, the Nested Logit model is suitable to be used to study the trip mode choice problem described by this paper.

2.1. The Descriptions of Travel Cost. The total travel time from zone i to zone j of customers involves the travel time on roadway and the waiting time which may be expressed as

$$
\begin{aligned}
t_c &= t_c^{ij} + W_i^{cc}, \\
t_d &= t_d^{ij} + W_i^{dc}, \\
t_T &= t_T^{ij} + W_i^{Tc}.
\end{aligned}
\tag{1}
$$

In the preceding equations t_c is the total time of cruising mode, t_d is the total time of dispatch mode, and t_T is the total time of other transit modes, while t_c^{ij} and t_d^{ij} are the travel time of taxis and t_T^{ij} is the travel time of other forms of transit from zone i to zone j. W_i^{cc} is the waiting time of customers that choose cruising mode in zone i and W_i^{dc} is the waiting time of customers choosing cruising mode in zone i, and W_i^{Tc} refers to the waiting time of customers choosing other modes of transit in zone i.
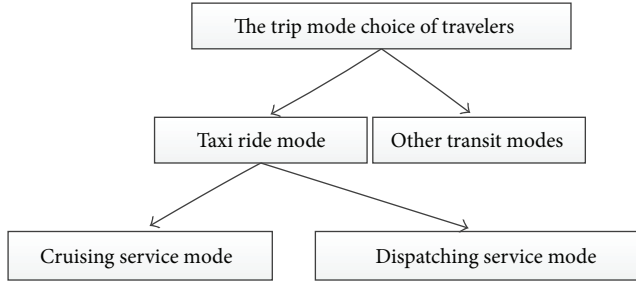
FIGURE 1: The trip mode choice process of travelers.

With U defined as trip cost the variable $U(t_l)$ indicates the cost brought by the perception time of mode l while $U(t_{\text{expected}})$ denotes the cost of the expected travel time relevant to a reference point and can be expressed as follows:

$$U(t_l) = \theta_{\text{trip}} t_l,$$

$$U(t_{\text{expected}}) = \theta_{\text{trip}} t_{\text{expected}}, \tag{2}$$

$$x = U(t_{\text{expected}}) - U(t_l).$$

In the model, x refers to the gain obtained by trip mode if $U(t_{\text{expected}}) \geq U(t_l)$, while x is the loss obtained by trip mode if $U(t_{\text{expected}}) < U(t_l)$.

A set reference point makes distinguishing between gain and loss of a trip with respect to CPT possible for prospective customers. Customers typically perceive gain or loss by comparing trip cost with the reference point cost. In this context, it was assumed that a reference point may be related to the actual travel time experienced, typically about 30 minutes, and was thus used for this study [20].

2.2. Cumulative Prospect Theory. CPT refers to the application of cumulative function to gains and losses. An uncertain prospect f is a function that assigns to each state of S a corresponding outcome. The prospect values are required to be arranged in increasing order to calculate the cumulative function. The prospect f can be regarded as a sequence of pairs (x_l, A_l), which yields x_l if A_l occurs, and A_l is a partition of S. The variable f^+ indicates the positive outcomes of f and f^- does the negative outcomes. For $f = (x_l, A_l)$ and $-m \leq l \leq n$, the relevant expressions are as follows:

$$V(f) = V(f^+) + V(f^-), \tag{3}$$

$$V(f^+) = \sum_{l=1}^{n} \pi_l^+ (f^+) v(x_l), \tag{4}$$

$$V(f^-) = \sum_{l=-m}^{-1} \pi_l^- (f^-) v(x_l), \tag{5}$$

where $V(f^+)$ is the projected gains and $V(f^-)$ is the projected losses. Further, if A_l is given by a probability distribution

$P(A_l) = P_l$, f can be regarded as an uncertain prospect (x_l, P_l). Here, $\pi_l^+(f^+)$ can be expressed as

$$\pi_n^+ = w^+(p_n) \quad l = n, \tag{6}$$

$$\pi_l^+ = w^+(p_l + \cdots + p_n) - w^+(p_{l+1} + \cdots + p_n) \tag{7}$$
$$0 \leq l \leq n-1.$$

With the assumption of continuous traffic flow on a given roadway the distribution function can be introduced into this model as shown in

$$p_l + \cdots + p_n = 1 - (p_{-m} + \cdots + p_{l-1}) = 1 - F(x_{l-1}),$$
$$p_{-m} + \cdots + p_l = F(x_l). \tag{8}$$

When combined with (8), (7) can be expressed as follows:

$$\pi_l^+ = w^+(1 - F(x_{l-1})) - w^+(1 - F(x_l))$$
$$= -\frac{dw^+(1 - F(x))}{dx} dx. \tag{9}$$

$\pi_l^-(f^-)$ can be also defined by

$$\pi_{-m}^- = w^-(p_{-m}) \quad l = -m, \tag{10}$$

$$\pi_l^- = w^-(p_{-m} + \cdots + p_l) - w^-(p_{-m} + \cdots + p_{l-1}) \tag{11}$$
$$1 - m \leq l \leq 0.$$

From (8), (11) can also be expressed as

$$\pi_l^- = \frac{dw^-(F(x))}{dx} dx. \tag{12}$$

Thus cumulative prospect theory may be defined as follows:

$$V = V^+ + V^-$$
$$= \int_{x_0}^{\infty} -\frac{dw^+(1 - F(x))}{dx} v(x)\, dx$$
$$+ \int_{-\infty}^{x_0} \frac{dw^-(F(x))}{dx} v(x)\, dx, \tag{13}$$

where w^+ and w^- are the functions for gains and losses, respectively [24], and are defined by

$$w^+(p) = \frac{p^r}{\left[p^r + (1-p)^r\right]^{1/r}},$$

$$w^-(p) = \frac{p^\delta}{\left[p^\delta + (1-p)^\delta\right]^{1/\delta}}. \tag{14}$$

From (4) and (5), it is known that $v(x)$ is defined as the value function with the following formulation:

$$v(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda(-x)^\beta & x < 0, \end{cases} \tag{15}$$

where x is the gain or loss; the variable x is defined as a gain if $x > 0$ and a loss if $x < 0$.

In these formulations, the parameters are $\alpha = \beta = 0.88$, $\lambda = 2.25$, $r = 0.61$, and $\delta = 0.69$ [24].

Based on this, the cumulative prospect theory of trip modes can be obtained so that a customer may choose the trip mode based on maximal cumulative prospect value.

2.3. The Relationship between Trip Mode Choices. Including cruising mode, dispatch mode, and other transit modes within the network the O-D pair (V_{ij}^{cc}, V_{ij}^{dc}, and V_{ij}^{T}) is used to indicate the cumulative prospect values, respectively.

By adopting the Nested Logit model, the probability P_{ij}^{cc} that a customer originating in zone i chooses cruising mode to travel to zone j is given as follows:

$$
\begin{aligned}
&P_{ij}^{cc} \\
&= \frac{\exp\left\{-\ln\left[\exp\left(\theta^f V_{ij}^{cc}\right) + \exp\left(\theta^f V_{ij}^{dc}\right)\right]\right\}}{\exp\left(V_{ij}^{T}\right) + \exp\left\{-\ln\left[\exp\left(\theta^f V_{ij}^{cc}\right) + \exp\left(\theta^f V_{ij}^{dc}\right)\right]\right\}} \quad (16) \\
&\cdot \frac{\exp\left(\theta^f V_{ij}^{cc}\right)}{\exp\left(\theta^f V_{ij}^{cc}\right) + \exp\left(\theta^f V_{ij}^{dc}\right)}.
\end{aligned}
$$

When the parameter θ^f is nonnegative, it may be used to reflect the degree of uncertainty in service from customers' perspective.

Meanwhile, the probability P_{ij}^{dc} of customers choosing dispatch mode is given as follows:

$$
\begin{aligned}
&P_{ij}^{dc} \\
&= \frac{\exp\left\{-\ln\left[\exp\left(\theta^f V_{ij}^{cc}\right) + \exp\left(\theta^f V_{ij}^{dc}\right)\right]\right\}}{\exp\left(V_{ij}^{T}\right) + \exp\left\{-\ln\left[\exp\left(\theta^f V_{ij}^{cc}\right) + \exp\left(\theta^f V_{ij}^{dc}\right)\right]\right\}} \quad (17) \\
&\cdot \frac{\exp\left(\theta^f V_{ij}^{dc}\right)}{\exp\left(\theta^f V_{ij}^{cc}\right) + \exp\left(\theta^f V_{ij}^{dc}\right)}.
\end{aligned}
$$

Therefore, the demand of customers choosing cruising service can be defined as follows:

$$
Q_{ij}^{cc} = Q_{ij}^{c} P_{ij}^{cc}, \quad (18)
$$

where Q_{ij}^{cc} is the customer demand for cruising mode and Q_{ij}^{c} is the total customer demand for taxis in a network from zone i to zone j.

Similarly, the customer demand for dispatch service may be defined as

$$
Q_{ij}^{dc} = Q_{ij}^{c} P_{ij}^{dc}, \quad (19)
$$

where Q_{ij}^{dc} is the customer demand for dispatch mode. Based on this, the total customer demand for taxi service is expressed by the following formula:

$$
Q_{ij}^{tc} = Q_{ij}^{cc} + Q_{ij}^{dc}, \quad (20)
$$

where Q_{ij}^{tc} is the total customer demand for taxi service from zone i to zone j.

3. Characteristics of Taxi Movement in a Network

3.1. Basic Assumptions of the Model. The basic assumptions of the model established in this paper are described as follows:

(1) Assume that all occupied taxis follow the shortest path choice principle.

(2) Once a taxi is occupied in zone i, a driver will choose the shortest route to go to zone j.

(3) Given an O-D pair, (i, j), for example, after a taxi has delivered customers to zone j, the taxi driver either stays in the same zone or moves to other zones in search of the next customer in an attempt at minimizing search time.

(4) Based on the context, all taxis in a network are divided into one of three types: occupied taxis from zone i to zone j, vacant taxis moving from zone j to zone i, or taxis searching for customers in zone i.

(5) The vacant taxi driver in zone i can choose between cruising mode and dispatch mode to provide service for customers.

In other words, taxis on roadways can be divided into two types (occupied or vacant) and beginning to search for customers after arriving at zone i; those taxis can decide to make service mode choice.

3.2. Introduction of Network Model Variables. Suppose that a road network with a taxi market is in a state of supply-demand equilibrium, defined as $G(V, A)$, where V is the set of nodes and A is the set of links. I and J are the set of origination and destination zones, respectively. $t(v_a)$ is expressed as the travel time on link a ($a \in A$) and v_a is expressed as link flow including taxis and sources of traffic flow. Here, suppose that a taxi is only occupied by a customer.

Assume that t_{ij}^{k} is the travel time on route k between O-D pair (i, j) and is expressed as follows:

$$
t_{ij}^{k} = \sum_{a \in A} t(v_a) \delta_{ij}^{ak}. \quad (21)
$$

Further, $\delta_{ij}^{ak} = 1$ if route k between O-D pair (i, j) uses link a and 0 otherwise; k is included in the set R_{ij} which is the set of routes between O-D pair (i, j).

Based on this context, t_{ij} is considered as the travel time via the shortest route from zone i to zone j:

$$
t_{ij} = \min\left(t_{ij}^{k}\right) \quad k \in R_{ij}. \quad (22)
$$

3.3. The Choice Relationship of Taxi Service Modes. In accordance with basic assumptions above, the probability that

a vacant taxi arriving at zone i chooses cruising mode to search for the next customer is given as follows:

$$p^{ct} = \frac{\exp\left(-\theta^t w_i^{ct}\right)}{\exp\left(-\theta^t w_i^{ct}\right) + \exp\left(-\theta^t w_i^{dt}\right)}, \tag{23}$$

where p^{ct} is the probability of taxis choosing cruising mode, while θ^t is a nonnegative parameter, w_i^{ct} is the search time of taxi drivers that chose the cruising mode, and w_i^{dt} is the search time of taxi drivers that chose dispatch mode.

Similarly, the probability of vacant taxis choosing dispatch mode to search for the next customer is given as follows:

$$p^{dt} = \frac{\exp\left(-\theta^t w_i^{dt}\right)}{\exp\left(-\theta^t w_i^{ct}\right) + \exp\left(-\theta^t w_i^{dt}\right)}, \tag{24}$$

where p^{dt} is the probability of taxis choosing dispatch mode.

3.4. Taxi Service to Time Relationship. There is a given demand matrix OD and the 1 h interval is studied. This paper assumes that the total number of taxis in a network including occupied and vacant ones is Num and it is a constant.

The total number of occupied taxis is given as $\sum_{i \in I} \sum_{j \in J} T_{ij}^{ot} t_{ij}$, with T_{ij}^{ot} defined as the number of occupied taxis in the studied 1-h interval from zone i to zone j; "ot" refers to occupied taxis.

Similarly, the total number of vacant taxis can be expressed as follows:

$$\sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} t_{ji} + \sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} p^{ct} w_i^{ct} + \sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} p^{dt} w_i^{dt}, \tag{25}$$

where T_{ji}^{vt} is the number of vacant taxis in the studied 1-h interval from zone j to zone i; "vt" refers to vacant taxis. In the above relationship, $\sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} t_{ji}$ is the number of vacant taxis on roadway; $\sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} p^{ct} w_i^{ct}$ is the number of vacant taxis choosing cruising mode to search for customer in zone i; $\sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} p^{dt} w_i^{dt}$ is the number of vacant taxis choosing dispatch mode in zone i.

Therefore, the total number of taxis in a network can be expressed as

$$\sum_{i \in I} \sum_{j \in J} T_{ij}^{ot} t_{ij} + \sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} t_{ji} + \sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} p^{ct} w_i^{ct}$$
$$+ \sum_{j \in J} \sum_{i \in I} T_{ji}^{vt} p^{dt} w_i^{dt} = \text{Num}. \tag{26}$$

3.5. Behavior Model of Taxi Drivers. The expected search time of drivers in one zone is assumed to be distributed in accordance with the Gumbel density function. Here w_i^t is the total searching time of vacant taxis in zone i and expressed as follows:

$$w_i^t = p^{ct} w_i^{ct} + p^{dt} w_i^{dt}. \tag{27}$$

The probability that a vacant taxi originating in zone j eventually meets a customer in zone i is given by

$$P = \frac{\exp\left\{-\theta\left(t_{ji} + p^{ct} w_i^{ct} + p^{dt} w_i^{dt}\right)\right\}}{\sum_{m \in I} \exp\left\{-\theta\left(t_{jm} + p^{ct} w_m^{ct} + p^{dt} w_m^{dt}\right)\right\}}, \tag{28}$$

$$i \in I, \ m \in I, \ j \in J,$$

where P is the probability that a taxi meets the next customer in zone i after having delivered a customer to zone j, θ is a nonnegative parameter and its explanation is referred to by Wong et al. [13].

3.6. The Relationship between Customer Wait Time and Taxi Search Time. According to Yang et al. [15], the function relationship between customer and taxi wait time can be described by applying the Cobb-Douglas model describing customer-taxi meeting function which is given by

$$W_i^{cc} = \frac{1}{0.01 \sum_{j \in J} T_{ji}^{vt} p^{ct} w_i^{ct}}, \tag{29}$$

$$W_i^{dc} = \frac{1}{10^4 \sum_{j \in J} T_{ji}^{vt} p^{dt} w_i^{dt}}, \tag{30}$$

where W_i^{cc} is the average waiting time of customers choosing cruising mode in zone i and W_i^{dc} is the average waiting time of customers choosing dispatch mode in zone i.

3.7. The Supply-Demand Equilibrium Relationship. When in an equilibrium state, the number of vacant taxis in a network should meet all customers' demands in their respective origination zones. That is to say, taxi service is available to every customer. Therefore, the supply-demand equilibrium relationship in a network is given as follows:

$$\sum_{i \in I} T_{ji}^{vt} = Q_j^{tc} = \sum_{i \in I} Q_{ij}^{tc}, \quad j \in J,$$
$$\sum_{j \in J} T_{ji}^{vt} = Q_i^{tc} = \sum_{j \in J} Q_{ij}^{tc}, \quad i \in I. \tag{31}$$

In terms of cruising mode, its supply-demand equilibrium relationship is given by

$$\sum_{j \in J} Q_{ij}^{cc} = \sum_{j \in J} T_{ji}^{vt} p^{ct}. \tag{32}$$

In terms of dispatch mode, its supply-demand equilibrium relationship is expressed by

$$\sum_{j \in J} Q_{ij}^{dc} = \sum_{j \in J} T_{ji}^{vt} p^{dt}. \tag{33}$$

4. A Mathematical Programming Model

4.1. The Model Descriptions. A joint taxi network model of trip distribution and assignment can be given as follows:

$$\text{minimize} \quad Z$$

$$= \sum_{a \in A} \int_0^{v_a} t_a(\omega)\, d\omega$$

$$+ \frac{1}{\theta} \sum_{j \in J} \sum_{i \in I} T_{ji}^{\text{vt}} \left(\ln T_{ji}^{\text{vt}} - 1 \right) \tag{34}$$

$$- \sum_{i \in I} \sum_{j \in J} \int_0^{Q_{ij}^{tc}} Q_{ij}^{-1}(\omega)\, d\omega$$

$$\text{subject to:} \quad \sum_{i \in I} T_{ji}^{\text{vt}} = Q_j^{tc} \tag{34a}$$

$$\sum_{j \in J} T_{ji}^{\text{vt}} = Q_i^{tc} \tag{34b}$$

$$\sum_{k \in R_{ij}} f_{ij}^k = T_{ij}^n + T_{ij}^{\text{ot}} + T_{ij}^{\text{vt}} \tag{34c}$$

$$v_a = \sum_{i \in I} \sum_{j \in J} \sum_{k \in R_{ij}} \delta_{ij}^{ak} f_{ij}^k \tag{34d}$$

$$f_{ij}^k \geq 0 \tag{34e}$$

$$T_{ji}^{\text{vt}} > 0, \tag{34f}$$

where T_{ij}^n is normal traffic flow in the studied 1-h interval from zone i to zone j and f_{ij}^k is the traffic flow on route k in the studied 1-h interval.

In terms of the established supply-demand equilibrium model, the following Lagrangian function can be formed:

$$L = \sum_{a \in A} \int_0^{v_a} t_a(\omega)\, d\omega + \frac{1}{\theta} \sum_{j \in J} \sum_{i \in I} T_{ji}^{\text{vt}} \left(\ln T_{ji}^{\text{vt}} - 1 \right)$$

$$- \sum_{i \in I} \sum_{j \in J} \int_0^{Q_{ij}^{tc}} Q_{ij}^{-1}(\omega)\, d\omega + \sum_{i \in I} \alpha_i \left(\sum_{j \in J} T_{ji}^{\text{vt}} - Q_i^{tc} \right)$$

$$+ \sum_{j \in J} \beta_j \left(\sum_{i \in I} T_{ji}^{\text{vt}} - Q_j^{tc} \right) \tag{35}$$

$$+ \sum_{i \in I} \sum_{j \in J} \mu_{ij} \left(T_{ij}^n + T_{ij}^{\text{ot}} + T_{ij}^{\text{vt}} - \sum_{k \in R_{ij}} f_{ij}^k \right).$$

After differentiating the Lagrangian function and applying Kuhn-Tucker conditions, the following relationship can be obtained in an equilibrium state:

$$T_{ji}^{\text{vt}} = \exp\left(-\theta \left(t_{ji} + \alpha_i + \beta_j \right) \right)$$

$$= \exp\left(-\theta t_{ji} \right) \exp\left(-\theta \alpha_i \right) \exp\left(-\theta \beta_j \right). \tag{36}$$

4.2. The Deduction Process of the Model. Firstly, (36) can be rewritten as the following form of gravity model:

$$T_{ji}^{\text{vt}} = A_i B_j \exp\left(-\theta t_{ji} \right), \quad j \in J, \; i \in I, \tag{37}$$

where $A_i = \exp(-\theta \alpha_i)$ and $B_j = \exp(-\theta \beta_j)$.

By using (34a) and (34b), it can be expressed as follows:

$$A_i = \frac{Q_i^{tc}}{\sum_{j \in J} B_j \exp\left(-\theta t_{ji} \right)}, \quad i \in I,$$

$$B_j = \frac{Q_j^{tc}}{\sum_{i \in I} A_i \exp\left(-\theta t_{ji} \right)}, \quad j \in J, \tag{38}$$

by first assuming $B_j = 1$ and then alternatively solving (38) until converged.

Secondly, according to (34a), (36) is also changed as follows:

$$\sum_{i \in I} T_{ji}^{\text{vt}} = \sum_{i \in I} \exp\left(-\theta \left(t_{ji} + \alpha_i + \beta_j \right) \right)$$

$$= \sum_{i \in I} \exp\left(-\theta \left(t_{ji} + \alpha_i \right) \right) \exp\left(-\theta \beta_j \right) = Q_j^{tc}, \tag{39}$$

$$j \in J.$$

Then, by combining (36) with (39), the following relationship is established:

$$T_{ji}^{\text{vt}} = Q_j^{tc} \frac{\exp\left(-\theta \left(t_{ji} + \alpha_i \right) \right)}{\sum_{m \in I} \exp\left(-\theta \left(t_{jm} + \alpha_m \right) \right)}. \tag{40}$$

By comparing (40) with (27), α_i can now be interpreted as the taxi searching time w_i^t in zone i. After making a set of deductions, w_i^t can be given as follows:

$$w_i^t = -\frac{\ln\left(\tau A_i \right)}{\theta}. \tag{41}$$

Here, the variable τ can be calculated by using

$$\tau = \exp\left\{ \frac{-\theta \left(\text{Num} - \sum_{i \in I} \sum_{j \in J} T_{ij}^{\text{ot}} t_{ij} - \sum_{j \in J} \sum_{i \in I} T_{ji}^{\text{vt}} t_{ji} \right) - \sum_{i \in I} \sum_{j \in J} Q_{ij}^{tc} \ln A_i}{\sum_{i \in I} \sum_{j \in J} Q_{ij}^{tc}} \right\}. \tag{42}$$

The theoretical process of (41) and (42) can be derived in detail [13]. By substituting the solution to A_i in (42), the taxi searching time w_i^t in zone i can be then obtained.

Furthermore, by combining (28) with (41), the following form is also established as follows:

$$w_i^t = -\frac{\ln(\tau A_i)}{\theta} = p^{ct} w_i^{ct} + p^{dt} w_i^{dt}. \quad (43)$$

4.3. The Existence of Equilibrium. In the taxi supply relationships, by combining (42) and (43), it can be seen that the taxi-searching time, w_i^t, is a continuous function of customer demand vector Q^{tc}. Furthermore, based on the assumption of the meeting function, (29) and (30), it can be discerned that customer wait time, W_i, varies continuously with taxi search time, w_i^t, in zone i. Thus, it is concluded that the customer wait time W_i is a continuous function of Q^{tc}; continuous mapping relationship can be expressed as $W_i = W_i(Q^{tc})$. In the customer demand relationships, it can be seen that customer demand Q^{tc} is a continuous function of the cumulative prospect vector V. Additionally, based on the relationship between cumulative prospect V and trip cost U it may be deduced that cumulative prospect V varies continuously with customer wait time W_i. Such a continuous mapping relationship is also regarded as $Q^{tc} = Q^{tc}(W_i)$. Thus, the relationship is $Q^{tc} = \Gamma(Q^{tc})$. In addition, the solution set of equilibrium model variables is a compact and convex set Ω and the relationship $\Gamma(Q^{tc}) \in \Omega$ exists. By applying Brouwer's fixed-point theorem, it is easily concluded that Γ has at least one fixed point in Ω. Therefore the existence of the equilibrium solution is guaranteed [25].

5. The Solution Algorithm

The iterative algorithm of taxi equilibrium problem of a network considering the trip mode choice model based on CPT is as follows.

Step 1. Initialize customer waiting time as follows.

Set an initial set of customer waiting times $W_i^{cc(0)}$ and $W_i^{dc(0)}$, $i \in I$, then let $n = n + 1$.

Step 2. Updated customer demand value is as follows.

Substitute $W_i^{cc(n)}$ and $W_i^{dc(n)}$ into the cumulative prospect function and update $V_{ij}^{cc(n)}$, $V_{ij}^{dc(n)}$, and V_{ij}^T. Then $Q_{ij}^{cc(n)}$ and $Q_{ij}^{dc(n)}$ are updated in accordance with (16) and (17).

Step 3. Updated taxi search time is as follows.

(1) Utilize the equilibrium model for taxis in a network.

By applying the iterative balancing algorithm, A_i, B_j, and T_{ji}^{vt} can be obtained. Substituting a set of variable values involving A_i, B_j, T_{ji}^{vt}, $Q_{ij}^{cc(n)}$, and $Q_{ij}^{dc(n)}$, and so forth into

(43) a formula involving $w_i^{ct(n)}$ and $w_i^{dt(n)}$ is established. The formula is defined by

$$p^{ct(n)} w_i^{ct(n)} + p^{dt(n)} w_i^{dt(n)} = -\frac{\ln(\tau A_i)}{\theta} \quad (44)$$

$$i = 1, 2, \ldots, 6.$$

(2) Update the supply-demand equilibrium relationship.

Substitute T_{ji}^{vt} and $Q_{ij}^{cc(n)}$ into (32) or substitute T_{ji}^{vt} and $Q_{ij}^{dc(n)}$ into (33). Then update the appropriate variables. From this condition, another formulation is expressed as (45) and it includes $w_i^{ct(n)}$ and $w_i^{dt(n)}$.

$$\sum_{j \in J} T_{ji}^{vt} p^{ct(n)} = \sum_{j \in J} Q_{ij}^{cc(n)} \quad i = 1, 2, \ldots, 6. \quad (45)$$

(3) Combining (44) with (45) makes it possible to solve for the values of $w_i^{ct(n)}$ and $w_i^{dt(n)}$.

Step 4. Update customer wait time as follows.

Substituting $w_i^{ct(n)}$ and $w_i^{dt(n)}$ into the relationship between taxi search time and customer wait time, (29) and (30), respectively, makes it permissible to update $W_i^{cc(n)}$ and $W_i^{dc(n)}$.

Step 5. Check the following.

The standard to stop iteration is defined by

$$\frac{\sqrt{\sum_i \left[W_i^{cc(n+1)} - W_i^{cc(n)}\right]^2}}{\sum_i W_i^{cc(n)}} \leq \varepsilon,$$

$$\frac{\sqrt{\sum_i \left[W_i^{dc(n+1)} - W_i^{dc(n)}\right]^2}}{\sum_i W_i^{dc(n)}} \leq \varepsilon, \quad (46)$$

where ε is a condition of iterative convergence.

If (46) is achieved, then stop. Otherwise, let $n = n + 1$ and go to Step 2.

6. A Numerical Example Analysis

A simple numerical example is given in order to check the convergence of the algorithm and present the resulting analysis of the equilibrium model.

6.1. Introduction of a Numerical Example. For the model a six-zone network is utilized with nodes representing traffic zones, arcs representing relationship of adjacent nodes, and arrows indicating road direction (Figure 2). Further, each node represents a potential origin or a potential destination for customers, as well as a searching location for vacant taxis.

The travel time function for the links is defined as follows [26]:

$$t_a = t_a^0 \left[1 + 0.15 \left(\frac{v_a}{C}\right)^4\right]. \quad (47)$$

TABLE 1: The parameters in the travel impedance functions.

Link	Start node	End node	t_a^0 (h)	C (veh/h)
1	1	2	0.25	250
2	1	4	0.20	200
3	2	1	0.25	300
4	2	3	0.15	200
5	2	5	0.30	250
6	4	1	0.25	200
7	4	5	0.15	150
8	5	4	0.20	200
9	5	6	0.25	250
10	5	2	0.30	300
11	3	2	0.15	150
12	3	6	0.15	150
13	6	5	0.20	200
14	6	3	0.20	200

TABLE 2: Matrix of the total customer demand of different zones.

Q_{ij}^c (person/h)	1	2	3	4	5	6	Total
1	0	50	20	20	10	15	115
2	40	0	15	25	20	10	110
3	20	10	0	50	10	25	115
4	10	20	30	0	20	15	95
5	25	20	15	20	0	20	100
6	15	30	25	25	20	0	115
Total	110	130	105	140	80	85	650

FIGURE 2: Network of the numerical example.

FIGURE 3: The taxi-searching time of dispatch mode versus taxi fleet size in six zones.

Values for the parameters are obtained from Table 1.

A matrix of the total customer demands for taxis and other transit modes from zone i to zone j is supposed in Table 2.

The parameters used in the example were $\theta_{\text{trip}} = 15$ yuan/h, $\theta^f = 0.3$, $\theta^t = 0.5$, $\theta = 0.3$, $\varepsilon = 0.03$, $W_i^{cc(0)} = W_i^{dc(0)} = 5$ min. Additionally, the 30-minute time used in this example was regarded as a reference point for different trip modes in the CPT function. When making the sensitivity analysis, this paper set 1000, 1100, 1200, 1300, and 1400 as the values of Num, respectively.

6.2. Result Analysis. Figure 3 shows the relationship of the taxi-searching times of dispatch mode versus the taxi fleet size in six zones. It can be seen that the taxi-searching times in zones are all positive and increase linearly with an increase in the taxi fleet size. This may be due to the fact that an increase in the taxi fleet size only enhances the availability of those taxis choosing dispatching service mode in taxi market but cannot decrease the searching time of dispatching service and achieve the improvement of dispatching service quality. That is to say, the changing trend of graphic in Figure 3 indicates that the demand for dispatch mode is less than its supply in taxi market. Specifically, taking a fixed taxi fleet size, for example, the taxi-searching times of dispatch mode in different zones are also different as shown in Figure 3. This is
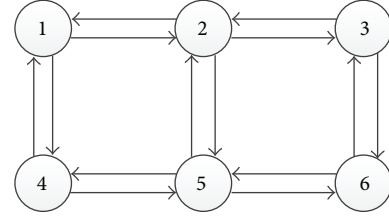
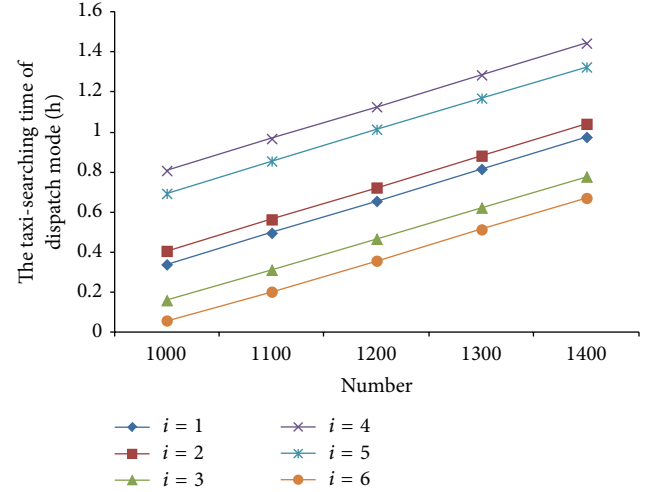mainly because both the number of customer which chooses the dispatch mode and the one of taxi which provides the dispatching service are different in six zones. Figure 3 shows that, in terms of the taxi-searching time of dispatch mode, zone 4 is maximal and zone 6 is minimal. In addition, it can be seen from Figure 3 that the maximal gap of taxi-searching time in different zones reaches at least 0.7 h considering the taxi fleet size. It can be shown that the characteristic of a traffic zone should be discussed in terms of service and operation of taxi market.

Figure 4 presents the relationship of the taxi-searching times of cruising mode versus the taxi fleet size in six zones. These taxi-searching times in zones are all positive and increase nonlinearly with an increase in the taxi fleet size. It is attributed to the fact that an increase in the taxi fleet size leads to more greater supply of cruising mode and more worse service quality. Specifically, taking a fixed taxi fleet size, for example, the taxi-searching times of cruising mode in different zones are also different as depicted in Figure 4. It can be shown that, in terms of the taxi-searching time of cruising mode, zone 4 is maximal and zone 3 is minimal. The maximal gap of taxi-searching time in different zones reaches at least 0.6 h considering the taxi fleet size. In addition, combining Figure 3 with Figure 4, it can be seen that the taxi-searching times of cruising mode in zones are greater than the ones of dispatch mode, which reflects that the service quality of
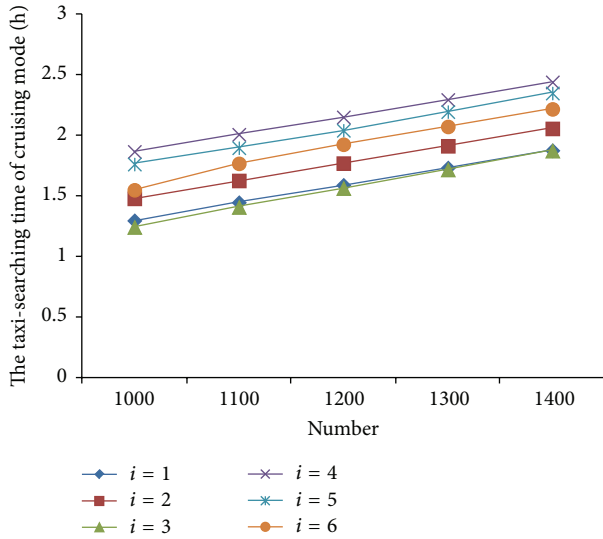
Figure 4: The taxi-searching time of cruising mode versus taxi fleet size in six zones.

dispatch mode is better than the one of cruising mode in taxi market.

Figure 5 depicts the relationship between the probability of different modes chosen and an increase in the taxi fleet size in six zones. Figure 5 shows that, in terms of these two service modes, their values have no great changes with an increase in the taxi fleet size. And furthermore, compared with the probabilities of dispatch mode chosen in six zones, the ones of cruising mode chosen are relatively less, which is mainly attributed to the policy of dispatch mode and its parameters adopted by this paper.

Figure 6 depicts the relationship of the total taxi-searching time versus the taxi fleet size in six zones. Figure 6 indicates that the total taxi-searching times in zones are all positive and increase linearly with an increase in the taxi fleet size. This is due to the fact that the increased taxis mainly search for customers instead of providing the occupied service. That is to say, an increase in the taxi fleet size can enhance taxi's availability in taxi market but cannot indicate the improvement of service quality. The complicated relationship between taxi supply and customer demand is totally determined by the market and network conditions. This conclusion is consistent with the achievement from Yang et al. [15]. And furthermore, taking a fixed taxi fleet size, for example, total taxi-searching times in different zones are also different as shown in Figure 6. This is mainly because customer demands and network conditions (e.g., the travel time to traverse OD pair) are different in six zones. Figure 6 shows that, in terms of the total taxi-searching time, zone 4 is maximal, zone 6 is minimal, and zone 3 is nearly equal to zone 6. And moreover, Figure 6 indicates that the maximal gap of total taxi-searching time in different zones reaches at least thirty minutes considering the taxi fleet size. This means the characteristic of a traffic zone directly influences taxi's availability and further the level of service quality in taxi market.

Figure 7 gives the relationship of the customer-waiting time in cruising mode versus the taxi fleet size in different zones. Figure 7 shows that the customer-waiting times in six zones are all positive and decrease nonlinearly with an increase in the taxi fleet size. This indicates that the marginal effect of advancing the taxi fleet size in improving service quality is diminishing. In terms of a given taxi fleet size, the customer-waiting times of cruising mode are different in six zones as shown in Figure 7. Figure 7 shows that, in terms of the customer-waiting time, zone 3 is maximal, zone 4 is minimal, and zone 4 is nearly equal to zone 5. And, moreover, Figure 7 indicates that the maximal gap of the customer-waiting time in different zones reaches at least three minutes. And, furthermore, this gap in six zones is decreasing with an increase in taxi fleet size, which means that the distribution and utilization of taxis in six zones will become similar with an increase in the taxi fleet size.

Figure 8 presents the relationship between the customer-waiting time of dispatch mode and an increase in the taxi fleet size in six zones. Figure 8 shows that the customer-waiting times in six zones are all positive and decrease nonlinearly with an increase in the taxi fleet size. In terms of a single taxi fleet size, the customer-waiting time of dispatch mode in zone 6 is maximal and the customer-waiting time of zone 4 is minimal. The maximal gap of this index in different zones reaches about six minutes. The tendency that the gap of customer-waiting time in dispatch mode in six zones is decreasing with an increase in the taxi number is then reflected.

7. Conclusions

This paper considered the effect of the integrated service mode including cruising and dispatch modes and the travel time uncertainty on taxi network equilibrium. To facilitate this aim the cumulative prospect theory was adopted to study the trip mode choice problem. Based on the elasticity of customer demand, a mathematical programming model was established to describe the supply-demand equilibrium relationship of taxi service on a network with respect to the integrated service mode. From this an iterative algorithm was designed to present the optimal solutions of cruising mode's parameters, dispatch mode's parameters, and taxis network's parameters in an equilibrium state. Finally, a numerical example was used to present the achievement of supply-demand equilibrium state and determine the outcomes of key parameters.

The results provided interesting insights and valuable recommendations for improving taxi market operation. Specifically, using taxi network equilibrium, the taxi-searching and customer-waiting time in six zones were obtained with an increase in the taxi fleet size. We found that the taxi-searching times of both dispatch mode and cruising mode in zones were positive and increased with an increase in the taxi fleet size, while the customer-waiting times of both dispatch mode and cruising mode in zones were also positive but decreased with an increase in the taxi fleet size. Moreover, the taxi-searching time of cruising mode was greater than

(a) The probability of cruising mode chosen

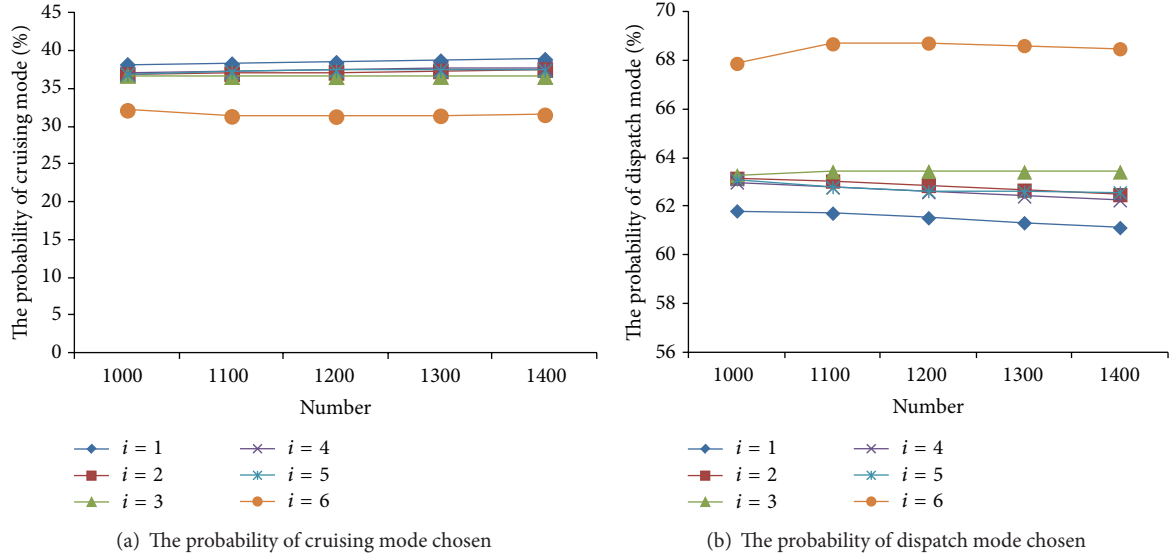(b) The probability of dispatch mode chosen

FIGURE 5: The probability of different modes chosen versus taxi fleet size in six zones.
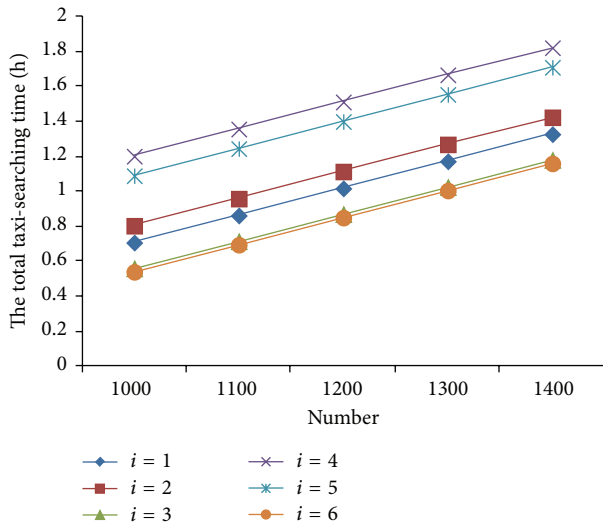

FIGURE 6: The total taxi-searching time versus taxi fleet size in six zones.
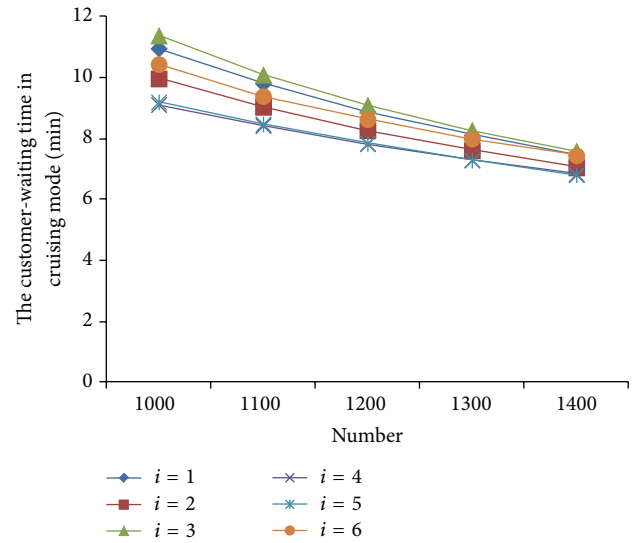

FIGURE 7: Customer-waiting time in cruising mode versus taxi fleet size in zones.

the one of dispatch mode, while the customer-waiting time of cruising mode was also longer than the one of dispatch mode in six zones. Further, the algorithm indicated that the probability of dispatch mode being chosen by a customer was greater than cruising mode in the six zones considered. Finally, based on the above results from different service modes, the total taxi-searching times of integrated taxi mode were also obtained. Thus, from the practical point of view, these results should enable policymakers to know current operation situations in taxi market and to further adjust and formulate the reasonable operation policies in terms of integrated service mode.

However, to make this mathematical model more suitable for reality, several relationships in the model still need

to be further improved. Firstly, the dispatch policy used in the model is a little simple so it cannot offer further insights into the interactions of cruising and dispatch modes. Therefore, applying a complicated dispatch policy will be more meaningful. Secondly, in the application of CPT, the reference points in those designed zones are only considered as constants instead of relevant functions, which need a suitable method to replace the current situation. And furthermore, an uncertainty of customers choosing trip mode should be reflected by adopting a continuous probability distribution function instead of the discrete function, which will be attempted to make further improvement.
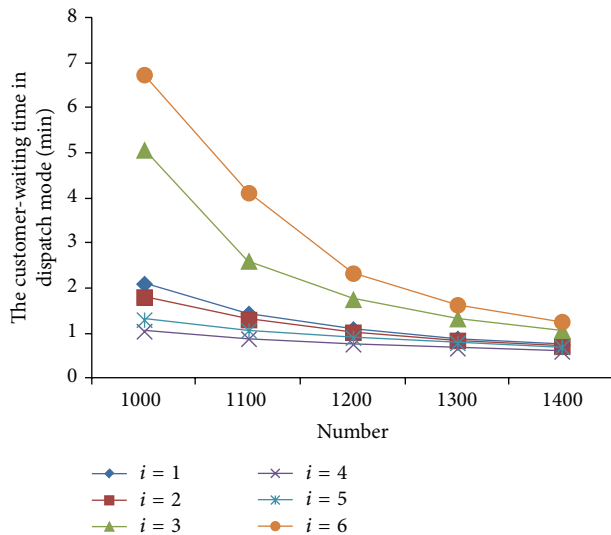
FIGURE 8: Customer-waiting time in dispatch mode versus taxi fleet size in zones.

Conflict of Interests

None of the authors has a financial relationship with a commercial entity that has an interest in the subject of this paper.

Acknowledgments

References

[1] X. Hu, N. Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 3, pp. 1193–1205, 2014.

[2] M. E. Beesley, "Regulation of taxis," *The Economic Journal*, vol. 83, pp. 150–172, 1973.

[3] M. E. Beesley and S. Glaister, "Information for regulating: the case of taxis. Royal economic society," *The Economic Journal*, vol. 93, no. 371, pp. 594–615, 1983.

[4] S. De Vany, "Capacity utilization under alternative regulatory restraints: an analysis of taxi markets," *The Journal of Political Economy*, vol. 83, no. 1, pp. 83–94, 1975.

[5] C. F. Daganzo, "An approximate analytic model of many-to-many demand responsive transportation systems," *Transportation Research*, vol. 12, pp. 325–333, 1978.

[6] R. D. Cairns and C. Liston-Heyes, "Competition and regulation in the taxi industry," *Journal of Public Economics*, vol. 59, no. 1, pp. 1–15, 1996.

[7] L. J. E. Fernandez, C. Joaquin de Cea, and M. Julio Briones, "A diagrammatic analysis of the market for cruising taxis," *Transportation Research Part E*, vol. 42, pp. 498–526, 2006.

[8] H. Yang, S. C. Wong, and K. I. Wong, "Demand-supply equilibrium of taxi services in a network under competition and regulation," *Transportation Research Part B: Methodological*, vol. 36, no. 9, pp. 799–819, 2002.

[9] H. Yang and S. C. Wong, "A network model of urban taxi services," *Transportation Research Part B: Methodological*, vol. 32, no. 4, pp. 235–246, 1998.

[10] H. Yang, Y. W. Lau, S. C. Wong, and H. K. Lo, "A macroscopic taxi model for passenger demand, taxi utilization and level of services," *Transportation*, vol. 27, no. 3, pp. 317–340, 2000.

[11] H. Yang, M. Ye, W. H. Tang, and S. C. Wong, "Regulating taxi services in the presence of congestion externality," *Transportation Research Part A: Policy and Practice*, vol. 39, no. 1, pp. 17–40, 2005.

[12] H. Yang, C. S. Fung, K. I. Wong, and S. C. Wong, "Nonlinear pricing of taxi services," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 337–348, 2010.

[13] K. I. Wong, S. C. Wong, and H. Yang, "Modeling urban taxi services in congested road networks with elastic demand," *Transportation Research Part B: Methodological*, vol. 35, no. 9, pp. 819–842, 2001.

[14] K. I. Wong and S. C. Wong, "A sensitivity-based solution algorithm for the network model of urban taxi services," in *Transportation and Traffic Theory in the 21th Century*, pp. 23–42, Elsevier Science, Adelaide, Australia, 2002.

[15] H. Yang, C. W. Y. Leung, S. C. Wong, and M. G. H. Bell, "Equilibria of bilateral taxi–customer searching and meeting on networks," *Transportation Research Part B: Methodological*, vol. 44, no. 8-9, pp. 1067–1083, 2010.

[16] H. Yang and T. Yang, "Equilibrium properties of taxi markets with search frictions," *Transportation Research Part B: Methodological*, vol. 45, no. 4, pp. 696–713, 2011.

[17] R. Arnott, "Taxi travel should be subsidized," *Journal of Urban Economics*, vol. 40, no. 3, pp. 316–333, 1996.

[18] M. von Massow and M. S. Canbolat, "Fareplay: an examination of taxicab drivers' response to dispatch policy," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2451–2458, 2010.

[19] J. Maria Salanora, M. Estrada, G. Aifadopoulou, and E. Mitsakis, "A review of the modelling of taxi services," *Procedia Social and Behavioral Sciences*, vol. 20, pp. 150–161, 2011.

[20] E. Avineri, "The effect of reference point on stochastic network equilibrium," *Transportation Science*, vol. 40, no. 4, pp. 409–420, 2006.

[21] E. J. van de Kaa, "Applicability of an extended prospect theory to travel behaviour research: a meta-analysis," *Transport Reviews*, vol. 30, no. 6, pp. 771–804, 2010.

[22] E. J. van de Kaa, "Prospect theory and choice behaviour strategies: review and synthesis of concepts from social and transport sciences," *European Journal of Transport and Infrastructure Research*, vol. 10, no. 4, pp. 299–329, 2010.

[23] Z. Li and D. Hensher, "Prospect theoretic contributions in understanding traveller behaviour: a review and some comments," *Transport Reviews*, vol. 31, no. 1, pp. 97–115, 2011.

[24] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[25] A. de La Fuente, *Mathematical Methods and Models for Economists*, Cambridge University Press, Cambridge, UK, 2000.

[26] U.S. Bureau of Public Roads, *Traffic Assignment Manual*, U.S. Bureau of Public Roads, Washington, DC, USA, 1964.

Research Article

Multistage CC-CV Charge Method for Li-Ion Battery

Xiaogang Wu,[1] Chen Hu,[1] Jiuyu Du,[2] and Jinlei Sun[3]

[1] *College of Electrical and Electronic Engineering, Harbin University of Science and Technology, Xue Fu Road 52, Harbin 150080, China*
[2] *State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing, China*
[3] *School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China*

Correspondence should be addressed to Xiaogang Wu; xgwu@hrbust.edu.cn

Charging the Li-ion battery with constant current and constant voltage (CC-CV) strategy at $-10°C$ can only reach 48.47% of the normal capacity. To improve the poor charging characteristic at low temperature, the working principle of charging battery at low temperature is analyzed using electrochemical model and first-order RC equivalent circuit model; moreover, the multistage CC-CV strategy is proposed. In the proposed multistage CC-CV strategy, the charging current is decreased to extend the charging process when terminal voltage reaches the charging cut-off voltage. The charging results of multistage CC-CV strategy are obtained at $25°C$, $0°C$, and $-10°C$, compared with the results of CC-CV and two-stage CC-CC strategies. The comparison results show that, at the target temperatures, the charging capacities are increased with multistage CC-CV strategy and it is notable that the charging capacity can reach 85.32% of the nominal capacity at $-10°C$; also, the charging time is decreased.

1. Introduction

With the advantages of zero pollution, high energy efficiency, and pluralistic energy sources, electric vehicle (EV) has been the new development point of motor industry [1–3]. Li-ion battery has been widely used in EV for its high energy density, long cycle life, and high safety level [4]. But the battery technology still cannot meet the EV demand of long travel distance, fast capacity recovery, and low temperature utilization [5]. At low temperature, battery chemical activity decreases, resistance increases, and capacity is decreased. Charging process is more difficult than the discharging process at low temperature [6, 7].

Much work has been done on charging strategies in recent years. In [8] a three-step charging method for Ni/MH battery was proposed to obtain the rapid charge. In [9], an optimum current charging strategy based on the boundary charging current curves was proposed. The boundary charging current curves were obtained by analysis of temperature rise and polarization voltage in charging process. The charging period was decreased and capacity was increased with the strategy. Reference [10] proposed a duty-varied voltage charging strategy that can detect and dynamically track the suitable duty of the charging pulse. Compared with conventional

CC-CV strategy, the charging speed was increased by 14%, and charging efficiency was increased by 3.4%. Reference [11] constructed a SOC estimation model and the CC-CV charging process was controlled by battery SOC. The charging capacity can be monitored to gain a higher level charging degree and avoid being overcharged. In [12], an Ant Colony System algorithm was used to select the optimum charging current among five charging states and the charging time was decreased and battery cycle life was extended by 25%. In [13], a Taguchi-based algorithm was used to obtain rapid charge. With the charging strategy, the battery capacity could reach to 75% in 40 min. In [14], a constant-polarization-based fuzzy-control charging method was proposed to adapt charging current acceptance with battery SOC stages. The charging strategy could shorten charging time with no obvious temperature rise. Ruan et al. and Zhao et al. [15, 16] studied the temperature characteristic of charging and discharging process. The temperature increased more in discharging process compared to the temperature increase in charging process. The pulse charging/discharging process was added before charging process so the battery could be preheated. The battery could start charging process at relatively high temperature and charging capacity was increased at low temperature.

TABLE 1: Equipment parameters.

Battery equipment	Maximum test current	20 A
	Maximum test voltage	5 V
	Test accuracy	0.1%
Temperature chamber	Maximum temperature	150°C
	Minimum temperature	−40°C
	Temperature tolerance	0.01°C
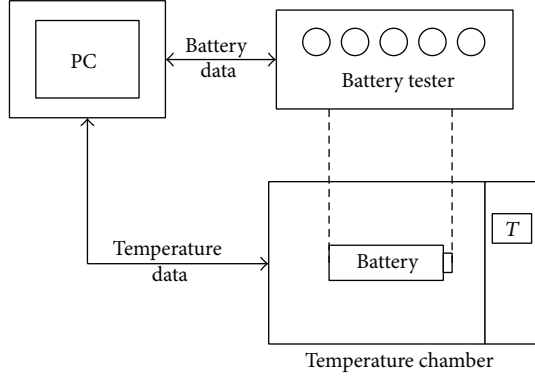
FIGURE 1: Experimental setup.

FIGURE 2: CC-CV strategy charging capacities at different temperature.

All the charging strategies increase the battery charging characteristic at different degrees proposed in [8–14]. But the charging performance at low temperature is not considered. Although the preheating charging strategy at low temperature proposed in [15, 16] can increase the charging capacity, the self-preheating process costs too much time and cannot work at low SOC condition. This paper analyzes the charging characteristic of a Li-ion battery at different temperature, uses electrochemical model and first-order RC equivalent circuit model to analyze the bad low temperature characteristic of Li-ion battery in theory, and proposes a multistage CC-CV strategy. The multistage CC-CV strategy is compared with CC-CV and two-stage CC-CV strategies at 25°C, 0°C, and −10°C.

2. Experimental

2.1. Battery and Equipment. The battery used is 18650 cylindrical Li-ion battery with normal capacity of 1.37 Ah, a normal voltage of 3.2 V, and a cut-off voltage of 3.6 V. The maximum charging and discharging rates are 1 C and 2 C, respectively. The positive electrode material is $LiFePO_4$, and negative electrode material is LiC_6. The battery tester is LD battery tester with 8 test channels and the test process can be programmed and monitored by computer. The battery was tested in a temperature chamber to ensure the temperature parameter to be constant. The detailed parameters of battery tester and temperature chamber are shown in Table 1. The experimental setup can be described as in Figure 1.

2.2. Experimental Process. The battery charging strategies tested in experiments were CC-CV, two-stage CC-CV, and multistage CC-CV. The test temperature points were 25°C, 0°C, and −10°C. The charging strategies are explained as follows.
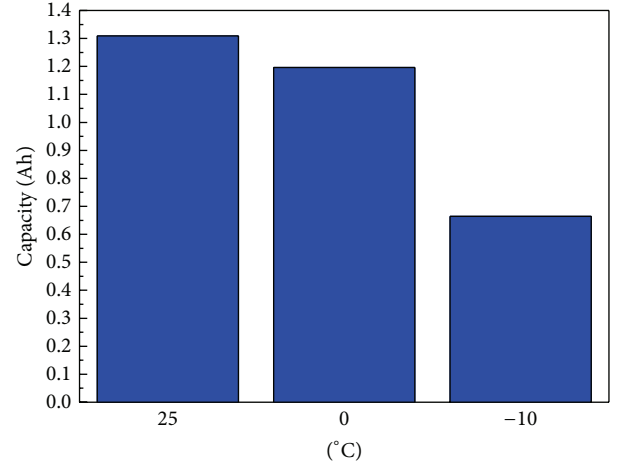
For the CC-CV strategy, the constant current process was charging at 0.3 C to the cut-off voltage of 3.6 V and the constant voltage process was charging at 3.6 V for 5 min.

For the two-stage CC-CV strategy, the first constant current process was charging at 1 C to the cut-off voltage of 3.6 V. Then in the second constant current process, the charging current was decreased to 0.5 C. Since the charging current was decreased, the terminal voltage was decreased below 3.6 V allowing the constant current process to be extended, until the terminal voltage reached the cut-off voltage once again. The constant voltage process was charging at 3.6 V for 5 min [17].

For the multistage CC-CV strategy, the constant current process was divided into ten stages. The maximum and minimum rates were 1 C and 0.1 C, respectively, and the charging current was decreased by 0.1 C when terminal voltage reached the cut-off voltage. The constant voltage process was charging at 3.6 V for 5 min.

3. Charging Characteristic of Battery at Low Temperature

3.1. Charging Capacity Characteristic at Different Temperature. The selected battery was charged by CC-CV strategy at 25°C, 0°C, and −10°C to obtain the charging capacity characteristic at low temperature. Before every charging process at different temperature, the battery was discharged empty at 25°C and kept for six hours to ensure the whole battery temperature to be uniform. As shown in Figure 2, the charging capacities at 25°C, 0°C, and −10°C are 1.309 Ah, 1.196 Ah, and 0.664 Ah, respectively. The charging capacity is decreased by 8.6% at 0°C and 49.3% at −10°C compared with that at 25°C. The charging capacity has a great decrease at −10°C.

3.2. OCV Characteristic at Different Temperature. The battery was tested by hybrid pulse power characteristic (HPPC) rule that is detailed in "Freedom CAR Battery Test Manual" [18] to obtain OCV, ohmic resistance (R_r), and polarization
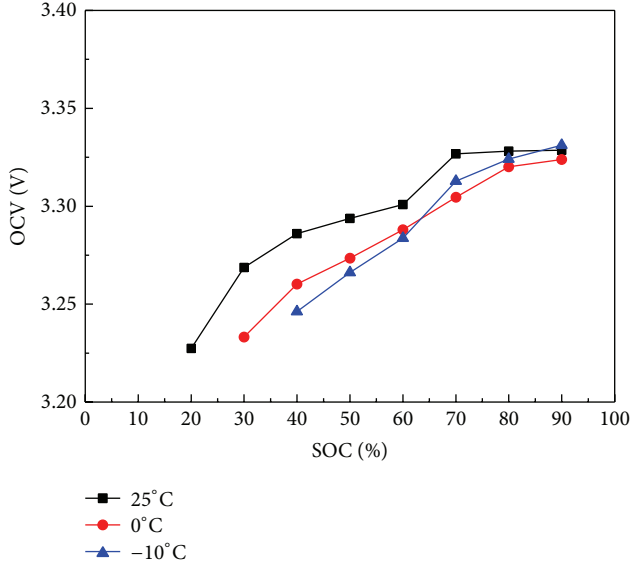
FIGURE 3: OCV curves at different temperature.

FIGURE 4: R_r curves at different temperature.

FIGURE 5: R_p curves at different temperature.

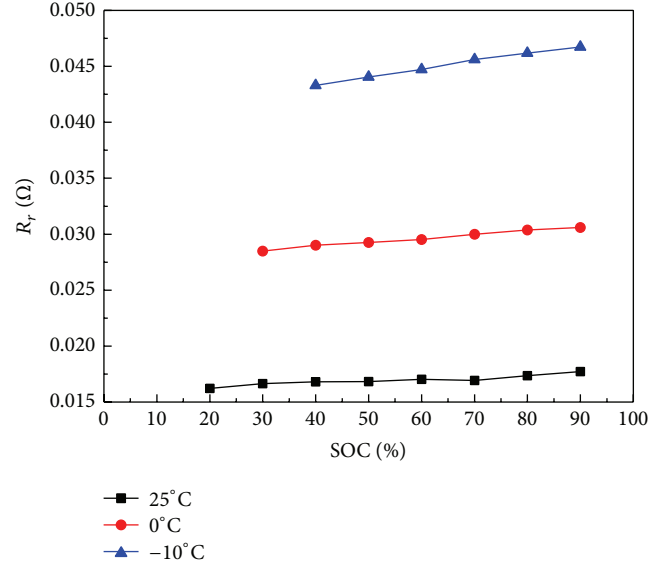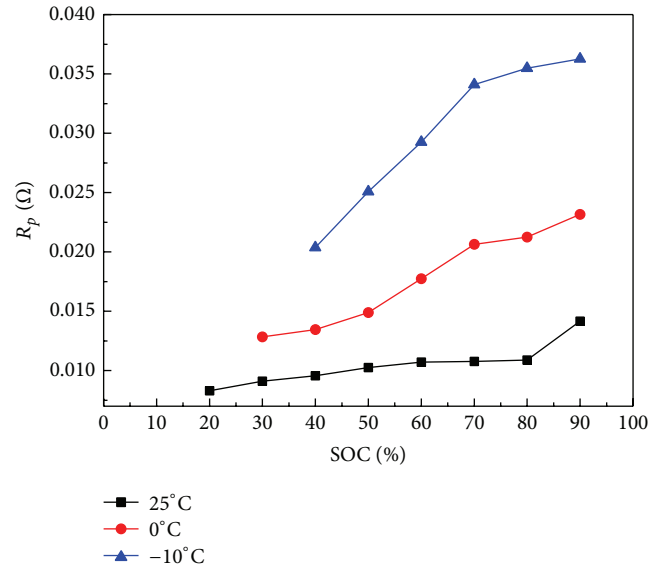resistance (R_p). SOC can be calculated by the following formula:

$$\text{SOC} = \text{SOC}_0 - \frac{1}{\text{AHC}} \int_0^t i \, d\tau, \tag{1}$$

where SOC_0 is the initial SOC of the battery, AHC is the normal capacity of the battery at $25°C$, and i is the discharge (positive i) or charge (negative i) current. As the OCV curves shown in Figure 3, the OCV reflects increasing tendency with temperature decreasing and the difference of OCV at different temperature is relatively more obvious at low SOC.
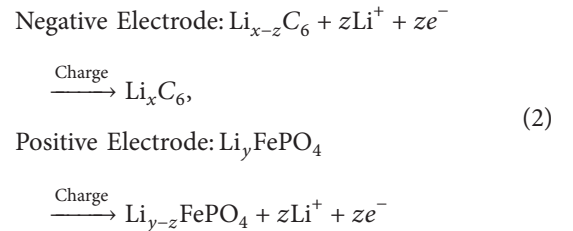
3.3. R_r and R_p Characteristic at Different Temperature. As shown in Figures 4 and 5, both R_r and R_p increase with temperature decreasing. R_r remains steady with SOC increasing, and the increase is nearly 258% at $-10°C$. R_p increases with SOC increasing and temperature decreasing, and the maximum increase in R_p is nearly 257% at $-10°C$ with 90% of SOC.

4. Electrochemical and First-Order RC Equivalent Circuit Model

4.1. Electrochemical Li-Ion Battery Model. Doyle et al. have proposed the porous electrode theory for the analysis of electrochemical process of Li-ion battery [19]. The one-dimensional geometry consists of negative/positive current collector, negative/positive electrodes, and separator. The negative current collector material is copper, and positive current collector material is aluminum. The positive electrode active material is LiFePO$_4$, and negative electrode active material is LiC$_6$. The separator is polyolefin porous membrane. The electrolyte is lithium salt dissolved in 1:1 or 2:1 liquid mixture of ethylene carbonate (EC) and dimethyl carbonate (DMC). The one-dimensional geometry example

of charging process is shown in Figure 6 [20] and the charging chemical equation is

$$
\begin{aligned}
&\text{Negative Electrode: Li}_{x-z}\text{C}_6 + z\text{Li}^+ + ze^- \\
&\qquad \xrightarrow{\text{Charge}} \text{Li}_x\text{C}_6, \\
&\text{Positive Electrode: Li}_y\text{FePO}_4 \\
&\qquad \xrightarrow{\text{Charge}} \text{Li}_{y-z}\text{FePO}_4 + z\text{Li}^+ + ze^-
\end{aligned}
\tag{2}
$$

In the charging process, the electrons move from the positive electrode to the negative electrode through the external circuit, and Li$^+$ moves from the positive electrode to the negative electrode through the separator in electrolyte.
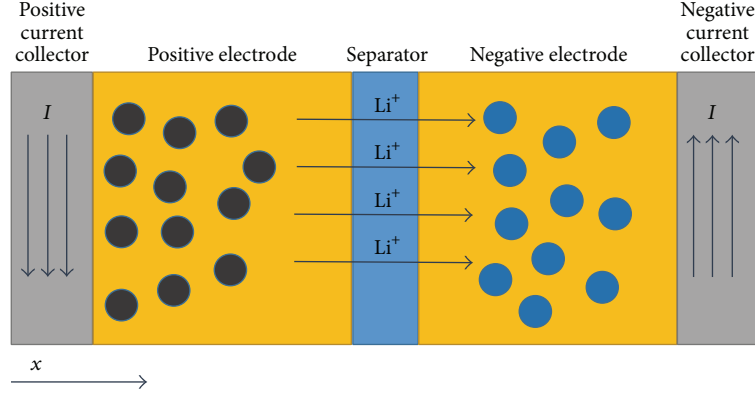
FIGURE 6: One-dimensional geometry example of charging process.

As the charging process is a chemical reaction, the reaction characteristic is influenced by concentration and Li$^+$ diffusion. The Li-ion concentration in electrolyte phase changes with time and can be described by Fick's second law along the x-coordinate shown in Figure 6 [21]:

$$\frac{\partial C_e}{\partial t} = \frac{\partial}{\partial x}\left(D_e \frac{\partial C_e}{\partial x}\right) + \frac{1 - t_+^0}{F} j^{\mathrm{Li}}, \tag{3}$$

where C_e is the concentration of Li-ion in electrolyte phase, D_e is Li$^+$ diffusion coefficient in electrolyte phase, t_+^0 is the transference number of lithium ions with respect to the velocity of the solvent, F is Faraday constant, and j^{Li} is charging transfer current density.

The distribution of Li-ion in solid state phase is also described by Fick's second law of diffusion in polar coordinates [21]:

$$\frac{\partial C_s}{dt} = \frac{D_s}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{C_s}{\partial r}\right), \tag{4}$$

where C_s is the concentration of Li-ion in solid, D_s is the Li$^+$ diffusion coefficient in solid state phase, and r is radius of spherical particle.

The Arrhenius formula shows the Li$^+$ diffusion coefficient D_s in solid state phase as shown below [21]:

$$D_s(T) = D_{\mathrm{ref}} \exp\left[\frac{E_{aD}}{R}\left(\frac{1}{T_{\mathrm{ref}}} - \frac{1}{T}\right)\right], \tag{5}$$

where E_{aD} is the activation energy for diffusion. R is the universal gas constant, D_{ref} is the reference diffusion coefficient at T_{ref}, T_{ref} is the reference temperature, and T is the temperature. Formula (5) shows that the diffusion coefficient decreases with the temperature decreasing. Reference [14] indicates that the solid state phase diffusion polarization dominates the total polarization and the solid state phase polarization is increased with diffusion coefficient decreasing. The increase of polarization results in higher polarization voltage compared with that of normal temperature, the terminal voltage increasing space during constant current charging process is decreased, and the charging capacity will be decreased.

The charging transfer current density can be obtained using the following Butler-Volmer formula [20]:

$$j^{\mathrm{Li}} = j_0 \left\{\exp\left[\frac{\alpha_a F}{RT}\eta\right] - \exp\left[\frac{\alpha_c F}{RT}\eta\right]\right\}, \tag{6}$$

where j_0 is the exchange current density, α_a and α_c are the transfer coefficients of anode and cathode, and η is the surface over potential, which can be obtained using the following formula [20]:

$$\eta = \phi_s - \phi_e - U_{\mathrm{ocv}}, \tag{7}$$

where ϕ_s is the solid phase potential, ϕ_e is the electrolyte phase potential, and U_{ocv} is the open circuit voltage.

j_0 can be described as shown below [20]:

$$j_0 = F k_0 C_e^{\alpha_a}\left(C_{s,\max} - C_{s,\mathrm{surf}}\right)^{\alpha_a} C_{s,\mathrm{surf}}^{\alpha_a}, \tag{8}$$

where k_0 is the reaction rate coefficient, $C_{s,\max}$ is the maximum Li-ion concentration in the electrodes, and $C_{s,\mathrm{surf}}$ is the Li-ion concentration on the active particles surface.

k_0 can be obtained using the following formula [20]:

$$k_0(T) = k_{0,\mathrm{ref}} \exp\left[\frac{E_{aR}}{R}\left(\frac{1}{T_{\mathrm{ref}}} - \frac{1}{T}\right)\right], \tag{9}$$

where E_{ar} is the reaction activation energy and $k_{o,\mathrm{ref}}$ is the reaction rate coefficient at T_{ref}. With the temperature decreasing, reaction rate coefficient is decreased. As formula (7) shows, k_0 is decreased with temperature decreasing. The charging reaction is impeded for the reaction rate coefficient decreasing. As the parameter is time-invariant, the charging obstruction can be considered as a resistive process. The increase of impedance also results in the terminal voltage increase and the decrease of charging capacity.

The electrochemistry model analysis of the charging process at low temperature shows that the main obstruction consists of polarization and impedance increase. This increase can be analyzed by the equivalent circuit model, the polarization can be modeled by capacitance and resistance in parallel, and the impedance can be modeled by resistance. A first-order RC equivalent circuit model is used in the next part.
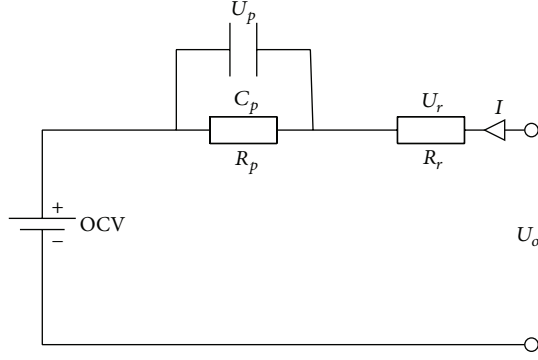
FIGURE 7: First-order RC equivalent circuit model.

FIGURE 8: Charging curves of CC-CV strategy at 25°C.

4.2. First-Order RC Equivalent Circuit Model. The first-order RC equivalent circuit model is used to analyze the charging process [21, 22]. As shown in Figure 7, R_r represents the ohmic resistance, U_r is the voltage on R_r, C_p and R_p, respectively, represent the polarization capacity and polarization resistance, U_p is the voltage on C_p and R_p, OCV is the open circuit voltage, U_o is the terminal voltage, and I is the charging current. The following formulas can be obtained:

$$C_p \frac{du_p(t)}{dt} + \frac{u_p(t)}{R_p} = i(t),$$

$$u_0(t) = u_{\text{ocv}}(t) + i(t) R_r + u_p(t). \tag{10}$$

With assumption of $i(0) = I$ and $u_p(0) = 0$, the following can be obtained:

$$u_p(t) = IR_p \left(1 - e^{-t/\tau}\right),$$

$$u_0(t) = u_{\text{ocv}}(t) + IR_r + u_p(t), \tag{11}$$

where $\tau = R_p C_p$.

It can be seen from formulas (10)-(11) that U_o is determined by OCV, R_p, R_r, and I. As is mentioned above, OCV changes little with temperature decreasing, while R_r and R_p increase significantly with temperature decreasing. The increase of R_p can be explained by the slow kinetics of electrochemical reaction influenced by temperature. The constant current process of CC-CV strategy is limited by cut-off voltage and the charging capacity mainly depends on the constant current process. At low temperature, R_r and R_p increase making U_r and U_p increase, and U_0 is higher than that at normal temperature. The cut-off voltage is reached earlier and the constant current process is stopped earlier [23]. The increasing of R_p and R_r depends on the battery design parameters and cannot be controlled during the charging process. The only parameter which can be controlled is the charging current. As proposed in [17], for a two-stage CC-CV strategy, the constant current charging process was divided into two stages. The first stage is charging battery with the maximum charging rate until the cut-off voltage is reached. The second stage charging current was decreased to half of the maximum charging rate, and the terminal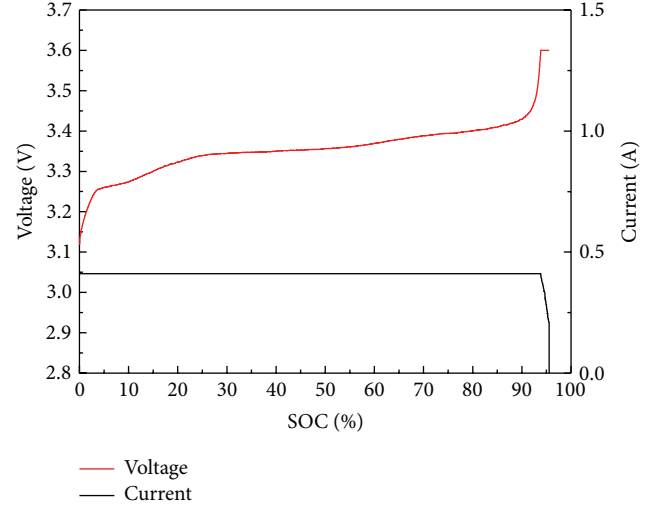 voltage can be decreased to extend the constant current charging process to increase capacity. According to the current decrease process of the two-stage CC-CV strategy, a multistage CC-CV strategy with more detailed current rates is proposed in this paper. Once the cut-off voltage is rapidly reached at a low temperature, the terminal voltage can be decreased with charging current decreasing, and the constant current charging process can be repeatedly extended to increase charging capacity. Meanwhile, the charging current is decreased from the maximum rate, and the multistage can automatically and degressively select the optimal charging current to use high charging rate as far as possible and shorten the charging period.

5. Result and Discussion

5.1. Different Charging Strategy Analysis at 25°C. Figures 8–10 show the terminal voltage curves with different charging strategies at 25°C. The terminal voltage of CC-CV strategy increases to 3.25 V at the low SOC range of 0%–10%, while the terminal voltages of two-stage CC-CV and multistage CC-CV strategies increase to near 3.4 V. The terminal voltage of CC-CV strategy increases to 3.4 V with SOC reaching 90% and has a huge increase to 3.6 V at the end of charging. The terminal voltages of two-stage CC-CV and multistage CC-CV strategies increase to 3.6 V with SOC of 85%. With current decreasing, the terminal voltage of two-stage CC-CV strategy decreases to 3.49 V and increases to 3.6 V again with SOC increasing of 7%. Unlike two-stage CC-CV strategy, the terminal voltage of multistage CC-CV strategy has more decreasing times to extend the charging SOC to a higher level.

Figure 11 shows the SOC curves of different charging strategies at 25°C. The charging capacities of CC-CV, two-stage CC-CV, and multistage CC-CV charging strategies are 1.309 Ah, 1.299 Ah, and 1.368 Ah, respectively. The capacities of two-stage CC-CV and multi-CC-CV strategies are higher than that of CC-CV strategy for current decreasing process. The multi-CC-CV has the highest charging capacity because the current decrease process of multistage CC-CV strategy
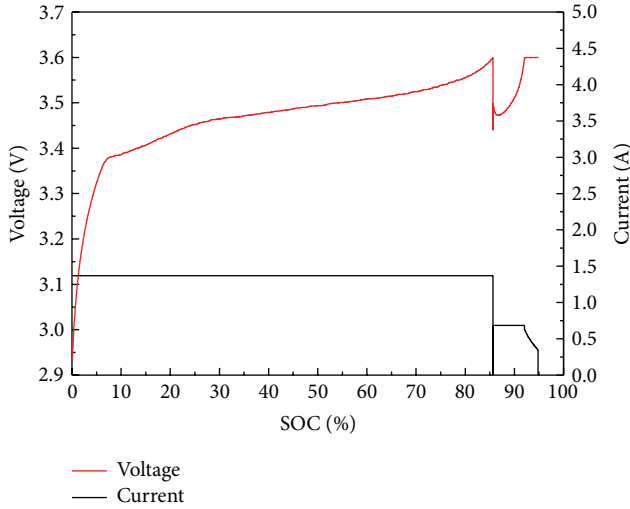
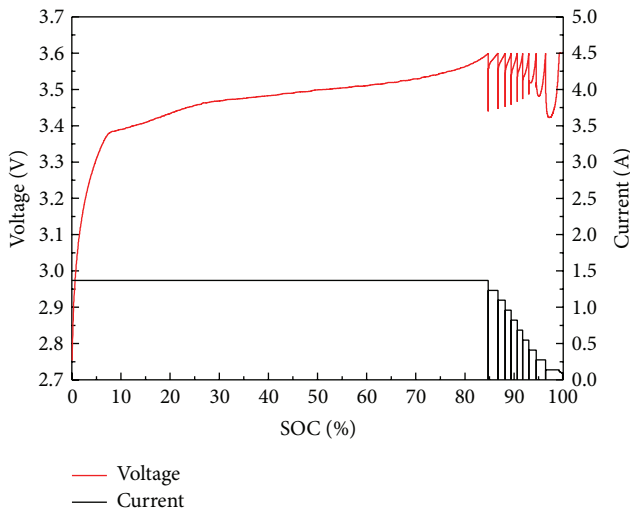FIGURE 9: Charging curves of two-stage CC-CV strategy at 25°C.

FIGURE 10: Charging curves of multistage CC-CV strategy at 25°C.

FIGURE 11: SOC curves of different charging strategies at 25°C.

has more gradients than two-stage CC-CV strategy. The charging periods of CC-CV, two-stage CC-CV, and multi-stage CC-CV charging strategies are 223 min, 67.4 min, and 94.7 min, respectively. It is obvious that the CC-CV charging strategy has the longest charging period for a low constant charging rate. Although the whole charging period of two-stage CC-CV is shorter than that of the multistage CC-CV, multistage CC-CV charging strategy has a larger charging capacity. The charging period of multistage CC-CV strategy is shorter than that of two-stage CC-CV strategy at the same charging SOC point 94.8%.

5.2. Different Charging Strategy Analysis at 0°C. As shown in Figure 12, unlike the terminal voltage at 25°C, the terminal voltage of CC-CV strategy increases to 3.35 V at low SOC range of 0%–10%. The terminal voltage increase slope during SOC range of 10%–80% is enhanced. The terminal voltage increase towards the cut-off voltage and sharp increase
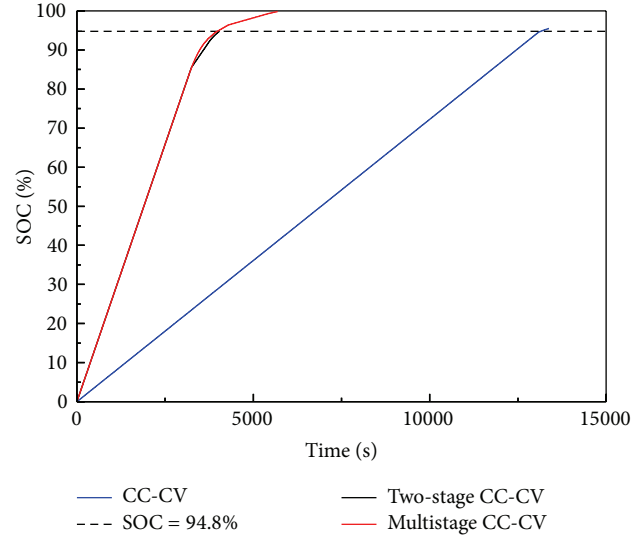
at 25°C with SOC range beyond 80% are vanished. The increase in terminal voltage indicates that the normal CC-CV charging process has been changed by the increase in internal resistance at low temperature.

As shown in Figures 13-14, the first charging stage of two-stage CC-CV strategy does not last long before the cut-off voltage is reached for the increase in internal resistance and the high charging rate. The second charging stage decreases the charging current rate by 0.5 C, and the terminal voltage decreases by 0.23 V and keeps on increasing until the cut-off voltage is reached. Unlike the two-stage CC-CV strategy, the current decreases by 0.1 C of the multistage CC-CV strategy. The terminal voltages of 1 C–0.7 C constant current charging stages increase rapidly with SOC below 22%. Terminal volt-ages of 0.6 C–0.4 C constant current charging stages increase slower with SOC between 22% and 73.4%. Terminal voltages of 0.3 C–0.1 C constant current charging stages increase rapidly again with SOC beyond 73.4%. The terminal voltage curve of multistage CC-CV indicates that multistage CC-CV strategy can automatically select the optimal charging current by cut-off voltage limiting and current decreasing.

The charging result at 0°C shows that the capacities of CC-CV, two-stage CC-CV, and multistage CC-CV charging strategies are 1.196 Ah, 0.758 Ah, and 1.246 Ah, respectively. Compared with the charging result at 25°C, the charging capacities of CC-CV, two-stage CC-CV, and multistage CC-CV charging strategies decrease by 8.2%, 39.5%, and 8.9%, respectively. As the main charging rate of two-stage CC-CV strategy is 0.5 C higher than 0.3 C of CC-CV strategy and the charging rate does not decrease further, the two-stage CC-CV strategy has the largest decrease in charging capacity decrease at 0°C. As multistage CC-CV strategy has 0.2 C and 0.1 C charging rate lower than 0.3 C of CC-CV strategy, the charging capacity of multistage CC-CV strategy is higher than that of CC-CV strategy.

Figure 15 shows the SOC curves of different charging strategies at 0°C. The charging periods of CC-CV, two-stage CC-CV, and multistage CC-CV charging strategies
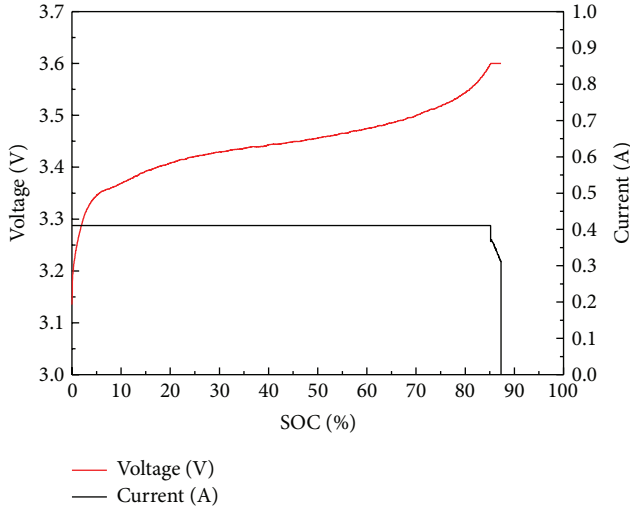
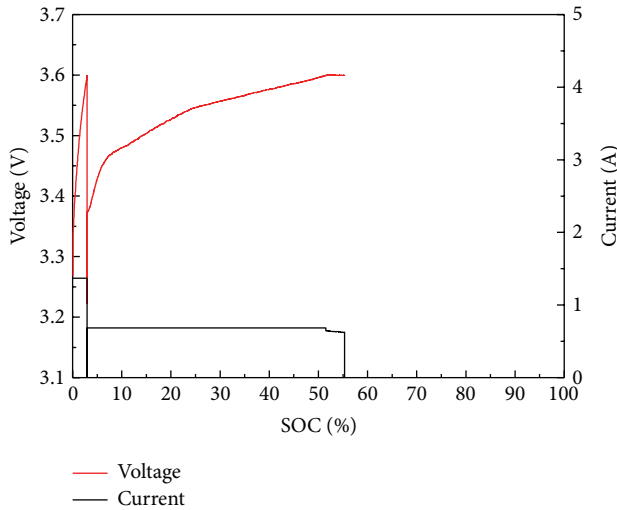FIGURE 12: Charging curves of CC-CV strategy at 0˚C.

FIGURE 15: SOC curves of different charging strategies at 0˚C.

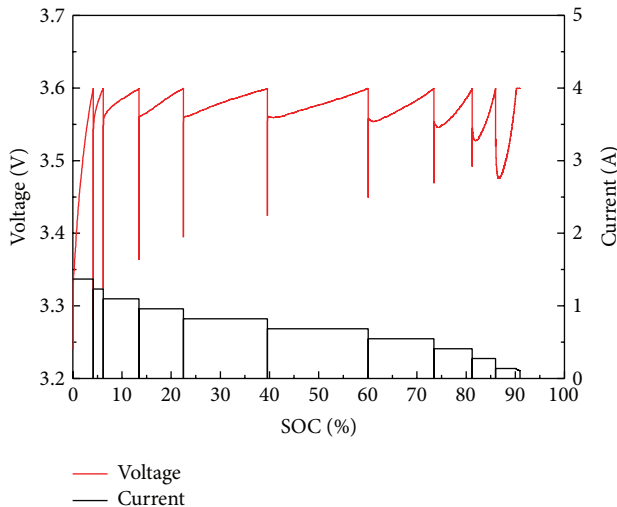FIGURE 13: Charging curves of two-stage CC-CV strategy at 0˚C.

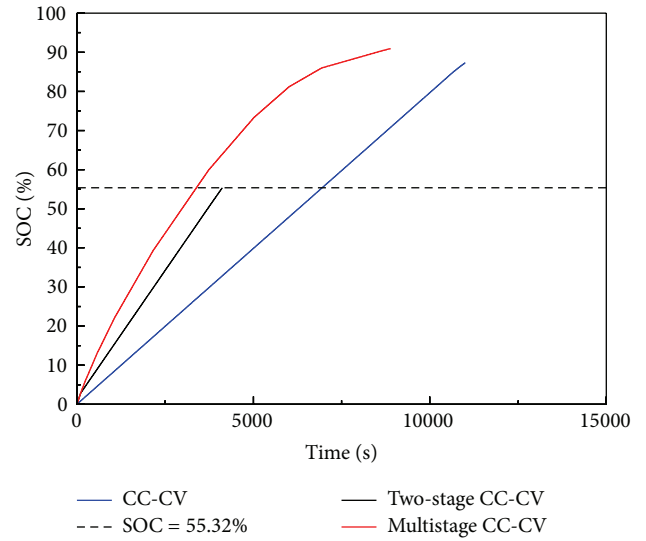FIGURE 14: Charging curves of multistage CC-CV strategy at 0˚C.

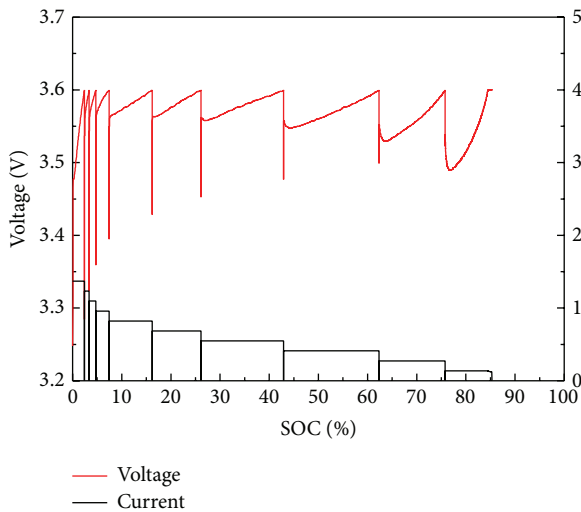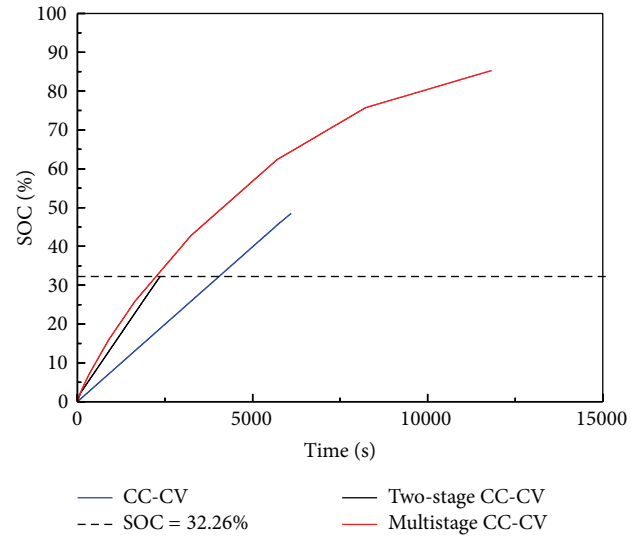are 183.4 min, 68.7 min, and 148 min. The curve tendency of multistage CC-CV charging strategy shows the obvious capacity increasing speed, although the speed is slowed down for the current decrease at later period. As the dotted line shows, the charging period of multistage CC-CV is shorter than that of two-stage CC-CV strategy with the same charging SOC of 55.32%. The multistage CC-CV still has the maximum charging capacity and minimum charging period at 0˚C.

5.3. Different Charging Strategy Analysis at −10˚C. Figures 16–18 show the terminal voltage curves with different charging strategies at −10˚C. The terminal voltage of CC-CV strategy reaches cut-off voltage at SOC point of 48.67%. The terminal voltage of the first stage of two-stage CC-CV strategy increases straightly towards the cut-off voltage and the second stage only extends the SOC to 32.26%. All the terminal voltages of charging current at 1 C–0.6 C of multistage CC-CV strategy increase rapidly to the cut-off voltage with SOC growth less than 10%. The terminal voltages of charging current at 0.5 C–0.1 C increase slower with near 75% of SOC growth. All the terminal voltage curves indicate that the voltage increases faster at the lower temperature and higher charging current rate.

The charging result at −10˚C shows that the capacities of CC-CV, two-stage CC-CV, and multistage CC-CV charging strategies are 0.664 Ah, 0.442 Ah, and 1.169 Ah, respectively. Compared with the charging result at 25˚C, the charging capacities of CC-CV, two-stage CC-CV, and multistage CC-CV charging strategies decrease by 47.08%, 62.56%, and 14.53%, respectively. It can be indicated that the charging capacity of CC-CV decreases badly and the first stage of two-stage CC-CV strategy oppositely becomes the capacity limit. The multistage CC-CV strategy can keep the charging capacity beyond 80% even at −10˚C.

Figure 19 shows SOC curves of different charging strategies at −10˚C, and the charging periods of CC-CV, two-stage
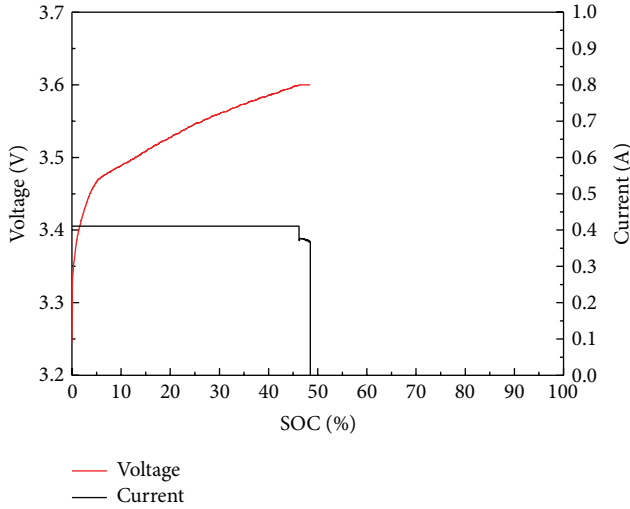
FIGURE 16: Charging curves of CC-CV strategy at −10°C.

FIGURE 19: SOC curves of different charging strategies at −10°C.
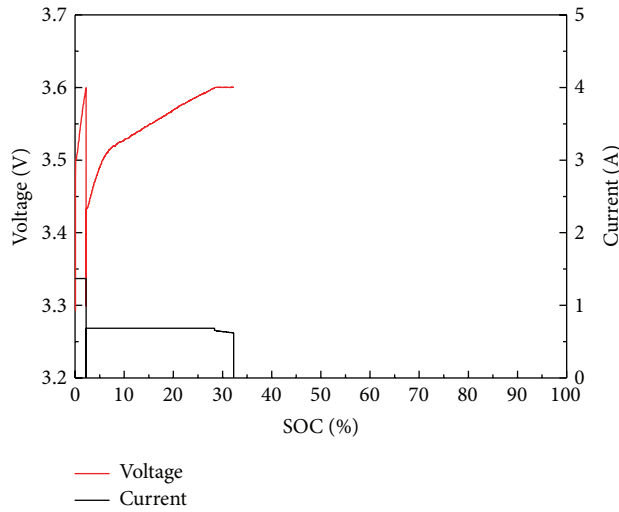
FIGURE 17: Charging curves of two-stage CC-CV strategy at −10°C.

CC-CV, and multistage CC-CV charging strategies are 101.7 min, 39.38 min, and 197.1 min, respectively. The curve of two-stage CC-CV charging strategy shows the obvious difficulty of capacity increasing at such temperature. Although the terminal voltage increasing slope of two-stage CC-CV strategy is close to that of multistage CC-CV strategy, the charging capacity is significantly different. The comparison of the SOC curves shows that the charging period of multistage CC-CV strategy is still the shortest at the same SOC point. The multistage CC-CV still has maximum charging capacity at −10°C.

5.4. Analysis of Multistage CC-CV Strategy. Figure 20 shows the capacity curves of different charging current rates of multistage CC-CV strategy at different temperature, and the high charging capacity corresponding charging current rate decreases with temperature decreasing. The charging capacity of 1 C is 1.162 Ah, beyond 80% of battery capacity, and the other charging rates only need to recover the rest of capacity at 25°C. While the high charging rate does not work well with temperature decreasing, the charging current rate with the maximum charging capacity of 0.28 Ah is 0.5 C at 0°C. The charging current rate with the maximum charging capacity of 0.266 Ah is 0.3 C at −10°C. The main capacity is charged with a range of charging current rates at low temperature. The multistage CC-CV can automatically select the optimal charging current rate for two reasons. (1) The cut-off voltage limit can stop the charging stage of the not optimal charging current rate. (2) The multistage has ten charging current rates from the maximum 1 C to the minimum 0.1 C ensuring the charging demands at different temperature points. The multistage CC-CV strategy is a wide temperature range charging strategy that keeps high charging capacity and low charging period.

6. Conclusion

It can be seen from the presentation above that the charging capacity of the CC-CV strategy can be only 48.47% of the

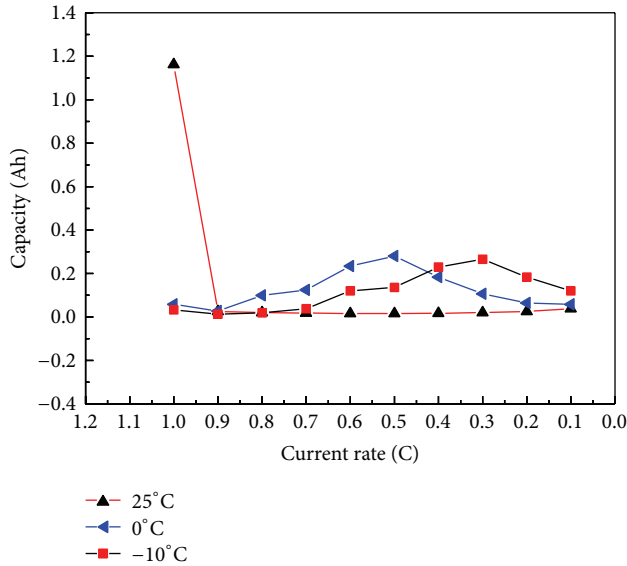FIGURE 18: Charging curves of multistage CC-CV strategy at −10°C.

FIGURE 20: Capacity curves of different charging current rates of multistage CC-CV strategy at different temperature.

normal capacity at −10°C. The charging process is analyzed by electrochemical Li-ion battery model and first-order RC equivalent circuit model. The increase in internal resistance is the main limitation of charging capacity at low temperature. The proposed multistage CC-CV strategy can extend the constant current charging process to obtain a larger capacity by decreasing the charging rate when the terminal voltage reaches the cut-off voltage. Experimental results indicate that the charging capacities with multistage CC-CV strategy at 25°C, 0°C, and −10°C are 1.368 Ah, 1.246 Ah, and 1.169 Ah, respectively. Compared with CC-CV and two-stage CC-CV strategies, the multistage CC-CV strategy has the largest charging capacities and the shortest charging periods at the target temperatures.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

References

[1] H. Zhang, X. Zhang, and J. Wang, "Robust gain-scheduling energy-to-peak control of vehicle lateral dynamics stabilization," *Vehicle System Dynamics*, no. 52, pp. 309–340, 2014.

[2] H. Zhang and J. Wang, "Vehicle lateral dynamics control through AFS/DYC and robust gain-scheduling approach," *IEEE Transactions on Vehicular Technology*, no. 19, 2015.

[3] F. Meng, H. Zhang, D. Cao, and H. Chen, "System modeling and pressure control of a clutch actuator for heavy-duty automatic transmission systems," *IEEE Transactions on Vehicular Technology*, vol. 99, pp. 1–9, 2015.

[4] D. Ansean, M. Gonzalez, M. V. Garcia, C. J. Viera, C. J. Anton, and C. Blanco, "Evaluation of LiFePO$_4$ batteries for electric vehicle applications," *IEEE Transactions on Industry Applications*, vol. 2, pp. 1855–1863, 2015.

[5] J. Fan and S. Tan, "Studies on charging lithium-ion cells at low temperatures," *Journal of the Electrochemical Society*, vol. 153, no. 6, pp. A1081–A1092, 2006.

[6] H.-S. Song, J.-B. Jeong, B.-H. Lee et al., "Experimental study on the effects of pre-heating a battery in a low-temperature environment," in *Proceedings of the IEEE Vehicle Power and Propulsion Conference (VPPC '12)*, pp. 1198–1201, IEEE, Seoul, Republic of Korea, October 2012.

[7] L. Liao, P. Zuo, Y. Ma et al., "Effects of temperature on charge/discharge behaviors of LiFePO$_4$ cathode for Li-ion batteries," *Electrochimica Acta*, vol. 60, pp. 269–273, 2012.

[8] T. Ikeya, N. Sawada, J.-I. Murakami et al., "Multi-step constant-current charging method for an electric vehicle nickel/metal hydride battery with high-energy efficiency and long cycle life," *Journal of Power Sources*, vol. 105, no. 1, pp. 6–12, 2002.

[9] Y. Gao, C. Zhang, Q. Liu, Y. Jiang, W. Ma, and Y. Mu, "An optimal charging strategy of lithium-ion batteries based on polarization and temperature rise," in *Proceedings of the IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific '14)*, pp. 1–6, IEEE, Beijing, China, September 2014.

[10] L.-R. Chen, "Design of duty-varied voltage pulse charger for improving Li-ion battery-charging response," *IEEE Transactions on Industrial Electronics*, vol. 56, no. 2, pp. 480–487, 2009.

[11] S. J. Huang, B. G. Huang, and F. S. Pai, "Fast charge strategy based on the characterization and evaluation of LiFePO$_4$ batteries," *IEEE Transactions on Power Electronics*, vol. 28, no. 4, pp. 1555–1562, 2013.

[12] Y.-H. Liu, J.-H. Teng, and Y.-C. Lin, "Search for an optimal rapid charging pattern for lithium-ion batteries using Ant Colony System algorithm," *IEEE Transactions on Industrial Electronics*, vol. 52, no. 5, pp. 1328–1336, 2005.

[13] Y.-H. Liu and Y.-F. Luo, "Search for an optimal rapid-charging pattern for Li-ion batteries using the Taguchi approach," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 12, pp. 3963–3971, 2010.

[14] J. Jiang, C. Zhang, J. Wen, W. Zhang, and S. M. Sharkh, "An optimal charging method for Li-ion batteries using a fuzzy-control approach based on polarization properties," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3000–3009, 2013.

[15] H. J. Ruan, J. C. Jiang, B. X. Sun, N. N. Wu, W. Shi, and Y. R. Zhang, "Stepwise segmented charging technique for lithium-ion battery to induce thermal management by low-temperature internal heating," in *Proceedings of the IEEE Transportation Electrification Conference and Expo (ITEC '14)*, Beijing, China, September 2014.

[16] X. W. Zhao, G. Y. Zhang, L. Yang, J. X. Qiang, and Z. Q. Chen, "A new charging mode of Li-ion batteries with LiFePO$_4$/C composites under low temperature," *Journal of Thermal Analysis and Calorimetry*, vol. 104, no. 2, pp. 561–567, 2011.

[17] D. Anseán, M. González, J. C. Viera, V. M. García, C. Blanco, and M. Valledor, "Fast charging technique for high power lithium iron phosphate batteries: a cycle life analysis," *Journal of Power Sources*, vol. 239, pp. 9–15, 2013.

[18] G. Hunt and C. Motloch, *Freedom Car Battery Test Manual for Power-Assist Hybrid Electric Vehicles*, Idaho National Engineering and Environmental Laboratory (INEEL), Idaho Falls, Idaho, USA, 2003.

[19] M. Doyle, J. Newman, A. S. Gozdz, C. N. Schmutz, and J.-M. Tarascon, "Comparison of modeling predictions with experimental data from plastic lithium ion cells," *Journal of the Electrochemical Society*, vol. 143, no. 6, pp. 1890–1903, 1996.

[20] M. Xu, Z. Q. Zhang, X. Wang, L. Jia, and L. X. Yang, "A pseudo three-dimensional electrochemical-thermal model of a prismatic LiFePO$_4$ battery during discharge process," *Energy*, vol. 80, pp. 303–317, 2015.

[21] B. Wu, V. Yufit, M. Marinescu, G. J. Offer, R. F. Martinez-Botas, and N. P. Brandon, "Coupled thermal-electrochemical modelling of uneven heat generation in lithium-ion battery packs," *Journal of Power Sources*, vol. 243, pp. 544–554, 2013.

[22] J. Kim, S. Lee, and B. H. Cho, "Complementary cooperation algorithm based on DEKF combined with pattern recognition for SOC/capacity estimation and SOH prediction," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 436–451, 2012.

[23] J. Kim, J. Shin, C. Chun, and B. H. Cho, "Stable configuration of a Li-ion series battery pack based on a screening process for improved voltage/SOC balancing," *IEEE Transactions on Power Electronics*, vol. 27, no. 1, pp. 411–424, 2012.

Research Article

Ultra-High-Speed Travelling Wave Protection of Transmission Line Using Polarity Comparison Principle Based on Empirical Mode Decomposition

Dong Wang,[1,2] Houlei Gao,[1,2] Sibei Luo,[1,2] and Guibin Zou[1,2]

[1]*Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education, Shandong University, Jinan 250061, China*
[2]*School of Electrical Engineering, Shandong University, Jinan 250061, China*

Correspondence should be addressed to Houlei Gao; houleig@sdu.edu.cn

The traditional polarity comparison based travelling wave protection, using the initial wave information, is affected by initial fault angle, bus structure, and external fault. And the relationship between the magnitude and polarity of travelling wave is ignored. Because of the protection tripping and malfunction, the further application of this protection principle is affected. Therefore, this paper presents an ultra-high-speed travelling wave protection using integral based polarity comparison principle. After empirical mode decomposition of the original travelling wave, the first-order intrinsic mode function is used as protection object. Based on the relationship between the magnitude and polarity of travelling wave, this paper demonstrates the feasibility of using travelling wave magnitude which contains polar information as direction criterion. And the paper integrates the direction criterion in a period after fault to avoid wave head detection failure. Through PSCAD simulation with the typical 500 kV transmission system, the reliability and sensitivity of travelling wave protection were verified under different factors' affection.

1. Introduction

According to the protection principle, travelling wave protection methods include travelling wave differential protection, travelling wave distance protection, travelling wave amplitude comparison protection, and travelling wave polarity comparison protection [1, 2].

Travelling wave differential protection principle is simple and clear. But travelling wave has attenuation characteristic. There may be large unbalance current in the transmission line to cause wrong operation. And it is also affected by the bus structure [3–5]. Travelling wave distance protection cannot protect the whole line and does not have direction discrimination ability. Same with differential protection, it is affected by bus structure, too [6–9]. Travelling wave amplitude comparison protection principle has improved much compared to other protection principles. But it is affected by the bus structures, fault inception angles, and different thresholds, too. The traditional travelling wave polarity comparison protection principle has a lot of advantages: high

operation speed, clear direction discrimination, and simple protection principle. But it is also affected by some things: fault initial angles, different bus structures, different fault locations, and even threshold. If those disadvantages can be overcome, a new travelling wave protection principle can be got [10–12].

The traditional travelling wave protection principle's application is limited by the transformer technology. The traditional current transformer (CT) and voltage transformer (VT) cannot transfer the travelling wave signal correctly. Currently, Rogowski based electronic current transformer (R-ECT) and capacitive divider electronic voltage transformer (C-EVT) have been able to transfer current travelling wave and voltage travelling wave accurately. And the output of C-EVT and R-ECT is the differential signal of the input. By integration circuit, the original signal can be regained exactly. So there is no transformer technology limit in the travelling wave protection principle anymore [13–16].

This paper compares the polarity relationship between voltage travelling wave and current travelling wave with

TABLE 1: Analysis of travelling wave polarity comparison protection.

Fault location	Superimposed voltage polarity	M side		N side		Schematic of fault superimposed state circuit
		Voltage	Current	Voltage	Current	
Internal fault	+ −	+ −	− +	+ −	− +	
External fault	+ −	+ −	− +	+ −	+ −	
External fault	+ −	+ −	+ −	+ −	− +	

different fault directions. Combined with empirical mode decomposition algorithm (EMD), the new travelling wave polarity comparison protection principle based on the integration of amplitude is derived. It not only uses the initial travelling wave but also uses the travelling wave after fault happens. So it is a reliable protection principle with obvious direction discrimination. By the way, this new protection principle is not affected by different initial angles, different grounding resistance, different bus structures, and different fault locations. To verify the characteristics of the new principle, a simulation based on PSCAD/EMTDC is carried on. And the PSCAD simulation proved that this new protection principle has the characteristics mentioned above indeed.

The high operation speed is a very important advantage for travelling wave protection. Considering the different parts of the new travelling wave's operation time, the new protection principle can determine if the fault is internal or external in 5 ms. Then it can send the signal to breaker to operate. So, it can be called an ultra-high-speed travelling wave protection.

2. Traditional Travelling Wave Polarity Comparison Protection

The direction element of the protection principle is a polarity comparison relay. It detects the initial voltage travelling wave and current travelling wave as comparison objects. When the voltage travelling wave and current travelling wave have opposite polarities of both sides, the internal fault can be determined. When the voltage travelling wave and current travelling wave have same polarities of any side, the external fault can be determined. The schematic of protection is shown in Table 1. (Superimposed voltage in the table appears at the moment that fault happens. It has same value and opposite polarity with the voltage on the transmission line just before the fault moment. And it is the voltage source in the circuit of the table indeed.)

3. Empirical Mode Decomposition

The empirical mode decomposition algorithm can distinguish the different scale fluctuations or trends in the signal gradually. And the result of EMD is a series of different characteristic scales data called intrinsic mode functions (IMF) [17–19].

The result of EMD can be described as

$$S(t) = \sum_{i=1}^{N} \text{IMF}_i(t) + R(t), \tag{1}$$

where $S(t)$ is the EMD result, $\text{IMF}_i(t)$ is i order intrinsic mode function, and $R(t)$ is trends signal.

Intrinsic mode function is a single component signal and it must meet the following two conditions: (1) difference between the number of extreme points and zero crossing points is not more than one over the entire length of the signal; (2) the envelope of IMF is symmetry about time axis.

The processes of EMD are described as the following steps.

(1) Find all the maxima of the original signal $S(t)$. Then the maxima envelope $E_+(t)$ can be calculated based on cubic spline interpolation. Similarly, the minimum envelope $E_-(t)$ can also be calculated. Then, the average envelope $M(t)$ can be defined as

$$M(t) = \frac{E_+(t) + E_-(t)}{2}. \tag{2}$$

(2) Let $S(t)$ be minus $M(t)$ to get a new signal $H_1^1(t)$:

$$H_1^1(t) = S(t) - M(t). \tag{3}$$

Then check if the following condition can be met:

$$\frac{\sum \left[H_1^k(t) - H_1^{k-1}(t)\right]^2}{\sum \left[H_1^{k-1}(t)\right]^2} \leq \varepsilon, \tag{4}$$

where k is the cycle number and the value of ε is between 0.2 and 0.3. This paper selects 0.3.

If (4) cannot be met, return to step (1).

If (4) can be met after k cycles, then $\text{IMF}_1(t)$ can be defined as

$$\text{IMF}_1(t) = H_1^k(t). \tag{5}$$

(3) Let original signal be minus $\text{IMF}_1(t)$ to get a residual signal $R(t)$ as

$$R(t) = S(t) - \text{IMF}_1(t). \tag{6}$$

Repeat steps (1) to (3) to get another intrinsic mode function $\text{IMF}_2(t)$. Repeat steps (1) to (3) until residual signal $R(t)$ is small enough or monotonic function.

Based on EMD, the first IMF of voltage travelling wave and current travelling wave can be calculated as Figure 1. The first figure and second figure are the voltage and current travelling wave before EMD, respectively. The third figure and fourth figure are the voltage and current travelling wave after EMD, respectively. As we can see, the similarity of the voltage and current travelling wave using EMD is more obvious than before. So it is more convenient to construct a protection principle using intrinsic mode function.

4. New Travelling Wave Polarity Comparison Protection

4.1. Derivation of Direction Criterion. As shown in Figure 2, fault component voltage appears between the fault location F in transmission line and the earth at time $t = 0$. Then, the transmission line will be charged. After a short time Δt, a short length of transmission line Δx is charged to $\Delta Q = Cu_0\Delta x$ (C is the capacitance value per unit length of transmission line). An electrical field E will appear surrounding this short transmission line. And the flow of current will form a magnetic field around the line. If Δx is small enough, the current i_0 can be described as

$$i_0 = \lim_{\Delta x \to 0} \frac{\Delta Q}{\Delta t} = \lim_{\Delta x \to 0} \frac{Cu_0\Delta x}{\Delta t} = Cu_0v, \tag{7}$$

where v is the speed of the wave and C is the capacitance value per unit length of transmission line.

Now the magnetic flux around Δx is $\Delta\Phi = Li\Delta x$. According to the law of electromagnetic induction, the electromotive force is described as

$$E = \lim_{\Delta x \to 0} \frac{\Delta\Phi}{\Delta t} = \lim_{\Delta x \to 0} \frac{LCu_0v\Delta x}{\Delta t} = LCu_0v^2, \tag{8}$$

where v is the speed of the wave, C is the capacitance value per unit length of transmission line, and L is the inductance value per unit length of transmission line.

Because the voltage on capacitance cannot change suddenly and Δx is small enough, E equals voltage u_0. Then the wave speed will be
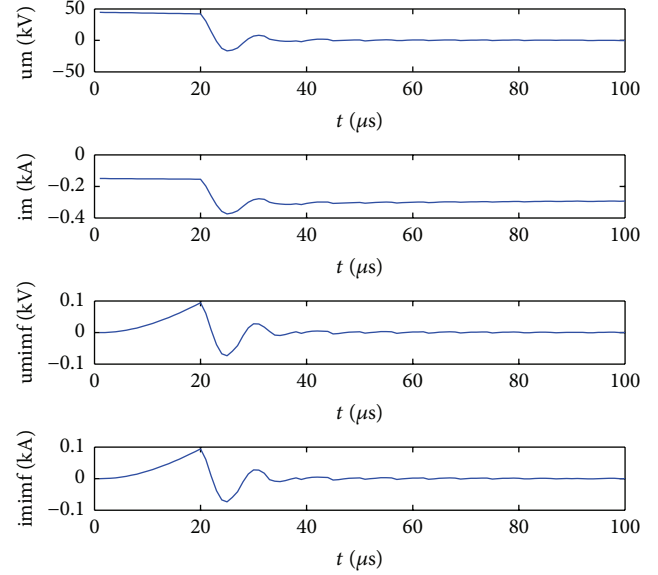
$$v = \frac{1}{\sqrt{LC}}. \tag{9}$$

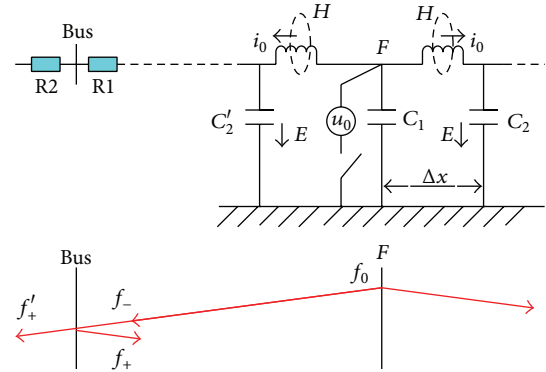FIGURE 1: The comparison of travelling wave and travelling wave after EMD.

FIGURE 2: Equivalent circuit of single phase line.

Take (9) to (7) to get

$$\frac{u_0}{i_0} = \sqrt{\frac{L}{C}}. \tag{10}$$

As we can see, the ratio of voltage and current is a constant value called wave impedance.

Define voltage amplitude conditioning factor k:

$$k = \frac{|i|}{|u|}. \tag{11}$$

And the value of k is approximately equal to $\sqrt{C/L}$.

Considering the initial polarity of voltage and current travelling wave, define a factor λ to identify the fault direction:

$$\lambda = \frac{\int_{t=t_F}^{t_F+d} ku(t)i(t)\,dt}{\int_{t=t_F}^{t_F+d} i^2(t)\,dt}, \tag{12}$$

where t_F is the arrival point of the travelling wave, d is the length of the integration time, and k is the voltage amplitude conditioning factor defined above.

The discretization of (12) can be described as

$$\lambda = \frac{\sum_{t=t_F}^{t_F+d} k u(t) i(t)}{\sum_{t=t_F}^{t_F+d} i^2(t)}. \tag{13}$$

Considering different fault directions, the value of λ can be calculated.

(1) If the fault happens as Figure 2 shows, it will be forward fault type for R1. Assume the transmission line is lossless. When the travelling wave arrived, the travelling wave signal of R1 will be

$$u = u_+ + u_- = \left(1 + k_{uf}\right) u_-,$$
$$i = i_+ + i_- = \left(1 + k_{if}\right) i_-, \tag{14}$$

where u is the voltage travelling wave, i is the current travelling wave, u_+ is the forward voltage travelling wave, u_- is the reverse voltage travelling wave, i_+ is the forward current travelling wave, i_- is the reverse current travelling wave, k_{uf} is the voltage reflection coefficient at the bus, and k_{if} is the current reflection coefficient at the bus. And the inequality relationship will be $0 \le |k_{uf}|, |k_{if}| \le 1$.

Now the amplitude conditioning factor can be calculated:

$$k = \frac{|i(t)|}{|u(t)|} = \frac{\left(1 + k_{if}\right) |i_-(t)|}{\left(1 + k_{uf}\right) |u_-(t)|}. \tag{15}$$

Because it is a forward direction fault for R1, the reverse voltage and current travelling wave have different polarities. So (15) can be simplified as

$$k = -\frac{\left(1 + k_{if}\right)}{\left(1 + k_{uf}\right)} \frac{i_-(t)}{u_-(t)}. \tag{16}$$

Take (14) and (16) to (13):

$$\lambda = \frac{\sum_{t=t_F}^{t_F+d} k u(t) i(t)}{\sum_{t=t_F}^{t_F+d} i^2(t)} = -1. \tag{17}$$

As we can see, the value of λ is a constant number -1 when it is a forward direction fault. And it is not affected by the reflection coefficient and the construction of the bus.

(2) If the fault happens as Figure 2 shows, it will be reverse fault type for R2. Assume the transmission line is lossless. When the travelling wave arrived, the travelling wave signal of R2 will be

$$u = u_+' = k_{uz} u_-,$$
$$i = i_+' = k_{iz} i_-, \tag{18}$$

where u is the voltage travelling wave, i is the current travelling wave, u_+' is the forward voltage travelling wave at R2, i_+' is the forward current travelling wave at R2, u_- is the
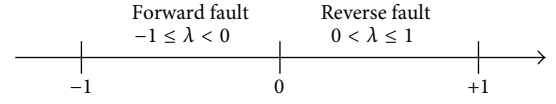

FIGURE 3: Fault direction discrimination schematic.

reverse voltage travelling wave at R1, i_- is the reverse current travelling wave at R1, k_{uz} is the voltage refractive coefficient at bus, and k_{iz} is the current refractive coefficient at bus. And there is an inequality relationship $k_{uz}, k_{iz} \ge 0$.

Now the amplitude conditioning factor can be calculated:

$$k = \frac{|i(t)|}{|u(t)|} = \frac{\left|k_{iz} i_-(t)\right|}{\left|k_{uz} u_-(t)\right|}. \tag{19}$$

Because it is a reverse direction fault for R2 and also a forward direction fault for R1, the reverse voltage and current travelling wave have different polarities. And the forward direction of R2 is opposite to R1. So (19) can be simplified as

$$k = \frac{|i(t)|}{|u(t)|} = \frac{k_{iz}}{k_{uz}} \frac{i_-(t)}{u_-(t)}. \tag{20}$$

Take (18) and (20) to (13):

$$\lambda = \frac{\sum_{t=t_F}^{t_F+d} k u(t) i(t)}{\sum_{t=t_F}^{t_F+d} i^2(t)} = 1. \tag{21}$$

As we can see, the value of λ is a constant number 1 when it is a reverse direction fault. And it is not affected by the reflection coefficient and the construction of the bus.

Taking a variety of errors in the actual system into account, the fault direction discrimination schematic is shown in Figure 3. When forward fault happens, the value of λ is less than zero. When reverse fault happens, the value of λ is greater than zero. If two fault direction discrimination results of both ends are forward fault, an internal fault can be determined. If one of the fault direction discrimination results of both ends is reverse fault, an external fault can be determined.

4.2. Protection Scheme. The protection scheme is shown in Figure 4. First of all, the three-phase voltage and current should be decoupled using Clark transformation. Then amplitude conditioning factor, defined above, can be calculated point by point in d length of time. After that, the value of λ can be calculated. Because the other end of the transmission line needs the value of λ to identify the fault section, the value of λ should send to another end though fiber path. Then the value of factor λ can be checked to identify the external fault type. If it is a forward fault for the relay, the value of factor λ' from another end will be received and checked to identify the external fault. If λ and λ' are both less than zero, an internal fault can be determined. At last, the breaker will clear the transmission line fault.

5. Simulation Analysis

5.1. Simulation Model in PSCAD. The 500 kV power transmission system is constructed in PSCAD/EMTDC as shown
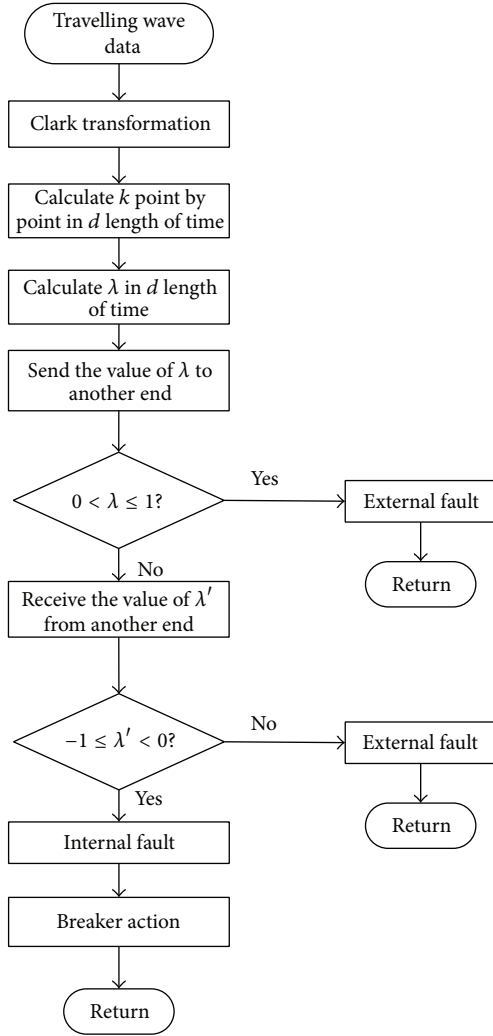
FIGURE 4: Flow chart of travelling wave protection.

TABLE 2: Transmission line parameters.

	R (Ω/km)	X (Ω/km)	G (S/km)	B (S/km)
Positive sequence	0.01798	0.29278	1×10^8	3.93905×10^{-6}
Negative sequence	0.01798	0.29278	1×10^8	3.93905×10^{-6}
Zero sequences	0.28662	1.08210	1×10^8	2.43767×10^{-6}

in Figure 5. The system includes three transmission lines whose lengths are 100 km, 200 km, and 100 km, respectively. R1 and R2 are two relays on the middle line. Now the new travelling wave polarity comparison protection can be studied by different fault locations and different fault types.

The transmission line uses frequency-dependent model and has uniform transposition. The transmission line parameters for per km length are shown in Table 2. The bus stray capacitance to ground is set to $C_S = 0.01\,\mu$F. Taking the past studies into account, the sampling rate is set to 1 MHz. The integration time d is 0.1 ms.

5.2. Typical Fault Examples. In order to verify the protection principle's operating characteristics, A phase to ground fault
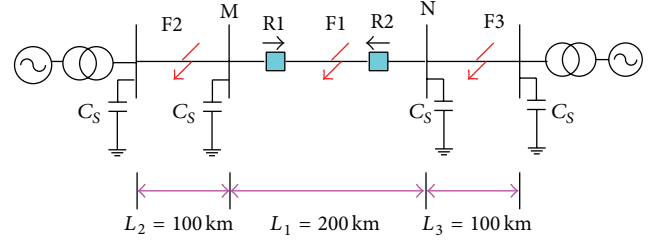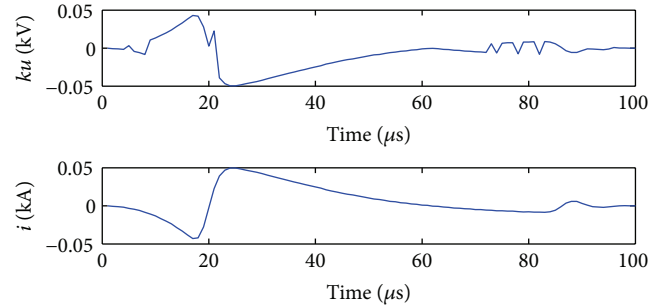

FIGURE 5: Model of 500 kV power transmission system.

FIGURE 6: Comparison chart of ku and i of M side.

is set located at F3. The initial fault angle is 45° and the ground resistance is 50 Ω. Using empirical mode decomposition algorithm, the first-order intrinsic mode function of voltage and current travelling wave of both sides can be calculated.

Take ku and i data in Figure 6 to (13) to calculate the $\lambda = -0.8710$. Then the forward fault of R1 can be determined.

Take ku and i data in Figure 7 to (13) to calculate the $\lambda = 0.9811$. Then the reverse fault of R2 can be determined.

As we can see, the fault discrimination results of R1 and R2 are corrected. Taking the protection scheme of Figure 4 into account, an external fault type can be determined.
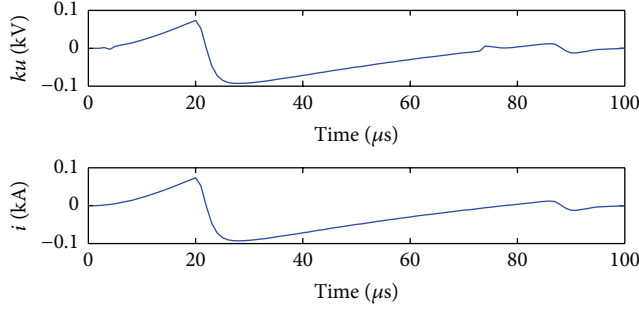
5.3. Relater Factors

5.3.1. Different Fault Location. Based on some different fault locations at F1 and F2, the fault discrimination factor λ of both M and N side is calculated.

Table 3 is the simulation results for different fault locations. The fault distance in the table is from the bus of M side to fault location. As can be seen, the fault principle based on EMD can identify fault direction correctly. Even at the beginning or end of the transmission line, it can still identify fault direction correctly.

5.3.2. Grounding Resistance. Based on some different grounding resistance at F1 (100 km away from the bus of M side) and F2 (10 km away from the bus of M side), the fault discrimination factor λ of both M and N side is calculated.

Table 4 is the simulation results for different grounding resistance. As can be seen, the fault principle based on EMD can identify fault direction correctly. With the increasing of grounding resistance, protection's sensitivity will not change.

FIGURE 7: Comparison chart of ku and i of N side.

FIGURE 8: Model of 500 kV power transmission system with different bus structures.

FIGURE 9: Schematic diagram of detection time.

5.3.3. Fault Initial Angle. Based on some different fault initial angel at F1 (100 km away from the bus of M side) and F2 (10 km away from the bus of M side), the fault discrimination factor λ of both M and N side is calculated.

Table 5 is the simulation results for different fault initial angles. As can be seen, the fault principle based on EMD can identify fault direction correctly. Even with small fault angles, it can still identify fault direction correctly. And, with the decreasing of fault initial angle, protection's sensitivity will reduce slowly.

5.3.4. Different Fault Types. Based on some different fault types at F1 (100 km away from the bus of M side) and F2 (10 km away from the bus of M side), the fault discrimination factor λ of both M and N side is calculated.

Table 6 is the simulation results for different fault types. As can be seen, the fault principle based on EMD can identify fault direction correctly.

5.3.5. Sampling Rate. Based on some different sampling rate and AG fault at F1 (100 km away from the bus of M side) and F2 (10 km away from the bus of M side), the fault discrimination factor λ of both M and N side is calculated.

Table 7 is the simulation results for different sampling rate. As can be seen, the fault principle based on EMD can identify fault direction correctly with the change of sampling rate.

5.3.6. Bus Structure. Traditional travelling wave protection principle is affected by the number of transmission lines connected to the bus. To verify the new EMD based protection principle, a new power transmission system is constructed in PSCAD as Figure 8.

Table 8 is the simulation results for different fault locations. As can be seen, the fault principle based on EMD can identify fault direction correctly with different bus structure.

6. Operation Time of Protection

The operation time of the travelling wave protection using polarity comparison principle based on EMD includes three parts: algorithm time, detection time, and propagation time. This new travelling wave protection principle can determine if the fault is inside or outside of the protection region in
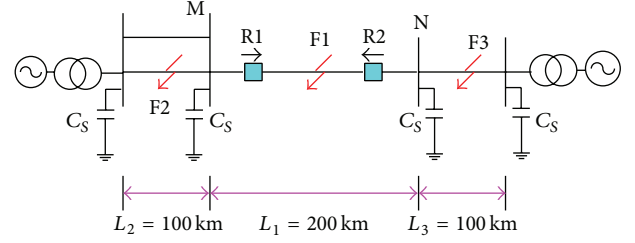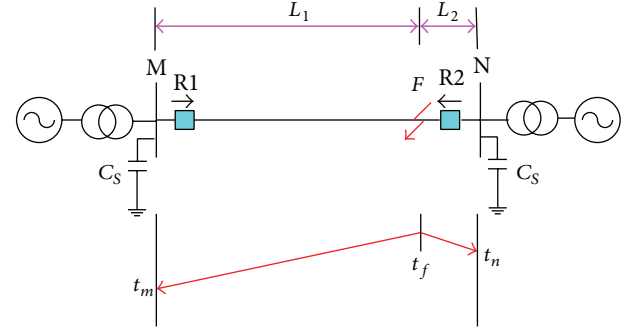
5 ms. Then it can send the signal to breaker to operate. So it can be called ultra-high-speed travelling wave protection. The following is the introduction and analysis of the three parts.

6.1. Algorithm Time. Algorithm time includes two parts: the integration time and calculation time of the principle. In this paper, integration time length (the factor d in (12) and (13)) is 0.1 ms. Considering the computing power of the protection unit now, the calculation time of algorithm is not longer than 0.5 ms. So, the algorithm time is not longer than 1 ms.

6.2. Detection Time. Detection time is the time difference of two sides' travelling wave arrival point. As we can see, the fault may happen everywhere in the transmission line. Then the arrival times of two sides are different, except that the fault happens in the middle of the line. As a protection principle which needs two sides' information to decide the operation of breaker, the time difference will delay the operation time. As shown in Figure 9, t_f, t_m, and t_n are the fault time, M side's arrival time, and N side's arrival time, respectively. And the time difference can be described as

$$\Delta t = t_m - t_n = \frac{L_1 - L_2}{v}. \tag{22}$$

And v is the travelling wave speed.

Because the transmission line is generally several hundred kilometers, this time is obviously not longer than 2 ms.

6.3. Propagation Time. After the direction discrimination of one side, as shown in Figure 10, the value of λ should transfer to another side. Propagation time is the time from one side to another side. As the length of transmission line is generally

Table 3: Simulation results for different fault locations.

Fault location	Fault distance/km	M side		N side		Results
		λ	Direction	λ	Direction	
F1	10	−0.8160	Forward	−0.5732	Forward	Internal
	100	−0.8763	Forward	−0.4071	Forward	Internal
	190	−0.9378	Forward	−0.7965	Forward	Internal
F2	10	0.9972	Reverse	−0.9603	Forward	External
	50	0.9980	Reverse	−0.3843	Forward	External
	90	0.9972	Reverse	−0.9603	Forward	External

Table 4: Simulation results for different grounding resistance.

Fault location	Grounding resistance/Ω	M side		N side		Results
		λ	Direction	λ	Direction	
F1	1	−0.4944	Forward	−0.8818	Forward	Internal
	100	−0.4829	Forward	−0.8802	Forward	Internal
	300	−0.4984	Forward	−0.8839	Forward	Internal
F2	1	0.9978	Reverse	−0.9438	Forward	External
	100	0.9963	Reverse	−0.8730	Forward	External
	300	0.8522	Reverse	−0.9033	Forward	External

Table 5: Simulation results for different fault angles.

Fault location	Initial angle/°	M side		N side		Results
		λ	Direction	λ	Direction	
F1	1	−0.3884	Forward	−0.3256	Forward	Internal
	45	−0.4291	Forward	−0.8732	Forward	Internal
	90	−0.4944	Forward	−0.8818	Forward	Internal
F2	1	0.9972	Reverse	−0.2356	Forward	External
	45	0.9997	Reverse	−0.9687	Forward	External
	90	0.9997	Reverse	−0.9700	Forward	External

Table 6: Simulation results for different fault types.

Fault location	Fault type	M side		N side		Results
		λ	Direction	λ	Direction	
F1	AG	−0.4944	Forward	−0.8818	Forward	Internal
	AC	−0.9600	Forward	−0.9602	Forward	Internal
	ABG	−0.4471	Forward	−0.8804	Forward	Internal
	ABCG	−0.9664	Forward	−0.7452	Forward	Internal
F2	AG	0.9972	Reverse	−0.9700	Forward	External
	AC	0.9991	Reverse	−0.6022	Forward	External
	ABG	0.8275	Reverse	−0.4659	Forward	External
	ABCG	0.9997	Reverse	−0.9586	Forward	External

Table 7: Simulation results for different sampling rate.

Fault location	Sampling rate/Hz	M side		N side		Results
		λ	Direction	λ	Direction	
F1	100 k	−0.9794	Forward	−0.6816	Forward	Internal
	500 k	−0.7345	Forward	−0.9637	Forward	Internal
	1 M	−0.5018	Forward	−0.8825	Forward	Internal
F2	100 k	0.9751	Reverse	−0.9807	Forward	External
	500 k	0.9994	Reverse	−0.9055	Forward	External
	1 M	0.9979	Reverse	−0.9684	Forward	External

TABLE 8: Simulation results for different fault locations.

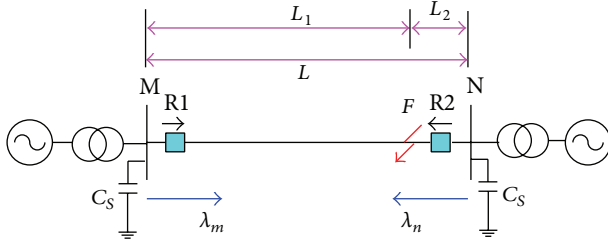| Fault location	M side		N side		Results
	λ	Direction	λ	Direction	
F1	−0.9731	Forward	−0.9603	Forward	Internal
F2	0.9980	Reverse	−0.9779	Forward	External
F3	−0.9718	Forward	0.9992	Reverse	External

FIGURE 10: Schematic diagram of propagation time.

several hundred kilometers, propagation time is no longer than 2 ms.

7. Conclusion

Comparing with the traditional polarity comparison travelling wave protection, the new travelling wave protection combines the relationship between amplitude and polarity. Based on empirical mode decomposition, the derivation of the direction criterion is finished. And this protection criterion not only uses the initial travelling wave front but also uses short time's (0.1 ms in the paper) travelling wave information after the initial travelling wave front. Through the integration of travelling wave, it can avoid the failure of the travelling wave's detection. So it can increase the reliability of the protection principle.

To verify the new protection principle, a simulation based on PSCAD is carried on. Taking the simulation results into account, this new protection principle is not affected by different fault locations, different fault types, different initial angels, different grounding resistance, and different bus structures. So it is a reliable travelling wave protection.

Operation speed is an important advantage for travelling wave protection. Because the new protection principle can send the operation signal to breaker in 5 ms, it can be called ultra-high-speed travelling wave protection.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the republication of this paper.

Acknowledgment

References

[1] G. Zou and H. Gao, "A traveling-wave-based amplitude integral busbar protection technique," *IEEE Transactions on Power Delivery*, vol. 27, no. 2, pp. 602–609, 2012.

[2] W. Chen, O. P. Malik, X. Yin, D. Chen, and Z. Zhang, "Study of wavelet-based ultra high speed directional transmission line protection," *IEEE Transactions on Power Delivery*, vol. 18, no. 4, pp. 1134–1139, 2003.

[3] G. Zou, S. Song, C. Xu, D. Liu, and H. Gao, "Fast busbar protection based on waveform integral of directional traveling waves," *Proceedings of the Chinese Society of Electrical Engineering*, vol. 34, no. 31, pp. 5677–5684, 2014.

[4] Z. Li and H. Hua, "Traveling wave protection based on wide area travelling wave information," *Proceedings of the CSEE*, vol. 34, pp. 6238–6245, 2014.

[5] J.-D. Duan, B.-H. Zhang, S.-B. Luo, and Y. Zhou, "Transient-based ultra-high-speed directional protection using wavelet transforms for EHV transmission lines," in *Proceedings of the IEEE/PES Transmission and Distribution Conference and Exhibition*, Dalian, China, August 2005.

[6] Q. H. Wu, J. F. Zhang, and D. J. Zhang, "Ultra-high-speed directional protection of transmission lines using mathematical morphology," *IEEE Transactions on Power Delivery*, vol. 18, no. 4, pp. 1127–1133, 2003.

[7] G. Zou, *Study on internal based travelling wave directional protection for transmission line [Ph.D. thesis]*, Shandong University, Jinan, China, 2009.

[8] M. Öhrström and L. Söder, "Fast protection of strong power systems with fault current limiters and PLL-aided fault detection," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1538–1544, 2011.

[9] G. B. Zou and H. L. Gao, "Extra high speed hybrid protection scheme for high voltage transmission line," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 83–90, 2014.

[10] H. Ha and Y. Yu, "Novel scheme of travelling wave based differential protection for bipolar HVDC transmission lines," in *Proceedings of the International Conference on Power System Technology (POWERCON '10)*, pp. 1–6, IEEE, Hangzhou, China, October 2010.

[11] J. D. Duan and B. H. Zhang, "Study of starting algorithm using travelling waves," *Proceedings of the Chinese Society for Electrical Engineering*, vol. 24, no. 9, pp. 30–36, 2004.

[12] G. B. Zou and H. L. Gao, "Algorithm for ultra high speed travelling wave protection with accurate fault location," in *Proceedings of the IEEE Power and Energy Society General Meeting*, pp. 20–24, Pittsburgh, Pa, USA, July 2008.

[13] M. Marracci, B. Tellini, C. Zappacosta, and G. Robles, "Critical parameters for mutual inductance between rogowski coil and primary conductor," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 2, pp. 625–632, 2011.

[14] W. B. Li, C. X. Mao, and J. M. Lu, "Study of Rogowski coils for measuring pulse currents of the high-power laser source," *International Journal of Power and Energy Systems*, vol. 2, pp. 96–101, 2005.

[15] Y. Liu, F. C. Lin, Q. Zhang, and H. Zhong, "Design and construction of a Rogowski Coil for measuring wide pulsed current," *IEEE Sensors Journal*, vol. 11, no. 1, pp. 123–130, 2011.

[16] H. Gao and R. Yang, "Analysis and test for electronic voltage transducer's transfer characteristics," *Journal of Beijing Jiaotong University*, vol. 38, no. 5, pp. 114–118, 2014.

[17] N. Huang and M. Wu, "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 459, no. 2037, pp. 2317–2345, 2003.

[18] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *The Royal Society of London. Proceedings—Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[19] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 460, no. 2046, pp. 1597–1611, 2004.