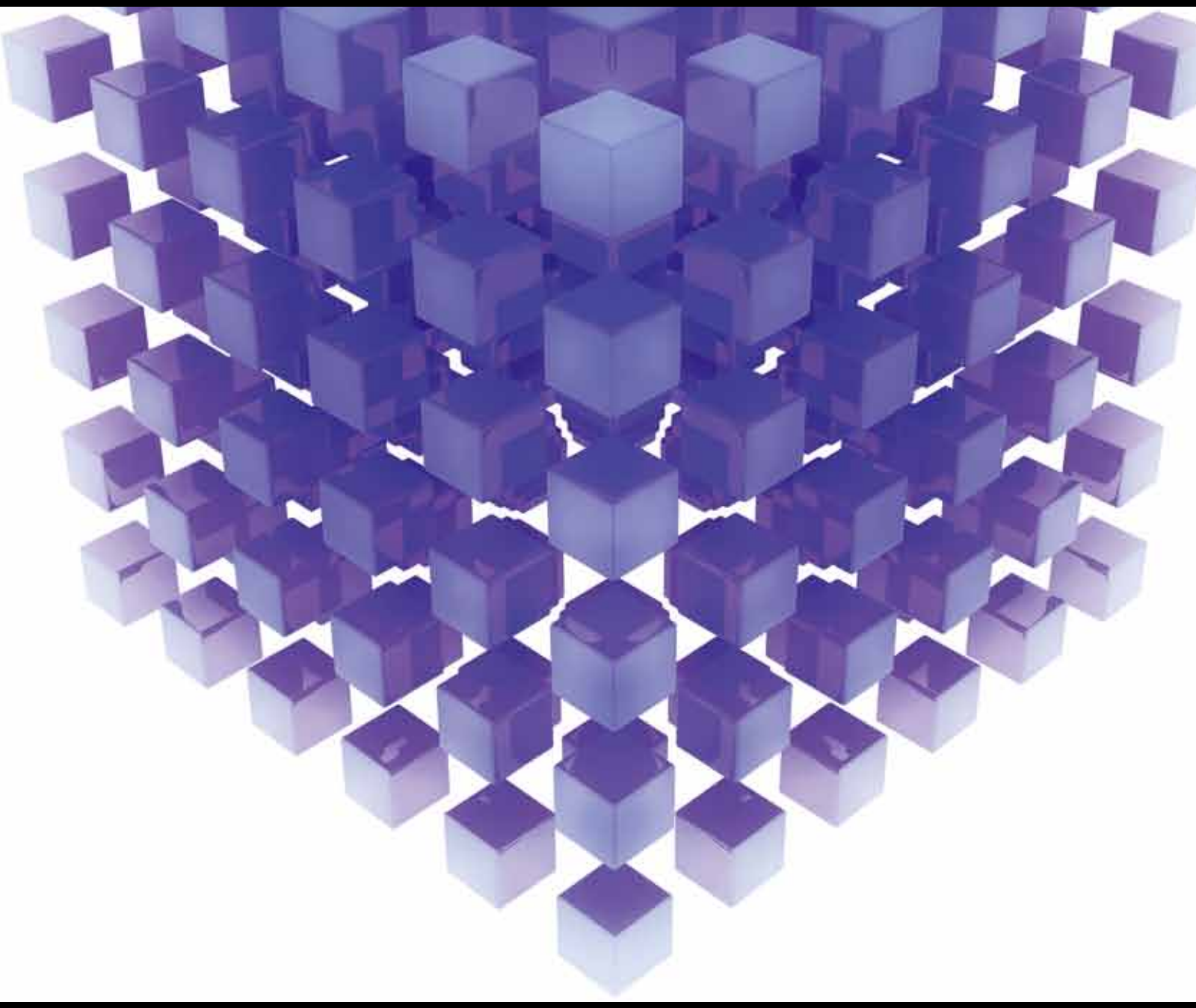
Mathematical Problems in Engineering

Transportation Modeling and Management

Guest Editors: Xiaoning Zhang, Henk van Zuylen, Yafeng Yin, Andy H. F. Chow, and Hu Shao



Transportation Modeling and Management

Mathematical Problems in Engineering

Transportation Modeling and Management

Guest Editors: Xiaoning Zhang, Henk van Zuylen, Yafeng Yin, Andy H. F. Chow, and Hu Shao

Editorial Board

Contents

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 381528, 4 pages
http://dx.doi.org/10.1155/2014/381528

Editorial
Transportation Modeling and Management

Xiaoning Zhang,[1] Henk van Zuylen,[2] Yafeng Yin,[3] Andy H. F. Chow,[4] and Hu Shao[5]

[1] *School of Economics and Management, Tongji University, Shanghai 200092, China*
[2] *Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands*
[3] *Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, USA*
[4] *Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK*
[5] *Department of Civil & Structural Engineering, The Hong Kong Polytechnic University, Hong Kong*

Correspondence should be addressed to Xiaoning Zhang; cexzhang@tongji.edu.cn

Received 19 October 2014; Accepted 19 October 2014; Published 24 December 2014

Due to the population growth and the rapid progress of urbanization, most big cities over the world face serious problems of traffic congestion, traffic safety, and vehicular emission. It is an urgent need to develop new models and new methods to describe and optimize the transportation system. This special issue is devoted to publishing the latest and significant results on scientific research in the mathematical methods of modeling and managing transportation systems.

After conducting rigorous peer review, this special issue has accepted 28 technical papers focusing on transportation modeling and management. We describe each one of them concisely as follows.

In "*A day-to-day route choice model based on reinforcement learning*," F. Wei et al. propose a day-to-day route choice model based on reinforcement learning and multiagent simulation. Travelers' memory, learning rate, and experience cognition are taken into account. Both the scenarios of link capacity degradation and random link capacity are used to illustrate the model's applications. Analyses and applications of the model demonstrate that the model is reasonable and useful for studying the day-to-day traffic dynamics.

In "*A schedule optimization model on multirunway based on ant colony algorithm*," Y. Jiang et al. develop a schedule optimization model for arrival-departure flights, to make full use of the slot of runway, reduce flight delay, and ensure fairness among airlines. The total delay cost and fairness among airlines are two objective functions. The ant colony algorithm is adopted to solve this problem and the result is more efficient and reasonable compared with FCFS (first come first served) strategy. Optimization results show that the flight delay and fair deviation are decreased by 42.22% and 38.64%, respectively.

In "*Grain emergency vehicle scheduling problem with time and demand uncertainty*," J. DongQing and Z. QunXiong study the grain transportation in relief and emergency supply chains. They take the grain emergency vehicle scheduling model between multiwarehouses as the research object. Under the emergency environment, the aim is to maximize the utilization of vehicles and minimize the delay. The randomness of the key parameters in grain emergency vehicle scheduling is determined through statistical analysis and the model is solved through robust optimization method. It is shown that the uncertainty of both time and demand has great influence on the efficiency of grain emergency vehicle scheduling problem.

In "*Analysis of time of day fare discounts on urban mass transit travel behavior, crowding, and waiting time*," X. Guo and H. Sun develop models to find an optimal discount fare and time intervals on morning peak hour. Two models have been analyzed to describe it with schedule delay because of the travel demand size. The first objective function is constructed on pressure equalization when the travel demand is small. The other objective function is to minimize total waiting time when the travel demand is large. In the end, numerical examples based on an artificial network are performed to characterize fare discount models.

In "*The bilevel design problem for communication networks on trains: model, algorithm, and verification*," Y. Tian et

al. propose a novel method to solve the problem of train communication network design. Firstly, they put forward a general description of such problem. Then, taking advantage of the bilevel programming theory, they create the cost-reliability-delay model (CRD model). They use a practical example to verify the accuracy and the effectiveness of the CRD model and further apply the novel method on a train with six carriages.

In "*Managing rush hour congestion with lane reversal and tradable credits*," Q. Li and Z. Gao develop models to reallocate the two-way road lane to cater for the asymmetric traffic flow. On the other hand, an effective tradable credit scheme is proposed to reduce the traffic demand and improve fairness for all travelers. A discrete bilevel programming model is established. The government at the upper level reallocates lanes on the two-way road to minimize the total system cost. The travelers at the lower level choose the optimal route on the basis of both travel time and credit charging for the lanes involved.

In "*Model of wagons' placing-in and taking-out problem in a railway station and its heuristic algorithm*," C. Guo and D. Lei develop models to decrease wagons' dwell time in railway stations and improve the efficiency of railway transportation. They take the locomotive running times between goods operation sites as weights, so the problem is described by a single machine scheduling problem, which is solved by a Hungarian algorithm.

In "*Developing a collaborative planning framework for sustainable transportation*," O. Ö. Özener considers a delivery network with multiple customers served by a single carrier. It executes a delivery plan with the minimum transportation cost and allocates the resulting costs and the emissions among the customers in a fair manner. In order to develop a mechanism that provides further reduction of the emissions, he studies a setting where the carrier takes the responsibility of the emissions and reflects the resulting inefficiencies while charging the customers.

In "*Storage space allocation of inbound container in railway container terminal*," L. Wang et al. consider the formulation and solution algorithm for storage space allocation problem of inbound containers in railway container terminal. The problem is formulated as two-stage optimization models, whose objectives are balancing the workload of inbound containers and reducing the overlapping amounts. An algorithm implement process based on rolling-horizon approach is designed to solve the proposed models. Real application shows that the proposed approach is effective to solve space allocation problem of inbound container and is significant for the operation and organization of railway container terminals.

In "*A bayesian combined model for time-dependent turning movement proportions estimation at intersections*," P. Jiao et al. present a back propagation neural network model to estimate dynamic turning movements, as well as the self-adaptive learning rate approach and the gradient descent with momentum method for solving. They design a revised sequential Kalman filtering algorithm. A field survey is implemented at an intersection in Beijing city to collect both time series of link counts and actual time-dependent turning movement flows, including historical and present data. The reported estimation results show that the Bayesian combined model is much more accurate and stable than other models.

In "*Multiagent based decentralized traffic light control for large urban transportation system*," Y. Xu et al. propose a novel decentralized multiagent based approach for massive traffic lights coordination to promote the large-scale green transportation. Within a local intersection, constraint optimizing agents are designed to efficiently search for joint activities of the lights.

In "*Modeling complex system correlation using detrended cross-correlation coefficient*," K. Dong et al. take traffic dynamic as an example of a complex system. By applying the detrended cross-correlation coefficient method to traffic time series, they find that the traffic fluctuation time series may exhibit cross-correlation characteristic. Further, they show that two traffic speed time series derived from adjacent sections exhibit much stronger cross-correlations than the two speed series derived from adjacent lanes. Similarly, they also demonstrate that the cross-correlation property between the traffic volume variables from two adjacent sections is stronger than the cross-correlation property between the volume variables of adjacent lanes.

In "*A two-stage model for project optimization in transportation infrastructure management system*," Z. Chen et al. propose a novel two-stage project optimization model, including budget allocation and project distribution. Moreover, the methods of dynamic programming (DP) and genetic algorithm (GA) are applied to obtain an effective solution. The findings indicate that the new optimization method can provide a satisfactory and reasonable maintenance schedule for transportation infrastructure maintenance agencies whose routine management will benefit from the newly proposed model.

In "*Exploring urban taxi drivers' activity distribution based on GPS data*," X. Hu et al. analyze the urban taxi driver's activity distribution characteristics from different temporal and spatial levels. In the time level, they identify the difference with taxi daily operation pattern (weekday versus weekends), continuous time in one day, passengers in vehicle time, and taxi drivers' operation frequency; in the space level, they explore the taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations. It is helpful for urban taxi drivers' operation and behavior pattern identification, as well as the contribution to the geographical activity space analysis.

In "*Measurement of international roughness index by using Z-axis accelerometers and GPS*," Y. Du et al. develop a measurement system equipped with Z-axis accelerometers and a GPS device. Using the self-designing measurement system based on the methodology proposed in this study, they perform a small-scale field test. They use a one-wheel linear model and two-wheel model to fit the variation of the Z-axis acceleration. The test results demonstrate that the low-cost measurement system has good accuracy and could enhance the efficiency of IRI measurement.

In "*Uncertain programming for network revenue management*," D. Mou and X. Wang present a chance-constrained

programming model based on the uncertainty theory for network revenue management, in which the fares and the demands both are uncertain variables rather than random variables. Based on the strategy of nested booking limits, a solution method of booking control is developed to solve the problem.

In "*Optimal coordinated strategy analysis for the procurement logistics of a steel group*," L. Deng et al. study the optimization of an internal coordinated procurement logistics system in a steel group and the decision on the coordinated procurement strategy by minimizing the logistics costs. Considering the coordinated procurement strategy and the procurement logistics costs, the aim of the optimization model was to maximize the degree of quality satisfaction and to minimize the procurement logistics costs.

In "*An integrative approach with sequential game to real-time gate assignment under CDM mechanism*," J.-q. Liu et al. study real-time airport gate assignment problem when small-scale or medium- to large-scale flight delays occur. Taking into account the collaborative decision making (CDM) of the airlines and the airport, slot assignment and gate assignment are integrated into mixed set programming (MSP), and a real-time gate assignment model is built and solved through MSP coupled with sequential game.

In "*Energy and environmental efficiency of China's transportation sector: a multidirectional analysis approach*," G. Bi present a nonradial DEA model with multidirectional efficiency analysis (MEA) involving undesirable outputs for the measurement of regional energy and environmental efficiency of China's transportation sector during the period 2006–2010. They not only evaluate the energy and environmental efficiency level and trend of China's transportation sector but also investigate the efficiency patterns of 30 regions and three major areas of China.

In "*Synthetic optimization model and algorithm for railway freight center station location and wagon flow organization problem*," X.-c. Liu et al. propose a two-stage method to optimize railway freight center stations location and wagon flow organization together. The location model is present with the objective to minimize the operation cost and fixed construction cost. Then, the second model of wagon flow organization is proposed to decide the optimal train service between different freight center stations. A heuristic algorithm that combined tabu search (TS) with adaptive clonal selection algorithm (ACSA) is proposed to solve those two models.

In "*SIRS model of passengers' panic propagation under self-organization circumstance in the subway emergency,*" H. Zhao et al. develop an improved SIRS model of passengers' panic spread in subway emergency with consideration of passengers' density, the characteristic of subway car with the confined space, and passengers' psychological factors. It helps the government and subway administration to master the panic spread mechanism and reduce the panic spread by improving measures and also provides certain reference significance for rail system construction, emergency contingency plans, and the construction and implementation of emergency response system.

In "*Berth allocation problem with quay crane assignment for container terminals based on rolling-horizon strategy*," L. Xiao and Z.-H. Hu develop a multiobjective optimization model which minimizes the total penalty costs. Then, the scheduling process is divided into a set of continual scheduling interval according to the dynamic arrival sequences. Meanwhile, rolling-horizon strategies for setting rolling and frozen windows and the parameter updating strategy are designed. The input parameters of the model in the next rolling window are updated according to the optimal results of each time window which have been obtained.

In "*Day-to-day scheduling travel time adjustment behavior and simulation*," H.-M. Li et al. propose a modeling framework to study the day-to-day scheduling travel time adjustment behavior on the basis of past experiences. Scheduling travel time is defined as the difference between the requested arrival time and the departure time. Mathematical equations are established to formulate every traveler's dynamic adjustment on his/her departure time.

In "*Traffic incident clearance time and arrival time prediction based on hazard models*," Y. b. Ji et al. develop some hazard based prediction models for both incident clearance time and arrival time. The best fitting distributions are drawn for both clearance and arrival time for 3 types of incident: crash, stationary vehicle, and hazard. The results show that Gamma, Log-logistic, and Weibull are the best fit for crash, stationary vehicle, and hazard incident, respectively. The obvious impact factors are given for crash clearance time and arrival time. The quantitative influences for crash and hazard incident are presented for both clearance and arrival. The model accuracy is analyzed at the end.

In "*Transportation network design considering morning and evening peak-hour demands*," H. Wang et al. propose an NDP model simultaneously considering both morning and evening peak-hour demands. The NDP problem is formulated as a bilevel programming model, where the upper level is to minimize the weighted sum of total travel time for network users travelling in both morning and evening commute peaks, and the lower level is to characterize user equilibrium choice behaviors of the travelers in two peaks. The proposed NDP model is transformed into an equivalent mixed integer linear programming (MILP), which can be efficiently solved by optimization solvers.

In "*Trafficability analysis at traffic crossing and parameters optimization based on particle swarm optimization method*," B. He and Q. Lu propose a mathematical model for the platoon's longitudinal movement. A systematic analysis of longitudinal control law is presented for the platoon of vehicles. The particle swarm optimization method is introduced to effectively optimize the parameters of platoon. The proposed method effectively finds the optimal parameters based on simulations and makes the spacing of platoon shorter.

In "*Investigating the in-vehicle crowding cost functions for public transit modes*," F. Qin conducts both quantitative and qualitative studies, focusing on remodeling the in-vehicle crowding cost functions for different transit modes. Three numerical case studies show that applying distinct in-vehicle crowding cost functions to different transit modes has implications not only for the cost structure of transit systems

and the magnitude of optimal service provisions but also for the presence of economies of scale in consumption.

In "*Potential field cellular automata model for pedestrian evacuation in a domain with a ramp*," X.-X. Jian and X. Zhang propose a potential field cellular automata model with a pushing force field to simulate the pedestrian evacuation in a domain with a ramp. They construct a cost potential depending on the ramp angle and introduce a function to evaluate the pushing force, which is related to the cost and the desired direction of pedestrian. The relationship between the slope of ramp and the pushing force is investigated. The changing of injured situations with the changing of the slope of ramp is also studied. When the number of pedestrians and the ramp angle arrive at certain critical points, the domino effect is simulated by this proposed model.

We think that the special issue presents many innovative and meaningful studies on transportation modeling and management related problems. We hope the published papers will provide a good foundation for future scientific research related to transportation management and optimization.

Acknowledgments

We also express our hearty thanks to all the contributors for sharing their excellent research works to this special issue and the anonymous reviewers for their time and effort in providing constructive and helpful reviews.

Xiaoning Zhang
Henk van Zuylen
Yafeng Yin
Andy H. F. Chow
Hu Shao

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 646548, 19 pages
http://dx.doi.org/10.1155/2014/646548

Research Article

A Day-to-Day Route Choice Model Based on Reinforcement Learning

Fangfang Wei, Shoufeng Ma, and Ning Jia

College of Management and Economic, Tianjin University, Tianjin 300072, China

Correspondence should be addressed to Ning Jia; jia_ning@tju.edu.cn

Received 10 April 2014; Accepted 28 August 2014; Published 30 September 2014

Academic Editor: X. Zhang

Day-to-day traffic dynamics are generated by individual traveler's route choice and route adjustment behaviors, which are appropriate to be researched by using agent-based model and learning theory. In this paper, we propose a day-to-day route choice model based on reinforcement learning and multiagent simulation. Travelers' memory, learning rate, and experience cognition are taken into account. Then the model is verified and analyzed. Results show that the network flow can converge to user equilibrium (UE) if travelers can remember all the travel time they have experienced, but which is not necessarily the case under limited memory; learning rate can strengthen the flow fluctuation, but memory leads to the contrary side; moreover, high learning rate results in the cyclical oscillation during the process of flow evolution. Finally, both the scenarios of link capacity degradation and random link capacity are used to illustrate the model's applications. Analyses and applications of our model demonstrate the model is reasonable and useful for studying the day-to-day traffic dynamics.

1. Introduction

In the past decades, day-to-day traffic dynamics (DDTD) have been developed substantially in the field of transportation, which are mainly used to study the traffic fluctuations and the evolution process, rather than the final or static equilibrium state [1]. Approaches used to study DDTD are very flexible because they allow a wide range of behavior rules, levels of aggregation, and types of models to be integrated in the same modeling framework [2]. Some researchers use deterministic models or stochastic models to study the day-to-day flow evolution. Smith and Friesz et al. proposed three dynamic systems by the deterministic traffic assignment processes [3, 4]. The flow evolution trajectory can be provided explicitly in deterministic models. Daganzo and Sheffi put forward the stochastic user equilibrium to model the evolution of the traffic pattern [5]. Stochastic models mainly focus on the probability distribution of flow states [6, 7]. In some literatures, traffic dynamics are modeled as continuous time processes [3, 4]. However, it is not reasonable to adjust traffic flow continuously due to travelers' activities constraints. Therefore, more appropriate methods appear by using discrete-time systems, in which travelers repeat route choice daily or weekly [2, 8]. The flow convergences of continuous time and discrete time are very different [9].

Another important branch of DDTD is the research about travelers' behaviors. The uncertainty inherent of travelers' behaviors increases the complexity of models about DDTD, which makes traditional models unable to represent complex DDTD well [10]. This motivates much research and efforts to solve the problem by using new methods. Among these new methods, agent-based techniques are very appropriate approaches. An agent knows its internal state, the evolution of system, the likely outcomes of its actions, and so forth, and it has some abilities, such as independent decision-making, beliefs, desires, and intentions. These characteristics ensure the scalability and robustness of agent-based models, which are important to portray the complexity of DDTD. Some scholars applied agent-based models to analyze travelers' behaviors. Rossetti modeled commuters' behaviors in the structure of beliefs, desires, and intentions (BDI) which is a classical structure in multiagent simulation. In their model, travelers could make decisions about route choice and departure time rationally [11]. Nakayama et al. proposed some models about travelers' route choice behaviors. They took into account the limitations of travelers' cognitive capabilities

and used microsimulation to examine the dynamic nature of travelers' route choice. They found that the network flow does not necessarily converge to the user equilibrium (UE) and travelers' cognitions of each route do not become homogeneous by learning [12–14]. Jha et al. assumed that each traveler used a disutility function to perceive travel time and schedule delay for evaluating the alternative travel choices, and then chose an alternative route according to the utility maximization principle [15]. Chen and Mahmassani modeled travelers' leaning process by using Bayesian theory and an agent-based simulation framework to study networks' dynamic properties [16]. Klügl and Bazzan used a simple heuristic model to stimulate the process of travelers' decision-making and studied the effect of traffic forecast [17]. They set up an agent-based model to stimulate travelers' route choice and found that travelers could avoid the Braess Paradox by learning [18]. Dia applied the agent-based approach to research drivers' behaviors under the influence of real-time information [19]. Liu and Huang used multiagent stimulation to investigate the differences of route update rules in two kinds of scenarios, which are traffic information unrelease and release, respectively [20]. Illenberger et al. proposed a model to study travelers' risk-sensitive route-choice behaviors [21]. Gao and Wang explored travelers' route choice under guidance information by using microsimulation [22].

Travelers' decision-marking is an important aspect in agent-based models. DDTD is not a process of one-short choice but a process of repeated decisions. Travelers can perceive travel time and the network characteristics from their travel experience. It is hard for them to make the best choice every time due to their bounded rationality and the uncertainty of environment. They obey a stochastic route choice rule. If the travel time of a path was short in the past days, its probability to be chosen is big in the current day. On the contrary, the probability is small. These characteristics are similar with the features of the reinforcement learning (RL) theory. Under RL, if an action yielded a high payoff in the past days, the probability assigned to it increases in the current round, or the behavior associated with the action gets reinforced [23]. RL asserts that decision makers' behaviors are moved towards the direction of random choice due to the environmental uncertainty [24]. These similarities make RL very suitable to analyze travelers' behaviors. However, it has been seldom used in this field. Ben-Elia proposed a learning-based model of route choice and found that information could improve travelers' learning rate and made them more prone to risk-seeking by a laboratory experiment [25]. Wahba and Shalaby proposed a departure time and route choice model based on the Markovian decision process and RL [26–28]. Zolfpour-Arokhlo et al. formulated a route planning system by combining Q value-based dynamic programming with RL. They created a priority route plan for vehicles by studying the weights of some components in road network environments such as road safety, traffic, and weather [29]. The applications of RL in route choice still need to be further researched.

In this paper, we propose a day-to-day route choice model based on RL. Travelers' memory level, learning rate, and cognition based on their travel experience are taken into account in the model. The decision-making and learning process of travelers follow the principles of the classical Bush-Mosteller (BM) model in RL.

The rest of the paper is organized as follows. Section 2 introduces some basic knowledge of BM model and then our day-to-day route choice model based on RL is formulated. Section 3 verifies the rationality of our model and analyzes its properties. Section 4 demonstrates the applications of our model in two traffic scenarios.

2. The Model

2.1. The Process of Travelers' Route Choice. Day-to-day route choice behavior is a repetitive decision process. Each traveler is an agent with the ability of learning and decision-making. They follow a stochastic route choice rule because of their bounded rationality and the environmental uncertainty. Travelers' route choice and travel time are not known by each other when there is no external information provided. Expected travel time (ET) and perceptive travel time (PT) are generated according to their experience. Travelers judge their choice by comparing their ET and PT after each travel and then update the probabilities that the paths will be chosen repeatedly. The characteristics of the process are similar with the features of the Bush-Mosteller (BM) model, which is a classical model in reinforcement learning (RL) and will be introduced in the next section. The process of travelers' route choice can be shown in Figure 1.

2.2. Model Formulation

2.2.1. The Bush-Mosteller Model. The Bush-Mosteller (BM) model is a classical learning model proposed by Bush and Mosteller in 1973. It consists of a learning algorithm and a stochastic decision rule. The consequences of a decision can create positive or negative stimulus, which updates the probability that the decision will be repeated [30]. Actions leading to satisfactory outcomes (i.e., outcomes that meet or exceed aspirations) will tend to be repeated in the future, whereas choices resulting in unsatisfactory experiences will be avoided. In BM model, each player focuses on his past choices and payoff but ignores all the choices and payoff of other players [31].

In the BM model, action updating takes place in two steps. Firstly, at each time t, the player i chooses an action c and calculates his stimulus by using the following formula:

$$S_c^t = \frac{u_c - A_i}{\sup \left| u(k) - A_i \right|}. \tag{1}$$

u_c is the payoff of the action c and A_i is the aspiration level of the player i at time t. $u(k)$ denotes any possible payoff if he chooses other strategies (assuming that players know all the potential payoff). The denominator in (1) represents the upper value of the set of possible differences between payoff

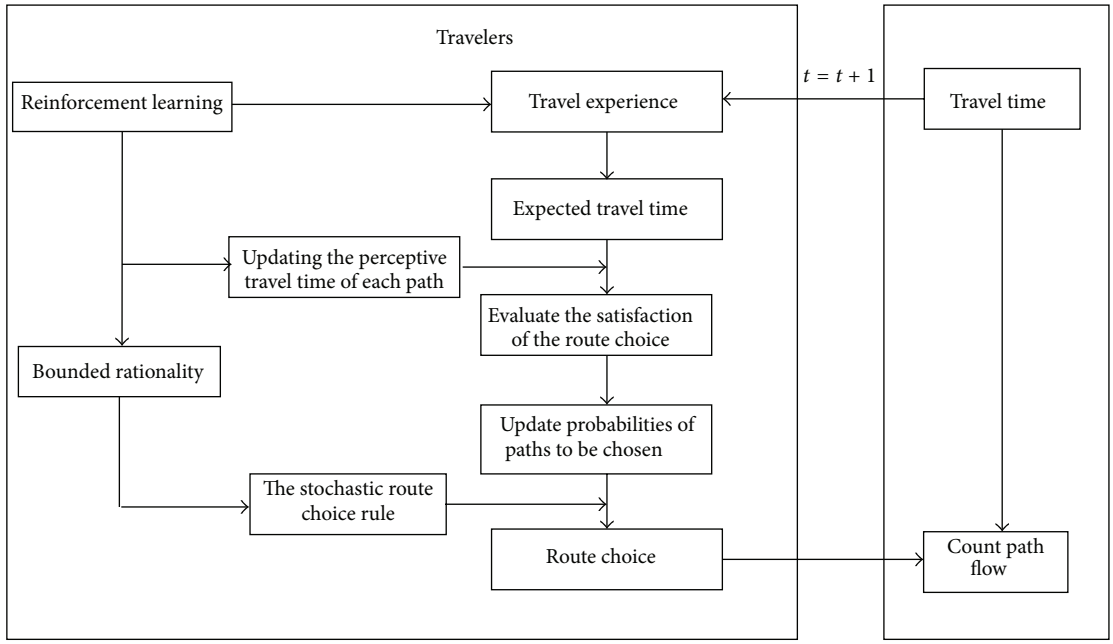


Figure 1: The process of travelers' route choice.

and aspiration. Secondly, the player i updates the selected probability of the action c as follows:

$$p_c^{t+1} = \begin{cases} p_c^t + (1 - p_c^t)\, ls_c^t, & \text{if } s_c^t \geq 0 \\ p_c^t + p_c^t ls_c^t, & \text{if } s_c^t < 0. \end{cases} \tag{2}$$

p_c^t is the selected probability of the action c at time t, l is the learning rate ($0 < l < 1$), and the changes of the probabilities for the actions not selected are derived from the constraint that probabilities always add up to unity.

Similarities make the BM model very suitable to research the day-to-day route choice behaviors. In the next section, a day-to-day route choice model will be proposed based on the BM model.

2.2.2. A Day-to-Day Route Choice Model. We consider a transportation network which is a fully-connected directed graph denoted as $G(N, L)$, consisting of a set of nodes N and a set of links L. Let W be the set of OD pairs, let d_w be the fixed travel demand between the OD pair $w \in W$, let R_w be the set of paths connecting the OD pair $w \in W$, let f_{kw}^t be the flow on the path $k \in R_w$ on day t, and let x_a^t be the flow on the link $a \in L$ on day t; d, f^t, and x^t denote demand, path flow, and link flow vectors, respectively; the link-path incidence matrix is denoted as B, then $x^t = Bf^t$; let C be the OD-path incidence matrix, then $d = Cf^t$. The link travel time vector and the path travel time vector are denoted as T^t and G^t, respectively, then $G^t = B'T^t$ [1].

Due to bounded rationality and environmental uncertainty, travelers obey a stochastic route choice rule: the bigger the chosen probability of a path, the more likely the path is selected by the traveler. Supposing the number of optional paths for the traveler i in the OD pair $w \in W$ is N. Let P_i be the set of the chosen probabilities of paths for him on day t, $P_i = \{p_1^t, p_2^t, \ldots, p_n^t\}$. A random number b can be generated by computer, if

$$\sum_{j=1}^{c-1} p_j^t < b \leq \sum_{j=1}^{c} p_j^t, \tag{3}$$

then the traveler i chooses the path c and the flow of the path adds one.

Each traveler only knows the travel time of the path he chooses on day t. However, they can form perceptive travel time (PT) of all the paths based on their travel experience. So a traveler's PT of a path can be used as its cost during his route choice process. Due to the particularity of route choice process, we make some modifications to (1) as follows:

$$S_i^t = \begin{cases} \dfrac{A_i^t - M_{ic}^t}{\max\{A_i^t - M_i^t(k)\}}, & \text{if } A_i^t - M_{ic}^t \geq 0 \\ \dfrac{A_i^t - M_{ic}^t}{\left|\min\{A_i^t - M_i^t(k)\}\right|}, & \text{if } A_i^t - M_{ic}^t < 0. \end{cases} \tag{4}$$

S_i^t is the stimulus of the traveler i when he chooses the path c on day t. A_i^t is his expected travel time (ET) and M_{ic}^t is his perceptive travel time (PT) of the path c. $M_i^t(k)$ is his any possible PT if he chooses other path. The two denominators in (4) donate the biggest benefit and the biggest loss that the traveler may get on day t. Piecewise function can reflect the degree of the traveler's satisfaction or dissatisfaction to his route choice better.

ET and PT are formed based on travelers' travel experience. Due to the limit of human's memory, travelers remember recent travel time more clearly than older ones. So the travel time on different days has different weight on their ET

and PT. The weighted average of all the travel time that a traveler has experienced can be used to calculate his ET and PT. Let ψ be the memory level of travelers ($0 < \psi \leq 1$), then ET and PT of the traveler i can be formulated as follows:

$$A_i^t = \frac{\sum_{j=1}^{t-1} \psi^{t-1-j} T_i^j}{\sum_{j=1}^{t-1} \psi^{t-1-j}}. \quad (5)$$

A_i^t is the ET of the traveler i on day t. T_i^j is the travel time that he has experienced on day j. A_i^t is formed before the travel on day t. Consider

$$M_{ik}^t = \frac{\sum_{j=1}^{t} \psi^{t-j} T_i^j \xi_{ik}^j}{\sum_{j=1}^{t} \psi^{t-j} \xi_{ik}^j}. \quad (6)$$

M_{ik}^t is the PT of the traveler i to the path k after the travel on day t. ξ_{ik}^j is a 0-1 variable. If he chooses the path k on day j, then $\xi_{ik}^j = 1$; else $\xi_{ik}^j = 0$. Because only one path can be chosen by him on day j, then

$$\sum_{k=1}^{N} \xi_{ik}^j = 1. \quad (7)$$

N is the number of paths that he can choose. So the PT of the path c which is chosen by him on day t can be formulated as

$$M_{ic}^t = \sum_{k=1}^{N} M_{ik}^t \xi_{ik}^t. \quad (8)$$

Because the traveler i does not know the actual travel time of the paths that he does not choose on day t, it is reasonable to use PT of these paths as their cost. Because PT is the weighted average of travel time that travelers have experienced, it can reflect the stability of a path's travel time in a long time. Considering the contingency of once travel time, it is not appropriate to use the actual travel time of the path chosen on day t as its evaluation criterion, therefore we also use its PT as its cost.

Stimulus on day t can be calculated by (4) to (8). Then the probability of the path c chosen by the travel i on day t can be updated by (2). In order to ensure all probabilities always add up to unity, the probabilities of those paths not chosen by the travel i on day t are updated as follows:

$$p_k^{t+1} = \begin{cases} p_k^t - p_k^t l S_i^t, & \text{if } S_i^t \geq 0 \\ \dfrac{p_k^t - p_k^t p_c^t l S_i^t}{1 - p_c^t}, & \text{if } S_i^t < 0 \end{cases} \quad k \in N,\ k \neq c. \quad (9)$$

S_i^t is the stimulus of the traveler i when he chooses the path c on day t. p_k^t is the chosen probability of the path k and p_c^t is the chosen probability of the path c.

3. Model Validation and Model Properties

In the previous section, we proposed a day-to-day route choice model. In this section, some numerical stimulations are made to validate the model by using three different types of test networks. The properties of the model are also analyzed.

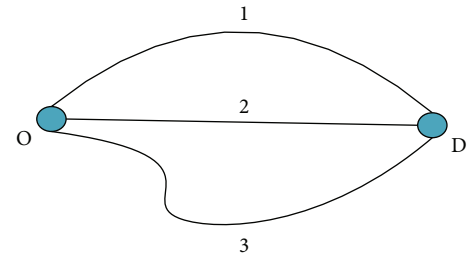


Figure 2: The first test network.

3.1. The Simulation Program and the Three Test Networks

3.1.1. The Simulation Program. The simulation program is briefly described as follows:

Step 0. (Parameters Initialization). Generate initial chosen probability of each path for every traveler and specify the value of ψ and l.

Step 1. Generate a random number of 0 to 1 for every traveler. Travelers choose routes according to the rule mentioned in (3).

Step 2. Count the flow of each route. Compute the travel time of each route.

Step 3. Calculate every traveler's expected travel time and perceptive travel time with (5) to (8).

Step 4. Compute the stimulus of every traveler with (4).

Step 5. Update the chosen probability of each path for every traveler with (2) and (9).

Step 6. Cycle to perform Step 1 to Step 5 within the prescribed number of days.

3.1.2. The Three Test Networks. Three test networks, which are used commonly in transportation literatures, are chosen as our numerical examples. The first test network is composed of three links as shown in Figure 2. The total OD demand is 300. Links' free flow travel time and capacities are $t_1^0 = 3.5$, $t_2^0 = 3$, and $t_3^0 = 2$; $c_1 = 90$, $c_2 = 100$, and $c_3 = 70$. The average link travel time function is assumed to be the BPR function. Consider

$$T_a = t_a^0 \left[\left(1 + 0.15 \left(\frac{x_a}{c_a} \right)^2 \right) \right], \quad (10)$$

t_a^0 is the free flow travel time of the link a, c_a is the link capacity, and x_a is the link flow.

The second test network is a 3×3 grid network, which is constituted by 9 nodes, 12 links, and 6 routes [1]. Node and

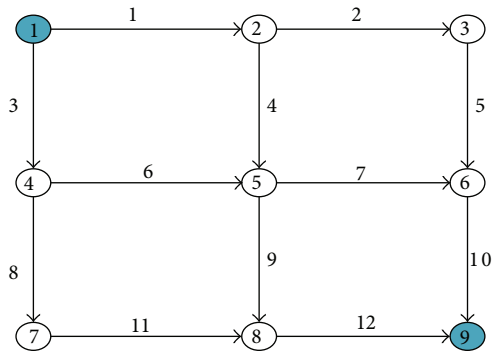


Figure 3: The second test network.

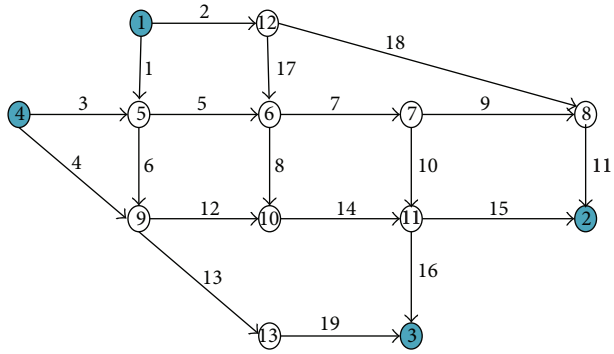


Figure 4: The third test network.

Table 1: Link characteristics of the second test network.

Link	t_a^0	c_a	Link	t_a^0	c_a
1	2.0	360	7	1.2	180
2	1.2	360	8	1.5	240
3	1.5	240	9	1.0	150
4	1.2	180	10	3.0	360
5	1.2	360	11	1.5	240
6	1.0	150	12	1.5	240

link numbers are shown in Figure 3. The total OD demand is 9000. The free flow travel time and capacities of the 12 links are given in Table 1.

The third test network named Nguyen and Dupuis' network has 13 nodes and 19 links as shown in Figure 4. There are four O-D pairs in the network, 1-2, 1-3, 4-2, and 4-3; the numbers of paths between those O-D pairs are 8, 6, 5, and 6, respectively, as shown in Table 2. The O-D demands are deterministic and given as $q_{1\text{-}2} = 660$, $q_{1\text{-}3} = 495$, $q_{4\text{-}2} = 412$, and $q_{4\text{-}3} = 495$ [32]. The link characteristics of the network are shown in Table 3.

3.2. Model Validation. As mentioned in our model, the memory level of travelers is ψ ($0 < \psi \leq 1$). The value of ψ can affect the weight of the past travel time on travelers' current expected travel time (ET) and perceptive travel time (PT). Travelers weight recent travel time more than older ones. The smaller the ψ-value, the less the travel time of the past days influence travelers' current ET and PT. With the increase of ψ-value, more travel time of the past days has influence on their ET and PT. When $\psi = 1$, it means that travelers can remember all travel time they have experienced and the travel time of each day has the same effect on ET and PT.

It is known that user equilibrium (UE) assumes that travelers are perfectly rational and know all information. So the conditions of reaching to UE are ideal. Obviously, $\psi = 1$ is also an ideal state in our model. If the flow distribution in our model can converge to UE under $\psi = 1$, then our model can be verified to be reasonable. The following are the stimulation results of the three test networks under the condition of $\psi = 1$, $l = 0.3$ (l is the learning rate).

In the first network, as shown in Figure 5, the flow fluctuations of the three paths become very weak on the 60th day and the network flow converges to equilibrium on the 210th day. The travel time of the three paths is approximately equal under the equilibrium. So the network flow converges to UE. Similarly, the flow of the second network converges to UE in Figure 6. The flow in the four OD pairs of the third network also converges to UE as shown in Figures 7, 8, 9, and 10, and there are almost no flow on the paths 6, 7, 8, 13, and 14 under UE.

As shown above, the three test networks can all converge to UE when $\psi = 1$ and $l = 0.3$. In order to study whether the learning rate can affect networks to converge to UE under the ideal state of $\psi = 1$, we set three different values of l: $l = 0.1$, $l = 0.5$, and $l = 0.9$. The second test network is used as the stimulation example. The corresponding day-to-day dynamics are shown in Figures 11 and 12. It can be seen that the network flows all converge to UE under different l when $\psi = 1$. And the convergence rate becomes faster with the increase of l. So the learning rate can impact the convergence rate but cannot influence networks to converge to UE when $\psi = 1$.

The results above not only illustrate that the network flow evolution based on our model can converge to UE when travelers' memory level is ideal ($\varphi = 1$), but also show that our model can demonstrate the process of flow and travel time evolution. And the network flow does not converge to UE smoothly but reach to UE after an oscillation period (as shown in Figure 6). These results validate the rationality of our model.

3.3. Model Properties. The rationality of our model has been validated in Section 3.2. In this part, its properties will be illustrated by analyzing the two parameters in the model (ψ and l). The second test network is used as the simulation example in this section.

3.3.1. Impact of Memory Level on Day-to-Day Traffic Dynamics. In order to see the impact of ψ on day-to-day traffic dynamics, we fix $l = 0.3$ and set three different values of ψ: $\psi = 0.1$, $\psi = 0.5$, and $\psi = 0.9$. The corresponding flow evolution and travel time evolution are shown in Figures 13 and 14. It can be seen clearly that the network flow does not converge to UE when $\psi = 0.1$ and $\psi = 0.5$, it converges to UE approximately as $\psi = 0.9$. And the flow fluctuations of

Table 2: The paths in the third test network.

OD pair	Serial number of paths	Link sequence	OD pair	Serial number of paths	Link sequence
1-2	1	1-5-7-9-11	4-2	15	3-5-7-9-11
	2	1-5-7-10-15		16	3-5-8-14-15
	3	1-5-8-14-15		17	3-5-7-10-15
	4	1-6-12-14-15		18	3-6-12-14-15
	5	2-18-11		19	4-12-14-15
	6	2-17-7-9-11			
	7	2-17-7-10-15			
	8	2-17-8-14-15			
1-3	9	1-5-7-10-16	4-3	20	3-5-7-10-16
	10	1-5-8-14-16		21	3-5-8-14-16
	11	1-6-12-14-16		22	3-6-12-14-16
	12	1-6-13-19		23	3-6-13-19
	13	2-17-7-10-16		24	4-12-14-16
	14	2-17-8-14-16		25	4-13-19

Table 3: Link characteristics of the third test network.

Link	t_a^0	c_a	Link	t_a^0	c_a	Link	t_a^0	c_a
1	7	300	8	13	250	15	9	300
2	9	200	9	5	250	16	8	300
3	9	200	10	9	300	17	7	200
4	12	200	11	9	500	18	14	300
5	3	350	12	10	550	19	11	200
6	9	400	13	9	200			
7	5	500	14	6	400			

$\psi = 0.1$ and $\psi = 0.5$ are strong, but that is weak relatively when $\psi = 0.9$.

The flow standard deviations (FSD) in the last 100 days can be used as the statistic to measure the flow fluctuations. We fix the value of l and increase the value of ψ from 0 to 1. Due to the length limitation of the paper, only four deterministic values of l are listed ($l = 0.3$, $l = 0.6$, $l = 0.7$, and $l = 0.9$). The corresponding changes of FSD are shown in Figure 15. After summarizing all simulation results, we find that when $l > 0.6$, the flow fluctuations weaken obviously with the increase of ψ, but when $l < 0.6$, only higher memory level ($\psi > 0.5$) can reduce the fluctuations obviously.

These phenomena above are consistent with the physical meaning of ψ: travelers can cognize travel time better with the increase of ψ. It is helpful to reduce the randomness of their route choice and make them do better choice. So a bigger ψ can weaken the flow fluctuations and promote the convergence of the network flow.

3.3.2. Impact of Learning Rate on Day-to-Day Traffic Dynamics. To see the impact of l on day-to-day traffic dynamics, we fix $\psi = 0.6$ and set three different values of l: $l = 0.1$, $l = 0.5$, and $l = 0.9$. As shown in Figures 16 and 17, we can see that the network flow converges to UE approximately only when $\psi = 0.6$ and $l = 0.1$. The flow fluctuations are strong under $\psi = 0.6$ and $l = 0.5$. When $\psi = 0.6$ and $l = 0.9$, the flow distribution deviates from the UE for several days after the network flow converges to UE, then comes back to UE again. And the process is repeated. This phenomenon is called "cyclical oscillation" by us. The cyclical oscillation found in the flow evolution is not an accidental phenomenon, which appears in a lot of stimulations with a high learning rate (such as $\psi = 0.5$, $l = 0.9$; $\psi = 0.8$, $l = 0.9$; $\psi = 0.9$, and $l = 0.85$). The corresponding flow evolutions can be seen in Figure 18.

We fix the value of ψ and increase the value of l from 0 to 0.99 for studying the impact of l on flow fluctuations. From Figure 19, we can see that when $0 \leq \psi \leq 0.5$, the flow fluctuations strengthen with the increase of l. But when $\psi \geq 0.5$, there are two inflection points (IP) in the process of FSD changes. With the increase of l, the flow fluctuations strengthen before the first IP, and then weaken until the second IP and strengthen again after the second IP. For example, when $\psi = 0.8$, the two IP are $l = 0.5$ and $l = 0.8$, respectively. when $l \leq 0.5$ and $l \geq 0.8$, the flow fluctuations strengthen with the increase of l. But when $0.5 \leq l \leq 0.8$, the flow fluctuations weaken with the increase of l.

The impact of l on day-to-day traffic dynamics can be explained: the larger the l, the more the paths' chosen probabilities are adjusted, which strengthens the flow fluctuations

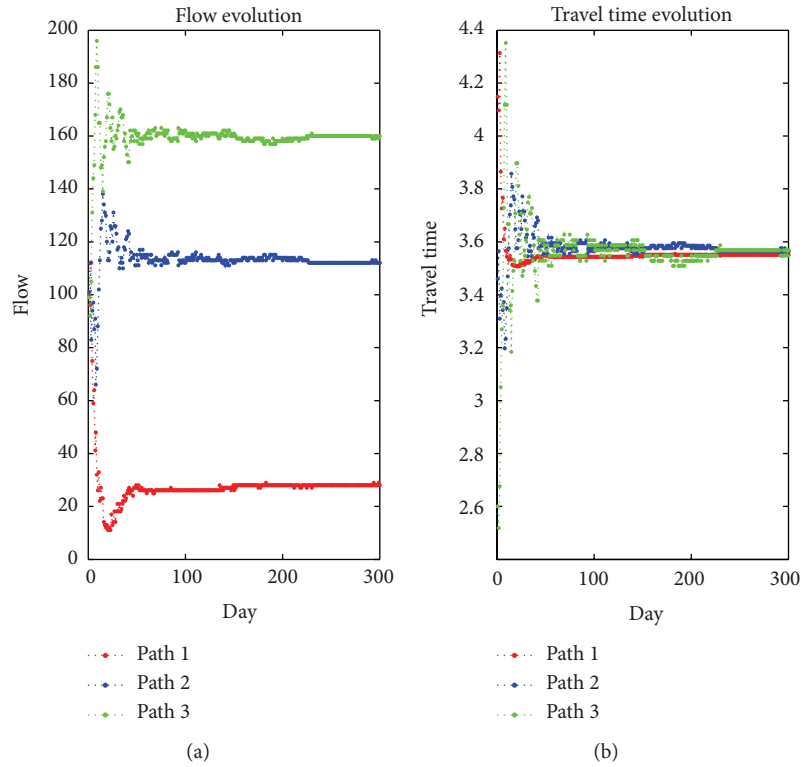


FIGURE 5: Flow evolution and travel time evolution of the first test network when $\psi = 1, l = 0.3$.

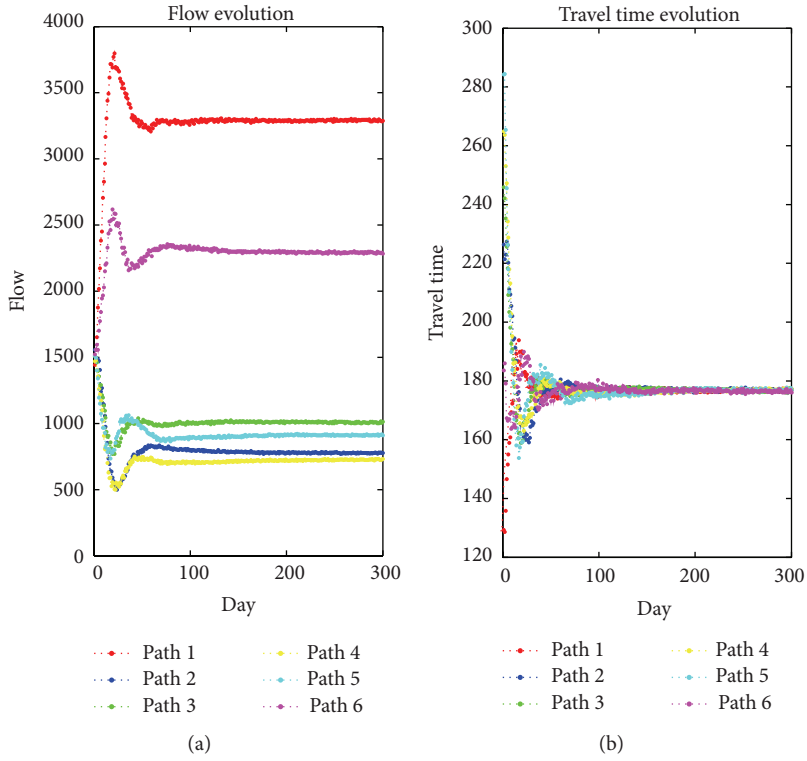


FIGURE 6: Flow evolution and travel time evolution of the second test network when $\psi = 1, l = 0.3$.

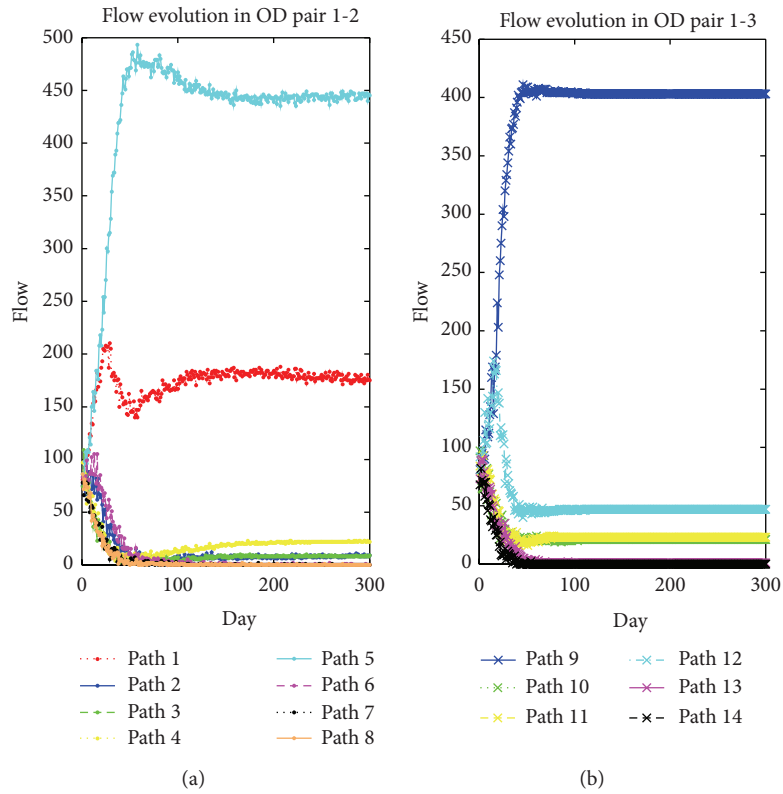


FIGURE 7: Flow evolution in OD pairs 1-2 and 1-3 of the third test network when $\psi = 1, l = 0.3$.

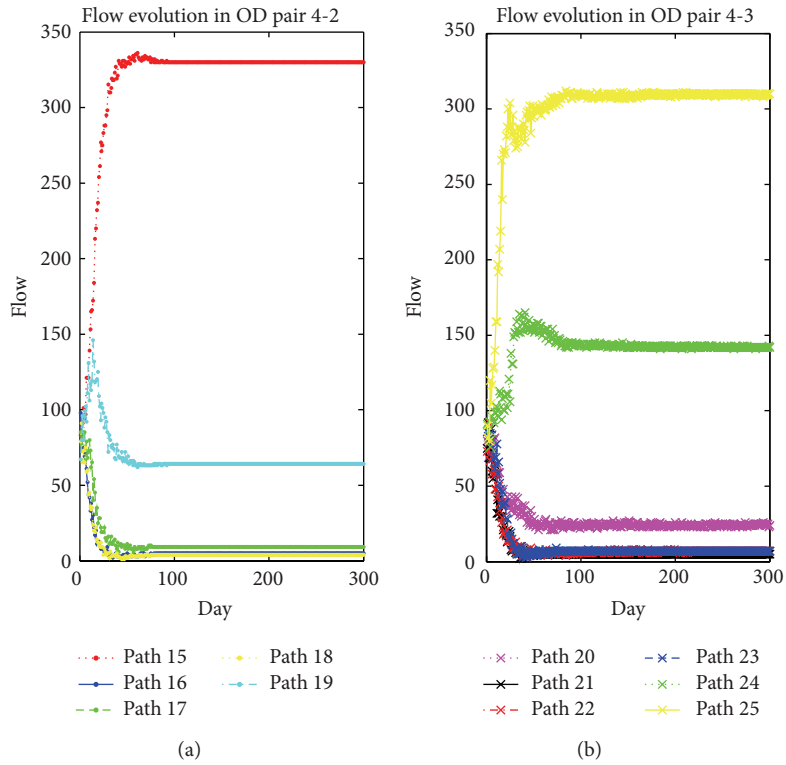


FIGURE 8: Flow evolution in OD pairs 4-2 and 4-3 of the third test network when $\psi = 1, l = 0.3$.

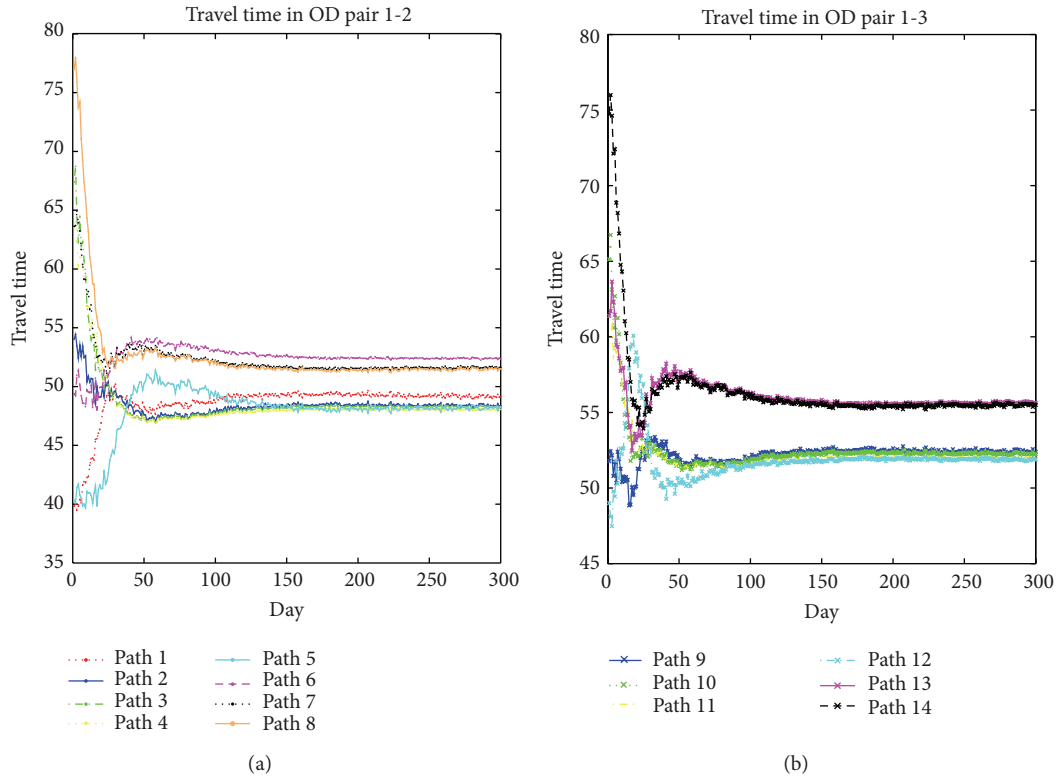


Figure 9: Travel time evolution in OD pairs 1-2 and 1-3 of the third test network when $\psi = 1, l = 0.3$.

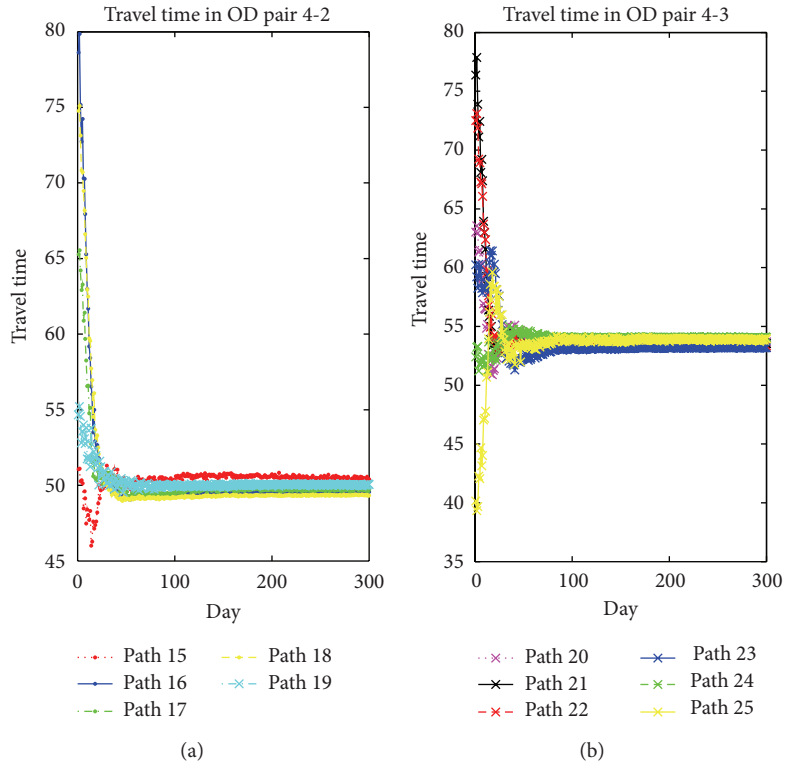


Figure 10: Travel time evolution in OD pairs 4-2 and 4-3 of the third test network when $\psi = 1, l = 0.3$.

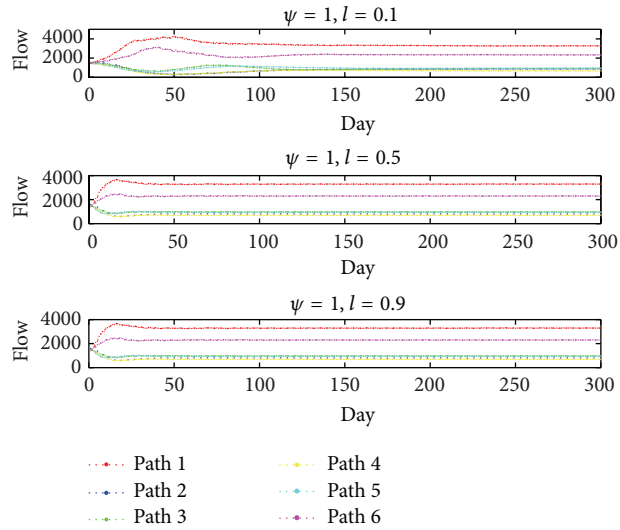


Figure 11: Flow evolution in the second test network with different l-value when $\psi = 1$.

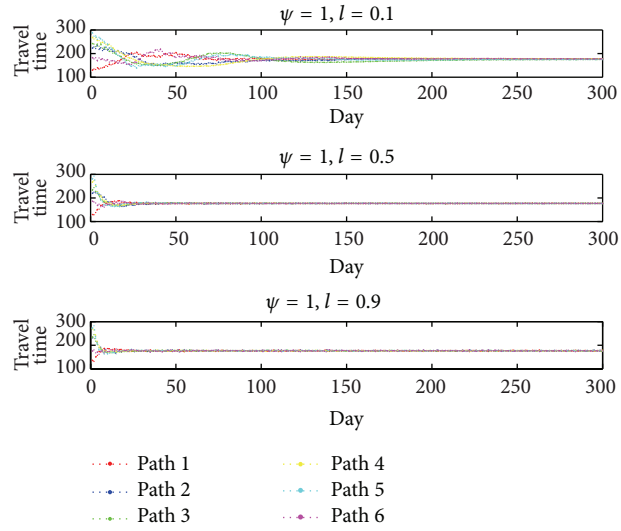


Figure 12: Travel time evolution in the second test network with different l-value when $\psi = 1$.

naturally. Memory level is helpful for weakening the flow fluctuations as mentioned in Section 3.3.1. So the effect of memory level and learning rate on day-to-day dynamics is opposite. Their opposite effect leads to the first IP in the process of FSD changes in Figure 19; the second IP is generated by the cyclical oscillation of flow evolution which appears under high learning rate.

4. Traffic Scenarios Analyses

In this section, in order to illustrate the applications of our model, two particular traffic scenarios are analyzed by using the third test network and some discussions are provided based on the numerical results. The two parameters of our model are fixed as $\varphi = 0.8$, $l = 0.3$ in this section.

4.1. Link Capacity Degradation Scenario. Assuming that the capacity of the first link in the third test network reduces 50% on the 1001th day because of road maintenance, and recovers on the 2001th day [33]. The day-to-day traffic dynamics are shown in Figures 20 and 21 (different color lines represent different paths, which can be seen in Figures 7 and 8). We can see that the network flow distributions in the four OD pairs all converge to UE approximately before the link capacity reduction. After the capacity reduction, the flow distributions in all OD pairs are changed. The changes in the OD pairs 1-2 and 1-3 are bigger than those in the OD pairs 4-2 and 4-3. These are consistent with our intuition: OD pairs 1-2 and 1-3 both contain the first link and, thus, should be more severely affected by the link capacity reduction. The first link does not belong to OD pairs 4-2 and 4-3, so the two OD pairs should be influenced gently. When the first link's capacity recovers, initial network equilibrium is restored. This is because when the capacity recovers, travelers can form similar cognition on the network with that formed before the capacity reduction. So the initial network equilibrium can be restored by revoking the link capacity degradation. Other point needs to be noticed: the first link's capacity reduction strengthens the flow fluctuations in the OD pairs 1-2 and 1-3. These suggest that the capacity reduction can reduce the stability of the network.

4.2. Links' Capacities as Random Variables Scenario. Links' actual capacities may be influenced by many factors, such as bad weather, traffic accidents, pedestrians' crossing street behaviors, traffic violations, and road maintenance. All these factors can reduce links' capacities. So links' actual capacities are no longer deterministic values but random variables [34].

Let c_a be the actual capacity of the link a, let h_a be the minimum capacity, and let G_a be the maximum capacity. Assume that every link's actual capacity follows a uniform distribution, which means $c_a \sim U(h_a, G_a)$.

In order to study the impact of links' actual capacities' random variations on the day-to-day traffic dynamics, the dynamics of deterministic capacities are used as reference. Let deterministic capacities be equal to the expected values of the actual capacities, let C_a denote the deterministic capacity of the link a, and let $E(c_a)$ denote the expected value of the actual capacity of the link a. Then

$$E(c_a) = \frac{h_a + G_a}{2} = C_a. \tag{11}$$

The mean and standard deviation of paths' flow and travel time in the last 100 days are used as comparison statistics. The third test network is used as the numerical example under the condition of $\psi = 0.8$, $l = 0.3$. Deterministic capacities use links' capacities in Table 3. Take the statistical results of the OD pairs 1-3 and 4-2 as examples to show the differences between deterministic capacities and random capacities on day-to-day dynamics. The results of

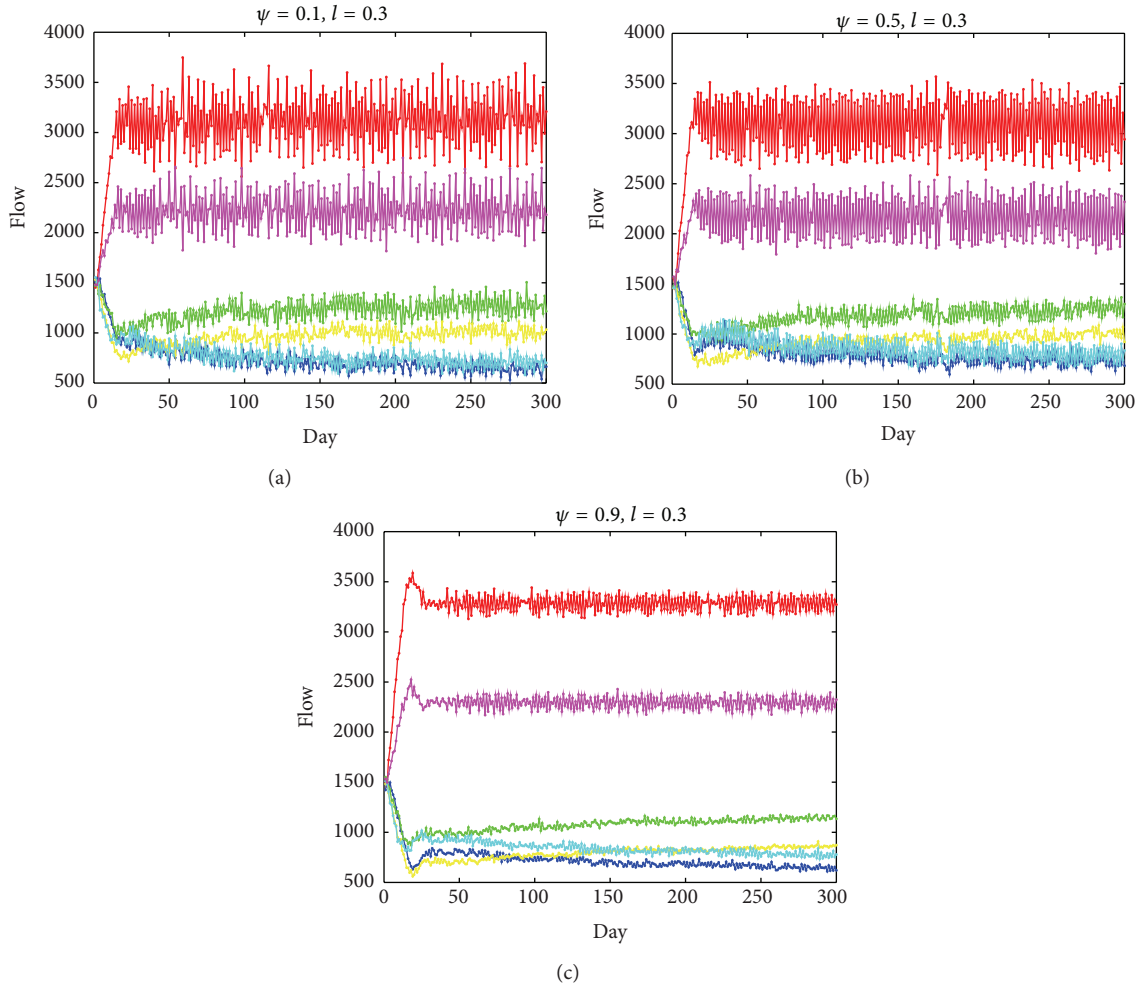


Figure 13: Flow evolution with different ψ-value when $l = 0.3$.

deterministic capacities and random capacities are shown in Tables 4 and 5, respectively.

Because paths' mean travel time equal approximately and their standard deviation are not big in Tables 4 and 5, the network flow under the two kinds of links' capacities both converge to UE approximately. But the UE of the two cases are different, because the mean path flow distributions are not the same by comparing Tables 4 and 5. So the random variations of links' capacities change the network's paths flow equilibrium. According to the mean flow of paths in Tables 4 and 5, links' mean flow under deterministic link capacities and random link capacities can be calculated as the following formula:

$$x_a^M = \sum_p \delta_{ap}^w f_p^M. \tag{12}$$

x_a^M and f_p^M represent the mean flow of the link a and the path p in the last 100 days. If the link a belongs to the path p, then $\delta_{ap}^w = 1$, else $\delta_{ap}^w = 0$. There are 19 links in the third test network. The calculation values of links' mean flow are shown in Figure 22. It can be seen clearly that the mean flow of links under the two kinds of links' capacities is almost equal. So the random variations of links' capacities almost do not change the equilibrium of links flow, but it can change the equilibrium of paths flow as mentioned above.

In addition, the random variations of links' capacities make the mean travel time of paths become longer, which can be explained as follows.

We assume $c_a \sim U(h_a, G_a)$ above, so the travel time of the link a is also a random variable. Let $\psi(c_a)$ be the probability density function of $c_a$and let T_a be the travel time of the link a. $E(T_a)$ denotes its expected travel time. Then

$$E(T_a) = \int_{h_a}^{G_a} T(c_a)\,\psi(c_a)\,dc_a = t_a^0 + 0.15 t_a^0 \frac{x_a^2}{h_a G_a}. \tag{13}$$

Because $h_a > 0$, $G_a > 0$ and $E(c_a) = (h_a + G_a)/2 = C_a$, then

$$E(T_a) > t_a^0 \left[\left(1 + 0.15\left(\frac{x_a}{C_a}\right)^2\right)\right]. \tag{14}$$

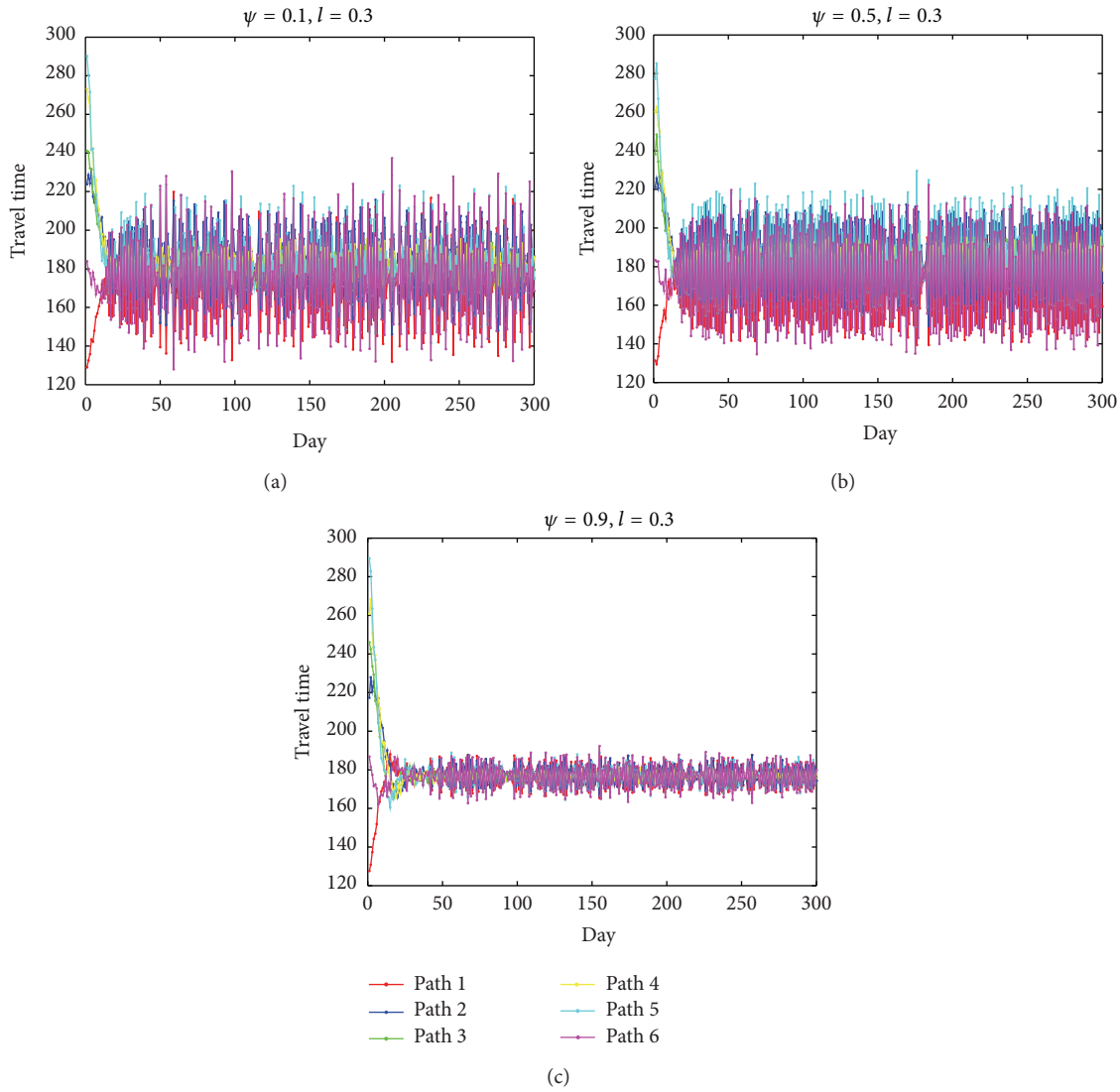


Figure 14: Travel time evolution with different ψ-value when $l = 0.3$.

Table 4: The statistical results of deterministic capacities.

OD pair	Serial number of paths	Link sequence	Mean flow	Flow standard deviation	Mean travel time	Time standard deviation
1-3	9	1-5-7-10-16	341.92	8.1509	51.975	0.5202
	10	1-5-8-14-16	34.67	3.6764	52.1231	0.352
	11	1-6-12-14-16	20.92	2.9565	52.1714	0.2675
	12	1-6-13-19	97.08	6.897	52.063	0.564
	13	2-17-7-10-16	0.24	0.5527	55.5151	0.3316
	14	2-17-8-14-16	0.17	0.4277	55.6632	0.3063
4-2	15	3-5-7-9-11	302.15	5.4926	49.7345	0.4098
	16	3-5-8-14-15	4.52	1.4318	49.9226	0.254
	17	3-5-7-10-15	37.09	3.4085	49.7746	0.4377
	18	3-6-12-14-15	1.19	1.2768	49.971	0.1818
	19	4-12-14-15	67.05	4.0982	49.7898	0.3518

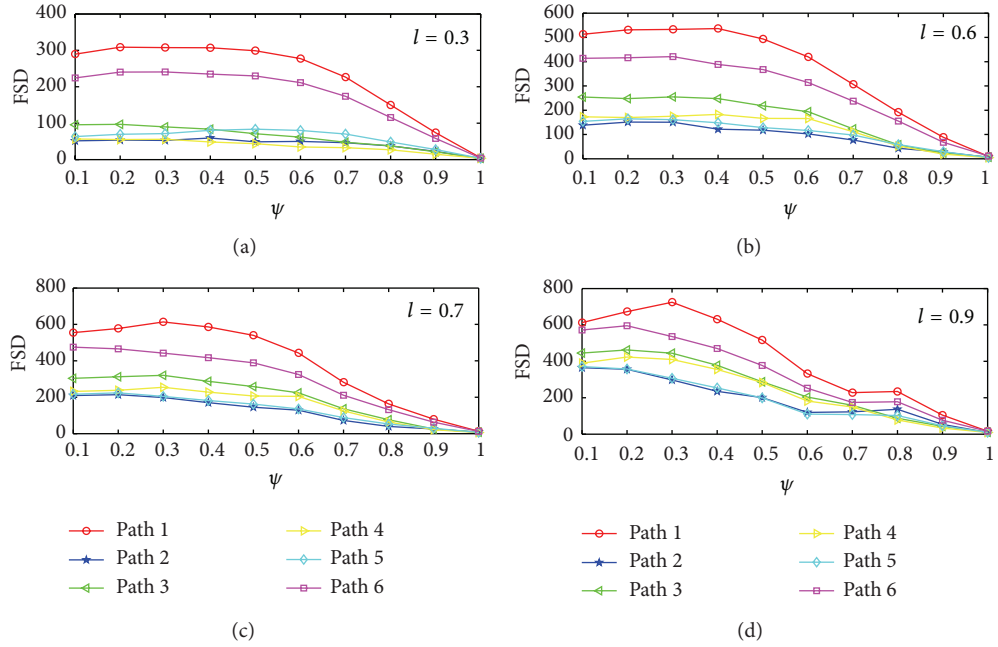


Figure 15: The changes of FSD with the increase of ψ under deterministic values of l.

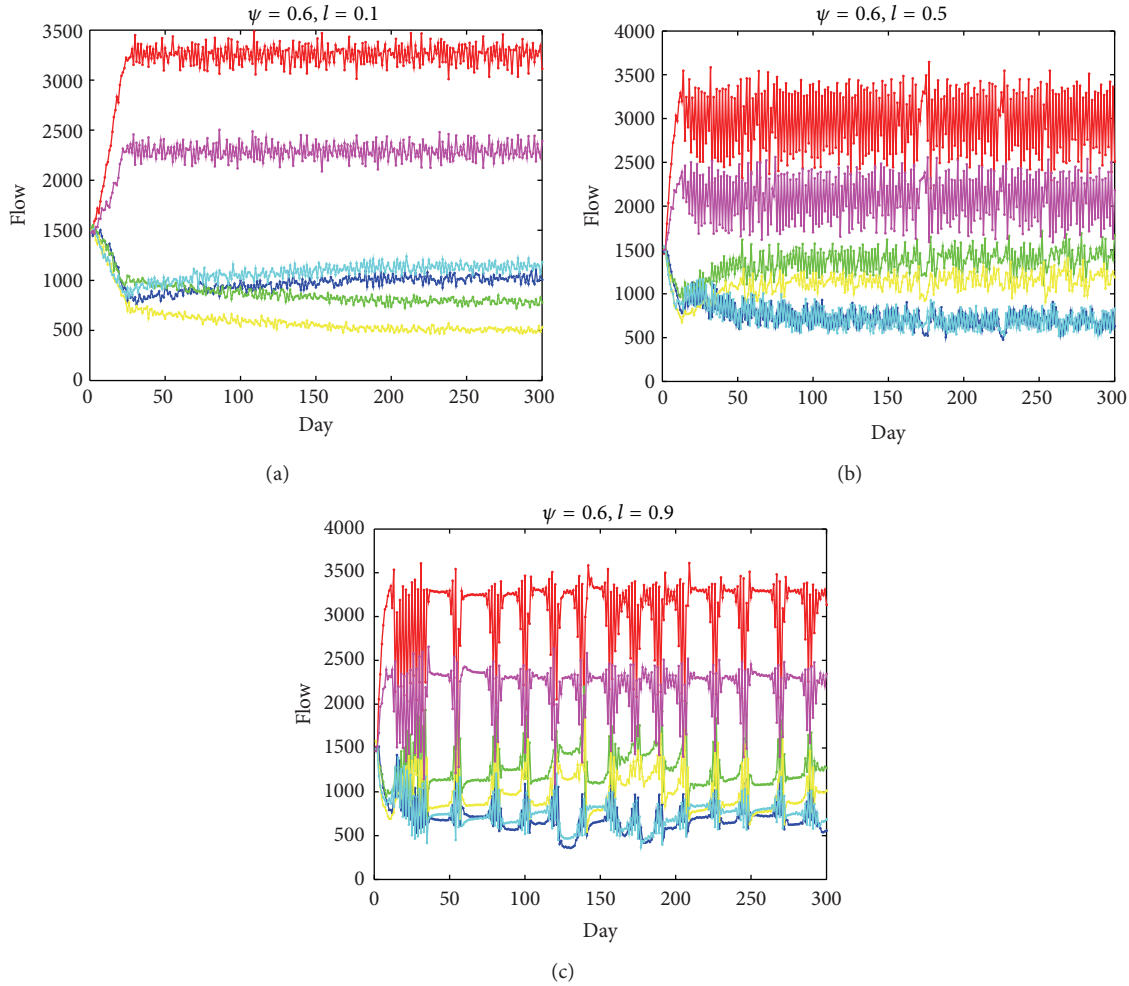


Figure 16: Flow evolution with different l-value when $\psi = 0.6$.

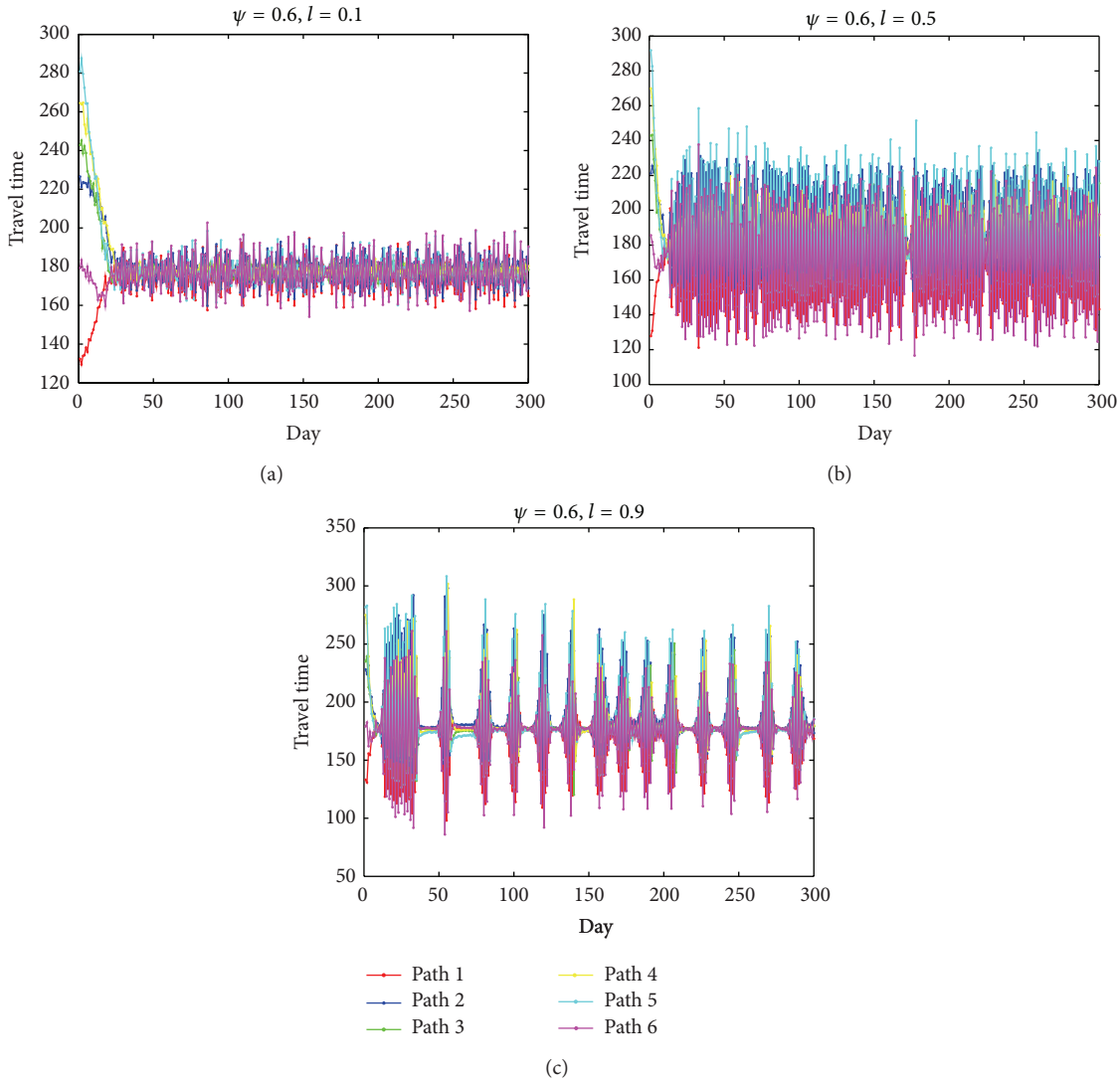


Figure 17: Travel time evolution with different l-value when $\psi = 0.6$.

Table 5: The statistical results of random capacities.

OD pair	Serial number of paths	Link sequence	Mean flow	Flow standard deviation	Mean travel time	Time standard deviation
1-3	9	1-5-7-10-16	291.11	16.9657	52.5273	2.8096
	10	1-5-8-14-16	48.7	6.3794	53.3368	2.4696
	11	1-6-12-14-16	37.06	5.9286	53.5686	2.1972
	12	1-6-13-19	117.4	11.8841	53.0173	2.2925
	13	2-17-7-10-16	0.64	0.7722	55.3073	2.5127
	14	2-17-8-14-16	0.09	0.2876	56.1167	2.1885
4-2	15	3-5-7-9-11	260.96	14.4564	49.7908	2.3642
	16	3-5-8-14-15	5.55	2.1385	51.2376	1.7302
	17	3-5-7-10-15	37.03	5.1707	50.4282	1.8506
	18	3-6-12-14-15	2.45	1.6659	51.4695	1.4769
	19	4-12-14-15	106.01	14.1461	50.3646	2.9827

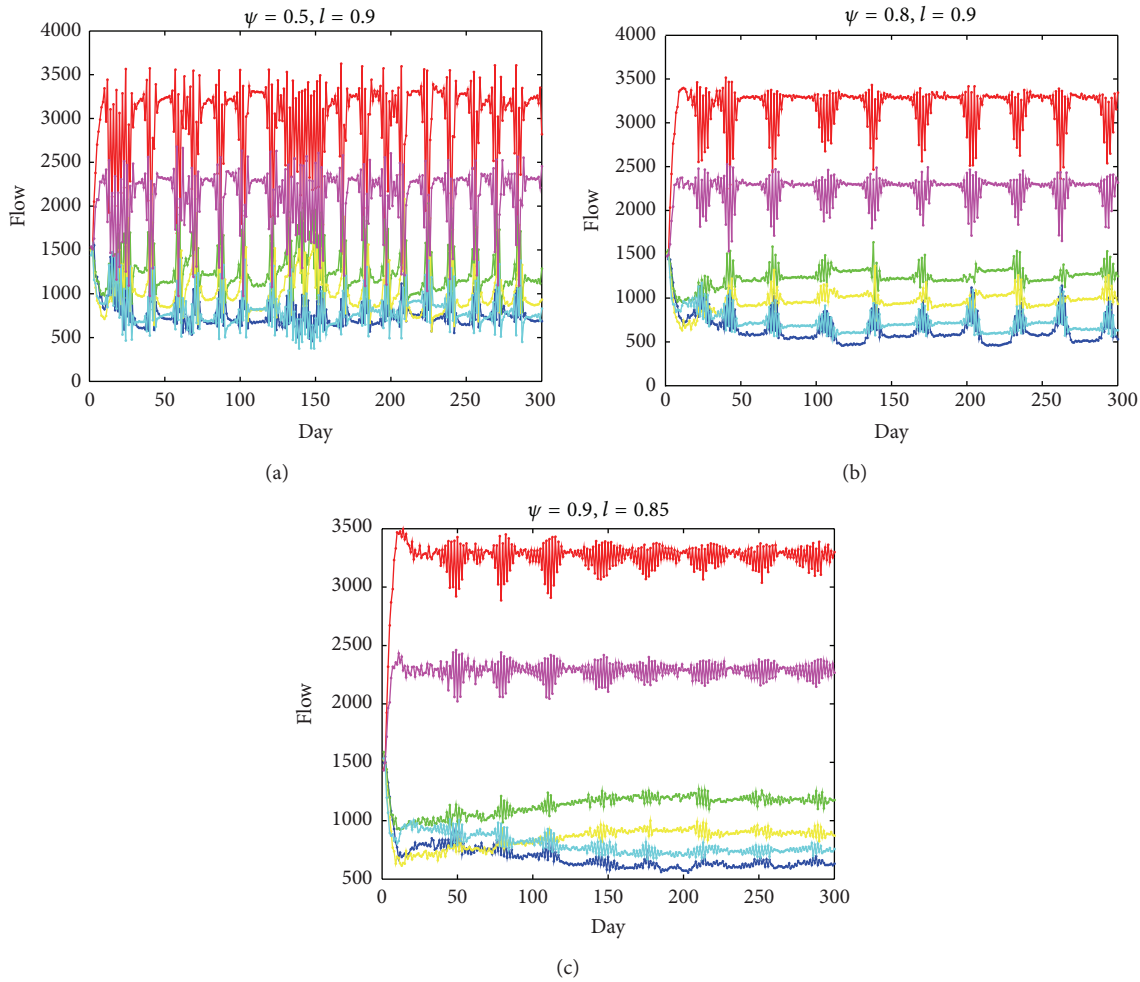


FIGURE 18: The cyclical oscillation in the flow evolution under some combinations of ψ and l.

The travel time of the path p can be represented as

$$T_{pw} = \sum_a \delta_{ap}^w T_a, \quad \forall p \in P_w,\ w \in W. \tag{15}$$

So the expected travel time of the path p is

$$E\left(T_{pw}\right) = \sum_a \delta_{ap}^w E\left(T_a\right). \tag{16}$$

Formula (14) means that when the deterministic capacities are equal to the expected values of the actual capacities, the expected travel time of links under actual capacities are longer than those under deterministic capacities. So the expected travel time of paths under actual capacities is also longer than those under deterministic capacities according to Formula (16). This is why the mean travel time of paths in Table 5 is longer than those in Table 4.

Another phenomenon is that when links' capacities change from deterministic values to random values, the flow of the paths which have the smallest mean travel time decreases by comparing Table 5 with Table 4. It is because the travel time standard deviations of these paths are bigger than most of other paths. Then the fluctuations of the paths' travel time are stronger, respectively. In other word, the stabilities of these paths' travel time are not better. So the flow on them decreases. It suggests that travelers not only consider the mean travel time of a path, but also consider the stability of the path travel time.

5. Conclusions

In this paper, we propose a day-to-day route choice model based on reinforcement learning. Travelers' memory level and learning rate are incorporated into the model. Travelers obey a stochastic route choice rule and use their experience-based cognition to evaluate their choice and then update the chosen probabilities of paths.

The rationality of the model is verified firstly. We assume that travelers can remember all the travel time they have experienced. Under the ideal hypothesis, the flow evolution based on our model in three different types of networks can all converge to UE. It is known that UE is equilibrium under ideal condition. So our model is verified to be reasonable.

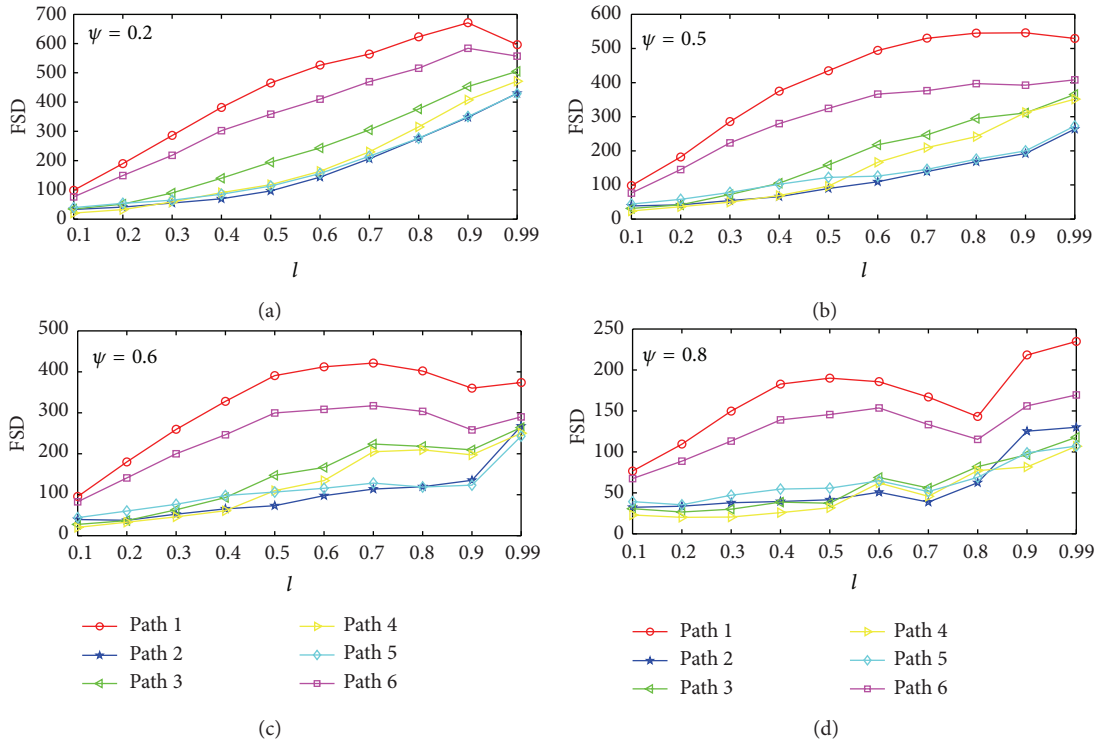


Figure 19: The changes of FSD with the increase of l.

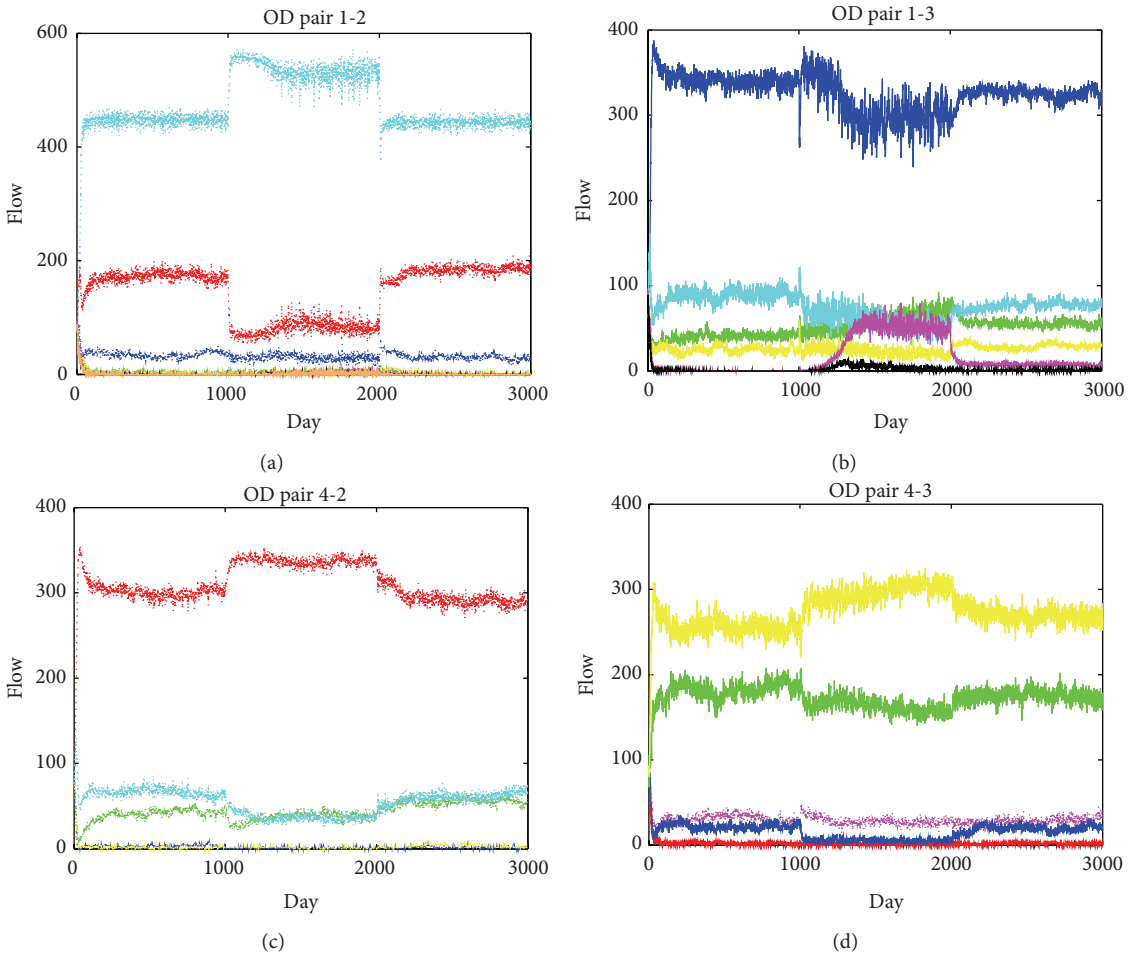


Figure 20: Flow evolution of the third test network in the first link's capacity reduction scenario.

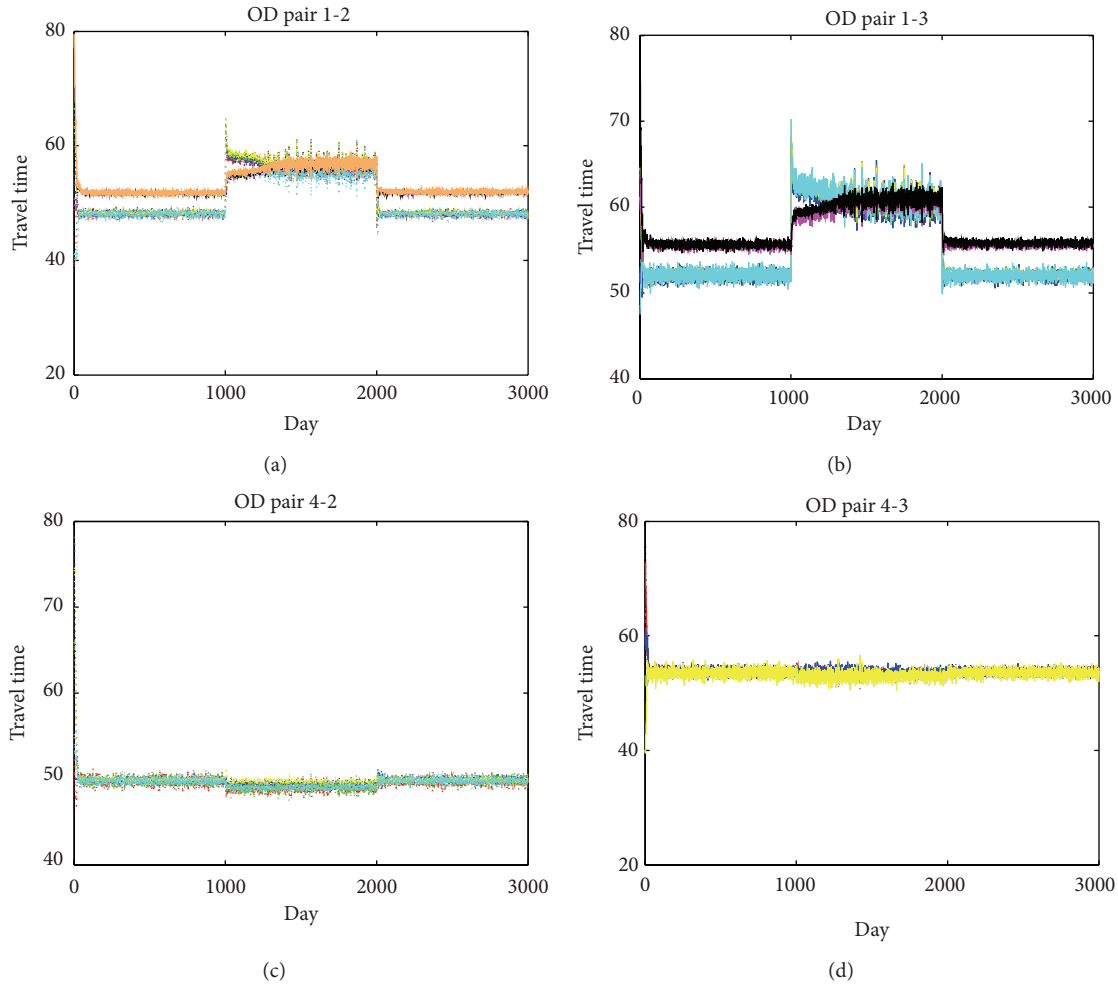


Figure 21: Travel time evolution of the third test network in the first link's capacity reduction scenario.

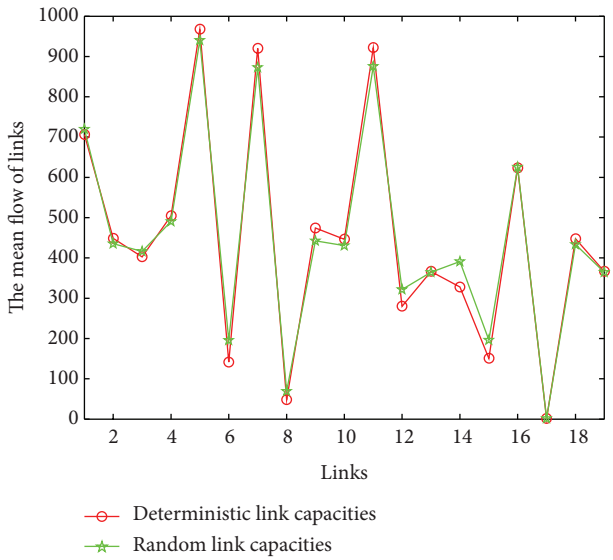


Figure 22: The mean flow of 19 links under deterministic link capacities and random link capacities.

Secondly, the properties of the model are analyzed. And we find that the network flow does not necessarily converge to UE under limited memory level. High learning rate can lead to cyclical oscillation in the process of flow evolution (as shown in Figure 18). Memory level can weaken the flow fluctuations, but learning rate can strengthen the fluctuations. When memory level is larger than 0.6, there are two inflection points in the process of flow fluctuations with the increase of learning rate (as shown in Figure 19). The first inflection point is caused by the opposite effect of memory level and learning rate on flow fluctuations. The second inflection point results from the cyclical oscillation which appears in flow evolution with high learning rate.

Two particular traffic scenarios are analyzed to illustrate the applications of our model. The first scenario is link capacity degradation. We find that the initial network equilibrium can be restored by revoking the link capacity degradation in our model. It is because when the capacity recovers, travelers can form similar cognition on the network with that formed before the capacity reduction. The second scenario is that links' capacities vary randomly because of bad

weather, accident, and so forth. The results in this scenario indicate the following: random variations of links' capacities can change the equilibrium of paths flow but cannot change the equilibrium of links flow, and these also increase the mean travel time of paths; when travelers choose route, they think about not only the mean travel time of paths, but also the stability of paths' travel time. Analyses and applications of our model demonstrate the model is reasonable and useful for studying the day-to-day traffic dynamics.

Some phenomena in the paper need to be further researched, such as the cyclical oscillation in the flow evolution as well as the influence of links' capacities' random variations on the network equilibrium. And in addition to experience, memory level, and learning rate, there are many other factors that can influence travelers' route choice, such as travel habits, travel purpose, and traffic information. The comprehensive influence of these factors on travelers' behaviors is also our future research. At the same time, empirical studies are necessary to investigate travelers' route choice behaviors.

Conflict of Interests

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

Acknowledgments

The Project is supported by National Natural Science Foundation of China (Grant nos. 70971094, 50908155, 71101102, and 71271150).

References

[1] X. He, X. Guo, and H. X. Liu, "A link-based day-to-day traffic assignment model," *Transportation Research B: Methodological*, vol. 44, no. 4, pp. 597–608, 2010.

[2] D. Watling and M. L. Hazelton, "The dynamics and equilibria of day-to-day assignment models," *Networks and Spatial Economics*, vol. 3, no. 3, pp. 349–370, 2003.

[3] M. J. Smith, "The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov," *Transportation Science*, vol. 18, no. 3, pp. 245–252, 1984.

[4] T. L. Friesz, D. Bernstein, N. J. Mehta, R. L. Tobin, and S. Ganjalizadeh, "Day-to-day dynamic network disequilibria and idealized traveler information systems," *Operations Research*, vol. 42, no. 6, pp. 1120–1136, 1994.

[5] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation Science*, vol. 11, no. 3, pp. 253–274, 1977.

[6] E. Cascetta, "A stochastic process approach to the analysis of temporal dynamics in transportation networks," *Transportation Research B*, vol. 23, no. 1, pp. 1–17, 1989.

[7] F. Yang and H. X. Liu, "A new modeling framework for travelers day-to-day route choice adjustment processes," Transportation and Traffic Theory 2007. Papers Selected for Presentation at ISTTT17, 2007.

[8] K. Parry and M. L. Hazelton, "Bayesian inference for day-to-day dynamic traffic models," *Transportation Research B: Methodological*, vol. 50, pp. 104–115, 2013.

[9] D. Watling, "Stability of the stochastic equilibrium assignment problem: a dynamical systems approach," *Transportation Research B: Methodological*, vol. 33, no. 4, pp. 281–312, 1999.

[10] D. Watling, "Urban traffic network models and dynamic driver information systems," *Transport Reviews*, vol. 14, no. 3, pp. 219–246, 1994.

[11] R. J. F. Rossetti, R. H. Bordini, A. L. C. Bazzan, S. Bampi, R. Liu, and D. V. Vliet, "Using BDI agents to improve driver modelling in a commuter scenario," *Transportation Research C: Emerging Technologies*, vol. 10, no. 5-6, pp. 373–398, 2002.

[12] S. Nakayama, R. Kitamura, and S. Fujii, "Drivers' learning and network behavior dynamic analysis of the driver-network system as a complex system," *Transportation Research Record*, vol. 1676, no. 1, pp. 30–36, 1999.

[13] S. Nakayama and R. Kitamura, "Route choice model with inductive learning," *Transportation Research Record*, vol. 1725, no. 1, pp. 63–70, 2000.

[14] S. Nakayama, R. Kitamura, and S. Fujii, "Drivers'route choice rules and network behavior: do drivers become rational and homogeneous through learning?" *Transportation Research Record*, no. 1752, pp. 62–68, 2001.

[15] M. Jha, S. Madanat, and S. Peeta, "Perception updating and day-to-day travel choice dynamics in traffic networks with information provision," *Transportation Research C: Emerging Technologies*, vol. 6, no. 3, pp. 189–212, 1998.

[16] R. Chen and H. S. Mahmassani, "Travel time perception and learning mechanisms in traffic networks," *Transportation Research Record*, vol. 1894, no. 1, pp. 209–221, 2004.

[17] F. Klügl and A. L. C. Bazzan, "Route decision behaviour in a commuting scenario: simple heuristics adaptation and effect of traffic forecast," *Journal of Artificial Societies and Social Simulation*, vol. 7, no. 1, 2004.

[18] A. L. C. Bazzan and F. Klugl, "Learning to behave socially and avoid the braess paradox in a commuting scenario," in *Proceedings of the 1st International Workshop on Evolutionary Game Theory for Learning in MAS*, 2003.

[19] H. Dia, "An agent-based approach to modelling driver route choice behaviour under the influence of real-time information," *Transportation Research C: Emerging Technologies*, vol. 10, no. 5-6, pp. 331–349, 2002.

[20] T.-L. Liu and H.-J. Huang, "Multi-agent simulation on day-to-day route choice behavior," in *Proceeding of the 3rd International Conference on Natural Computation (ICNC '07)*, vol. 5, pp. 492–498, Haikou, China, August 2007.

[21] J. Illenberger, G. Flotterod, and K. Nagel, "A model of risk-sensitive route-choice behavior and the potential benefit of route guidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 384–389, 2011.

[22] F. Gao and M.-Z. Wang, "Route choice behavior model with guidance information," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 6, pp. 64–69, 2010.

[23] R. Lahkar and R. M. Seymour, "Reinforcement learning in population games," *Games and Economic Behavior*, vol. 80, pp. 10–38, 2013.

[24] J. G. Cross, "A stochastic learning model of economic behavior," *The Quarterly Journal of Economics*, vol. 87, no. 2, pp. 239–266, 1973.

[25] E. Ben-Elia and Y. Shiftan, "Which road do I take? A learning-based model of route-choice behavior with real-time information," *Transportation Research A: Policy and Practice*, vol. 44, no. 4, pp. 249–264, 2010.

[26] M. M. Wahba, *Microsimulation Learning-based Approach to Transit Assignment*, University of Toronto, 2008.

[27] M. Wahba and A. Shalaby, "Large-scale application of MILATRAS: case study of the Toronto transit network," *Transportation*, vol. 38, no. 6, pp. 889–908, 2011.

[28] M. Wahba and A. Shalaby, "Learning-based framework for transit assignment modeling under information provision," *Transportation*, vol. 41, no. 2, pp. 397–417, 2014.

[29] M. Zolfpour-Arokhlo, A. Selamat, S. Z. Mohd Hashim, and H. Afkhami, "Modeling of route planning system based on Q value-based dynamic programming with multi-agent reinforcement learning algorithms," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 163–177, 2014.

[30] M. W. Macy and A. Flache, "Learning dynamics in social dilemmas," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, supplement 3, pp. 7229–7236, 2002.

[31] S. Lasaulce and H. Tembine, *Game Theory and Learning for Wireless Networks: Fundamentals and Applications*, Academic Press, 2011.

[32] H. Xu, Y. Lou, Y. Yin, and J. Zhou, "A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing," *Transportation Research B: Methodological*, vol. 45, no. 2, pp. 311–328, 2011.

[33] X. Guo and H. X. Liu, "Bounded rationality and irreversible network change," *Transportation Research B: Methodological*, vol. 45, no. 10, pp. 1606–1618, 2011.

[34] Q. Wang and W. Xu, "A user equilibrium model based on cumulative prospect theory for degradable transport network," in *Proceeding of the 4th International Joint Conference on Computational Sciences and Optimization (CSO '11)*, pp. 1078–1082, Yunnan, China, April 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 368208, 11 pages
http://dx.doi.org/10.1155/2014/368208

Research Article

A Schedule Optimization Model on Multirunway Based on Ant Colony Algorithm

Yu Jiang, Zhaolong Xu, Xinxing Xu, Zhihua Liao, and Yuxiao Luo

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 210016, China

Correspondence should be addressed to Yu Jiang; jiangyu07@nuaa.edu.cn

Received 11 April 2014; Revised 7 July 2014; Accepted 18 July 2014; Published 28 September 2014

Academic Editor: Hu Shao

In order to make full use of the slot of runway, reduce flight delay, and ensure fairness among airlines, a schedule optimization model for arrival-departure flights is established in the paper. The total delay cost and fairness among airlines are two objective functions. The ant colony algorithm is adopted to solve this problem and the result is more efficient and reasonable when compared with FCFS (first come first served) strategy. Optimization results show that the flight delay and fair deviation are decreased by 42.22% and 38.64%, respectively. Therefore, the optimization model makes great significance in reducing flight delay and improving the fairness among all airlines.

1. Introduction

With the rapid development of the Chinese civil aviation industry, the number of flights increases sharply and hub airports change their single runway to multirunway. The air traffic control managers put first come first served (FCFS) to use to schedule the arrival-departure flights, which can lead to the waste of air resources and make the terminal area congestion more serious. Therefore, the conflict between demand and supply is more and more sharp Hu and Paolo [1]. Meanwhile, with the implementation of collaborative decision making (CDM) mechanism in airport resources management, flight scheduling problem should account for not only flight delay but also equity among airlines. A more efficient and reasonable model or algorithm is needed to approach the problem. Therefore, the air traffic congestion will be alleviated and the total operation cost of airlines will be reduced by approaching the scheduling and optimization of arrival-departure flights.

In recent years, the capacity of most airports cannot satisfy the rapid increase demand because of the increase in number of flights and severe weather. At present, ground delay procedure (GDP) is the main method to be used to approach the contradiction between capacity supply and demand [2]. With the improvement of airport resource management technique, a modified GDP strategy, collaborative ground delay program (CDM-GDP) has been implemented in some hub airports. The modified program can improve the utilization efficiency of airport resources observably [3]. The basic process of CDM-GD is that the air traffic control departments allocate landing slots to airlines; then the airlines can merge and cancel the flights according to the allocated information of slots and feed the adjustment information back to the air traffic control management, and then the air traffic control departments make the final decision after receiving the feedback information from airlines [4]. Many scholars at home and abroad have done some research in CDM-GDP. Hoffman et al. [5] discussed several tools applied to CDM system and a new algorithm ration-by-schedule (RBS) was proposed. Vossen et al. [6] described a new allocation procedure based on FCFS in CDM strategy to schedule the arrival-departure flights. They established a equity allocation mechanism, but the efficiency of algorithm should be improved. Mukherjee and Hansen [7] put forward a dynamic stochastic integer programming (IP) model for the airport ground holding problem, but the equity was ignored for small airlines. Ball et al. [8] presented a new ration-by-distance (RBD) algorithm showing that the equity and efficiency were improved at a certain extent. Hu and Su [9] modeled the ground-holding management system, which

provides the theoretical basis and method for the actual traffic management, but they ignored the influence of limited capacity for taking off flights. Ma et al. [10] designed a scheme of approaching queue and optimization schedule of flights. But the satisfaction function is just a local optimization and it is too complicated to make an adjustment in the simulation. Zhou et al. [11] proposed an effectiveness-fairness (E-E) standard aimed at arrival slot time allocated method by analyzing and simulating the traffic flow model in CDM-GDP. The simulation results showed that the single priority decreased the total delay cost more efficiently than the double priority and the fair factors were also taken into account. Zhou et al. [12] proposed an evaluation function which was used to evaluate the priority of delay cost coefficient on the basis of existing slot allocation algorithm. The method used in this paper was more flexible and effective compared with the traditional method on the total delay cost and equity, but it lacked the flexibility in slot time allocation for arrival flights. Zhang and Hu [13] proposed a multiobjective optimization model based on the principles of effectiveness-efficiency-equity trade-offs. But the research lacked further research into the slot time reassignment in CDM GDP based on the real information of aircraft. Zhan et al. [14] adopted the Ant Colony Algorithm (ACA) and receding horizon control (RHC) to optimize the scheduling from the robustness and effectiveness of the queue model on arrival flights. But the instability of the algorithm should be improved and the real scheduling of multirunway should be taken into account. Andrea D'Ariano et al. [15] studied the problem of flight sorting at congested airports. The research regarded the flight scheduling problem as an extension of workshop scheduling problem, but the optimized results were suboptimal feasible solutions. Helmke et al. [16] presented an integrated approach to solve mixed-mode runway scheduling problem by using mixed-integer program techniques. However, further researches into the flight scheduling problem under multirunway with mixed operation were required. Samà et al. [17] presented the real-time scheduling flight in order to reschedule the flight based on receding horizon control strategy and conflict detection. However, the optimization approaches requiring frequent retiming and rerouting in consecutive time horizons decreased the scheduling robustness. Hancerliogullari et al. [18] researched into the aircraft sequencing problem (ASP) under multirunway with mixed operation mode. They put greedy algorithm to simulate the model. However, the fairness among airlines was not taken into consideration.

All in all, though the optimization results of most researches at home and abroad satisfied the scheduling of flight, the studies pay little attention to real-time flight information. Most of the studies consider collaborative ground delays program of approach flight without analyzing departure flight. In fact, if the delay of departure flight is dealt with unreasonably, it can lead to unfairness among arrival-departure flights and the increase of flight delay. In the paper, a multiobjective optimization model is established based on multirunway arrival-departure flight. The maximum delay of departure flight is limited to ensure the fairness between arrival-departure flights according to real-time flight information. Meanwhile, a fair runway slot allocation mechanism is established with the objective of minimizing the cost caused by airline delays. As Ant Colony Algorithm has unique advantages in continuous dynamic optimization, the paper introduces it to simulate and validate the model with the expectation of reducing the loss of airline delays, improving runway utilization, and ensuring the fairness among airlines.

2. Model

2.1. Description. The queue of arrival-departure flights in multirunway airport is a continuous dynamic process, and it changes with the real-time information of flights. The schedule optimization of arrival-departure flight in multi-runway airport can be described as follows: within a time window, a number of flights belonging to different airlines are waiting for landing or taking off. The managers in airport should make a reasonable allocation schedule (such as arrival-departure time, sequence, and operation runway) for all flights to minimize the total delay cost in the study period under the condition of safe operation of flight and airport resources and to balance the cost of total delay among airlines. The paper selects the research time period in rush hour in hub airport to study the airport surface operation. After modeling and simulation, the results can be applied in the management of airport surface operation in any type of airports.

2.2. Assumptions

(1) The parallel runways studied in the paper operate independently.

(2) In the research time period, the runway capacity cannot meet the demands of flights.

(3) All the arrival flights do not delay when they are in the take-off airport, and they can arrive at the destination terminal aerospace studied on time.

(4) The basic information (such as flight plans and other information of all flights) within the studied period is known.

(5) Each arrival-departure flight can only be assigned to one time slot in the studied period.

2.3. Definition

F^A: Set of arrival flight, $F^A = \{f_1^A, f_2^A, \ldots, f_m^A\}$

F^D: Set of departure flight, $F^D = \{f_1^D, f_2^D, \ldots, f_n^D\}$

F: Set of arrival-departure flights, $F^A \cup F^D = F$

H: Set of airlines, $H = \{H_1, H_2, \ldots, H_p\}$

S: Set of arrival-departure flights slots, $S = \{s_1, s_2, \ldots, s_W\}$

R: Set of runways in the airport, $R = \{1, 2, \ldots, r\}$

$f_{rti} = \begin{cases} 1 & \text{if flight } i \text{ takes off or lands on runway } r \\ 0 & \text{else} \end{cases}$

$f_{H_a}^i$: Flight i belongs to airline H_a

$s_g^{f_{H_a}^{Ai}} = \begin{cases} 1 & \text{if flight slot } s_g \text{ is assigned to an arrival flight } f_{H_a}^{Ai} \\ 0 & \text{else} \end{cases}$

$s_h^{f_{H_a}^{Di}} = \begin{cases} 1 & \text{if } s_h \text{ is assigned to adeparture flight } f_{H_a}^{Di} \\ 0 & \text{else} \end{cases}$

C: Total flight delay cost

C_{H_a}: Total flight delay cost of airline H_a

PC: The sum of absolute deviation of flight delay cost

$\alpha_r = \begin{cases} 1 & \text{if any flight lands on runway } r \\ 0 & \text{else} \end{cases}$

$\beta_r = \begin{cases} 1 & \text{if any flight takes off on runway } r \\ 0 & \text{else} \end{cases}$

ORT_{rf_b}: The end time of flight f_b taking off from or landing on runway r after optimization

STf_{rf_c}: The original time of flight f_c taking off from or landing on at runway r after optimization

S_{rbc}^{A}: The minimum safety interval of continuous landing on runway r

S_{rbc}^{D}: The minimum safety interval of continuous taking off on runway r

S_{rbc}^{AD}: The minimum time interval when a departure flight follows an arrival flight on runway r

S_{rbc}^{DA}: The minimum time interval when an arrival flight follows a departure flight on runway r

$TJTY_{\max}$: The maximum delay time when an arrival flight lands in advance compared to scheduled time

$TJZY_{\max}$: The maximum delay time when an arrival flight lands later than the scheduled time

$TLTY_{\max}$: The maximum delay time when a departure flight takes off in advance compared to the scheduled time

$TLZY_{\max}$: The maximum delay time when a departure flight takes off later than the scheduled time

$ETf_{H_a}^{Ai}$: The estimated arrival time of flight f_i^A which belongs to H_a

$ETf_{H_a}^{Dj}$: The estimated departure time of flight f_j^D which belongs to H_a

$STf_{H_a}^{Ai}$: The actual arrival time of flight f_i^A belonging to H_a after optimization schedule

$STf_{H_a}^{Dj}$: The actual departure time of flight f_j^D belonging to H_a after optimization schedule

$C_{H_a}^{Ai}$: The unit time delay cost of arrival flight f_i^A belonging to H_a after optimization schedule

$C_{H_a}^{Dj}$: The unit time delay cost of departure flight f_j^D belonging to H_a after optimization schedule.

2.4. Objective Function. Two objectives are taken into consideration in the paper: total delay cost and the fairness among airlines. The total delay cost of all flights is used to reduce the light delay and improve the utilization of runway. The fairness is used to balance the equity among all the airlines and to protect the benefit of small airlines. So a multiobjective function based on it is modeled.

2.4.1. The Objective Function of Delay Cost. Different wake vortex separations between different arrival-departure flights are different according to the types of aircraft. Therefore, we can improve the capacity of runway and reduce total delay time by adjusting the arrival-departure order of all flights. ICAO aircraft wake turbulence separation criteria are specified in Table 1.

The optimized target of delay cost in the paper is to minimize the total delay of all arrival-departure flights, which is based on improving the capacity of runway. The objective function of delay cost can be described as follows:

$$\begin{aligned} \min C &= \min \sum_{r=1}^{r} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{a=1}^{p} \left[C_{H_a}^{Ai} \left(STf_{H_a}^{Ai} - ETf_{H_a}^{Ai} \right) \alpha_r s_g^{f_{H_a}^{Ai}} \right. \\ &\quad \left. + C_{H_a}^{Dj} \left(STf_{H_a}^{Dj} - ETf_{H_a}^{Dj} \right) \beta_r s_h^{f_{H_a}^{Di}} \right]. \end{aligned} \tag{1}$$

2.4.2. The Objective Function of Fairness. Delay cost is related to the aircraft type. In general, large airlines are preferred to small aircraft types. If the research only takes the delay cost as a single function, it is likely to lead to serious unfairness among airlines, especially to small airlines. Therefore, absolute deviation of delay cost is introduced to ensure the fairness among airlines.

Definition of standard flight: assume some type of flight to be a standard flight and all other flights can be transformed into it according to aircraft type and delay cost. For example, if we take a large aircraft as a standard flight and the number equals 1, then a light aircraft may be transformed as 0.6 and a heavy aircraft as 1.8. If a standard flight is denoted by $\overline{f}_B$, $\lambda_{H_a}^i$ is defined as the number of standard flights after transformation from flight i; then the relation expression can be demonstrated as follows:

$$\lambda_{H_a}^i = \frac{f_{H_a}^i}{\overline{f}_B}. \tag{2}$$

In order to ensure fairness among airlines, the researchers first transform all flights belonging to different airlines into standard flights. Then the researchers can calculate the average delay cost of standard flight by using the total delay cost of all flights. In the same way, the researcher can get average delay cost of each airline and the total absolute deviation of delay cost. It is obvious that the lower the total absolute deviation is, the more fairness we can balance among airlines. The fairness optimization objective function of airlines is as follows:

$$\min PC = \sum_{a=1}^{p} \left| \frac{C}{\sum_{a=1}^{p} \sum_{H_a} \lambda_{H_a}^i} - \frac{C_{H_a}}{\sum_{H_a} \lambda_{H_a}^i} \right|, \tag{3}$$

where $C_{H_a} / \sum_{H_a} \lambda_{H_a}^i$ is the average delay cost of flights belonging to airline H_a; $C / \sum_{a=1}^{p} \sum_{H_a} \lambda_{H_a}^i$ is the average delay cost of all flights.

Table 1: Aircraft wake turbulence separation criteria of ICAO.

Type of flight	Trailing					
	The minimum time interval/s			The minimum distance interval/km		
	Small	Large	Heavy	Small	Large	Heavy
Leading						
Slight	98	74	74	6	6	6
Large	138	74	74	10	6	6
Heavy	167	114	94	12	10	8

2.5. Constraints. Consider

$$\sum_{g=1}^{W} s_g^{f_{H_a}^{Ai}} = 1. \tag{4}$$

An arrival flight occupies one slot resource and each slot resource can be assigned to a certain arrival flight. Constraint (4) is the constraint of slot resource allocation for arrival flights:

$$\sum_{h=1}^{W} s_h^{f_{H_a}^{Di}} = 1. \tag{5}$$

Similarly, a departure flight occupies one slot resource and each slot resource can be assigned to a certain departure flight. Constraint (5) is the constraint of slot resource allocation for departure flights:

$$f_{rti} \le 1. \tag{6}$$

For the safety of flight operation, each flight can only occupy one runway and only one or none aircraft can occupy the runway at the same time. A constraint of slot resource allocation of runways is expressed as constraint (6):

$$\begin{gathered} \sum f_{rti} \le r \\ \alpha_r + \beta_r \le 1 \quad \alpha_r \in \{0,1\} \quad \beta_r \in \{0,1\}. \end{gathered} \tag{7}$$

According to constraint (6), at a certain time, the number of flights occupying runways will not be greater than the number of runways. Meanwhile, in order to ensure safety in a certain slot, only one flight is arriving or leaving at runway r. The runway resource controlling constraint is shown as constraint (7):

$$TJTY_{\max} \le STf_{H_aDj} - ETf_{H_aDj} \le TJZY_{\max}. \tag{8}$$

To ensure the safety of arrival flights, the actual arrival time should meet the maximum delay time ($TLTY_{\max}$ and $TLZY_{\max}$) after optimization. Constraint (8) is to ensure the time of arrival flight:

$$TJTY_{\max} \le STf_{H_aDj} - ETf_{H_aDj} \le TJZY_{\max}. \tag{9}$$

Constraint (9) is to ensure the time of departure flight. Extending the departure flight delay can decrease the service level in most airports and it is necessary to limit the amount of delay time. Like the constraint for arrival flights, the actual departure time should meet the maximum delay time ($TJTY_{\max}$ and $TJZY_{\max}$) after optimization. Consider

$$ORT_{rf_b} - STf_{rf_c} \ge \begin{cases} S_{rbc}^{A} & \text{continue arrivals} \\ S_{rbc}^{D} & \text{continue departures} \\ S_{rbc}^{AD} & \text{departure follows arrival} \\ S_{rbc}^{DA} & \text{arrival follows departure.} \end{cases} \tag{10}$$

The minimum safety interval in different conditions which is related to the order of arrival-departure flight should be taken into account. Constraint (10) is the minimum flight safety interval constraint of runway r.

3. Ant Colony Algorithm Design

Ant Colony Algorithm (ACA) is a metaheuristic algorithm, which uses a heuristic method to search for the space that may be related to feasible solutions. In ACA, the ant can select the path comprehensively based on pheromones and heuristic factors of the environment. It can release the pheromones after traveling the path of the network. In the algorithm, an individual ant can identify and release all the pheromones. All pheromones from the ant colony are used to complete the whole and complex optimization process. ACA has the characteristics of self-organization and distributed computing. Therefore it is able to make a global search. It can effectively avoid local solutions to an extent. Meanwhile, ACA can get the optimized solution faster than other traditional algorithms.

3.1. Algorithm Description

3.1.1. Single Runway Flight Scheduling of ACA. Single runway scheduling problem can be transformed to a TSP which takes each flight as a node and the interval time between flights as the path length of nodes. The problem can be solved by traditional ACA, and the results are satisfactory. The model is built as follows: the node f_i in the network is the element of flight set F and the distance f_{ij} between nodes f_i and f_j is the time interval. When the algorithm begins, the ant k heads from a virtual starting node f_0. The starting node is set up to ensure that all the ants in ACA can start at the same node. The ant k traverses all the nodes of network, so a flight sequence is obtained. We can calculate wait time and delay cost in the queue according to the sequence. The ACA for single runway flight scheduling is shown in Figure 1.

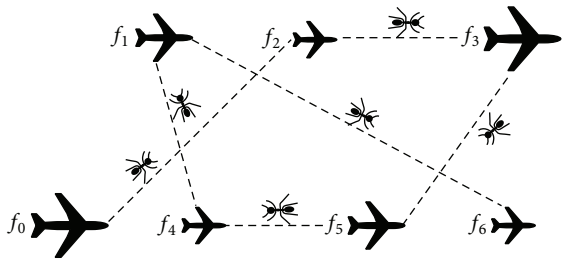


Figure 1: The ACA for single runway schedule model.

3.1.2. Multirunway Flight Scheduling of ACA. In order to adapt to the multirunway flight scheduling model, the ACA for single runway model should be modified. In the multirunway flight scheduling model, a node f_i may contain several subnodes $r_n \in R$. The distance $l_{i_m j_n}$ between the subnode r_m of f_i and the subnode r_n of node f_j is the minimum safety interval. The ant k heads from a virtual starting node f_0 and travels all nodes of the network. When the ant arrives at a node f_i, it selects a subnode r_n to get a sequence contained runway number. The ACA for multirunway flight scheduling is shown in Figure 2.

3.1.3. The State Transition Equation. The amount of information on each path and the heuristic information can decide the transition direction of ant k ($k = 1, 2, \ldots, m$). It records the traveled nodes by search table tabu_k ($k = 1, 2, \ldots, m$). $p^k_{i_m j_n}(t)$ is denoted as the state transition probability of ant k changing direction from the subnode r_m to the subnode r_n at time t. Formula (11) is as follows:

$$p^k_{i_m j_n}(t) = \begin{cases} \dfrac{\left[\tau_{i_m j_n}(t)\right]^\alpha * \left[\eta_{j'_n, j_n}(t)\right]^\beta}{\sum_{\kappa \in \text{allowed}_k} \left(\sum_{n \in R} \left[\tau_{i_m \kappa_n}(t)\right]^\alpha * \sum_{n \in R} \left[\eta_{\kappa'_n, \kappa_n}(t)\right]^\beta\right)}, & \text{if } \kappa \in \text{allowed}_k \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

$\tau_{i_m j_n}(t)$ is the pheromone concentration on path $l_{i_m j_n}$ at time t. $\text{allowed}_k = \{F - \text{tabu}_k\}$ means that the ant k can choose the node which is the node that never traversed next step. α is the pheromone heuristic factor, which decides how the pheromones have impact on path choosing; β is the expected heuristic factor, which decides the degree of attention of visibility when ants make a choice.

$\eta_{j'_n, j_n}(t)$ is an expected factor, which is evaluated as

$$\eta_{j'_n, j_n}(t) = \frac{1}{r_{j'_n, j_n}}, \tag{12}$$

where $r_{j'_n, j_n}$ is the minimum safety interval of the flight f_j and former flight f'_j.

3.1.4. The Updating Strategy of Pheromone. Updating pheromone of all nodes is needed when all the iterations are completed. With the increasing of pheromone concentration, the residual pheromone evaporates in proportion. In order to get better optimization results, only the best ant can release pheromone of iteration. Therefore, the pheromone updating can be adjusted as the following rules:

$$\begin{aligned} \tau_{i_m j_n}(t+1) &= \rho \tau_{i_m j_n}(t) + \Delta \tau^{\text{best}}_{i_m j_n}, \\ \Delta \tau_{i_m j_n} &= \begin{cases} \dfrac{Q}{f\left(k^{\text{best}}\right)}, & i \in k^{\text{best}} \\ 0, & i \notin k^{\text{best}}, \end{cases} \end{aligned} \tag{13}$$

where ρ is the volatilization coefficient of pheromone; Q is the amount of pheromone; $\Delta \tau_i$ is the total incremental of pheromone in this circulation of node i.

3.2. The Design of ACA. The design of ACA in the simulation is as follows.

Step 1. Set parameters. We set $\alpha = 2$, $\beta = 1.5$, $\rho = 0.7$, and $Q = 120000$.

Step 2. Get the flight information. We get flight information (including the type of flight, the estimated time of arrival or departure, and so forth) and other known data by reading the files.

Step 3. Initialize the pheromone and expectations of paths of solution space and empty the tabu list.

Step 4. Set $NC = 0$ (NC is the iteration). Generate the initial k ants on the virtual node f_0.

Step 5. Ants select a node orderly according to formula (11).

Step 6. If all ants complete a traversal, turn to Step 7, otherwise turn to Step 5.

Step 7. If the searching results of all ants meet the constraints, then reduce the pheromone increment when updating pheromone.

Step 8. Calculate the target value of all ants and record the best ant solutions.

Step 9. Update the pheromone of each node according to formula (13).

Step 10. If $NC < NC_{\max}$, without stagnating, delete the ant and set $NC \rightarrow NC + 1$; then reset the data and turn to Step 5; otherwise output the optimal results and the calculation is over.

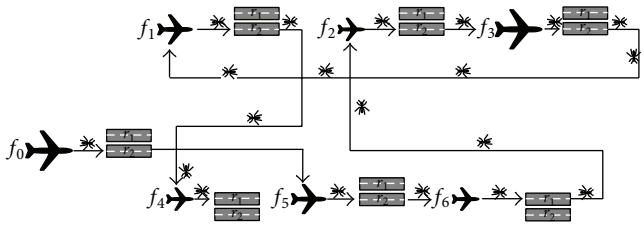


Figure 2: The ACA for multirunway schedule model.

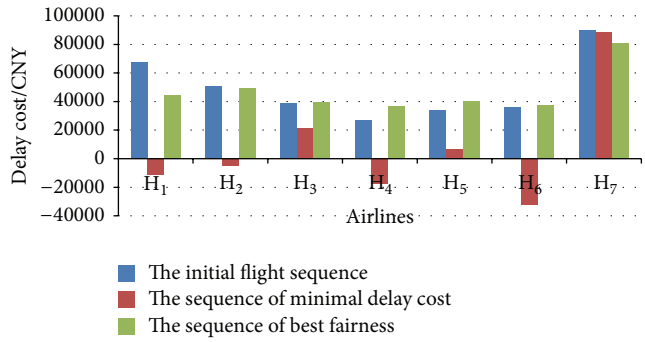


Figure 3: The delay cost comparison of airlines on three flight sequences.

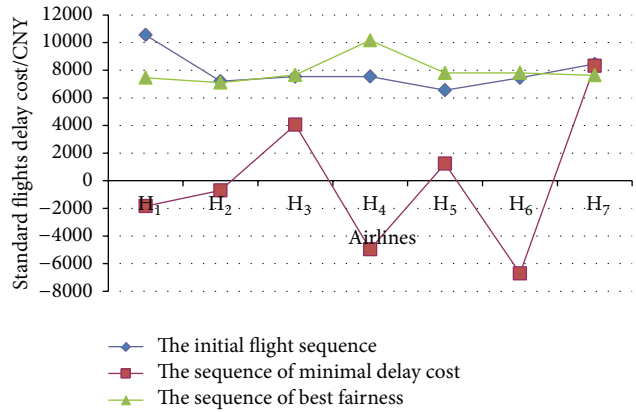


Figure 4: The delay cost comparison of airlines standard flights on three flight sequences.

4. Simulation and Vertification

In the paper, we put C program to use to simulate flight schedule problem on multirunway with mixed operation mode. The core algorithm of the program is the ACA design. In order to achieve the objectives of delay cost and fairness, we first take the delay cost as the objective and we take fairness as a constraint and we can get the flight sequence with minimal delay cost. After that, we set fairness as the objective and the delay cost as a constraint, and we get a flight sequence with the best fairness. Finally, the initial flight sequence and these two flight sequences are compared. A peak hour is selected from a certain large airport of China and the authors chose flights in the busiest 15 minutes from that peak hour. Two parallel runways run independently. There are 38 flights (belonging to 7 airlines) to be scheduled. After optimization, the initial flight information and optimized results are shown in Table 2.

In the paper, some real operation data is selected from a certain hub airport and ACA is designed to solve the model. As shown in Table 2, flight delay is serious due to unreasonable slot assignment; it can lead to unfair competition among airlines before optimization. The total delay cost of three types of flight sequences is listed in Table 3 and Figure 3.

The initial sequence is the initial flight sequence before optimization. The minimal delay cost and fairness among airlines are solved according to object function (1) and function (2). The minimal delay sequence means that we take the objective function (1) as the main objective function and objective function (2) as a constraint in optimization. The best fairness sequence means that we set objective function (2) as the main objective function and the objective function (1) as a constraint in optimization. Then the paper contrasts the three flight data among airlines.

The detailed information of delay cost of standard flight is listed in Table 4 and Figure 4. After transforming flights into standard flights, we can compare delay cost and fairness directly.

From Figures 3 and 4, we can draw the conclusion that the sequence of minimal delay cost can decrease the delay cost obviously after optimization. The sequence of best fairness can improve the fairness among all the airlines obviously. But from Figure 4, we can see that the fairness among 7 airlines decreases when delay cost is minimal; the delay cost of 7 airlines improves obviously when the fairness is best.

In order to reduce the delay cost of airlines and increase the fairness among airlines, we view the minimum delay cost as objective and control the range of flight delay cost variation. The simulation steps are as follows.

(1) First, the fairness is not taken into account and we get the minimum delay cost. We statistic the total delay and the delay cost deviation.

(2) Then limit the range of airline flight delay cost deviation by a large number of data. We get the simulated data of total delay and delay cost deviation of flights in the cases $PC < 50000$, $PC < 30000$, $PC < 25000$, $PC < 20000$, $PC < 15000$, respectively.

(3) Finally, we make an analysis of simulation data and study the relationship between the delay cost and fairness. The relation curve of delay cost and delay cost deviation is shown in Figure 5.

As shown in Figure 5, there is a trend relationship between delay cost and fairness. When the delay cost decreases, the fairness among airlines is not satisfied. Reducing the delay deviation of flights may lead to the increase in total delay cost. In order to make balance of the relationship between the delay cost and fairness and make a better

Table 2: The initial flight delay data.

Flight number	Airline	Type	Unit delay cost of departure	Estimated time of departure	Actual time of departure	Departure runway	The delay cost of actual departure
F001	H_1	S	1.2	0:00:00	0:00:00	0	0
F002	H_2	S	1.1	0:00:00	0:01:36	0	105.6
F003	H_2	L	6.2	0:00:00	0:02:42	0	1004.4
F004	H_3	M	2.4	0:00:00	0:00:00	1	0
F005	H_4	M	2.5	0:05:00	0:04:20	0	−100
F006	H_4	S	1.4	0:05:00	0:08:29	1	292.6
F007	H_5	M	2.4	0:05:00	0:09:35	1	660
F008	H_6	M	2.6	0:05:00	0:08:42	0	577.2
F009	H_6	L	5.7	0:10:00	0:17:41	0	2627.7
F010	H_3	M	2.4	0:10:00	0:19:07	1	1312.8
F011	H_1	M	2.3	0:10:00	0:20:23	1	1432.9
F012	H_2	L	6.2	0:10:00	0:18:59	0	3341.8
F013	H_7	M	2.7	0:10:00	0:20:37	0	1719.9
F014	H_5	S	1.3	0:15:00	0:25:11	1	794.3
F015	H_2	M	2.6	0:15:00	0:25:15	0	1599
F016	H_7	M	2.7	0:15:00	0:26:17	1	1827.9
F017	H_3	S	1	0:15:00	0:27:15	0	735
F018	H_1	L	5.1	0:15:00	0:27:33	1	3840.3
Flight number	Airline	Type	Unit delay cost of approach	Estimated time of approach	Actual time of approach	Approach runway	The delay cost of actual approach
F019	H_7	L	62.6	0:00:00	0:01:14	1	4632.4
F020	H_7	S	25.6	0:01:12	0:04:01	1	4326.4
F021	H_7	M	42.8	0:02:32	0:05:15	1	6976.4
F022	H_5	L	63.2	0:03:56	0:04:16	0	1264
F023	H_4	M	41.5	0:04:02	0:06:29	1	6100.5
F024	H_6	M	41.6	0:04:31	0:06:10	0	4118.4
F025	H_7	S	25.6	0:05:06	0:11:53	1	10419.2
F026	H_3	M	42	0:05:34	0:09:56	0	11004
F027	H_7	M	42.8	0:05:42	0:11:10	0	14038.4
F028	H_1	M	42.4	0:06:54	0:13:07	1	15815.2
F029	H_7	L	62.6	0:07:01	0:12:24	0	20219.8
F030	H_7	L	62.6	0:07:34	0:14:21	1	25478.2
F031	H_4	M	41.5	0:08:03	0:16:15	1	20418
F032	H_3	S	24.7	0:08:43	0:15:11	0	9583.6
F033	H_5	L	63.2	0:09:12	0:17:29	1	31410.4
F034	H_3	M	42	0:09:51	0:16:25	0	16548
F035	H_6	M	41.6	0:10:13	0:21:37	1	28454.4
F036	H_1	M	42.4	0:10:14	0:21:51	0	29552.8
F037	H_2	L	63	0:11:06	0:22:51	1	44415
F038	H_1	M	42.4	0:12:53	0:24:09	0	16900

Table 3: Comparison on delay cost of flights from three flight sequences.

Airline sequence	H_1	H_2	H_3	H_4	H_5	H_6	H_7
				Total delay cost			
Initial	67541.2	50465.8	39183.4	27176.1	34128.7	36070.3	89638.6
Minimal delay	−11009	−4900.6	21132.1	−17837.5	6450.6	−32170.6	88220
Best fairness	44734.4	49719.8	39821.5	36614.3	40545.6	37500.4	80947.4

Table 4: Comparison on delay cost of standard flights from three flight sequences.

Airline sequence	H_1	H_2	H_3	H_4	H_5	H_6	H_7
				Total delay cost			
Initial	11256.87	7209.4	7535.27	7548.92	6563.21	7453.69	8456.47
Minimal delay	−1834.83	−700.086	4063.865	−4954.86	1240.5	−6702.21	8322.642
Best fairness	7455.733	7102.82	7657.981	10170.64	7797.23	7812.583	7636.547

Note: H_1 to H_7 refer to standard flights in Table 4.

Table 5: Comparison on total delay of five flight sequences for flights.

Airline sequence	H_1	H_2	H_3	H_4	H_5	H_6	H_7
				Total delay cost			
Minimal delay	−11009	−4900.6	21132.1	−17837.5	6450.6	−32170.6	88220
Optimization 1	19309.8	5431.1	17095	5972.5	8288.5	28083	58219.5
Optimization 2	33832.6	48721.1	41518.3	17449.4	19336.1	33148.5	53602.4
Optimization 3	16693.8	13682.5	20322.1	15976.9	14430.4	6743.8	33800.8
Best fairness	44734.4	49719.8	39821.5	36614.3	40545.6	37500.4	80947.4

Table 6: Comparison on delay cost of five flight sequences for standard flights.

Airline sequence	H_1	H_2	H_3	H_4	H_5	H_6	H_7
				Total delay cost			
Minimal delay	−1834.83	−700.086	4063.865	−4954.86	1240.5	−6702.21	8322.642
Optimization 1	3218.3	775.8714	3287.5	1659.028	1593.942	5850.625	5492.406
Optimization 2	5638.767	6960.157	7984.288	4847.056	3718.481	6905.938	5056.83
Optimization 3	2782.3	1954.643	3908.096	4438.028	2775.077	1404.958	3188.755
Best fairness	7455.733	7102.829	7657.981	10170.64	7797.231	7812.583	7636.547

Note: H_1 to H_7 refer to standard flights in Table 6.

sequence of flights, we can make flight sequence by limiting the delay deviation of flights. So it not only reduces the total delay cost of airlines but also takes care of the fairness among airlines.

From the simulation results, we can get a number of flight sequences by controlling the sum of absolute deviation of flight delay cost. Five sequences (minimal delay cost, optimization 1, optimization 2, optimization 3, and the best fairness) to make a comparison of total delay cost of airlines are shown in Table 5 and Figure 6. Table 6 and Figure 7 show the same comparison by transforming flights into standard flights.

The total delay cost and the fairness among airlines have been improved obviously after optimization. In actual operation, the decision makers can get several optimized flight sequences by controlling the range of flight delay cost deviation and by selecting a preferred one according to real-time information.

In Table 7, we list the optimal flight sequence in which both the delay cost and fairness are acceptable. The optimized results show that the total delay cost reduces greatly and the fairness is also acceptable when compared with the initial flight sequence. The standard flight delay cost deviation of initial flight sequence and optimized flight sequence is

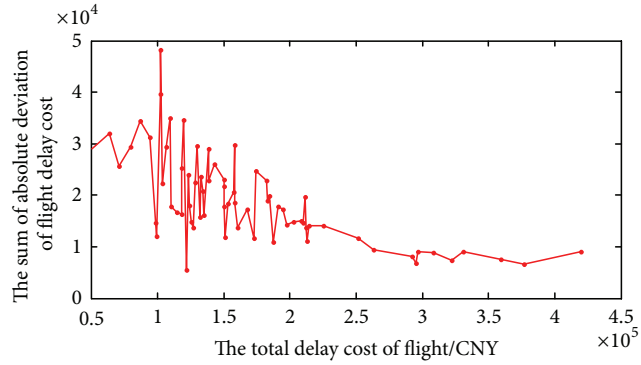


Figure 5: The trend relationship between delay cost and fairness.

calculated based on Table 7. The results are shown in Table 8 and Figure 8. The histogram shows the contrast of flight delay deviation between the initial flight sequence and the optimized flight sequence. From the histogram we can find that the delay cost of airlines declines 42.22% at least after optimization. The sum of delay deviation declines 38.64%. So the schedule model and solution have not only reduced the total delay cost significantly but also ensured the fairness among all the airlines.

Table 7: The comparison between initial flight delay cost and the optimized flight delay cost.

Flight number	Type	Unit delay cost of departure	Estimated time of departure	Actual time of departure	Departure runway	Delay cost of actual departure	Optimal Departure time	Departure runway after optimization	Departure delay cost after optimization
H_1F001	S	1.2	0:00:00	0:00:00	0	0	0:18:30	0	1332
H_2F002	S	1.1	0:00:00	0:01:36	0	105.6	0:08:50	1	583
H_2F003	L	6.2	0:00:00	0:02:42	0	1004.4	0:11:58	0	4451.6
H_3F004	M	2.4	0:00:00	0:00:00	1	0	0:03:10	0	456
H_4F005	M	2.5	0:05:00	0:04:20	0	−100	0:14:54	0	1485
H_4F006	S	1.4	0:05:00	0:08:29	1	292.6	0:22:20	1	1456
H_5F007	M	2.4	0:05:00	0:09:35	1	660	0:11:12	1	892.8
H_6F008	M	2.6	0:05:00	0:08:42	0	577.2	0:27:30	1	3510
H_6F009	L	5.7	0:10:00	0:17:41	0	2627.7	025:44	0	5380.8
H_3F010	M	2.4	0:10:00	0:19:07	1	1312.8	0:26:16	1	2342.4
H_1F011	M	2.3	0:10:00	0:20:23	1	1432.9	0:22:56	0	1784.8
H_2F012	L	6.2	0:10:00	0:18:59	0	3341.8	0:24:10	0	5270
H_7F013	M	2.7	0:10:00	0:20:37	0	1719.9	0:16:50	1	1107
H_5F014	S	1.3	0:15:00	0:25:11	1	794.3	0:19:08	1	322.4
H_2F015	M	2.6	0:15:00	0:25:15	0	1599	0:13:44	1	−197.6
H_7F016	M	2.7	0:15:00	0:26:17	1	1827.9	0:25:02	1	1625.4
H_3F017	S	1	0:15:00	0:27:15	0	735	0:21:42	0	402
H_1F018	L	5.1	0:15:00	0:27:33	1	3840.3	0:10:24	0	−1407.6
Flight number	Type	Unit delay cost of approach	Estimated time of approach	Actual time of approach	Approach runway	Delay cost of actual approach	Optimal Approach time	Approach runway after optimization	Approach delay cost after optimization
H_7F019	L	62.6	0:00:00	0:01:14	1	4632.4	0:06:00	0	22536
H_7F020	S	25.6	0:01:12	0:04:01	1	4326.4	0:23:56	1	34918.4
H_7F021	M	42.8	0:02:32	0:05:15	1	6976.4	0:02:28	1	−171.2
H_5F022	L	63.2	0:03:56	0:04:16	0	1264	0:13:16	0	35392
H_4F023	M	41.5	0:04:02	0:06:29	1	6100.5	0:03:42	1	−830
H_6F024	M	41.6	0:04:31	0:06:10	0	4118.4	0:06:50	1	5782.4
H_7F025	S	25.6	0:05:06	0:11:53	1	10419.2	0:15:44	1	16332.8
H_3F026	M	42	0:05:34	0:09:56	0	11004	0:07:54	0	5880
H_7F027	M	42.8	0:05:42	0:11:10	0	14038.4	0:00:00	1	−14637.6
H_1F028	M	42.4	0:06:54	0:13:07	1	15815.2	0:12:28	1	14161.6
H_7F029	L	62.6	0:07:01	0:12:24	0	20219.8	0:00:00	0	−26354.6
H_7F030	L	62.6	0:07:34	0:14:21	1	25478.2	0:04:56	1	−9890.8
H_4F031	M	41.5	0:08:03	0:16:15	1	20418	0:09:56	1	4689.5
H_3F032	S	24.7	0:08:43	0:15:11	0	9583.6	0:20:44	1	17808.7
H_5F033	L	63.2	0:09:12	0:17:29	1	31410.4	0:04:26	0	−18075.2
H_3F034	M	42	0:09:51	0:16:25	0	16548	0:01:54	0	−20034
H_6F035	M	41.6	0:10:13	0:21:37	1	28454.4	0:09:08	0	−2704
H_1F036	M	42.4	0:10:14	0:21:51	0	29552.8	0:01:14	1	−22896
H_2F037	L	63	0:11:06	0:22:51	1	44415	0:16:10	0	19152
H_1F038	S	25	0:12:53	0:24:09	0	16900	0:20:06	0	10825

Total: the total cost of the initial flight sequence delay is 343446.5/CNY; the total cost of optimized flight sequence delay is 102681.0/CNY.

Table 8: The total delay cost and deviation of standard flight.

Airline sequence	H_1	H_2	H_3	H_4	H_5	H_6	H_7	Deviation
				Total delay cost				
Initial	11256.9	7209.4	7535.2	7548.92	6563.21	7453.7	8456.5	10456.5
Optimized	2782.3	1954.6	3908.1	4438.0	2775.1	1405.0	3188.8	5487.0

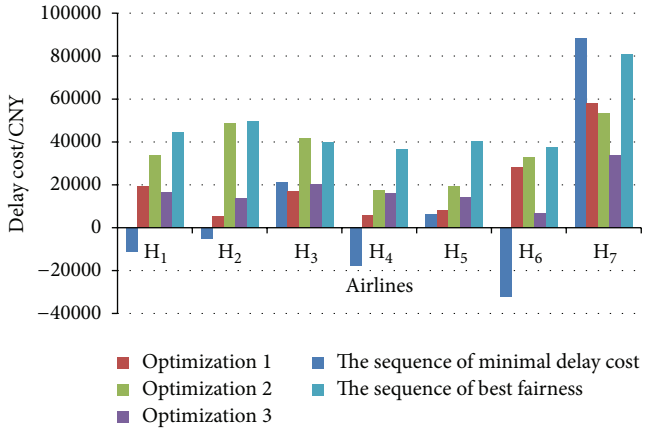


Figure 6: Comparison of the delay cost of airlines.

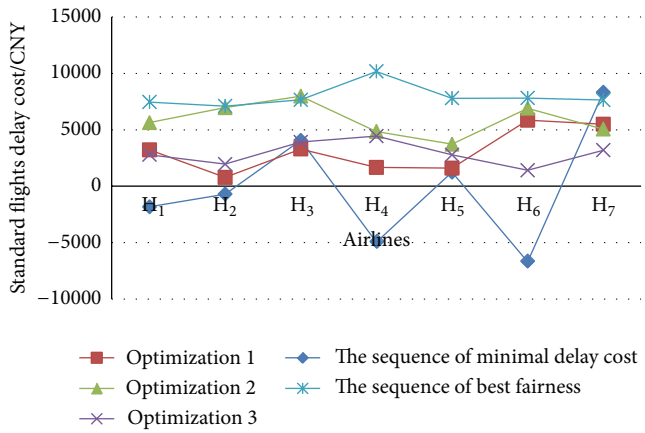


Figure 7: Comparison of the delay of five flight sequences for airline standard flights.

5. Conclusions

In the paper, a mixed multirunway operation flight scheduling optimization model based on multiobjective is proposed. Two objectives are considered: the total delay cost and fairness among airlines are two objective functions. The ACA is introduced to solve the model. The simulation results show that the total delay cost decreases significantly and the fairness among airlines is also acceptable. Meanwhile, ACA used in the paper solves the model with great efficiency.

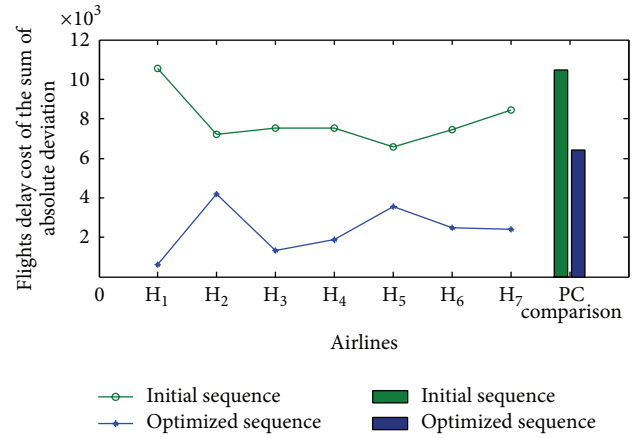


Figure 8: Contrast between the total delay cost and deviation of standard flight.

Conflict of Interests

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China, the Civil Aviation Administration of China (no. U1333117), China Postdoctoral Science Foundation (no. 2012M511275), and the Fundamental Research Funds for the Central Universities (no. NS2013067).

References

[1] X. Hu and E. di Paolo, "Binary-representation-based genetic algorithm for aircraft arrival sequencing and scheduling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 301–310, 2008.

[2] M. C. Wambsganss, "Collaborative decision making in air traffic management," in *New Concepts and Methods in Air Traffic Management*, pp. 1–15, Springer, Berlin, Germany, 2001.

[3] M. Wambsganss, "Collaborative decision making through dynamic information transfer," *Air Traffic Control Quarterly*, vol. 4, no. 2, pp. 107–123, 1997.

[4] K. Chang, K. Howard, R. Oiesen, L. Shisler, M. Tanino, and M. C. Wambsganss, "Enhancements to the FAA ground-delay program under collaborative decision making," *Interfaces*, vol. 31, no. 1, pp. 57–76, 2001.

[5] R. Hoffman, W. Hall, M. Ball, A. R. Odoni, and M. Wambsganss, "Collaborative decision making in air traffic flow management," NEXTOR Research Report RR-99-2, UC Berkley, 1999.

[6] T. Vossen, M. Ball, R. Hoffman et al., "A general approach to equity in traffic flow management and its application to mitigating exemption bias in ground delay programs," *Air Traffic Control Quarterly*, vol. 11, no. 4, pp. 277–292, 2003.

[7] A. Mukherjee and M. Hansen, "A dynamic stochastic model for the single airport ground holding problem," *Transportation Science*, vol. 41, no. 4, pp. 444–456, 2007.

[8] M. O. Ball, R. Hoffman, and A. Mukherjee, "Ground delay program planning under uncertainty based on the ration-by-distance principle," *Transportation Science*, vol. 44, no. 1, pp. 1–14, 2010.

[9] M. H. Hu and L. G. Su, "Modeling research in air traffic flow control system," *Journal of Nanjing University Aeronautics & Astronautics*, vol. 32, no. 5, pp. 586–590, 2000.

[10] Z. Ma, D. Cui, and C. Chen, "Sequencing and optimal scheduling for approach control of air traffic," *Journal of Tsinghua University*, vol. 44, no. 1, pp. 122–125, 2004.

[11] Q. Zhou, J. Zhang, and X. J. Zhang, "Equity and effectiveness study of airport flow management model," *China Science and Technology Information*, vol. 4, pp. 126–116, 2005.

[12] Q. Zhou, X. Zhang, and Z. Liu, "Slots allocation in CDM GDP," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 32, no. 9, pp. 1043–1045, 2006.

[13] H. H. Zhang and M. H. Hu, "Multi-objection optimization allocation of aircraft landing slot in CDM GDP," *Journal of Systems & Management*, vol. 18, no. 3, pp. 302–308, 2009.

[14] Z. Zhan, J. Zhang, Y. Li et al., "An efficient ant colony system based on receding horizon control for the aircraft arrival sequencing and scheduling problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, pp. 399–412, 2010.

[15] A. D'Ariano, P. D'Urgolo, D. Pacciarelli, and M. Pranzo, "Optimal sequencing of aircrafts take-off and landing at a busy airport," in *Proceeding of the 13th International IEEE Conference on Intelligent Transportation Systems (ITSC 10)*, pp. 1569–1574, Funchal, Portugal, September 2010.

[16] H. Helmke, O. Gluchshenko, A. Martin, A. Peter, S. Pokutta, and U. Siebert, "Optimal mixed-mode runway scheduling—mixed-integer programming for ATC scheduling," in *Proceedings of the IEEE/AIAA 30th Digital Avionics Systems Conference (DASC '11)*, pp. C41–C413, IEEE, Seattle, Wash, USA, October 2011.

[17] M. Samà, A. D'Ariano, and D. Pacciarelli, "Optimal aircraft traffic flow management at a terminal control area during disturbances," *Procedia—Social and Behavioral Sciences*, vol. 54, pp. 460–469, 2012.

[18] G. Hancerliogullari, G. Rabadi, A. H. Al-Salem, and M. Kharbeche, "Greedy algorithms and metaheuristics for a multiple runway combined arrival-departure aircraft sequencing problem," *Journal of Air Transport Management*, vol. 32, pp. 39–48, 2013.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 971861, 8 pages
http://dx.doi.org/10.1155/2014/971861

Research Article

Grain Emergency Vehicle Scheduling Problem with Time and Demand Uncertainty

Jiang DongQing[1] and Zhu QunXiong[2]

[1] *School of Science, Beijing University of Chemical Technology, Beijing 100029, China*
[2] *College of Information Science & Technology, Beijing University of Chemical Technology, Beijing 100029, China*

Correspondence should be addressed to Jiang DongQing; jiangdq@mail.buct.edu.cn

Received 8 February 2014; Revised 12 July 2014; Accepted 15 July 2014; Published 20 August 2014

Academic Editor: X. Zhang

Grain transportation plays an important role in many relief and emergency supply chains. In this paper, we take the grain emergency vehicle scheduling model between multiwarehouses as the research object. Under the emergency environment, the aim of the problem is to maximize the utilization of vehicles and minimize the delay time. The randomness of the key parameters in grain emergency vehicle scheduling, such as time and demand, is determined through statistical analysis and the model is solved through robust optimization method. Besides, we apply the numerical examples for experimental analysis. We compare the robust optimization model with classic model to illustrate the differences and similarities between them. The results show that the uncertainty of both time and demand has great influence on the efficiency of grain emergency vehicle scheduling problem.

1. Introduction

In recent years, a lot of large-scale public emergencies occurred frequently, such as the outbreak of SARS in 2003, the Wenchuan earthquake in China in 2008, and the Fukushima nuclear explosion in Japan in 2011, which not only resulted in huge losses, but also left a grieving memory in people's mind. With the rapid development of modern society, social problems of the population, resources, and environment become increasingly acute, leading kinds of public emergencies to occur more frequently. What is worse, the affected scope becomes much wider. Therefore, it is very significant to carry out the grain emergency scheduling effectively, which can protect the safety of people's life in emergencies. How to improve the capability of grain emergency scheduling for the government and enterprises has also become an important topic.

Vehicle scheduling problem is a kind of combinatorial optimization problem widely used in the fields of transportation, logistics, distribution, and so forth. In the process of vehicle service, we may encounter a variety of unexpected events, which will lead to the uncertainty of information, that is, the emergency vehicle scheduling problem, while, in this paper, we will mainly research the grain emergency vehicle scheduling problem. It requires us to adjust the determined parameters according to certain rules, with the uncertainties fully considered. In this case, the deterministic theory and method can no longer deal with these problems. We need to study a new set of theories and methods of vehicle scheduling problem with uncertain information.

At present, the research on emergency logistics is relatively small, while most scholars study the qualitative problem, such as emergency mechanism, the support of emergency materials, and the enactment of emergency laws and regulations. Studies on grain emergency vehicle scheduling are even less. The distribution and transportation of emergency materials especially grain is an important part of the research of emergency logistics, so is the implementation phase of the emergency logistics. Ordinary grain vehicle scheduling problem mainly considers the cost savings. However, the grain emergency vehicle scheduling pays more attention to improving the utilization rate of vehicle efficiency and reducing the delay time with all the constraints met.

In contrast, in this paper we will establish a grain emergency vehicle scheduling model, in which time and demand

are uncertain. Under the emergency environment, the aim of the problem is to maximize the utilization of vehicles and minimize the delay time. We assume that both the time and demand uncertainty belongs to a box uncertainty. After determining the uncertainty of time and demand, we establish the vehicle scheduling model with parameters determined and undetermined. Then we use the branch and bound algorithm and the robust optimization method to solve the problems. Finally, we list the numerical examples for experimental analysis. The results show that the uncertainty of both time and demand has great influence on the efficiency of grain emergency vehicle scheduling problem. Through the analysis of the data, we will provide the theoretical basis for the practical decision.

The rest of this paper is organized as follows. Section 2 is a brief review of the relevant literature on grain emergency vehicle scheduling. In Section 3, we present a model for grain emergency vehicle scheduling problem. We establish the vehicle scheduling model with parameters determined and undetermined. Finally, we assume that both the time and demand uncertainty belongs to a box uncertainty. We describe a computational study in Section 4. We compare the robust solution against the deterministic solution and then illustrate the effect of various system parameters on the objective function, such as time and demand. We conclude the paper with Section 5 and point out the direction for further research.

2. Literature Review

Because of the existence of uncertain factors and multiple objectives, such as the number of vehicles, the latest arrival time and other constraints, the modeling and solving of the grain emergency vehicle scheduling problem is very complicated. At present, there are few studies on this problem, which cannot compare with its importance and current application requirements. Besides, most literature studies the framework of commercial vehicle scheduling problem, without considering some special needs generated by the specific scene of the emergency. For example, the objective function of most literature is to minimize the total cost, while the grain emergency vehicle scheduling problem is to satisfy the needs and reduce the delay time.

Problems where a set of vehicles with finite capacity have to be scheduled at minimum cost are known as the vehicle scheduling problem. This class of problem was introduced by Dantzig and Ramser [1]. Rathi et al. [2] adopt the traditional optimization algorithm for the assignment problem with LP model, but the result is easy to fall into locally optimal solution. This method has strong limitations for the emergency material distribution problem. F. Fiedrich et al. [3] study the problem with time, the quantity of relief goods, and other resources being limited. In their studies, they improve the quality of emergency relief through the effective use of resource, with the minimum of deaths as the objective function. Finally, they put forward the optimal planning model of a plurality of disaster area distribution and emergency relief materials transportation after the earthquake.

Recently, Morales [4] uses robust optimization for the vehicle routing problem with stochastic demands (VRPSD). It is assumed that vehicles replenish at the depot, the worst-case value for the recourse action is computed by finding the longest path on an augmented network, and the problem is solved using a tabu search heuristic.

Using a robust optimization idea to solve uncertain linear optimization problem was first proposed by Soyster [5]. Although its method is too conservative, it is a kind of new train for the study of uncertainty optimization problem, and it is an optimizing condition based on the so-called worst-case, which has caused many scholars' attention and has established foundation for the rapid development of the robust optimization theory. J. M. Mulvey et al. [6] for the first time put forward the concept of robust optimization and applied it to repast problem, power capacity expansion, matrix balance, image reconstruction, and so on.

The robust optimization assumes that uncertain parameters belong to a given bounded uncertainty set. Ben-Tal and Nemirovski [7] propose that a LP with uncertain parameters belonging to a polyhedral uncertainty set has a robust counterpart problem. It is a LP whose size is polynomial in the size of the original problem, and the uncertainty can result in NP-hard problems. With the robust optimization ideas spreading unceasingly, the foundation work is considered by Ben-Tal and Nemirovski [8, 9], L. EI Ghaoui et al. [10, 11], and Bertsimas et al. [12–14].

Most research has no detailed introduction on scheduling model and scheduling route planning. At the same time, because of information under uncertainty, the demand of emergency supplies and transportation time cannot be accurately predicted. Hence scheduling model needs to be further researched and developed under uncertainty.

In contrast, in this paper we mainly use robust optimization for the grain emergency vehicle scheduling problem. Our paper differs from most of the existing literature in the following ways: (1) the objective function is to maximize the utilization of vehicles and minimize the delay time, while most of the literature aims to minimize the transportation cost or unmet demand; (2) we assume that both time and demand uncertainty belongs to a box uncertainty, in contrast with most of the stochastic models.

3. Grain Emergency Vehicle Scheduling Model

In this section, we first identify the grain emergency vehicle scheduling model. The objective function is to maximize the utilization of vehicles and minimize the delay time. Then we present the uncertainty set and the robust counterpart model. Finally, we propose the time and demand uncertainty belonging to a box uncertainty set.

3.1. Identifying the Grain Emergency Vehicle Scheduling Model. Now, consider an emergency vehicle routing problem among several grain warehouses. Problem can be described as follows: there is a batch of relief grain from different warehouses for different demand nodes. Each warehouse has a number of the same type of cars, which start from the warehouse issued,

leave immediately after arriving at the demand nodes, and eventually return to the warehouse. Besides, each car can only go to a demand node up to one time. If the time of arrival in the demand node is greater than the latest served time, this will cause delay in time. The objective function is to maximize the utilization of vehicles and minimize the delay time.

Parameters. The parameters are as follows:

m: the total number of warehouses,

n: the total number of demand nodes,

q: the total number of vehicles,

A: the set of m grain warehouses and n demand nodes,

C: the set of m grain warehouses,

D: the set of n demand nodes,

K: the set of q vehicles,

k_i: the number of vehicles available in warehouse i,

s_i: the amount of relief supplies provided by warehouse i,

c_k: the capacity of vehicles k ($k = 1, \ldots, q$),

r_k: utilization rate of vehicle k ($k = 1, \ldots, q$),

dl_i: the latest time relief grain arrived at demand node i, after which more wounded may appear ($i = 1, \ldots, n$),

t_{ijk}: the travel time between node i and node j for vehicle k,

d_i: the demand at each node i ($i = 1, \ldots n$),

γ: the weight of the utilization of vehicles.

Decision Variables. The decision variables are as follows:

$$x_{ijk} = \begin{cases} 1, & \text{if vehicle } k \text{ goes from node } i \text{ to node } j; \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

y_{ik}: the amount of relief supplies transported to node i by vehicle k,

T_{ik}: the time vehicle k arrives at node i. We ignore the waiting and loading time after the vehicles arrive at demand nodes, regarding them as 0 or a very small constant,

δ_{ik}: the delay time of vehicle k for delivering goods at node i. If vehicle k arrives at node I before dl_i, it is equal to 0. Otherwise, it is more than 0.

Objective Function. Consider the following:

$$\max z = \gamma * \sum_{k \in K} \sum_{i \in D} \frac{y_{ik}}{c_k} - \sum_{k \in K} \sum_{i \in D} \delta_{ik}, \quad (2)$$

$$\sum_{k \in K} \sum_{j \in D} x_{ijk} < k_i \quad \forall i \in C, \quad (3)$$

$$\sum_{k \in K} \sum_{j \in A} x_{ijk} \geq 1 \quad \forall i \in D, \quad (4)$$

$$\sum_{i \in C} \sum_{j \in D} \sum_{k \in K} x_{ijk} = \sum_{j \in D} \sum_{i \in C} \sum_{k \in K} x_{jik}, \quad (5)$$

$$\sum_{i \in A} x_{ijk} = \sum_{i \in A} x_{jik} \leq 1 \quad \forall j \in D,\ k \in K, \quad (6)$$

$$x_{ijk} + x_{jik} \leq 1 \quad \forall i \in D,\ j \in D,\ k \in K, \quad (7)$$

$$T_{ik} = \delta_{ik} = 0 \quad \forall i \in C,\ k \in K, \quad (8)$$

$$M * \left(x_{ijk} - 1\right) < T_{ik} + t_{ijk} - T_{jk} < M * \left(1 - x_{ijk}\right) \quad \forall i \in C, \quad j \in D, \quad k \in K, \quad (9)$$

$$M * \left(x_{ijk} - 1\right) < T_{ik} + t_{ijk} - T_{jk} < M * \left(1 - x_{ijk}\right) \quad \forall i \in D, \quad j \in D, \quad k \in K, \quad (10)$$

$$0 < T_{ik} - \delta_{ik} < dl_i * \sum_{j \in A} x_{ijk} \quad \forall i \in D,\ k \in K, \quad (11)$$

$$\delta_{ik} < \sum_{j \in A} x_{ijk} * M \quad \forall i \in D,\ k \in K, \quad (12)$$

$$s_i - \sum_{k \in K} \sum_{j \in D} x_{ijk} y_{jk} > 0 \quad \forall i \in C, \quad (13)$$

$$\sum_{i \in D} y_{ik} < c_k \quad \forall k \in K, \quad (14)$$

$$y_{ik} < c_k * \sum_{j \in A} x_{ijk} \quad \forall i \in D,\ k \in K, \quad (15)$$

$$\sum_{k \in K} y_{ik} + U_i - d_i \geq 0 \quad \forall i \in D, \quad (16)$$

$$x_{ijk} = \{0, 1\} \quad (\forall i \in A, j \in A, \text{but } i, j \text{ do not belong to } C \text{ at the same time}), \quad (17)$$

$$y_{ik}, \delta_{ik}, T_{ik} > 0 \quad \forall i \in D,\ k \in K. \quad (18)$$

Notice that M is a positive number, arbitrarily large. In this model, the objective function (2) seeks to maximize the utilization of vehicles and minimize the delay time. Constraint (3) ensures that the number of vehicles which started from warehouse is less than the number of vehicles existing in warehouses. Constraint (4) ensures that there is at least one vehicle serving each demand node. Constraint (5) represents that a vehicle from the warehouse should finally return to

the warehouse. Constraint (6) requests that the vehicles must leave as soon as possible after arriving, and each vehicle arrives at a demand node no more than once. Constraint (7) ensures that a vehicle leaving from one demand node could not return to the same node again. Constraint (8) limits the arrival time and delay time at warehouses to 0. Constraints (9) and (10) impose the restriction on the arrival time and travel time between two demand nodes. Constraint (11) represents the relationship between the arrival time, delay time, and the latest arrival time of a demand node. Constraint (12) limits the delay time to 0, when the vehicle does not go through the demand node. Constraint (13) ensures that the total amount of relief supplies provided by warehouse i is less than the capacity of warehouse i. Constraint (14) ensures that the loading amount of each vehicle is less than the capacity of vehicle k. Constraint (15) limits the amount of relief supplies to 0, if the vehicle does not go through the demand node. Constraint (16) limits the constraints among the demand of grain, the grain delivered, and the unmet demand. Constraints (17) and (18) impose the restriction on the decision variables.

3.2. Robust Grain Emergency Vehicle Scheduling Problem. Due to the highly unpredictable reasons, emergencies may cause high uncertainty on traffic. This will seriously affect the vehicle travel time, and it is very different with the general predictive value. For example, earthquake or other natural disasters may cause damage to the road; people's panic may cause excessive traffic congestion, resulting in traffic time being prolonged. On the other hand, due to effective emergency measures, the transportation time will be shortened if the relief supplies are given priority to. Besides, the demand of grain at each point, usually related to the number of the wounded, is difficult to estimate accurately. We can only make a rough estimate according to the historical information. Therefore, for grain emergency vehicle scheduling problem we now consider the transport time and demand uncertainty.

After the occurrence of unexpected events, the travel time from the warehouse to the demand node and the demand of each point cannot be determined accurately, neither can its probability distribution. So we use a robust method, making the travel time and demand belong to a bounded set U. As a result, this model is more suitable for realistic situation.

In consideration of the computability of the robust counterpart problem, which is transformed from the deterministic model, we choose a closed bounded convex set as the set of uncertainty. For example, $U_D = \{\mathbf{d} \mid \mathbf{d}^0 + \rho \sum_{s=1}^{m} y_s \widehat{\mathbf{d}^s}, y \in Y\}$, where $\rho\,(\rho > 0)$ is the level of uncertainty, $\mathbf{d}^0$ is the nominal value of demand, $\widehat{\mathbf{d}^s}$ is the scenario vectors, and $\mathbf{y}_s$ is the weight of scenario vectors.

Specifically, U_D can be described as the following three sets:

$$\text{convex hull: } U_{D_1} = \left\{\mathbf{d} \mid \mathbf{d}^0 + \rho\sum_{s=1}^{m} y_s\widehat{\mathbf{d}^s}, \mathbf{y} \in R^m, \ \mathbf{y} \geq 0, \sum_{s=1}^{m} y_s \leq 1\right\},$$

$$\text{box: } U_{D_2} = \left\{\mathbf{d} \mid \mathbf{d}^0 + \rho\sum_{s=1}^{m} y_s\widehat{\mathbf{d}^s}, \mathbf{y} \in R^m, \ \|\mathbf{y}\|_\infty \leq 1\right\},$$

$$\text{ellipsoidal: } U_{D_3} = \left\{\mathbf{d} \mid \mathbf{d}^0 + \rho\sum_{s=1}^{m} y_s\widehat{\mathbf{d}^s}, \ \mathbf{y} \in R^m, \mathbf{y}^T\mathbf{Q}\mathbf{y} \leq \Omega^2\right\}, \tag{19}$$

where $\mathbf{Q}$ is a positive definite matrix. The weight of scenario vectors $\mathbf{y}_s$ belongs to a bounded set $\mathbf{Y}_i$ for $i = 1, 2, 3$.

We now propose the robust counterpart problem for grain emergency vehicle scheduling problem with both time and demand uncertainty belonging to a box uncertainty set U. Recall that we consider the problem only with uncertainty in constraints (9), (10), and (16).

Proposition 1. *If the uncertainty set is U_{D_2}, the robust corresponding model to the model above can be obtained through the use of constraint* (20) *instead of constraints* (9) *and* (10)*, where t_{ijk}^0 is the nominal value of travel time:*

$$\begin{aligned}
&\left[M * \left(x_{ijk} - 1\right) + \rho\sum_{s=1}^{m}\left|\widehat{t_{ijk}^s}\right|\right] \\
&\quad < T_{ik} + {t_{ijk}}^0 - T_{jk} \\
&\quad < \left[M * \left(1 - x_{ijk}\right) - \rho\sum_{s=1}^{m}\left|\widehat{t_{ijk}^s}\right|\right] \\
&\qquad \forall i \in C, \quad j \in D, \quad k \in K, \\
&\left[M * \left(x_{ijk} - 1\right) + \rho\sum_{s=1}^{m}\left|\widehat{t_{ijk}^s}\right|\right] \\
&\quad < T_{ik} + {t_{ijk}}^0 - T_{jk} \\
&\quad < \left[M * \left(1 - x_{ijk}\right) - \rho\sum_{s=1}^{m}\left|\widehat{t_{ijk}^s}\right|\right] \\
&\qquad \forall i \in D, \quad j \in D, \quad k \in K.
\end{aligned} \tag{20}$$

Proof. Using the definition of Y_2 we can write $\sup_{y\in\mathbf{Y}_2}\mathbf{y}^T\mathbf{D}_j$ and its dual as the following pair of LPs:

$$\begin{aligned}
&\text{(Primal) max } \ \mathbf{y}^T\mathbf{D}_j \quad \text{s.t. } \mathbf{y} \leq \mathbf{e}, \mathbf{y} \geq -\mathbf{e}, \\
&\text{(Dual) min } \ \mathbf{e}^T\left(\boldsymbol{\alpha} + \boldsymbol{\beta}\right), \quad \text{s.t. } \boldsymbol{\alpha} - \boldsymbol{\beta} = \mathbf{D}_j, \boldsymbol{\alpha}, \boldsymbol{\beta} \geq 0.
\end{aligned} \tag{21}$$

□

It is simple to verify that the optimal solution to the dual problem will satisfy $\alpha_k^* + \beta_k^* = \rho\sum_{s=1}^{m}|\widehat{t_{ijk}^s}|$ for every $k = 1, \ldots, q$. Therefore, the dual optimal objective value is $\rho\sum_{k=1}^{q}\sum_{s=1}^{m}|\widehat{t_{ijk}^s}|$. Enforcing the robust feasibility condition

on (9) and (10) with the above optimal dual objective value we obtain constraints (20).

Similarly, we can replace constraint (16) by constraint (22):

$$\sum_{k\in K} y_{ik} + U_i - d_i^0 \geq \rho \sum_{s=1}^{m} \widehat{d_i^s} \quad \forall i \in D. \tag{22}$$

4. Experimental Analysis

From the part above, we can see that the transformed robust model is still 0-1 mixed integer programming. Although the 0-1 mixed integer programming could not be solved by polynomial time algorithm, small scale problems can still use the exact algorithm such as the branch and bound method. We first use the branch and bound algorithm to solve the deterministic model with many warehouses.

Assume that there are three warehouses and eight demand nodes. The total amount of relief grain provided by each warehouse is 1100, 1800, and 2700, and the number of vehicles available in each warehouse is 1, 2, and 3. The capacity of vehicle is 1000. The transportation time between two demand nodes is assumed to be symmetric. Then we generate 8 integers randomly between [100, 1000] as the demand at each node, and another 8 integers between [1, 10] as the latest time relief supplies arrived at demand node. Last, generate integers between [1, 10] as the travel time between every two nodes (see Table 1, unit: hours).

To solve this deterministic problem, we use the linear interactive and general optimizer (LINGO). The results are as follows (see Table 2).

From Table 1, we can see that, with the weight becoming larger, both the total amount of utilization rate of vehicle and the total delay time will be increased. Therefore, the size of the weight depends on which factor the decision maker pays more attention to.

Next, we will analyze the influence of travel time and demand on the optimal vehicle route. Here we assume that the weight of the utilization of vehicles is 1.

When the travel time changes slightly, such as the travel time between warehouse 2, and the demand node 1 changes to 5 hours, the optimal route becomes 3-6-7-4-11-1 and 3-5-9-10-3-8-2. The route has changed greatly. Therefore, we can conclude that the optimal solution is very sensitive to changes in travel time.

When we change the demand of the demand node 2 from 450 to 500, the optimal route becomes 3-6-5-8-2, 3-6-7-4-11-1, and 3-5-9-10-2. The result is different with the former one. Thus, the optimal solution is sensitive to changes in both travel time and demand.

Now we have a feasibility test on the original result. Then we find that when the demand or travel time changes, the original optimal route becomes infeasible.

Through the experiment above, we can see that when the travel time between demand nodes or the demand at each node is slightly changed, the final optimal route will also be changed (see Table 3). This indicates that the optimal solution is very sensitive to the travel time and demand. Due to various factors, when the incident occurs, we cannot know the exact and accurate demand and travel time between demand nodes. Thus, there is probably some deviation from the original estimates. If we still calculate the route according to the original estimate, this may cause larger utilization rate of vehicle or increased delay time. Besides, even the requirements of feasibility cannot be met, resulting in more serious consequences.

Then we first have a test on the robust counterpart of grain emergency vehicle scheduling problem. Here we choose the box constraints for testing. We consider a box uncertainty set for the travel time. It is given by $t_{ijk} \in [t_{ijk}^0 - \rho\widehat{t_{ijk}}, t_{ijk}^0 + \rho\widehat{t_{ijk}}]$, where t_{ijk}^0 are the nominal values and $\widehat{t_{ijk}}$ are the deviation values. We use the data from Table 1 as the nominal values of travel time (when $\gamma = 10$).

Then we generate integers randomly between [0, 5] as the deviation values of travel time between nodes.

Next, let us consider the changes of objective function when the travel time is uncertain. To solve this model with time uncertainty, we use the linear interactive and general optimizer (LINGO). The optimal solutions under different values of ρ are as follows (see Table 4).

When the level of uncertainty ρ changes, the optimal routes change as well. Although the delay time changes, the utilization rate of vehicle does not change. With a box uncertainty set for the travel time, we use the optimal solution for the nominal value to have the feasibility test. We conclude that the optimal routes seeking out from the nominal value are sometimes infeasible. Through the experiment, it can be obtained that the optimal solution of deterministic model is sensitive to the travel time. When the travel time deviates from the nominal value, the optimal solution for the nominal value is not feasible sometimes, but the optimal solution obtained by the robust model is not sensitive to the travel time so as to maintain the robust solutions.

Secondly, we use LINGO to calculate the optimal solution with uncertain demand under different values of ρ. The results are as follows (see Table 5).

When the level of uncertainty ρ changes, the optimal routes change as well. With the change of demand, the utilization rate of vehicles and the delay time change as well, which eventually leads to the inconsistency of objective function. From the results we can see that, with the level of uncertainty increasing, the utilization rate of vehicles reduced and delay time increased, which also result in the objective function being reduced. So we can conclude that we should try to improve the accuracy of demand forecasting and reduce the fluctuation of demand.

With a box uncertainty set for the demand, we use the optimal solution for the nominal value to have the feasibility test. We conclude that the optimal routes seeking out from the nominal value are sometimes infeasible. Through the experiment, it can be obtained that the optimal solution of deterministic model is sensitive to the travel time. When the travel time deviates from the nominal value, the optimal solution for the nominal value is not feasible sometimes, but the optimal solution obtained by the robust model is not sensitive to the travel time so as to maintain the robust solutions.

Table 1: The travel time between every two nodes.

	W1	W2	W3	DN1	DN2	DN3	DN4	DN5	DN6	DN7	DN8
W1	—	1	2	4	3	6	3	5	9	7	6
W2	1	—	3	2	5	7	6	3	4	8	1
W3	2	3	—	4	2	1	2	7	5	6	8
DN1	4	2	4	—	5	8	1	6	3	7	1
DN2	3	5	2	5	—	2	4	2	1	5	3
DN3	6	7	1	8	2	—	1	5	4	8	1
DN4	3	6	2	1	4	1	—	3	7	4	9
DN5	5	3	7	6	2	5	3	—	9	4	3
DN6	9	4	5	3	1	4	7	9	—	2	8
DN7	7	8	6	7	5	8	4	4	2	—	7
DN8	6	1	8	1	3	1	9	3	8	7	—

W: warehouse; DN: demand node.
The demand generated randomly: 700, 450, 900, 300, 700, 450, 750, and 600.
The latest arrival time generated randomly: 0, 0, 0, 5, 4, 7, 6, and 9.

Table 2: Vehicle scheduling under different weight.

Weight	$\gamma = 1$	$\gamma = 10$	$\gamma = 100$
Vehicle 1	—	1-5-3	1-5-9-1
Vehicle 2	—	2-11-4-1	2-4-1
Vehicle 3	—	2-11-6-1	2-11-1
Vehicle 4	3-6-11-4-7-1	3-7-8-3	3-6-8-1
Vehicle 5	3-5-8-1	3-5-9-10-1	3-6-7-3
Vehicle 6	3-6-5-9-10-3	3-6-9-10-2	3-10-1
Utilization rate of vehicle			
Detail	$r_1 = 0\%, r_2 = 0\%,$ $r_3 = 0\%, r_4 = 70\%,$ $r_5 = 100\%, r_6 = 100\%$	$r_1 = 35\%, r_2 = 100\%,$ $r_3 = 80\%, r_4 = 100\%,$ $r_5 = 100\%, r_6 = 70\%$	$r_1 = 90\%, r_2 = 70\%,$ $r_3 = 60\%, r_4 = 100\%,$ $r_5 = 90\%, r_6 = 75\%$
Total	**2.70**	**4.85**	**4.85**
Delay time			
Detail	0	$\delta_{4,2} = 2, \delta_{5,1} = 3,$ $\delta_{6,3} = 2, \delta_{10,5} = 1,$ $\delta_{11,2} = 1, \delta_{11,3} = 1$	$\delta_{4,2} = 2, \delta_{5,1} = 3,$ $\delta_{8,4} = 1, \delta_{9,1} = 4,$ $\delta_{11,3} = 1$
Total	**0**	**10**	**11**
Objective function	2.70	47.5	474

Table 3: Random deviation of travel time.

	W1	W2	W3	DN1	DN2	DN3	DN4	DN5	DN6	DN7	DN8
W1	0	1	2	2	0	3	1	0	4	3	1
W2	1	0	3	1	3	2	1	2	1	5	0
W3	2	3	0	1	2	1	2	4	2	1	2
DN1	2	1	1	0	2	4	1	1	2	3	1
DN2	0	3	2	2	0	2	0	2	1	0	3
DN3	3	2	1	4	2	0	1	2	1	4	1
DN4	1	1	2	1	0	1	0	0	2	1	3
DN5	0	2	4	1	2	2	0	0	5	1	2
DN6	4	1	2	2	1	1	2	5	0	0	5
DN7	3	5	1	3	0	4	1	1	0	0	1
DN8	1	0	2	1	3	1	3	2	5	1	0

Table 4: Vehicle routes under different ρ with travel time uncertainty.

Uncertainty	$\rho = 1.5$	$\rho = 2.0$	$\rho = 3.0$
Vehicle 1	1-4-1	1-6-1	1-6-1
Vehicle 2	2-8-1	2-4-2	2-10-2
Vehicle 3	2-10-1	2-10-1	2-4-2
Vehicle 4	3-4-11-6-7-1	3-9-11-7-1	3-5-9-1
Vehicle 5	3-7-8-5-11-1	3-5-8-10-1	3-8-10-4-5-9-2
Vehicle 6	3-6-9-5-3	3-8-11-10-1	3-6-11-4-8-7-1
Utilization rate of vehicle			
Detail	$r_1 = 70\%, r_2 = 70\%,$ $r_3 = 75\%, r_4 = 70\%,$ $r_5 = 100\%, r_6 = 100\%$	$r_1 = 90\%, r_2 = 70\%,$ $r_3 = 55\%, r_4 = 100\%,$ $r_5 = 100\%, r_6 = 70\%$	$r_1 = 90\%, r_2 = 75\%,$ $r_3 = 50\%, r_4 = 70\%,$ $r_5 = 100\%, r_6 = 100\%$
Total	**4.85**	**4.85**	**4.85**
Delay time			
Detail	$\delta_{4,1} = 1, \delta_{8,2} = 0.1,$ $\delta_{10,3} = 0.5,$	$\delta_{4,2} = 0.1, \delta_{6,1} = 0.1,$ $\delta_{10,3} = 0.1$	$\delta_{4,3} = 0.1, \delta_{6,1} = 0.1,$
Total	**1.6**	**0.3**	**0.2**
Objective function	46.9	48.3	48.2

Table 5: Vehicle routes under different ρ with travel demand uncertainty.

Uncertainty	$\rho = 1.5$	$\rho = 2.0$	$\rho = 3.0$
Vehicle 1	—	—	—
Vehicle 2	2-4-2	2-11-2	2-11-1
Vehicle 3	2-11-6-1	2-11-6-1	2-4-1
Vehicle 4	3-6-9-1	3-5-9-10-1	3-5-8-1
Vehicle 5	3-5-8-1	3-5-8-3	3-6-7-4-11-2
Vehicle 6	3-7-10-1	3-6-7-4-11-8-3	3-9-10-3
Utilization rate of vehicle			
Detail	$r_1 = 0\%, r_2 = 62.5\%,$ $r_3 = 100\%, r_4 = 72.5\%,$ $r_5 = 100\%, r_6 = 90\%$	$r_1 = 0\%, r_2 = 35\%,$ $r_3 = 100\%, r_4 = 100\%,$ $r_5 = 70\%, r_6 = 100\%$	$r_1 = 0\%, r_2 = 35\%,$ $r_3 = 100\%, r_4 = 100\%,$ $r_5 = 65\%, r_6 = 100\%$
Total	**4.25**	**4.05**	**4.00**
Delay time			
Detail	$\delta_{4,3} = 2, \delta_{11,2} = 1,$	$\delta_{6,3} = 2, \delta_{11,2} = 2,$ $\delta_{11,3} = 1$	$\delta_{4,2} = 2, \delta_{6,3} = 2,$ $\delta_{11,3} = 1,$
Total	**3**	**5**	**5**
Objective function	46.9	39.5	35.5

The examples above provide some management measures for the decision makers. Firstly, it is very important for them to make a trade-off between the utilization rate of vehicles and the delay time. As shown in Table 2, the weight has great influence on the objective function. The managers should determine the weight according to different situations and the importance of the two factors. Secondly, they should locate the warehouses properly to make the distance shorter. It is a good way to reduce the delay time. In addition, they should also use some technology to make a better prediction of travel time and demand. This will also reduce the complexity of the model.

5. Conclusions

Emergency logistics is a new research field, which has good application background and broad perspective. This paper mainly studies the grain emergency vehicle scheduling optimization problems. The research is summarized as follows.

(1) We analyze the characteristics of grain emergency logistics and the significance of grain emergency vehicle scheduling problem based on the analysis and reference of the relevant theory.

(2) According to the specific characteristics of emergency logistics, we established the mathematical model of

grain emergency vehicle scheduling model. On the basis of research on grain emergency logistics vehicle scheduling problem, the question is discussed, and corresponding mathematical models are established.

(3) Then we consider the uncertainty in travel time and demand and present the robust counterpart model. We propose the uncertainty belonging to a box uncertainty set. Our study shows that the robust optimization is an effective method to solve the vehicle scheduling problem with uncertainty, because it is not necessary to know the amount of uncertainty distribution function in advance.

(4) We use LINGO to solve the two models. Besides, we use an example to analyze the model. Example shows that the robust method is feasible and effective in solving grain emergency vehicle scheduling problem and optimization. Our computational results show that the chance constrained model can be more or less efficient than the robust model depending on the problem parameters and uncertainty assumptions. Lastly, we put forward some management measures to improve the logistics system when uncertainties occur.

However, future work is still needed to solve the grain emergency vehicle scheduling model. Further work may consider additional uncertainty sets such as convex hull and ellipsoidal uncertainty. These sets has been proved to solve to larger problems while restricting the uncertainty away from worst-case scenarios.

Conflict of Interests

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

References

[1] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959.

[2] A. K. Rathi, R. S. Solanki, and R. L. Church, "Allocating resources to support a multicommodity flow with time windows," Oak Ridge National Lab, Oak Ridge, Tenn, USA, 1990.

[3] F. Fiedrich, F. Gehbauer, and U. Rickers, "Optimized resource allocation for emergency response after earthquake disasters," *Safety Science*, vol. 35, no. 1–3, pp. 41–57, 2000.

[4] J. C. Morales, *Planning robust freight transportation operations [Dissertation]*, 2006.

[5] A. L. Soyster, "Convex programming with set-inclusive constraints and applications to inexact linear programming," *Operations Research*, vol. 21, pp. 1154–1157, 1973.

[6] J. M. Mulvey, R. J. Vanderbei, and S. A. Zenios, "Robust optimization of large-scale systems," *Operations Research*, vol. 43, no. 2, pp. 264–281, 1995.

[7] A. Ben-Tal and A. Nemirovski, "Robust solutions of uncertain linear programs," *Operations Research Letters*, vol. 25, no. 1, pp. 1–13, 1999.

[8] A. Ben-Tal and A. Nemirovski, "Robust solutions of linear programming problems contaminated with uncertain data," *Mathematical Programming B*, vol. 88, no. 3, pp. 411–424, 2000.

[9] A. Ben-Tal and A. Nemirovski, "Robust optimization—methodology and applications," *Mathematical Programming B*, vol. 92, no. 3, pp. 453–480, 2002.

[10] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, 1997.

[11] L. El Ghaoui, F. Oustry, and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 33–52, 1999.

[12] D. Bertsimas and M. Sim, "Robust discrete optimization and network flows," *Mathematical Programming*, vol. 98, no. 1–3, pp. 49–71, 2003.

[13] D. Bertsimas, D. Pachamanova, and M. Sim, "Robust linear optimization under general norms," *Operations Research Letters*, vol. 32, no. 6, pp. 510–516, 2003.

[14] D. Bertsimas and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 686705, 6 pages
http://dx.doi.org/10.1155/2014/686705

Research Article

Analysis of Time of Day Fare Discounts on Urban Mass Transit Travel Behavior, Crowding, and Waiting Time

Xiao Guo and Huijun Sun

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Huijun Sun; hjsun1@bjtu.edu.cn

Received 20 February 2014; Revised 18 July 2014; Accepted 18 July 2014; Published 12 August 2014

Academic Editor: Hu Shao

Every morning, commuters select the regularly dispatched urban mass transit for traveling from a residential area to a workplace. This paper aims to find an optimal discount fare and time intervals on morning peak hour. As a direct and flexible traffic economic instrument, fares can influence commuters' behavior. Therefore, fare discount has been proposed to regulate traffic flow in different time. Two models have been analyzed to describe it with schedule delay because of the travel demand size. The first objective function is constructed on pressure equalization when the travel demand is small. The other objective function is to minimize total waiting time when the travel demand is large. In the end, numerical examples based on an artificial network are performed to characterize fare discount models.

1. Introduction

Problems of sustainability in transportation are already quite noticeable in almost all large cities. Due to congestion, more waiting time is wasted on the roads, and commuters are more likely to get stressed and frustrated, which in turn causes accidents. Forecasts of traffic condition are less reliable, so punctual arrival becomes difficult and travelers suffer losses from early departure or late arrival [1]. As an integral part and essential feature of cities, public transportation plays an important role in developed and developing countries. Providing efficient public transportation systems has been recognized as a potential way of alleviating traffic congestion, improving mobility, mitigating air pollution, and reducing energy consumption. It also provides a convenient way to connect residential areas and CBD [2].

The crowding on the platform has been generated on rush hours. Commuters usually decide the departure time to minimize the individual travel cost. In early literature, the studies about public transit were based on the traditional marginal cost pricing theory [3]. Mohring [4, 5] developed a microeconomic foundation for public transportation services with fixed demand, which was quite useful for long-term public transit service planning in a static sense. However, this model cannot describe a commuter's time-of-use behavior on rush hours.

Some studies consider the time dimension to describe a commuter's behavior. Various time-varying pricing models and algorithms have been proposed to determine the tolls for system performance optimization under given physical and economic pricing constraints. This first work is done by Vickrey [6] who develops a dynamic toll model with deterministic queuing theory, which leads to an equilibrium of costs on all commuters. Subsequently, this model has been extended by many others [7, 8], and the concept of schedule delay cost has been widely used in proposing various commuting equilibrium models [9–13]. In 1982, Small [14] accomplished an important empirical study and calibrated the disutility function of schedule delay using the survey data in the United States.

In a transit system, it is generally unpractical to provide the capacity for all commuters to arrive on a train during peak hours in congested cities; so some of commuters may have to wait until a train arrives that they can board. They probably have to stand during the peak period and suffer discomfort due to congestion on platform. Kraus and Yoshida [15] and Kraus [16] analyze the economic analysis of commuters' time-of-use decision and examine the optimal pricing and

service in an urban mass transit system. Huang et al. [17, 18] introduce an in-vehicle congestion function in modeling the choice of urban mass transit runs. Li and Hensher [19] review public transport crowding valuation research. In those papers, a commuter's cost contains the cost of waiting time at the origin transit stop and is similar to the cost of queuing time in the bottleneck model.

Some papers talk about the transit frequency when they consider waiting time on rush hours [20, 21]. Yu et al. [21] define that it is constant. Thus, they cannot describe the changing of commuters' waiting time. In this paper, waiting time is redefined to describe the congestion phenomenon. Waiting time is different from that as we discussed above. It is a measure of a scalar congestion. Meanwhile, it is also different from the waiting time in traditional bottleneck model. In bottleneck model, waiting time has relationship to departure rate, and it is different due to queue length. In our model, waiting time is not only related to the departure time, but also related to the congestion in a train. According to our assumption, this concept can explain the reason of congestion through our model.

On the other side, transit service level is low in most cities of China. Some passengers have to stay on platform, and there exist seats on off-peak hours. Fare can influence commuters' behavior as a direct and flexible traffic economic instrument. Therefore, this paper proposes fare discount model to enhance the transit service level. Time-varying transit fares have been investigated to solve congestion when timetable is fixed. Low fare would attract some commuters who are in the peak hour transfer to the lower price period. They will change some commuters' behavior and alleviate the congestion in public transit system.

This paper investigates time-varying railroad fares discount models. There exist two kinds of models because of travel demand. The small travel demand is named when the max number of passengers in a train is smaller than the capacity. According to assumptions, the waiting time is a fixed value to everyone. Thus, the objective function is constructed on pressure equalization. It is the total sum of squared differences between the ridership and the number of seats in trains. The second situation is discussed in the condition of large travel demand. The objective function is to minimize total waiting time as described in former. This paper explores the consequences of fare adjustment in the urban railroad mode. The results show that the structure of the ticket fare discount does indeed influence the behavior of the commuters.

The paper is organized as follows. Section 2 introduces generalized travel cost for urban railroad with the schedule delay penalty. In Section 3, two different models have been considered to alliviate the congstion. In Section 4, numerical analyses are provided to illustrate the models. Section 5 concludes the paper.

2. Basic Assumptions and Generalized Travel Cost Function

In this paper, a railroad connecting the residents to the CBD has been considered. In the morning, there are N commuters who have to make their trips to the CBD for work by this public transit. Due to the physical constraint of the train's capacity, most commuters necessarily arrive earlier or later than the work starting time, and some commuters have to stay on platform waiting for the following trains. In such a case, commuters have to incur the costs of schedule delay and the costs of waiting time. Some assumptions have been described in the next section before generalized travel cost fuctions are displayed.

2.1. Basic Assumptions. To facilitate the presentation of the essential ideas of this paper, all commuters are assumed to be identical in perceiving the time value and schedule delay penalty. The following basic assumptions are made in this paper.

(1) The commuters who are in a train have the same travel cost. That is to say, those commuters are assumed to have the same waiting time when they are on the same train.

(2) Commuters must obey the first-come-first-served basis. All commuters can board train when the number of commuters is smaller than the capacity of the train. Only the former s (s represents the capacity of a train.) commuters can board train when the number of commuters on platform is bigger than s. The rest of them need to wait the following trains.

2.2. Generalized Travel Cost Function. This paper aims to provide an explaination of how fare discounts affect cummuters' travel behavior on rush hours. The cost experienced by a railroad commuter incldues the in-vehicle time, the waiting time on platform, the schedule delay of arrival at workplace, the urban railroad fare, and the body discomfort. In this paper, the waiting time has been redefined to describe the congestion phenomenon on the platform. The schedule delay is assumed to be fixed in each train when the timetable is fixed. The congestion discomfort is generated by body congestion in the train carriage, which is a monotonically increasing function of the ridership.

Travel Time. Firstly, let $[0, A]$ be the morning peak operating time interval. According to each urban railroad's departing time, $[0, A]$ can be divided into n subintervals $(a_{i-1}, a_i]$, where $a_0 = 0$, $a_n = A$, and $i \in \{1, 2, \ldots, n\}$. a_i is the departing time of the ith urban railroad. Let $w(t)$ be the waiting time when a commuter arrives at platform at time t, and the departing time of commuters in the same vehicle obeys the uniform distribution. $[0, A]$ can be divided into n subintervals $(t_{i-1}, t_i]$, where $t_0 = 0, t_n = A$, and $i \in \{1, 2, \ldots, n\}$. t_{i-1} is the time of a commuter who arrives at platform and boards a vehicle as the first passenger. t_i is the time of a commuter who arrives at platform and boards the vehicle as the last passenger. According to the former assumption, the waiting time of commuters in the same vehicle can be easily calculated $w_i = a_i - (t_{i-1} + t_i)/2$, $i \in \{1, 2, \ldots, n\}$. Constraint $t_i \le a_i$, $i \in \{1, 2, \ldots, n\}$, can ensure that no one can get on vehicle after

the vehicle's leaving platform. Let η_1 be the unit cost of travel time. Then the travel time costs are $\eta_1(T+w_i), i \in \{1, 2, \ldots, n\}$.

Schedule Delay. Schedule delay means the time with respect to the work starting time (assumed to be same for all commuters in the same vehicle). Let t^* be the work starting time; if a commuter arrives at workplace before t^*, the schedule delay is $t^* - (a_i + T)$; if a commuter arrives at workplace on t^*, then $t^* - (a^* + T) = 0$, where $a^* = t^* - T$ is the perfect departure time; if a commuter arrives at workplace after t^*, the schedule delay is $(a_i + T) - t^*$. Let η_2 be the unit cost of schedule delay time-early and let η_3 be the unit cost of schedule delay time-late. An early schedule delay is $\eta_2[t^* - (a_i + T)]$, and a late schedule delay is $\eta_3[(a_i + T) - t^*]$.

Congestion Effect. The congestion effect is generated by body congestion in urban railroad carriage. $g(x_i)$ represents the body congestion cost of unit in-carriage time which is a monotonically increasing function of the number of commuter x_i in the urban railroad [17]. Let $g(x_i) = G(x_i/x_0)$, where G represents a parameter ($G > 0$) and x_i denotes the number of commuters in a urban railroad depart at time a_i, $i = \{1, \ldots, n\}$. x_0 represents the number of seats.

The Urban Railroad Fare. The urban railroad fare is represented by p.

Integrating the travel time, schedule delay, travel comfort, and travel fare, a commuter's generalized travel cost can be formulated as follows:

$$C(t) = \begin{cases} \eta_1 \left(T + a_i - \dfrac{t_{i-1} + t_i}{2} \right) + \eta_2 \left[t^* - (a_i + T) \right] \\ \quad + G\dfrac{x_i}{x_0} + p \qquad \text{if } a_i + T \le t^* \\ \eta_1 \left(T + a_i - \dfrac{t_{i-1} + t_i}{2} \right) + \eta_3 \left[(a_i + T) - t^* \right] \\ \quad + G\dfrac{x_i}{x_0} + p \qquad \text{if } a_i + T \ge t^*. \end{cases} \tag{1}$$

3. The Schedule Delay Penalty Models

This section describes two formulations for the problem of finding a user-equilibrium strategy assignment. The first model describes the schedule delay penalty model when the travel demand is small. In this model, some factors can be determined by assumptions.

3.1. The Fare Discount Model under Equal Pressure. In this section, the small travel demand has been considered in this schedule delay penalty model, in which the ridership in each train is smaller than s. This means all commuters waiting on platform can get on the vehicle. The case can be described as $t_i = a_i$, $i = 1, \ldots, n-1$. So the waiting time can be caculated accurately $w(t) = \gamma \cdot \Delta t$, where γ represents the ratio of average waiting time and Δt represents headway. γ is approximately $1/2$ if commuter arrivals at station follow a uniform distribution, while vehicles arrive at station deterministically. The travel time cost is $\eta_1(T + (1/2)\Delta t)$. The times of commuters from the residential area to the nearby station and the end urban railroad to the workplace are fixed. Thses two values do not affect the calculation results. Both of them can be calculated into T.

The aim is to alleviate traffic congestion through reducing the ridership in peak hour and shifting commuters from the peak to the off-peak period. When the edges' fare is smaller than the peak hour's, some commuters will be attracted to change their departure time and transfer to the edges. This paper hopes to calculate the discount intervals and the size of the fare discount under equal pressure. This process can be described as follows.

We set the discount time intervals $[a_0, a_k]$ and $[a_l, a_n]$, where $a_k \le t^* - T$, $a_l \ge t^* - T$. Let the discounted railroad fare be $p\theta$, where $0 \le \theta \le 1$. The discounted intervals and discounted railroad fare would be calculated to minimize the variance between the ridership and the number of seats in a train.

According to the assumptions, the travel cost can be divided into four classes. The first class is before t^*, and the travel cost contains fare discount; the second class is before t^*, and the travel cost contains fare discount; the third class is after t^*, and the travel cost contains fare discount; the last class is after t^*, and the travel cost contains fare discount.

The travel cost of the first class could be described by the following formula:

$$C_i(t) = \eta_1 \left(T + \frac{1}{2}\Delta t \right) + \eta_2 \left[t^* - (a_i + T) \right] + G\frac{x_i}{x_0} + p\theta, \quad t \in [a_0, a_k]. \tag{2}$$

The travel cost of the second class could be described by the following formula:

$$C_i(t) = \eta_1 \left(T + \frac{1}{2}\Delta t \right) + \eta_2 \left[t^* - (a_i + T) \right] + G\frac{x_i}{x_0} + p, \quad t \in (a_k, a^*]. \tag{3}$$

The travel cost of the third class could be described by the following formula:

$$C_i(t) = \eta_1 \left(T + \frac{1}{2}\Delta t \right) + \eta_3 \left[(a_i + T) - t^* \right] + G\frac{x_i}{x_0} + p, \quad t \in (a^*, a_l]. \tag{4}$$

The travel cost of the last class could be described by the following formula:

$$C_i(t) = \eta_1 \left(T + \frac{1}{2}\Delta t \right) + \eta_3 \left[(a_i + T) - t^* \right] + G\frac{x_i}{x_0} + p\theta, \quad t \in (a_l, a_n]. \tag{5}$$

All the commuters have been divided into four classes. At equilibrium state, all commuters incur the same travel cost

no matter what time they leave home. This model can be described as the following programming problem:

$$\min\left(\sum_{i=1}^{n}(x_i - x_0)^2\right) \tag{6}$$

s.t. (2)–(5)

$$\sum_{1}^{n} x_i = N. \tag{7}$$

$$C_1(t) = C_2(t) = \cdots = C_n(t). \tag{8}$$

The objective function (6) means to minimize the total sum of squared differences between the ridership and the number of seats in trains. Constraints (2)–(5) state the travel cost in different trains. Constraint (7) represents the flow conservation. Constraint (8) guarantees the travel costs of each commuter are equal at equilibrium.

3.2. The Fare Discount Model to Minimize the Total Waiting Time. In this section, the railroad fare discount has been considered when the travel demand is large. Like the former, the discount intervals and railroad fare have been analyzed. Given intervals, we set the discount intervals $[a_0, a_k]$ and $[a_l, a_n]$, where $a_k < t^* - T$, $a_l > t^* - T$, and $a_k < a_l$. Let the discount ticket fare be $p\theta$, where $0 \le \theta \le 1$. This paper hopes to calculate the discount intervals and the size of the fare discount to minimizing the total waiting time. According to the assumptions, the travel cost can be divideed into four classes. The first class is before t^*, the travel cost with discount; the second class is before t^*, the travel cost without discount; the third class is after t^*, the travel cost without discount; the last class is after t^*, the travel cost with discount.

The travel cost of the first class could be described by the following formula:

$$C_i(t) = \eta_1\left(T + a_i - \frac{t_{i-1} + t_i}{2}\right) + \eta_2\left[t^* - (a_i + T)\right] + G\frac{x_i}{x_0} + p\theta, \quad t \in [a_0, a_k]. \tag{9}$$

The travel cost of the second class could be described by the following formula:

$$C_i(t) = \eta_1\left(T + a_i - \frac{t_{i-1} + t_i}{2}\right) + \eta_2\left[t^* - (a_i + T)\right] + G\frac{x_i}{x_0} + p, \quad t \in (a_k, a^*]. \tag{10}$$

The travel cost of the third class could be described by the following formula:

$$C_i(t) = \eta_1\left(T + a_i - \frac{t_{i-1} + t_i}{2}\right) + \eta_3\left[(a_i + T) - t^*\right] + G\frac{x_i}{x_0} + p, \quad t \in (a^*, a_l]. \tag{11}$$

The travel cost of the last class could be described by the following formula:

$$C_i(t) = \eta_1\left(T + a_i - \frac{t_{i-1} + t_i}{2}\right) + \eta_3\left[(a_i + T) - t^*\right] + G\frac{x_i}{x_0} + p\theta, \quad t \in (a_l, a_n]. \tag{12}$$

All the commuters have been divided into four classes. In general, the equilibrium condition among commuters can be statd as follows: every commuter is unable to find another departure time to reduce the total waiting time. This model can be described as the following programming problem:

$$\min\sum_{i=1}^{n} x_i\left(a_i - \frac{t_{i-1} + t_i}{2}\right) \tag{13}$$

s.t. (9)–(12)

$$0 \le x_i \le s \quad i = 1, \ldots, n \tag{14}$$

$$a_0 < t_1 < t_2 < \cdots < t_{n-1} < a_n \tag{15}$$

$$t_i \le a_i \quad i = 1, \ldots, n \tag{16}$$

$$\sum_{1}^{n} x_i = N \tag{17}$$

$$C_1(t) = C_2(t) = \cdots = C_n(t). \tag{18}$$

The objective function (13) means to minimize the total waiting time. Constraints (9)–(12) state the travel cost in each urban railroad. Constraint (14) guarantees the number of commuters in each urban railroad should be less than the capacity. Constraint (15) states that the commuters should obey the first-come-first-served basis. Constraint (16) states the time which the last commuter gets on urban railroad before the departure time of the urban railroad. Constraint (17) represents the flow conservation. Constraint (18) guarantees the travel costs of each commuter are equal at equilibrium.

4. Numerical Results

In this section, the models are simulated by LINGO 11. A set of numerical results is presented to illustrate the effects of considerable factors. It is assumed that there are 24 trains in 2 hours. Table 1 provides parameter values used for computational experiments. In this paper, η_1, η_2, and η_3 should obey $\eta_2 < \eta_1 < \eta_3$ [14].

Firstly, the schedule delay penalty model with small travel demand has been studied. The travel demand equals 13000. According to the assumptions in this study, commuters choose their departure time to minimize the total travel cost. When $\theta = 0.45$, the objective function reaches the optimal. If the discount is larger or smaller than 0.45, the objective function cannot get optimal. For example, the fare discount has no effect when $\theta = 0$ in all time of the rush hours. Table 1 describes the change of the discount with the commuters' choice.

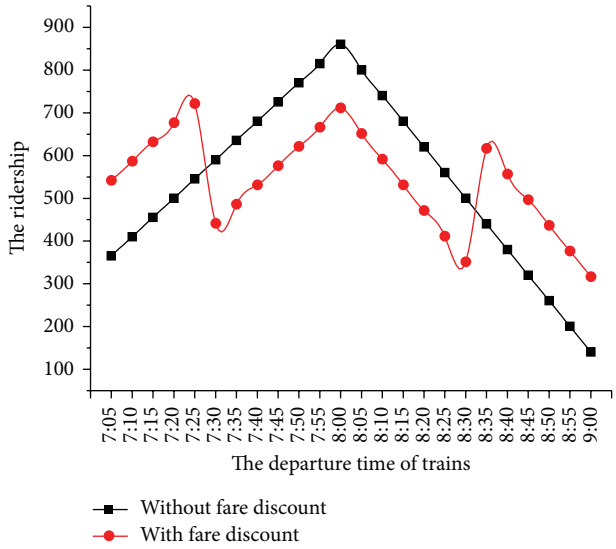


Figure 1: The result of fare discount model under equal pressure.

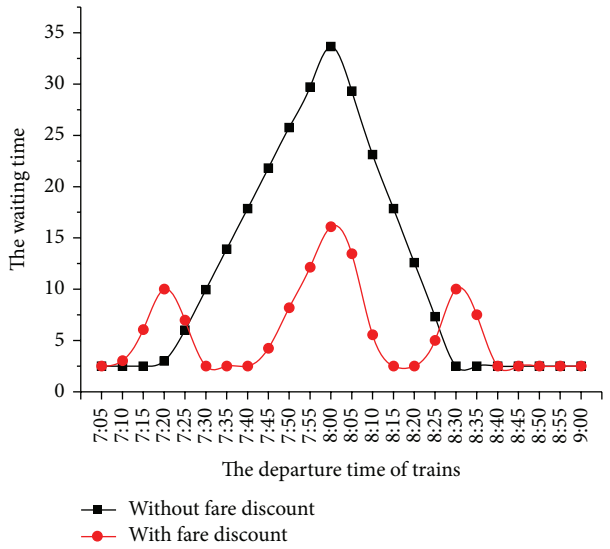


Figure 2: The result of fare discount model to minimize the total waiting time.

From Figure 1, we find the ridership in the no discount time interval has reduced 1939. The travel cost has been reduced from 5.72 RMB to 5.17 RMB, and the rate of the discount is 0.45. That means when the discount is the 0.45 of the ticket fare, there are 21.59% of the commuters who change their departure time. Without discount, we find that there are free seats in each vehicle in the last 10 minutes. After the discount, there is no free seat in each vehicle. The congestion in the middle time interval has been alleviated. That is to say, the system resources have been effectively utilized.

Different from the schedule delay penalty model with small travel demand, this model has a large travel demand which equals 32000. From Figure 2, we find the max waiting time has been reduced from 33.65 minutes to 16.08 minutes. The total waiting time has been reduced from 383443.2 minutes to 184076.8 minutes, when the rate of the discount is 0.52. If the discount rate is larger than 0.52, the fare discounts for the trasportantion system have no effective.

Table 1: Parameter values used in computational experiments.

Parameter	Symbol	Value
The travel time	T	20 minutes
The waiting time	Δt	5 minutes
The unit cost of travel time	η_1	0.038 RMB per min
The unit cost of schedule delay time-early	η_2	0.03 RMB per min
The unit cost of schedule delay time-late	η_3	0.04 RMB per min
The number of seats	x_0	240
Ticket price	p	2 RMB
A parameter of body congestion	G	0.8

For seeking the minimum travel cost, some commuters reach at station early. Long waiting times have been explained through this numerical result. This means the discount of urban railroad fare is an effective approach to solve congestion in the platform. The congestion in the middle time interval has been alleviated and the system resources have been effectively utilized.

According to travel demand size, the schedule delay penalty models can be divided into two kinds. This paper finds both models alleviate the congestion in the platform and vehicle. The special model is suitable in the small travel demand size. Different from the speical model, the schedule delay penalty model with large travel demand is more feasible in real life.

5. Conclusions

In this paper, the schedule delay model has been constructed in urban railroad where the urban railroad fare adjustment is considered. To determine fare adjustments, the paper constructs two user-equilibrium models duo to travel demand size. The first model is a special case of the second model when there is a small travel demand. The second model is the main model which considers big travel demand. Commuters are loaded on a first-come-first-served basis and must wait for the following vehicles. Two numerical examples are provided.

The first example is based on a small travel demand. The number of commuters has reduced and alleviated the congestion in the no discount time interval if there is ticket fare adjustment. Different from the second model, this model puts more emphasis on the rational use of resources and avoids the waste. That is to say, the system resources would be effectively utilized when the fare discount can be implemented.

The second example is based on a large travel demand, in which numerical analysis is performed. From the numerical results, some interesting observations are made with practical implementations. Firstly, the discount should only be used when the travel demand size is sufficiently large. In the small size, the former model has been enough. Secondly, the

max waiting time has reduced nearly a half. This means the price discount is an effective approach to solve congestion in platform and carriage. Thirdly, under the discount of the railroad fare, commuters have saved the travel cost significantly.

Finally, both models have alleviated the congestion in the middle time interval with discount, and the system resources have been effectively utilized. This is our government desired result. The presented model and concept in this study can be further explored in either theoretical direction or practical applications. In reality, the population is heterogeneous due to income level, perception on general travel cost's composition, and so forth and can be divided into multiple classes. A multiclass model and serviceability analyses would be more flexible and feasible to be applied in different regions with diverse population configurations.

Conflict of Interests

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

Acknowledgments

This paper is partly supported by National Basic Research Program of China (2012CB725400) and NSFC (71271023 and 71322102).

References

[1] H. Yang and X. Wang, "Managing network mobility with tradable credits," *Transportation Research Part B*, vol. 45, no. 3, pp. 580–594, 2011.

[2] S. I. Y. Chien and C. F. M. Tsai, "Optimization of fare structure and service frequency for maximum profitability of transit systems," *Transportation Planning and Technology*, vol. 30, no. 5, pp. 477–500, 2007.

[3] W. S. Vickrey, "Some implications of marginal cost pricing of public utilities," *American Economis Review*, vol. 45, no. 2, pp. 605–620, 1955.

[4] H. Mohring, "Optimization and scale economies in urban bus transportation," *American Economic Review*, vol. 62, no. 4, pp. 591–604, 1972.

[5] H. Mohring, *Transportation Economics*, Ballinger, Cambridge, Mass, USA, 1976.

[6] S. Vickrey, "Theory and transportation investment," *American Economic Review*, vol. 59, no. 2, pp. 414–431, 1969.

[7] R. Arnott, A. de Palma, and R. Lindsey, "A structural model of peak-period congestion: a traffic bottleneck with elastic demand," *The American Economic Review*, vol. 83, no. 1, pp. 161–179, 1993.

[8] T. Yao, M. M. Wei, B. Zhang, and T. Friesz, "Congestion derivatives for a traffic bottleneck with heterogeneous commuters," *Transportation Research B*, vol. 46, no. 10, pp. 1454–1473, 2012.

[9] F. Xiao, W. Shen, and H. Michael Zhang, "The morning commute under flat toll and tactical waiting," *Transportation Research B*, vol. 46, no. 10, pp. 1346–1359, 2012.

[10] Q. Tian, H. J. Huang, and W. H. K. Lam, "Modeling the crowding effects of transit system on commuters' departure time choice behaviors," in *Proceedings of the 9th International Conference on Industrial Management*, pp. 386–393, 2008.

[11] E. J. Gonzales and C. F. Daganzo, "Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing," *Transportation Research B: Methodological*, vol. 46, no. 10, pp. 1519–1534, 2012.

[12] A. de Palma and S. Proost, "Imperfect competition and congestion in the city," *Journal of Urban Economics*, vol. 60, no. 2, pp. 185–209, 2006.

[13] R. Zhong, A. Sumalee, and T. Maruyama, "Dynamic marginal cost, access control, and pollution charge: a comparison of bottleneck and whole link models," *Journal of Advanced Transportation*, vol. 46, no. 3, pp. 191–221, 2012.

[14] K. A. Small, "The scheduling of consumer activities: work trips," *The American Economic Review*, vol. 72, no. 3, pp. 467–479, 1982.

[15] M. Kraus and Y. Yoshida, "The commuter's time-of-use decision and optimal pricing and service in urban mass transit," *Journal of Urban Economics*, vol. 51, no. 1, pp. 170–195, 2002.

[16] M. Kraus, "A new look at the two-mode problem," *Journal of Urban Economics*, vol. 54, no. 3, pp. 511–530, 2003.

[17] H. J. Huang, Q. Tian, H. Yang, and Z. Y. Gao, "Modeling the equilibrium bus riding behavior in morning rush hour," in *Proceedings of the 9th Annual Conference of the Hong Kong Society of Transportation Studies*, pp. 434–442, Hong Kong, China, 2004.

[18] H.-J. Huang, Q. Tian, H. Yang, and Z.-Y. Gao, "Modal split and commuting pattern on a bottleneck-constrained highway," *Transportation Research E: Logistics and Transportation Review*, vol. 43, no. 5, pp. 578–590, 2007.

[19] Z. Li and D. A. Hensher, "Crowding and public transport: a review of willingness to pay evidence and its relevance in project appraisal," *Transport Policy*, vol. 18, no. 6, pp. 880–887, 2011.

[20] A. Schoebel and S. Scholl, "Line planning with minimal traveling time," in *Proceedings of 5th Workshop on Algorithmic Methods and Models for Optimization of Railroads*, ATMOS, Palma de Mallorca, Spain, 2005.

[21] B. Yu, Z. Yang, and J. Yao, "Genetic algorithm for bus frequency optimization," *Journal of Transportation Engineering*, vol. 136, no. 6, pp. 576–583, 2010.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 840619, 9 pages
http://dx.doi.org/10.1155/2014/840619

Research Article

The Bilevel Design Problem for Communication Networks on Trains: Model, Algorithm, and Verification

Yin Tian,[1] Hong-hui Dong,[1] Li-min Jia,[1] Yong Qin,[1] and Si-yu Li[2]

[1] *State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China*
[2] *Shenzhen Graduate School, Peking University, Shenzhen 518055, China*

Correspondence should be addressed to Li-min Jia; jialm@vip.sina.com

Received 11 February 2014; Revised 8 July 2014; Accepted 15 July 2014; Published 5 August 2014

Academic Editor: X. Zhang

This paper proposes a novel method to solve the problem of train communication network design. Firstly, we put forward a general description of such problem. Then, taking advantage of the bilevel programming theory, we created the cost-reliability-delay model (CRD model) that consisted of two parts: the physical topology part aimed at obtaining the networks with the maximum reliability under constrained cost, while the logical topology part focused on the communication paths yielding minimum delay based on the physical topology delivered from upper level. We also suggested a method to solve the CRD model, which combined the genetic algorithm and the Floyd-Warshall algorithm. Finally, we used a practical example to verify the accuracy and the effectiveness of the CRD model and further applied the novel method on a train with six carriages.

1. Introduction

With the development of network communication technique, an increasing number of novel networks are proposed to substitute for the traditional train communication network (TCN). No matter what functions those systems perform, all of them similarly have a communication network, which is the common, core characteristic of those networks. The topology of such a network is decisive; once improper, the reliability will decrease and the delay will increase. That influences the operation of the whole network.

For a network that is applied to train control or safety monitoring, reliability and delay are two decisive indicators. Information transferred on such networks contains control command (e.g., brake and accelerate) or safety-monitoring information (e.g., speed and axle temperature). If the information transmission fails due to either link brakes or information delays, the performance of the whole train will be seriously affected. Two types of solutions can solve this problem at network-designing stage: one is to add links so as to increase the reliability, and the other is to devise reasonable paths for data transmission to reduce the network delay.

To optimize reliability and delay, most of the existing algorithms deal with them separately. But simultaneous optimization better conforms the demand of automatized design and improves efficiency. Starting from that point, this paper proposes a new algorithm, which helps to design the physical topology structure under cost constraint, so as to achieve maximum reliability. Meanwhile, based on such reliability, this algorithm can devise reasonable data transmission paths and obtain the minimum delay.

Currently, there are many algorithms to deal with the communication network design problem. But most of them only discussed parts of the issue, such as the equipment for specific purpose [1, 2] or other characteristics of networks [3, 4], and failed to discuss systematically the demand of network design as a whole. Some algorithms emphasized the topology structure design with constrained reliability [5, 6]. Some, under cost limitation, explored the best physical topology possessing the maximum reliability [7] but did not consider the requirements of delay. Several algorithms are dedicated in improving real time of networks with known physical topology, by way of substituting communication protocols and certain equipment [8]. And some studies applied top-down

design to devise train networks but limitedly focused on the hardware [9].

Although some research did try to optimize simultaneously the physical and logical topology [10], they proposed neither systematical models nor corresponding solutions. Usually, such simultaneous optimization is treated as a multiobjective process [11], but that will complicate the solving and make it hard to ensure the ultimate optimal solution.

The present paper takes advantage of the bilevel programing theory [12] to optimize simultaneously under cost constraint, that is, address the reliability and the real time under cost limitation at the same time. The upper level is physical topology design, defined as improving performance of the network through changing the connections between nodes. The lower level is physical topology design, defined as obtaining the highest efficiency by optimizing the transmission paths. Treat the upper-level problem as a discrete network design problem (DNDP) [13] and the lower-level problem as a minimum delay problem (MDP).

The rest of the paper is structured as follows. In Section 2, we point out the problem of bilevel programming in train communication networks and propose CRD model. In Section 3, the solution of the CRD model is illustrated and explained. In specific, such solution is based on genetic algorithm and Floyd-Warshall algorithm, and the key issue is how to deliver data effectively between these two algorithms. In Section 4, we use a numerical example to verify the accuracy of the CRD model and further apply it to solve a practical design problem. In Section 5, conclusion and prospects are provided.

2. A Programming Model for the Discrete Network Design Problem on Trains

2.1. A Basic Description of the Design Problem of Train Communication Networks. The design problem of train communication networks (DPTCN) is to seek the most appropriate network through adjusting the physical and logical topology of nodes and links. As communication networks between trains are bidirectional, here we model the physical topology with a graph Gp; that is, $Gp = [N, L, A]$, where N stands for the nodes in a network, L represents the links therein, and A is the adjoint matrix describing the relationship between nodes and links of Gp. The optimization of the physical topology of train communication networks means finding the optimal solution gp_{opt} of Gp, with constraints of $Cp = [cp1, cp2, cp3, \ldots]$ and $gp_{\text{opt}} \in Gp$. Correspondently, use another graph Gl to model the logical topology; that is, $Gl = [N, L, A, L^*, A^*]$, in which L^* illustrates the actual links required by data stream, and A^* is the adjoint matrixes of the network data stream. The optimization of the logical topology is to get the optimal solution gl_{opt} of Gl, with limitations that $Cl = [cl1, cl2, cl3, \ldots]$ and $gl_{\text{opt}} \in Gl$.

Depending on different design demands of communication networks on trains, there will be various constraints and objectives of DPTCN. Nodes and link elements depend on the characteristics of the particular problem of interest. For example, if we are going to build an economical control network on a train, then we have to, under cost constraint, design a network that has maximum reliability and minimum delay. If, on the other hand, we are going to establish a safety-monitoring network that requires reasonable reliability and is insensitive to cost, what we need to do is to devise a network, under reliability constraint, costing least and possessing the maximum bandwidth. This paper focuses on the first situation and builds the cost-reliability-delay model (CRD model) to achieve the goal.

2.2. Basic Ideas of the Bilevel Programming Model for the Discrete Network Design Problem. As for the train communication networks in train control systems, reliability and delay are two vital indicators. To ensure smooth transmission of information between carriages, every carriage has a node responsible for communication. In the traditional train communication networks such as WTB, all the nodes are arranged in a line; that is, each node connects only with nodes in the next carriage. This is mainly due to the limited transmission distance of the industrial field bus. In practical application, such topology severely restricts networks' performance. One of the defects is that, if one of the links brakes, the whole network will be shut down. Another disadvantage is that data in either end of the network will be extremely crowded. However, the gradually introduced Ethernet can be helpful to overcome these defects. We may take advantage of the Ethernet's longer transmission distance and try to break through the traditional internode connections. With effective connection, one can optimize both reliability and delay.

In fact, DPTCN with limited delay is a DNDP in communication area. It can be described as a leader-follower game. The leaders are the designers of networks, and the followers are the control signal and the safety-monitoring information that can freely choose communication links. The leaders are able to design the physical topology and further to influence the logical topology of information stream, so that one can obtain the most appropriate network. This kind of interaction game can be presented by the following bilevel programming problem:

$$\begin{aligned} (P) \quad & \max_{u} \quad F(x, u) \\ & \text{s.t.} \quad G(x, u) \le 0, \end{aligned} \tag{1}$$

where $x = x(u)$ is implicitly defined by

$$\begin{aligned} (L) \quad & \min_{u} \quad f(x, u) \\ & \text{s.t.} \quad g(x, u) \le 0. \end{aligned} \tag{2}$$

The bilevel programming model contains two submodels. (P) is the upper-level question, that is, the physical topology planning issue. (L) is the lower-level question, that is, the logical topology planning issue. In (P), F and u are the objective function and the decision vector, respectively; and G is the constraint of the decision vector. In (L), f and x are the objective function and the decision vector, respectively; and g is the constraint of the decision vector.

One should treat the equipment of train communication network as consisting of two kinds of elements, nodes and

links, and establish the CRD model. In such a model, the physical part changes the existing links through optimization program, so as to satisfy the redundancy requirement and lower the construction cost. On the other hand, the logical topology part concerns how to distribute information stream so as to obtain the minimum transmission time. Information from physical topology design is transmitted to logical topology through a transfer function $x = x(u)$. Based on that information, logical topology designs reasonable paths for data transmission.

2.3. The Physical Topology Optimization. The present paper focuses on the optimization problem that, under cost constraint, changes existing links to improve reliability. Generally, reliability issues of communication networks assume that the reliability of nodes and links is random, independent, known, and static [14]. Because communication on trains is bidirectional, one can use graphs with nonoriented links to describe the network. $G = (N, L, A)$ is a network without parallel link, that is, a nonredundant network, and there is no outlier. Then, the optimization problem can be described as

$(U1)$

$$\max \quad R(x) = \left\{ \sum_{\Omega} \left[\prod_{l \in L'} P\left(l_j\right) \right] \cdot \left[\prod_{l_{ij} \in (L \setminus L')} \left(1 - P\left(l_j\right)\right) \right] \right\} \cdot \left[\prod_{i=1}^{N} P\left(n_i\right) \right], \tag{3}$$

$$\text{s.t.} \quad \sum_{i=1}^{N} \sum_{j=1}^{N} c\left(l_j\right) d_j u_j + \sum_{i=1}^{N} c\left(n_i\right) \le C(x), \tag{4}$$

$$P\left(l_j\right) = F\left[c\left(l_j\right)\right], \tag{5}$$

$$P\left(n_j\right) = G\left[c\left(n_j\right)\right]. \tag{6}$$

$R(x)$ is the reliability of the whole network; $P(l_j)$ is the reliability of link l_j; $P(n_i)$ is the reliability of node n_i; Ω is the assembly of all operational states of the network, and, here, $\Omega = gp_{\text{opt}}$; $C(x)$ is the maximum cost of the whole network; $c(l_j)$ is the unit cost (per distance) of link l_j; d_j is the distance of link l_j; $c(n_i)$ is the cost of node n_i; L is the set amount of given links; N is the set amount of given nodes; F is the function relationship between link reliability and unit cost; G is the function relationship between node reliability and cost. At any time instant, only some links of G might be operational. A state of G is a subgraph (N, L') when L' is the set amount of operational links. If $l_j \in L'$, then $u_j = 1$ or $u_j = 0$.

In the upper level, network designers' objective is to obtain the maximum reliability through changing links. Constraint (4) ensures that the overall cost of construction does not exceed the specified maximum cost. Constraints (5) and (6) describe the function relationship between instrument's reliability and cost.

2.4. The Logical Topology Optimization. After fixing the physical topology, designers need to optimize the logical topology based on different design objectives. This paper pursues minimum transmission delay. Based on that, the optimization of information stream's logical topology is actually to obtain the minimum delay of the specific physical topology designed by upper level. Such logical optimization problem can be presented as

$$(L1) \quad \min \quad T(x) = \sum_{\Phi} \left[\sum_{l \in L''} t\left(l_j\right) + \sum_{n \in N'} t\left(n_i\right) \right], \tag{7}$$

$$\text{s.t.} \quad \Phi \in \Omega, \tag{8}$$

$$P\left(l_j\right) = f\left[t\left(l_j\right)\right], \tag{9}$$

$$P\left(n_j\right) = g\left[c\left(n_j\right)\right]. \tag{10}$$

$T(x)$ is the overall delay; $t(l_j)$ is the delay of link l_j; $t(n_i)$ is the delay of node n_i; when transmitting data from one node to another arbitrary one, the transmission path is a subset of G, described as (N', L''); Φ is the assembly of all data transmission paths, and, here, $\Phi = gl_{\text{opt}}$; f is the function relationship between link delay and link unit price; g is the function relationship between node delay and node cost.

In the lower level, designers devise the way in which information will be transmitted. Constraint (8) ensures the logical topology optimization process is based on the specific physical topology designed by upper level. Constraints (9) and (10) describe the function relationship between network equipment delay and cost.

3. A Solution Algorithm for the Bilevel Train Communication Network Design Problem

The present paper uses genetic algorithms to solve the discussed problem. The reason for choosing that algorithm is that, firstly, the optimization question in bilevel programming is hard to deal with [15] and, secondly, the optimization of communication network's reliability, delay, and cost is also a NP-complete problem [14]. Genetic algorithms are capable of handling these questions. It has been applied to optimizations of various nonlinear problems. A great amount of researches suggest that genetic algorithms can effectively deal with the optimization of networks [16–18]. But none of the referred studies discussed physical and logical topology simultaneously.

Here, a solution for the CRD model based on the genetic algorithm is proposed: maximize the reliability under limited cost, then transmit the result network to the lower level, and finally obtain the logical topology with the minimum delay. Generally, the reliability and the cost of equipment are proportional, while the delay and the cost are inversely proportional. These relationships provide a possibility to convert the cost constraint of lower level into upper-level's constraint through certain price. The solution process can be described as follows.

Step 1. Set original parameters, including the amount of nodes, the distance between nodes, the maximum cost, the node unit price, the node reliability, the link unit price, and the link reliability.

Step 2. Based on the original parameters, create original gene, design the physical connection between nodes via genetic algorithm, and generate the optimum solution under limited cost. After this, the physical topology design is complete.

Step 3. Determine whether the physical topology meets actual requirement. If so, go to Step 4. If not, record such structure into the unsuitable solution database and go back to Step 2 to find other optimum solutions.

Step 4. Convert the gene of the best solution into adjoint matrix that stands for the physical topology and deliver it to the logical topology module.

Step 5. Find the communication style with the minimum delay between arbitrary nodes. After this, the logical topology design is complete.

Step 6. Determine whether the logical topology meets actual requirement. If so, the overall solution is complete. If not, record such connection style into the unsuitable solution and seek the reason. If it is due to the logical topology, then go back to Step 4. If not, go back to Step 2.

There are several points that need attention. First of all, the logical topology designed here is based on an assumption that the network bandwidth is much larger than the amount of data within the network. This is because the maximum bandwidth of existing TCN is 1.5 Mbps, but it will increase to 100 MBps with application of the Ethernet. Even if the existing control instruction is increased by 10 times, network bandwidth is also an order of magnitude larger than the data traffic. Secondly, the criterion in Step 3 indicates constraints other than cost, for example, no outlier is permitted, length limitations caused by the cables themselves. Thirdly, the criterion in Step 5 is a complementary constraint to prevent the assumption that "the network bandwidth much larger than the amount of data" fails. If the received data of a node exceed the actual throughput, then such topology will be discarded.

Because this solution process is modular, adjustment on specific constraint is convenient and will not affect other modules. There are two key issues: one is the design of genetic algorithm and the other is the method to obtain the minimum delay. These two points will be specifically illustrated as follows.

3.1. The Genetic Algorithm for Solving Physical Topology Optimization Problems. The most outstanding feature of genetic algorithms is that they implement multiple iterations, eliminate bad genes, and filter out the optimal solution. The coding structure of gene and the determination of fitness function, as well as crossover and mutation rule, pose a significant impact on the results and efficiency of genetic algorithms. This paper defines various modules of the genetic algorithm applicable to this study. As there has been a great amount of research successfully using genetic algorithms for network reliability analysis, this paper shifts the focus to how to deliver the genetic algorithm's result to the lower level for further calculation.

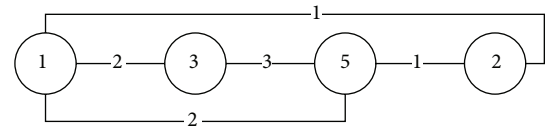


Figure 1: A simple train communication network.

3.1.1. Gene-Coding Structure. As for solving the CRD model, one of the key issues is how to deliver the information between the two levels, and an effective gene-coding structure is decisive to such delivery. The first step of the gene-coding structure is to determine the length of gene. For a train communication network that has N_d sets of nodes, the relationship of its link amount N_l and node amount N_d can be illustrated in

$$N_d - 1 \le N_l \le \frac{(N_d + 1) N_d}{2}. \tag{11}$$

Therefore, to describe the reliability of the whole network, all the genes referred to in this paper have $0.5(N_d+1)N_d+N_d$ bits. $|0.5(N_d+1)N_d|$ stands for the link reliability and $|N_d|$ for the node reliability.

For different nodes and links, one can use different integers to describe their reliability. For example, 1 means the most reliable equipment and 2 less reliable equipment, and so forth. If there are N kinds of equipment, and each bit of every gene ranges among $[0, N]$ in which 0 means there is no link at all. For example, Figure 1 shows a network with 4 nodes and 5 links.

The numbers in Figure 1 stand for the reliability of corresponding equipment. Notice that the train communication network has a feature that differs from other communication networks: nodes in the network are linearly arrayed. This structure is the same as WTB in TCN [19].

Different from traditional methods of solving reliability problems with application of the genetic algorithm, this paper not only pursues the optimal solution, but also focuses on the delivery of the network structure from the upper level to the lower level. Because the gene-coding structure must represent the network structure, which is usually described by adjoint matrixes, the gene-coding structure evolves from adjoint matrixes. Figure 2 shows the relationship between adjoint matrix a and gene x.

The first six numbers represent the gene standing for six links, and the last four numbers the four nodes in Figure 1 from left to right. Because train communication networks are full-duplex networks, their adjoint matrixes are symmetric.

3.1.2. The Fitness Function. The objective of optimization here is, under limited cost, to find the most reliable physical

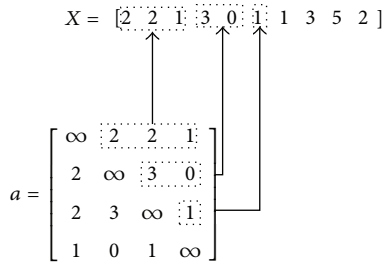


Figure 2: The relationship between adjoint matrix and gene.

topology. Therefore, two elements are necessary in the fitness function: cost and reliability. In genetic algorithms' operation, good solutions on optimal boundary are often the result of breeding between a feasible solution and an infeasible solution. As a result, when setting the fitness function, simply excluding the infeasible solution should be avoided. The right way is to set an effective penalty function that reduces the proportion of infeasible solutions in the population. Concluding from practical experience, the following fitness function has good computational efficiency:

$$\text{Fit}(x) = \begin{cases} R(x), & c(x) \le C,\ \text{count}(\text{rea}(x)) \neq 0, \\ \lambda, & \text{others.} \end{cases} \tag{12}$$

In this function, C is the maximum cost. count(rea(x)) $\neq$ 0 is the amount of 0 in x's reachability matrix rea(x). This constraint ensures that there will be no outlier in the network. $R(x)$ is system reliability. $c(x)$ is the actual cost of the system. λ is the penalty factor and $\lambda < E(R(x))$. According to actual conditions, the recommended value of λ ranges from 0.05 to 0.15.

3.1.3. Genetic Operators. The efficiency of genetic algorithms depends on the population size, the selection method, the crossover and mutation operators, and the stopping criteria. After a large number of experiments, the most appropriate genetic algorithm parameters for this paper have been found: the population size is 400 and limit the range of each gene in the population to be an integer from 1 to N. Choose stochastic uniform as selection method and use one-point crossover with a crossover probability of 0.8. Select mutation-uniform with rate of 0.03. The termination criterion is 500 generations.

3.2. The Algorithm for Solving Logical Topology Optimization Problems. Assume that, in train communication network, the minimum delay only depends on the shortest path between nodes, and such a path has no relation with link length. In other words, the time for delivering a message is only affected by the number of nodes such as information passed. The reasonableness of this assumption is justified in the following.

Table 1: Node delay.

Frame size	Average delay (μs)
64	82.54
128	88.75
256	99.97
512	123.88
1024	169.84
1280	199.52
1518	210.69

Table 2: Link delay.

Length	Average delay (μs)
10	0.49
30	0.49
60	0.50
90	0.53

It is well known that, no matter which architecture of monitoring network is, the transmission delay T can be similarly described as

$$T = T_{\text{td}} + T_{\text{pd}} + T_{\text{qd}}, \tag{13}$$

where T_{td} is the transmission delay, T_{pd} is the processing delay, T_{sd} is the sending delay, and T_{qd} is the queuing delay. Among the four parameters, T_{td} is relevant to links, while the remaining three relate to nodes. One can test the delay of links and nodes separately. The tested node is a dedicated Ethernet switch for trains that conforms to EN50155 standard [20]. The tested link is an Unshielded Twisted Pair 6/100Base-TX that conforms to TIA/EIA-568B.2-1 standard [21]. The test process exactly complies with RFC-2544 standard [22]. The node delay is shown in Table 1 and the link delay is presented in Table 2.

The delay in Table 1 refers to the time period during which the first bit of data frame enters a node and the last bit leaves the node. When testing, the node's carrier was 50% of its limit. The result reflects nodes' average delay under various data frames.

The delay in Table 1 refers to the time period during which the first bit of data frame enters to a link and the last bit leaves the link. The testing communication rate was 100 Mbps. The result reflects links' average delay under various link lengths.

Concluding from the testing results, nodes' delay is very trivial when compared with links' delay. Therefore, the transmission delay can be ignored. Additionally, because the sending delay and the queuing delay are products of forwarding, the optimization of delay concerns how many time specific information has been forwarded, that is, how many nodes it has passed.

As the objective of this paper is to find the shortest communication path from one node to another in train communication system, which is a kind of all-pairs shortest path problem, we apply the Floyd-Warshall algorithm. Such an algorithm had been greatly useful in solving the shortest path problems of networks [23–25]. The key issue here is

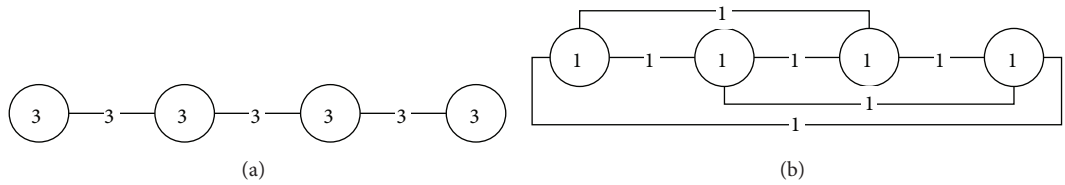


Figure 3: Sample networks with four nodes.

Table 3: The reliability and cost of link types.

Type	Reliability	Unit cost
1	0.95	20
2	0.85	12
3	0.75	8

Table 4: The reliability and cost of node types.

Type	Reliability	Cost
1	0.95	2500
2	0.90	1750
3	0.85	1500

how to convert the result optimal gene of upper level into the adjoint matrix required by solving the Floyd-Warshall algorithm. Here, reuse the relationship between gene x and adjoint matrix a in 3.1 to achieve the conversion.

4. Numerical Examples

In the present section, two examples will be posed to verify the accuracy and the effectiveness of the CRD model, and Example 2 will solve a practical problem in specific.

Example 1. First, assume there are three kinds of links and three types of nodes, and they have different reliability and unit cost. The determination of parameters is shown in Tables 3 and 4.

There are four nodes in the network aligned arrayed linearly, and the node intervals are 40 meters. Concluding from the calculations, to build a network that has the lowest reliability, as illustrated in Figure 3(a), the minimum cost will be 6,960. In contrast, to establish one that has the highest reliability, as presented in Figure 3(b), the maximum cost will be 18,000. Notice that the broken line in Figure 3(b) does not mean the actual length increases, and it is merely for clear illustration.

Set the cost limitation as 6,960 and 18,000; the penalty factor $\lambda = 0.1$. Run the proposed algorithm and see whether the outcome is accurate. The results are shown in Figure 4.

The figure presents the outcome of the genetic algorithm and the shortest communication path result of logical topology optimization. As can be seen from the figure, when the cost constraint is 6,960, we get the optimal result at about the 100th iteration, and the optimal solution is about −0.2202. When the cost constraint is 18,000, we get the optimal result at about the 10th iteration and the optimal solution is about −0.8147. The numbers and dashes under Figure 4 stand for the optimal path results. For example, "A-B-C-D" means the optimal path for delivery from node A to D is to pass node B and node C. It can be reasonably concluded that both levels of the proposed bilevel programming algorithm can yield feasible solutions.

Example 2. There is a train with six carriages, and one node of each carriage needs to be connected with nodes in other carriages. This structure has been widely applied in actual metro trains [26]. Because each carriage is an average 25 meters long and considering the actual wiring type, wiring between carriages requires 30-meter-long cable. Due to signal attenuation, the maximum link length in the network cannot exceed 100 meters. The cost and the reliability are given in Tables 1 and 2. Set the cost constraint as 14,500 at minimum and 22,000 at maximum. Under these conditions, find the optimal communication style with the highest reliability and the shortest delay.

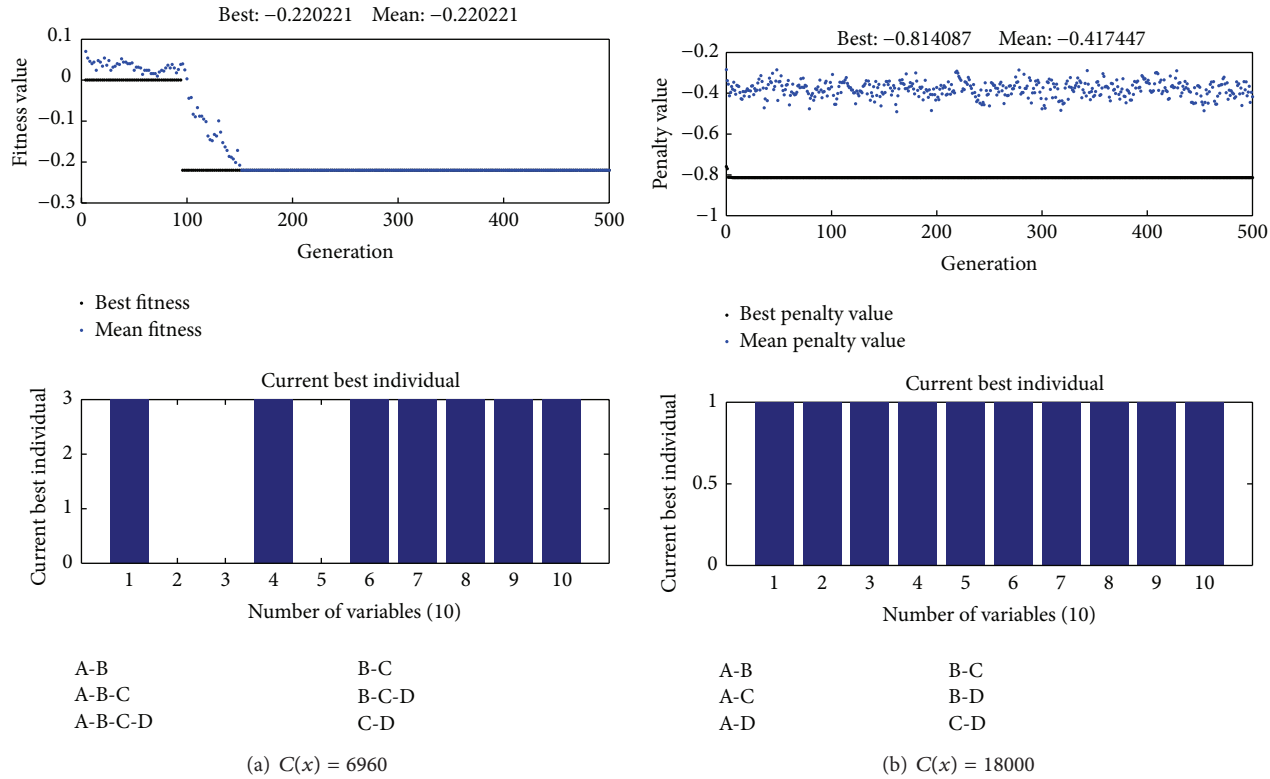
Run the CRD model and obtain the optimal solutions of nodes' connection, as shown in Table 5 and Figure 5.

Compare the optimization results of these two sets of conditions: when the cost increases, the reliability of train communication network increases. Meanwhile, while the number of links increases, the delay of networks decreases. Concluding from the specific analysis, the right way to obtain networks with reasonably high reliability as well as low delay is to add links so as to form redundant connections, rather than rely on choosing equipment with high reliability.

It can be concluded from the actual example that the CRD model is able to optimize simultaneously the physical and the logical topology of networks, changing the traditional methods and reducing parameters' repeat imputation as well as enhancing working efficiency. Seeing from the actual example verification, the CRD model can find most reliable links under cost constraint and the transmission method with minimum delay, so that it satisfies the ultimate optimization objective. Therefore, the CRD model is feasible and effective to solve practical problems. However, we did not consider the variation of reliability when a link passes carriages, which should be an additional condition for logical topology design. Although this problem does not affect the cost, reliability, and delay in the proposed algorithm, it definitely influences the choice of equipment during construction. And this will be the direction for future research.

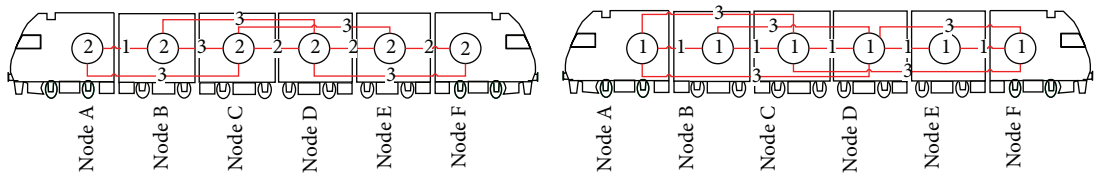
Table 5: The optimal solutions of Example 2.

Optimal solution x	Highest cost $C(x)$	Reliability $R(x)$	The optimal logical link
[1 3 0 0 0 3 3 0 0 2 3 0 2 3 2 2 2 2 2 2 2]	14500	0.4841	A-B; A-C; A-C-D; A-C-E; A-C-E-F; B-C; B-D; B-D-E; B-D-F; C-D; C-E; C-E-F; D-E; D-F; E-F;
[1 3 3 0 0 1 3 0 0 1 0 3 1 3 1 1 1 1 1 1 1]	22000	0.7249	A-B; A-C; A-D; A-D-E; A-D-F; B-C; B-D; B-D-E; B-D-F; C-D; C-D-E; C-F; D-E; D-F; E-F;



(a) $C(x) = 6960$ (b) $C(x) = 18000$

Figure 4: The results of Example 1.



(a) $C(x) = 14500$ (b) $C(x) = 22000$

Figure 5: The physical links of Example 2.

5. Conclusion

In this paper, a general model is proposed to describe the problem of train communication network design. Such design task is divided into two processes: the physical topology optimization and the logical topology optimization. Designing train communication network under cost constraint is the focus of this paper. Taking advantage of the bilevel programming theory, this paper proposes CRD model, which is able to balance the cost, real time, and

delay, to address the design problem. The upper level of CRD model discusses the way to obtain the network with maximum reliability under cost constraint. The lower level explores the method to get the minimum delay for the physical topology designed by upper level. The application of CRD model shows that the cost of network designing is proportional to reliability while inversely proportional to delay. What's more, this paper puts forward a solution for the CRD model that combines the genetic algorithm and the Floyd-Warshall algorithm. With proper configuration, such solution can comprehensively reflect genes of networks, as well as effectively incorporating the two algorithms referred. Finally, the accuracy and the effectiveness of the CRD model and its solution are verified by a practical example. Another actual application of the CRD model is made on a metro train. The result justifies that the CRD model has certain guiding significance and can provide theoretical foundation to construction projects. Emphasis of future researches will be on the weight of nodes in network reliability.

Conflict of Interests

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

Acknowledgments

The authors would like to thank the reviewers for their detailed reviews and constructive comments, which have helped improve the quality of this paper. This work was supported by the National High Technology Research and Development Program of China ("863" Program) (Grant nos. 2011AA110505 and 2011AA11A102).

References

[1] J. C. Moreno, E. J. Laloya, and J. Navarro, "Line redundancy in MVB-TCN devices: a control unit design," in *Proceedings of the IEEE Mediterranean Electrotechnical Conference (MELECON '06)*, pp. 789–794, IEEE, May 2006.

[2] J. C. Moreno, E. Laloya, and J. Navarro, "A link-layer slave device design of the MVB-TCN bus (IEC 61375 and IEEE 1473-T)," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 6 I, pp. 3457–3468, 2007.

[3] M. Bertoluzzo, G. Buja, and D. Pradal, "Experimental analysis of a time-triggered network designed for train communications," in *Proceedings of the 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON '07)*, pp. 2604–2609, IEEE, November 2007.

[4] H. Zhiwu, Z. Sheng, G. Weihua, and L. Jianfeng, "Research and design of protocol analyzer for multifunction vehicle bus," in *Proceedings of the 7th World Congress on Intelligent Control and Automation (WCICA '08)*, pp. 8353–8357, Chongqing, China, June 2008.

[5] E. Szlachcic, "Fault tolerant topological design for computer networks," in *International Conference on Dependability of Computer Systems (DepCos-RELCOMEX '06)*, 2006.

[6] S.-T. Cheng, "Topological optimization of a reliable communication network," *IEEE Transactions on Reliability*, vol. 47, no. 3, pp. 225–233, 1998.

[7] F. Shao, X. Shen, and P. Ho, "Reliability optimization of distributed access networks with constrained total cost," *IEEE Transactions on Reliability*, vol. 54, no. 3, pp. 421–430, 2005.

[8] H. Unzueta, J. Jiménez, J. L. Martín, J. Andreu, and C. Cuadrado, "An emulator to develop the Wire Train Bus protocol stack," in *Proceeding of the 32nd Annual Conference on IEEE Industrial Electronics (IECON '06)*, pp. 3721–3726, Paris, France, November 2006.

[9] J. Jiménez, J. L. Martin, C. Cuadrado et al., "A top-down design for the train communication network," in *IEEE International Conference on Industrial Technology*, vol. 2, pp. 1000–1005, 2003.

[10] T. Fencl, P. Burget, and J. Bilek, "Network topology design," *Control Engineering Practice*, vol. 19, no. 11, pp. 1287–1296, 2011.

[11] C.-S. Wang and C.-T. Chang, "Integrated genetic algorithm and goal programming for network topology design problem with multiple objectives and multiple criteria," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 680–690, 2008.

[12] Z. Gao, J. Wu, and H. Sun, "Solution algorithm for the bi-level discrete network design problem," *Transportation Research B. Methodological*, vol. 39, no. 6, pp. 479–495, 2005.

[13] H. Poorzahedy and M. A. Turnquist, "Approximate algorithms for the discrete network design problem," *Transportation Research B: Methodological*, vol. 16, no. 1, pp. 45–55, 1982.

[14] F. Altiparmak, B. Dengiz, and A. E. Smith, "Reliability optimization of computer communication networks using genetic algorithms," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4676–4680, October 1998.

[15] H. Farvaresh and M. M. Sepehri, "A branch and bound algorithm for bi-level discrete network design problem," *Networks and Spatial Economics*, vol. 13, no. 1, pp. 67–106, 2013.

[16] J. Weng, X. Zou, and X. Qian, "Reliability optimization of overlay multicast topology with efficient genetic algorithm," in *Proceedings of the 8th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM '12)*, IEEE, September 2012.

[17] K. Liu, Z. Chen, A. Abraham, W. Cao, and S. Jing, "Degree-constrained minimum spanning tree problem using genetic algorithm," in *Proceedings of the 4th World Congress on Nature and Biologically Inspired Computing (NaBIC '12)*, pp. 8–14, Mexico City, Mexico, November 2012.

[18] E. L. Da Silva, H. A. Gil, and J. M. Areiza, "Transmission network expansion planning under an improved genetic algorithm," *IEEE Transactions on Power Systems*, vol. 15, no. 3, pp. 1168–1174, 2000.

[19] H. Kirrmann and P. A. Zuber, "The IEC/IEEE train communication network," *IEEE Micro*, vol. 21, no. 2, pp. 81–92, 2001.

[20] E. Idt, *Railway Applications-Electronic Equipment Used on Rolling Stock*, 50155, 2007.

[21] A. C. H. Cable, TIA/EIA-568 Category 6 solid conductor unshielded twisted pair (UTP), 23 AWG, 100 ohm; 4 individually twisted pairs; covered with blue jacket and complying with all relevant parts of and addenda to latest edition of TIA/EIA-568 and UL 444.1. locations other than in plenums, provide NFPA.70.

[22] S. Bradner and J. McQuaid, "RFC 2544. Benchmarking methodology for network interconnect devices," 1999.

[23] S. Pallottino, "Shortest–path methods: complexity, interrelations and new propositions," *Networks*, vol. 14, no. 2, pp. 257–267, 1984.

[24] P. Hofner and B. Moller, "Dijkstra, floyd and warshall meet kleene," *Formal Aspects of Computing*, vol. 24, no. 4–6, pp. 459–476, 2012.

[25] A. Aini and A. Salehipour, "Speeding up the Floyd-Warshall algorithm for the cycled shortest path problem," *Applied Mathematics Letters*, vol. 25, no. 1, pp. 1–5, 2012.

[26] B. S. Zhang and S. X. Lu, "Numerical simulation and analysis of compartment fire in subway train," *Applied Mechanics and Materials*, vol. 166, pp. 2726–2730, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 132936, 5 pages
http://dx.doi.org/10.1155/2014/132936

Research Article

Managing Rush Hour Congestion with Lane Reversal and Tradable Credits

Qing Li[1,2] and Ziyou Gao[1]

[1] *School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[2] *Department of Statistics, College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China*

Correspondence should be addressed to Qing Li; i_liqing@163.com

Received 11 April 2014; Revised 21 June 2014; Accepted 1 July 2014; Published 4 August 2014

Academic Editor: X. Zhang

Within the morning and evening rush hour, the two-way road flows are always unbalanced in opposite directions. In order to make full advantage of the existing lanes, the two-way road lane has to be reallocated to play the best role in managing congestion. On the other hand, an effective tradable credit scheme can help to reduce the traffic demand and improve fairness for all travelers. So as to alleviate the commute congestion in urban transportation network, a discrete bilevel programming model is established in this paper. In the bilevel model, the government at the upper level reallocates lanes on the two-way road to minimize the total system cost. The traveler at the lower level chooses the optimal route on the basis of both travel time and credit charging for the lanes involved. A numerical experiment is conducted to examine the efficiency of the proposed method.

1. Introduction

In many big cities, a large number of people commute from residential areas to their workplaces in the central business district (CBD) in the morning. It leads to the phenomenon that the lanes are congested in the direction from residential areas to CBD while the other lanes in the opposite direction are free. But in the evening, all of that will be reversed. In order to adjust the asymmetric amounts of traffic flow in the two directions of one road section, the lane reversal is employed as a traffic control technique. It is an effective way to make full use of existing road resource and can increase traffic supply.

Lane reversal has been used for many years in managing road congestion. In recent years, lane reversal has been introduced in Chinese big cities, such as Beijing, to relieve morning and evening commute congestion. In the last year, Chaoyang Road, a big thoroughfare on the city's east side, has been tested as the first lane reversal in Beijing to allow traffic to travel in either direction depending on certain conditions. According to the practice, the lane reversal can help to alleviate traffic congestion to some extent.

The implement of lane reversal is a network design problem (NDP). A plenty of research works are focused on the effectiveness, feasibility, and safety of implementing lane reversal [1–5]. The development of techniques, applications, and the engineering practices of lane reversal also attracts much attention [6]. In practice, improper use of lane reversal may lead to a worse condition of the whole road network. Congestion on the road with lane reversal may be relieved, but other parts of the network will be more congested. Therefore, the lane reversal must be put into use on the basis of adjustment of the whole urban road network [7, 8]. A lot of research on NDP is concerned with uncertain demand [9–12]. For simplicity, the commute traffic demand is assumed to be fixed demand. Apparently the number of reallocated lanes is integer. As a result, the implement of lane reversal is a discrete network design problem (DNDP). In essence, the lane reversal contributes to increasing road supply.

In addition to increasing supply, another efficient way to manage congestion is reducing traffic demand. It is widely acknowledged that congestion pricing is helpful in traffic demand management, but the equity debates confined its adoption all over the world. Without taking into account

commuters' income level, trip purposes, and valuation of time, congestion pricing scheme might increase individuals' travel costs. Though it can improve the system performance, it often has to face public resistance.

According to the latest research on congestion control, tradable credit scheme is proved to be a fairer, more effective, and more practical congestion management scheme. It should be noticed that Yang and Wang first explored a tradable credit scheme in a general transportation network equilibrium context with homogeneous and heterogeneous travelers, respectively [13, 14]. Under their credit scheme, the social planner is assumed to develop an initial distribution of the credits to all eligible travelers and link-specific charges to travelers on that link. Credits can be traded freely among travelers. It is the competitive market rather than the planner that can determine the price of the credit. An optimization model subject to a total credit consumption constraint is formulated to find the equilibrated credit price. With an appropriate distribution of credits among travelers and correct selection of link-specific rates, the results of the traditional congestion pricing scheme can be duplicated. Furthermore, the combined tradable credit scheme is shown to be both system-optimal and a Pareto-improvement in a revenue-neutral manner. The same context of research has been done on the tradable credits scheme employed in managing rush hour travel choice and bottleneck congestion [15–18].

In this paper, a discrete bilevel programming model is constructed for managing rush hour congestion caused by underutilization of the existing road resource. The proposed model employs tradable credit scheme and lane reversal for increasing traffic supply and decreasing traffic demand, respectively. At the upper level, the government chooses optimal lanes to be reallocated to minimize the total system costs. Taking into account the generalized travel cost including travel time and link-specific charges for using the links, the travelers at the lower level will choose the optimal route to minimize it.

This paper is organized as follows. In Section 2, a discrete bilevel programming model is established with lane reversal and tradable credit scheme. A chaotic algorithm is employed to solve the proposed model. In Section 3, numerical experiments and analysis results are illustrated with a basic network. In Section 4, conclusion of this paper is presented.

2. Discrete Bilevel Programming Model with Lane Reversal and Tradable Credits Scheme

Consider a two-way network $G = (N, A)$ with a set N of nodes and a set A of directed links. Link in one direction of a road section is denoted by $a \in A$ and link in the opposite direction is $a' \in A$. Let W and R_w denote the set of O-D pairs and the set of all routes between an O-D pair $w \in W$. The travel demand for each O-D pair $w \in W$, denoted by d_w, $d_w > 0$, is given and fixed.

To all directed links in the road network, n_a and $n_{a'}$ denote the number of lanes on link $a \in A$ and $a' \in A$. For simplicity, let $n_a = n_{a'}$ before lane reversal. The capacity of each lane on link $a \in A$ is assumed to be equal to unity, which is denoted by c_a. That is to say, the traffic capacity of link $a \in A$ is $n_a \cdot c_a$ before lane reversal. After lane reversal, u_a is used to denote the number of lanes in the opposite direction on link $a' \in A$ to be added to link $a \in A$. If $u_a > 0$, it means that u_a lanes on the opposite direction will add to link $a \in A$ and its capacity is equal to $(n_a + u_a) \cdot c_a$. On the other hand, if $u_a < 0$, it means that u_a lanes on link $a \in A$ will be added to link $a' \in A$ on the opposite direction. Relative to other means for network capacity enhancement, the investment of lane reversal can be ignored.

For simplicity, consider a separable link travel cost function $t_a(v_a, u_a)$, which is assumed to be nonnegative, continuously differentiable, convex, and monotonically increasing with respect to the amount of aggregate traffic flow v_a on link $a \in A$ and u_a, the number of lanes to be reallocated. In general, $\partial t_a(v_a, y_a)/\partial y_a$ is assumed to be continuous too. Assume also that travelers are homogeneous, which means that they have the same value of time (VOT).

Tradable credit scheme is characterized by its initial distribution and the charging scheme. To minimize complexity, the initial distribution schemes considered here will be O-D specific for a given and fixed demand d_w. $k = \{k_w\}$ is used to denote the credit distribution scheme. Here k_w is the amount of credit distributed to each traveler over the O-D pair $w \in W$. Let K denote the total amount of credits for all links which is predetermined, and obviously $K = \sum_{w \in W} k_w d_w$. The credits are month specific so as no one can benefit from the trade of credit. Analysis here is restricted to link-specific credit charging. Let $\kappa = \{\kappa_a, a \in A\}$ denote the charging scheme, where κ_a is the credit charge for using link $a \in A$. Then (K, κ) will be used to represent a credit charging scheme κ with a total number of credits K for all links.

Let f_r^w denote the traffic flow on route $r \in R_w$ between O-D pair $w \in W$. f is a path flow vector $f = (f_r^w, r \in R_w, w \in W)$, and Ω_f represents the set of feasible path flow patterns defined as follows:

$$\Omega_f = \left\{ f \mid f_r^w \geq 0, \sum_{r \in R_w} f_r^w = d_w, r \in R_w, w \in W \right\}. \quad (1)$$

Let $v = (v_a, a \in A)$ denote the link flow vector, and Ω_f represents the set of feasible link flow patterns defined as follows:

$$\Omega_v = \left\{ v \mid v_a = \sum_{w \in W} \sum_{r \in R_w} f_r^w \delta_{a,r}, f \in \Omega_f, a \in A \right\} \quad (2)$$

and $\delta_{a,r} = 1$ if route r uses link a and 0 otherwise.

It has been proved that not all (K, κ) can guarantee the existent feasible network flow patterns. The amount of credits might not be big enough for supporting all travelers going through the network even if all of them choose the least-credit paths. In order to ensure the existence of feasible network flow patterns, the feasible set of credit schemes denoted by Ψ as follows:

$$\Psi = \left\{ (K, \kappa) \mid \exists f \in \Omega_f \text{ such that } \sum_{a \in A} v_a \kappa_a \leq K, v \in \Omega_v \right\} \quad (3)$$

Ψ is assumed to be nonempty.

The following bilevel programming model is to minimize the sum of total system costs, while the travelers choose the optimal route for minimizing the generalized travel costs including both travel time and link-specific credit charges for using the links. The model can be described by

$$\text{(BLP)} \quad \min_{U} \text{SC} = \sum_{a \in A} v_a t_a \left(v_a, u_a\right) \tag{4}$$

subject to

$$\begin{gathered} u_a \in \left\{-n_a, -\left(n_a - 1\right), \ldots, -1, 0, 1, \ldots, n_a - 1, n_a\right\}, \quad a \in A \\ u_a + u_{a'} = 0, \quad a \in A,\ a' \in A, \end{gathered} \tag{5}$$

where $V = (v_a(u_a))$ is the solution of the next problem

$$\min_{V} \sum_{a \in A} \int_0^{v_a(u_a)} t_a \left(\theta, u_a\right) d\theta \tag{6}$$

subject to

$$\begin{gathered} \sum_{a \in A} v_a \kappa_a \le K, \\ \sum_{r \in R_w} f_r^w = d_w, \quad r \in R_w,\ w \in W, \\ v_a = \sum_{w \in W} \sum_{r \in R_w} f_r^w \delta_{a,r}, \quad a \in A, \\ f_r^w \ge 0, \quad \forall w \in W,\ r \in R_w. \end{gathered} \tag{7}$$

The model is a mixed integer nonlinear programming program. The decision variable at upper level is integer while variable at lower level is real number. Chaos algorithm here is adopted to solve the proposed model [7]. The steps of the algorithm are as follows.

Step 1. Assume chaos variable is denote by random number $y_a^0 \in [0, 1]$, $a \in A$. Let the initial optimal solution $u_a^0 = 0$, $a \in A$, $\text{SC}^0 = +\infty$. Check set is denoted by Φ, and it is null set. Counter $m = 1$.

Step 2. Chaos variable y_a^m is generated by the following equation:

$$y_a^m = 4 y_a^{m-1} \left(1 - y_a^{m-1}\right), \quad a \in A. \tag{8}$$

Step 3. Carrier can be produced by the following equation:

$$\widetilde{u}_a^m = -n_a - \varepsilon_1 + \left(3n_a + \varepsilon_2\right) y_a^m, \quad a \in A, \tag{9}$$

where ε_1 and ε_2 are all small enough positive numbers, with $\varepsilon_1 < \varepsilon_2$. Let u_a^m be equal to rounding $\widetilde{u}_a^m$, $a \in A$. If $u_a^m \notin \Phi$, $a \in A$, add u_a^m to Φ; then turn to Step 4; otherwise, turn to Step 2.

Step 4. For given u_a^m, $a \in A$, solve the users equilibrium (UE) assignment model at the lower levels (6)-(7). The solution is the link flow vector $v^m = (v_a^m, a \in A)$ at UE.

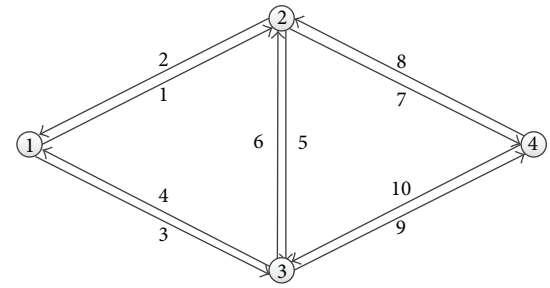


Figure 1: Original structure of the test network before lane reversal.

Table 1: Input data for the test network.

	$t_a(v_a, u_a) = t_a^0(1 + 0.15(v_a/C_a)^4)$			
Link a	t_a^0	C_a	n_a	c_a
1 and 2	4.0	40	2	20
3 and 4	6.0	40	2	20
5 and 6	2.0	60	3	20
7 and 8	5.0	40	2	20
9 and 10	3.0	40	2	20

Table 2: Transportation condition before lane reversal and tradable credit scheme.

Link a	Link flow v_a	Link utilization rate $\chi_a = v_a/C_a$
1 and 2	52.5322 and 31.5193	1.3133 and 0.7880
3 and 4	47.4678 and 28.4807	1.1867 and 0.7120
5 and 6	15.7200 and 13.4320	0.2620 and 0.2239
7 and 8	46.8122 and 28.0873	1.1703 and 0.7022
9 and 10	53.1878 and 31.9127	1.3297 and 0.7978
System cost	$\text{SC} = 1.8240 \times 10^3$	

Step 5. Solve the model at the upper level and get SC^k. If $\text{SC}^m < \text{SC}^0$, let $\text{SC}^0 = \text{SC}^m$, $u_a^0 = u_a^m$, $a \in A$.

Step 6. If the termination condition is met, output the optimal solution u_a^0, $a \in A$; otherwise, turn to Step 2.

3. Numerical Experiments

In this paper, a basic two-way road network, as shown in Figure 1, is employed to validate the efficiency of the proposed model. The network consists of 4 nodes, 10 links, and 2 O-D pairs. One O-D pair is from node 1 to node 4 and the other is from node 4 to node 1. In the morning rush hour, the traffic demand of the two O-D pairs is 100 and 60 in an hour.

The link cost function $t_a(v_a, u_a)$ used here is classical BPR function. Initial capacity of link $a \in A$, which has n_a lanes before lane reversal and each lane's capacity is c_a, is denoted by $C_a = n_a c_a$. The number of C_a, c_a, n_a, and the link free flow travel time t_a^0, $a \in A$, are all illustrated in Table 1.

The traffic flows under UE before lane reversal can be calculated and are shown in Table 2. It shows that the two-way road traffic flows are unbalanced in opposite directions. A parameter $\chi_a = v_a/C_a$ is introduced here to estimate the

Table 3: Results after lane reversal.

Lane reversal					SC
u_1 and u_2	u_3 and u_4	u_5 and u_6	u_7 and u_8	u_9 and u_{10}	
0 and 0	1 and −1	0 and 0	1 and −1	0 and 0	1680

Table 4: Transportation condition after lane reversal.

Link a	Link flow x_a	Link utilization ratio $\chi_a = v_a/K_a$
1 and 2	48.9004 and 41.3251	1.2225 and 1.0331
3 and 4	51.0996 and 18.6749	0.8517 and 0.9337
5 and 6	00.0000 and 22.9740	0.0000 and 0.3829
7 and 8	48.9004 and 18.3510	0.8150 and 0.9176
9 and 10	51.0996 and 41.6490	1.2775 and 1.0412
System cost	SC = 1.67997×10^3	

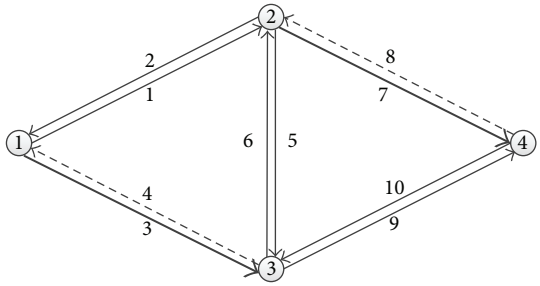


Figure 2: The structure of the test network after lane reversal.

utilization ratio of link $a \in A$. As is shown Table 2, the utilization ratios of links 1, 3, 7, and 9 are bigger than 1. That means that these links have been overused. Otherwise, the utilization ratios of the links in opposite directions are far smaller than them. It can be derived from Table 2 that the utilization ratio of link 2 is slightly bigger than half of that of link 1. And the same happens between link 10 and link 9.

Solve the models (4)–(7) with the use of the algorithm in Section 2. The amounts of credit charging for using links are set as follows: $\kappa_1 = \kappa_5 = 1$, $\kappa_8 = 2$, and $\kappa_a = 0$ for other links $a \in A$. The parameter $\theta = 0.05$ transfers money into equivalent time value. The strategy of the lane reversal and the system cost can be obtained as shown in Table 3.

It is shown in Table 3 that one lane of link 4 and one lane of link 8 have been reversed in their directions and added to links 3 and link 7 respectively. Other lanes of other links do not have any changes. It can be found with the system cost illustrated in Table 4 that the scheme of lane reversal and tradable credits is efficient for managing rush hour congestion.

The transportation condition after lane reversal is shown in Figure 2. The lanes of links 3 and 4 and links 7 and 8 have been adjusted. The traffic flows after lane reversal can be calculated and are shown in Table 4. From the utilization ratio it can be concluded that most of the links have been made full use of. And the system cost is lower than that before lane reversal.

4. Conclusions

With the development of economics and city scales, the road becomes more and more congested in morning and evening rush hour. In order to achieve better effect in solving the congestion problem, traffic demand and increasing road supply are all considered at the same time in this paper. A bilevel programming model is proposed to deal with the two-way road unbalance usage problem. In order to make full advantage of the existing lanes, the two-way road lanes have to be reallocated to play the best role in managing congestion. An effective tradable credit scheme is also employed to help to alleviate the commute congestion with lane reversal in urban transportation network. The models and the algorithm are demonstrated with the basic two-way road network example.

In the future research, the heterogeneous users should be considered. Users with different job and income may have different value of time, so it will be helpful for transportation planning to simulate the real situation.

Conflict of Interests

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

Acknowledgments

This study was jointly supported by the National Basic Research Program of China (2012CB725401) and the National Natural Science Foundation of China (71271205 and 71322102).

References

[1] J. Hemphill and V. H. Surti, "A feasibility study of a facility for a Denver street corridor," *Transportation Research Record*, vol. 514, pp. 29–32, 1974.

[2] R. J. Caudill and N. M. Kuo, "Development of an interactive planning model for contraflow lane evaluation," *Transportation Research Record*, vol. 906, pp. 47–54, 1983.

[3] Y. Jiang and L. Bao, "Study on setting of variable lanes near intersection between one-way and two-way traffic," *Journal of Shanghai Jiaotong University*, vol. 45, no. 10, pp. 1562–1566, 2011.

[4] F. Ni and Z. Liu, "The optimal location of variable message sign," *Information and Control*, vol. 32, no. 5, pp. 395–398, 2003.

[5] X. Gong and S. Kang, "Study and application of traffic direction changing algorithm for urban tide traffic situation," *Journal of Transportation Systerns Engineering and Information Technology*, vol. 6, no. 6, pp. 33–40, 2006.

[6] P. B. Wolshon and L. Lambert, *Convertible Roadways and Lanes: A Synthesis of Highway Practice*, NCHRP Synthesis 340, National Cooperative Highway Research Program, Transportation Research Board, Washington, DC, USA, 2004.

[7] H. Zhang and Z. Gao, "Optimaization approach for traffic road network design problem," *Chinese Journal of Management Science*, vol. 15, no. 2, pp. 86–91, 2007.

[8] X. Liang, C. Hu, C. Ma, and B. Mao, "Empirical study on variable lanes design of chaoyang north street in Beijing," in *Proceedings of the 30th Chinese Control Conference (CCC '11)*, pp. 5527–5531, July 2011.

[9] H. Shao, W. H. K. Lam, and M. L. Tam, "A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand," *Networks and Spatial Economics*, vol. 6, no. 3-4, pp. 173–204, 2006.

[10] H. Shao, Q. Tian, X. Yuan, and G. Wang, "Risk-taking path choice behaviors under ATIS in transportation networks with demand uncertainty," *Journal of Southeast University*, vol. 24, pp. 43–48, 2008.

[11] H. Shao, W. H. K. Lam, Q. Meng, and M. L. Tam, "Travel time reliability-based traffic assignment problem," *Journal of Mangement Sciences in China*, vol. 12, no. 5, pp. 27–35, 2009.

[12] H. Shao, W. H. K. Lam, A. Sumalee, and A. Chen, "Journey time estimator for assessment of road network performance under demand uncertainty," *Transportation Research C: Emerging Technologies*, vol. 35, pp. 244–262, 2013.

[13] H. Yang and X. Wang, "Managing network mobility with tradable credits," *Transportation Research B: Methodological*, vol. 45, no. 3, pp. 580–594, 2011.

[14] X. Wang, H. Yang, D. Zhu, and C. Li, "Tradable travel credits for congestion management with heterogeneous users," *Transportation Research E: Logistics and Transportation Review*, vol. 48, no. 2, pp. 426–437, 2012.

[15] Y. M. Nie and Y. Yin, "Managing rush hour travel choices with tradable credit scheme," *Transportation Research B: Methodological*, vol. 50, pp. 1–19, 2013.

[16] F. Xiao, Z. Qian, and H. M. Zhang, "Managing bottleneck congestion with tradable credits," *Transportation Research B: Methodological*, vol. 56, pp. 1–14, 2013.

[17] L. Tian, H. Yang, and H. Huang, "Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users," *Transportation Research E*, vol. 54, pp. 1–13, 2013.

[18] G. Wang, Z. Gao, M. Xu, and H. Sun, "Models and a relaxation algorithm for continuous network design problem with a tradable credit scheme and equity constraints," *Computers & Operations Research*, vol. 41, pp. 252–261, 2014.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 493809, 8 pages
http://dx.doi.org/10.1155/2014/493809

Research Article

Model of Wagons' Placing-In and Taking-Out Problem in a Railway Station and Its Heuristic Algorithm

Chuijiang Guo and Dingyou Lei

School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, China

Correspondence should be addressed to Chuijiang Guo; guochuijiang@sina.com

Received 13 March 2014; Revised 1 July 2014; Accepted 1 July 2014; Published 23 July 2014

Academic Editor: X. Zhang

Placing-in and taking-out wagons timely can decrease wagons' dwell time in railway stations, improve the efficiency of railway transportation, and reduce the cost of goods transportation. We took the locomotive running times between goods operation sites as weights, so the wagons' placing-in and taking-out problem could be regarded as a single machine scheduling problem, $1\left|p_{ij}\right|C_{\max}$, which could be transformed into the shortest circle problem in a Hamilton graph whose relaxation problem was an assignment problem. We used a Hungarian algorithm to calculate the optimal solution of the assignment problem. Then we applied a broken circle and connection method, whose computational complexity was $O(n^2)$, to find the available satisfactory order of wagons' placing-in and taking-out. Complex problems, such as placing-in and transferring combined, taking-out and transferring combined, placing-in and taking-out combined, or placing-in, transferring, and taking-out combined, could also be resolved with the extended algorithm. A representative instance was given to illustrate the reliability and efficiency of our results.

1. Introduction

After a wagon has been loaded in a loading station, it should be taken out from the goods yard (or railway siding) and coupled to its departure train by a shunting locomotive. When the train arrives at an unloading station, some wagons are removed and placed on their operation sites.

The total time of the process of transporting a wagon from origin to destination can be split into three parts: time spent in the loading station and unloading station, transit time in the marshalling station, and time on the run. According to Chinese statistics, stay time in loading and unloading stations occupies more than 70 percent of the total time [1]. Therefore, organizing wagons' placing-in and taking-out timely can decrease wagons' dwell time in loading and unloading stations, improve the efficiency of railway transportation, and reduce the cost of goods transportation.

The placing-in and taking-out problem has puzzled scholars since the 1950s, and a number of helpful studies have been published. Lei et al. [1] formulated a mathematical model of optimizing operation for placing-in and taking-out wagons in branch-shaped sidings and depicted it as a typical traveling salesman problem (TSP). A combination of a genetic algorithm and an ant colony algorithm called GACA was presented. Li and Du [2] built a graph and a scheduling theory model of the fetching and delivering wagon problem in a branch-shaped railway siding and suggested two fast and simple algorithms. Shi et al. [3] and Yu and Li [4] turned placing-in and taking-out on branch-shaped sidings into a searching Hamilton circle with minimum weights. Heuristic algorithms were used to resolve the problem. Du and Li [5] built a model of scheduling theory for fetching and delivering vehicles on radial sidings and provided a fast algorithm.

Another related problem is TSP with pick-up and delivery, which includes the additional constraint that delivery customers must be visited before any pick-up customers. Nenad and colleagues presented a variable neighborhood search approach for solving the one-commodity pick-up and delivery traveling salesman problem. They adapted a collection of neighborhood structures, k-opt, double-bridge,

and insertion operators, mainly used for solving the classical traveling salesman problem [6]. Renaud and colleagues proposed three mathematical models for the pick-up and delivery problem with shuttle routes and a branch-and-cut-and-price algorithm to solve it [7]. Rais and colleagues described mixed integer-programming formulations for the problem with and without time windows for services [8]. Another helpful research is cited in the reference list [9, 10].

The published research on the placing-in and taking-out wagons problem mostly establishes its model on certain ideal conditions. For example, some scholars considered wagons' placing-in and taking-out as two separate operations, which were difficult to apply to actual situations. In addition, the accuracy of the heuristic algorithms provided deserves further consideration. In this paper, we take the locomotive running time between goods operation sites as weights, regard the wagons' placing-in and taking-out problem as a single machine scheduling problem $1|p_{ij}|C_{\max}$, and transform it into the shortest circle problem in a Hamilton graph, whose relaxation problem is an assignment problem. We use a Hungarian algorithm to calculate the optimal solution of the assignment problem. Then, we use a broken circle and connection method to find a satisfactory order of wagons' placing-in and taking-out. In addition, we study complex problems: placing-in and transferring combined or taking-out and transferring combined; placing-in and taking-out combined; and placing-in, transferring, and taking-out combined.

The remainder of the paper is organized as follows. The problem description is shown in Section 2. In Section 3, we provide a single machine scheduling problem for wagons' placing-in separately operation and transform it into the shortest circle problems in a Hamilton graph, and a broken circle and connection method is used to resolve it. We extend our research to other operation patterns in Section 4. A representative numerical example is presented in Section 5. Section 6 concludes.

2. Problem Description

Railway sidings and freight yards are places where goods are loaded and unloaded. In this paper, we refer to them as goods operation sites. Depending on their layout in a railway station, we can divide them into three types: radial, branch-shaped, and mixed. In a radial station, after the shunting locomotive has sent a wagon-group to one goods operation site, it must go back to the railway station before running on to the next site. Accordingly, every wagon-group reaches its corresponding site at a different time and returns to the railway station at a different time too. For a branch-shaped station, the shunting locomotive does not need to return to the station before placing-in another wagon-group. That is to say, all groups' arrival times are different, but their return times are the same. As regards the mixed goods operation site, the wagon operation shares some common features with the two types above.

In the loading station or the unloading station, a shunting locomotive is assigned to certain tasks, which include placing-in wagons, taking-out wagons, and transferring wagons between the station or operation sites. Because of the diversity of assignment operations and several combinations, the pattern of placing-in and taking-out wagons is summarized as placing-in separately, taking-out separately, placing-in and taking-out combined, placing-in and transferring combined, taking-out and transferring combined, and placing-in, transferring, and taking-out combined.

Considering that trains leave the station on time as much as possible, this paper discusses how the dispatcher optimizes the operation sequence of related sites to minimize the total running time of the locomotive. Only in this way can we reduce the possibility of train delay to a meaningful extent, decrease the cost of locomotive operation to a minimum, and cut down wagons' dwell time in the station.

3. Wagons' Placing-In Separately or Taking-Out Separately

Wagons' placing-in separately means that a shunting locomotive couples all wagons, sends them to the operation sites one by one, puts the wagons in the right place, and then goes back to the station alone. Regardless of whether the wagons waiting to be sent are loaded or empty, the operation process of wagons placing-in separately comprises the following steps: selecting wagon-group, running to the sites, placing wagons, and returning the station. Taking-out separately means that the shunting locomotive goes to goods sites to retrieve the required wagons one by one and goes back to the railway station. The operational process of wagons' taking-out separately comprises the following steps: running to the sites, taking-out wagons, returning the station. It can be concluded from the above that the two operational processes have some identical features, and, hence, we can unite them in one mathematical model.

3.1. Mathematical Model. In this paper, we will discuss the wagons' placing-in problem in a branch-shaped railway station. To fit the actual circumstance of a railway station, the following conditions below are set.

(1) There is only one shunting locomotive in the railway station. The locomotive should send wagons to relevant operation sites, visit every site only once, and return to the railway station after its work has been accomplished.

(2) Numbers of the wagons waiting for placing-in (or taking-out) are known beforehand.

(3) The locomotive running times between all sites are obtained from actual data.

(4) The number of wagons coupled by the locomotive has no influence on the locomotive running times between operation sites.

In what follows, we describe a series of tasks involving four goods sites in a railway station. Several wagon-groups should be sent to corresponding goods sites.

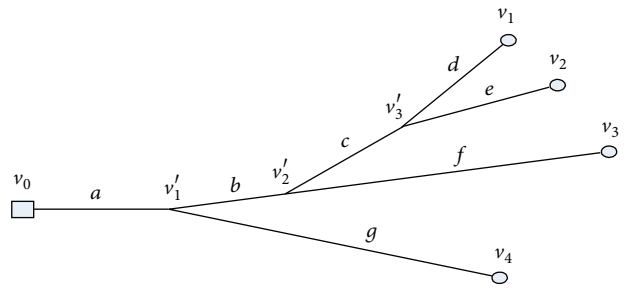


Figure 1: Layout station with branch-shaped goods operation sites.

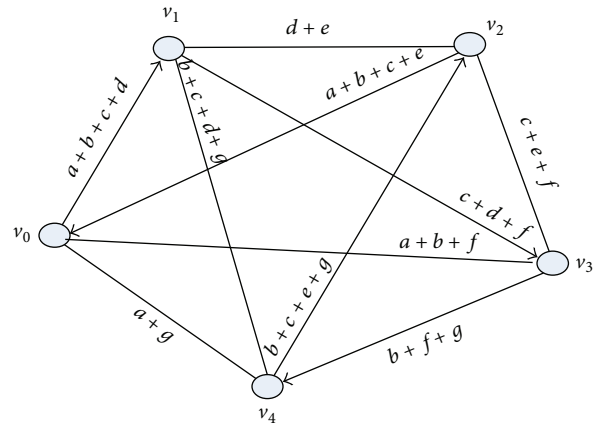


Figure 2: Hamilton graph.

In Figure 1, point v_0 is considered as the railway station and points v_1, v_2, v_3, v_4 are related goods sites. v_1', v_2', v_3' are representations of turnout between rail lines. Notations $a \sim g$ refer to locomotive running times between points (operation sites, turnouts, or the station), which can be gained from actual measurement data. Although there is a slight time discrepancy between the times the locomotive runs back and forth on the same line, we regard them as the same for the sake of simplicity.

If we think of the shunting locomotive as a machine, the operation of placing-in wagon-groups to the site i as processing workpiece J_i and wagons placing-in completion is considered to be workpiece J_i processing accomplishment; the wagons' placing-in problem can be regarded as a single machine scheduling problem $1|p_{ij}|C_{\max}$. If workpiece J_i is the predecessor of workpiece J_j, then the processing time of J_j is t_{ij} $(i \neq j, i, j = 0, 1, 2, 3, 4)$. Accordingly, $i = 0$ means that there is no predecessor workpiece. In other words, the shunting locomotive starts from the station en route to its first operation site. Our objective is to optimize the shunting locomotive operation to achieve minimum operation time after it has accomplished all its tasks. Therefore, the problem of wagons' placing-in separately can be regarded as a single machine scheduling problem $1|p_{ij}|C_{\max}$. p_{ij} is machine processing time, which corresponds to t_{ij} in this paper. $C_{\max}$ is completed time of the last workpiece, which is the time in which the shunting locomotive returns to the railway station. Scheduling problem $1|p_{ij}|C_{\max}$ can be solved by transforming it into the well-known problem of the shortest circle in a Hamilton graph.

We can convert wagons' placing-in separately into the problem of the shortest circle in a Hamilton graph $G = [V, A, C]$. In the Hamilton graph G shown in Figure 2, V signifies the point set $\{v_0, v_1, \ldots, v_4\}$, A is presentation of edge set $\{(v_i, v_j) \mid v_i, v_j \in V\}$, and C expresses shunting locomotive running time set $\{t_{ij} \mid i, j \neq 0, 1, \ldots, 4\}$.

We regard vertex v_0 as the railway station, which is the starting vertex of Hamilton graph G. We consider related operation sites of a series of tasks, which should be visited only once by the locomotive, as vertices of Hamilton. If we can find a circle which goes through all related operation sites only once and its origin and destination are v_0, a Hamilton circle will be achieved. Shunting locomotive running time between any two sites is shown in Figure 1. Corresponding weights of Hamilton circle are the aggregation of locomotive running time on the lines. For example, $v_0 \to v_1 \to v_3 \to v_4 \to v_2 \to v_0$ is a Hamilton circle which has path length of $2a + 4b + 4c + 2d + 2e + 2f + 2g$.

Since all vertices of Hamilton graph G in Figure 2 are connected, it is a complete graph. On the other hand, it is an undirected graph on condition that the time the locomotive spends running back and forth between two sites is considered to be the same. A Hamilton graph composed of N goods operation sites (including the station) has $(N-1)!$ Hamilton circles. It is obvious that it is not easy to obtain the optimal solution with the enumeration method. Hence, it is essential to choose a simple method to select the optimal scheme from the many schemes available.

The problem of the shortest circle in a Hamilton graph is a classic NP-complete combinatorial problem and therefore there is no known polynomial time algorithm (unless $P = \mathrm{NP}$) that is able to solve all instances of the problem. Consequently, heuristic algorithms are used to provide solutions that are of high quality but not necessarily optimal. Since the 1990s, various approaches have been proposed to solve the problem, such as branch-and-bound, cutting planes, 2-opt, simulated annealing, neural network, and tabu search. Some of these methods are exact algorithms and the others are near-optimal or approximate algorithms [11–17]. In this paper, we provide a broken circle and connection method to find a satisfactory circle in a Hamilton graph easily, that is, the satisfactory available order of wagons' placing-in and taking-out.

In a Hamilton circle, as seen in Figure 2, an arrow's head connects with another arrow's tail. In other words, a vertex is the starting-point of one arrow and the end-point of another. Some characteristics of assignment problem in integer programming problems afflict the Hamilton circle. Assigning v_i to v_j as its source point is equivalent to task i being assigned to worker j to do. Accordingly, the weights of edge correspond to the quantity of resource costs.

Introduce the variable x_{ij} and define it as follows:

$$x_{ij} = \begin{cases} 1, & \text{if } v_i \text{ is assigned to } v_j \text{ as its source point} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Table 1: Time coefficient of assignment problem.

Start-point	End-point				
	v_0	v_1	v_2	v_3	v_4
v_0	M	$a+b+c+d$	$a+b+c+e$	$a+b+f$	$a+g$
v_1	$a+b+c+d$	M	$d+e$	$c+d+f$	$b+c+d+g$
v_2	$a+b+c+e$	$d+e$	M	$c+e+f$	$b+c+e+g$
v_3	$a+b+f$	$c+d+f$	$c+e+f$	M	$b+f+g$
v_4	$a+g$	$b+c+d+g$	$b+c+e+g$	$b+f+g$	M

According to the above analysis, the Hamilton graph of wagons' placing-in separately can be relaxed as an assignment problem, and its linear model is shown as follows:

$$\min z = \sum_{i=0}^{n} \sum_{j=0}^{n} t_{ij} c_{ij}, \tag{2}$$

$$\text{s.t.} \sum_{i=0}^{n} x_{ij} = 1, \quad j \in V, \tag{3}$$

$$\sum_{j=0}^{n} x_{ij} = 1, \quad i \in V, \tag{4}$$

$$x_{ij} = 0 \text{ or } 1, \quad i, j \in V. \tag{5}$$

Equation (2) is the objective function which indicates that we want to minimize the whole locomotive running time of a series of placing-in wagon operations. Constraints (3) and (4) signify that the shunting locomotive goes forward to an operation site and returns only once, respectively. Constraint (5) denotes that x_{ij} is a binary variable.

The transformed time coefficient of Hamilton graph G in Figure 2 is shown in Table 1. All vertices of graph G lack a ring. On the other hand, because the shortest circle problem is a minimization problem, we suppose that c_{ij} $(i = j)$ is equal to M (M is a sufficient positive number), in order to ensure that no vertex can be assigned to itself as source point.

The assignment problem, which should be pointed out, is only a relaxation problem of the Hamilton graph discussed. The shortest circle in a Hamilton graph must be a connected graph which requires all arrows' heads and tails to be connected rather than divided. Therefore, the complete form of the Hamilton graph model should include the following two constraints too:

$$\sum_{i,j \in V} x_{ij} \le |S| - 1, \qquad S \subset V, \quad 2 \le |S| \le n-2, \tag{6}$$

where S in the above constraints is a proper subset of set V. $|S|$ denotes the number of elements. Constraints (6) are subtour elimination constraints which prohibit the formation of subtours, that is, tour subsets of fewer than n vertices.

3.2. Solving Method. Solving the model of wagons' placing-in separately in the railway station should be performed in two steps. First, a classical algorithm of the assignment problem, the Hungarian algorithm, is used to calculate the lower bound z_*. Second, with the help of a broken circle and connection method, we can obtain a satisfactory circle in a Hamilton graph, which corresponds to the satisfactory wagons' placing-in order for related operation sites.

3.2.1. Assignment Problem Method. As a classical algorithm of assignment problems, the Hungarian algorithm can obtain a global optimal solution. With an increasing coefficient matrix dimension, the algorithm's running time becomes a little longer, but it can reduce computational efforts much more than any enumeration method. In this paper, the Hungarian algorithm is selected to resolve the relaxation problem of wagons' placing-in separately—the assignment problem.

3.2.2. Broken Circle and Connection Method. The solution obtained with the Hungarian algorithm, which should be pointed out, is only a solution to the relaxation problem. If it has formed a Hamilton circle, the optimal solution regarding the shortest cycle of the Hamilton graph is obtained. If it is still a N $(N > 1)$ bipartite graph, then we break the circles and connect them as follows.

Assuming that the point set of one bipartite graph is $U = \{u_1, u_2, \ldots, u_i, u_j, \ldots, u_m\}$ and that of another is $V = \{v_1, v_2, \ldots, v_k, v_l, \ldots, v_m\}$, we select randomly two point pairs (u_i, u_j) and (v_k, v_l) and exchange their heads and tails. Corresponding weight increments can be computed as follows:

$$\Delta t = t_{il} + t_{kj} - t_{ij} - t_{kl}. \tag{7}$$

Compare weight increments of all point pairs, select the point pairs which have minimal increments, and connect them together. In this case, an N bipartite graph is transformed into an $N-1$ bipartite graph. If this process continues, we gradually obtain a connected graph. Finally, a satisfactory circle of a Hamilton graph can be obtained.

3.2.3. Solution Procedure. According to the analysis above, we perform a computational procedure for the wagons' placing-in separately problem as follows.

Step 1. Transform wagons' placing-in order problem in the railway station into the shortest circle in Hamilton graph G.

Step 2. Compute the relaxation problem to obtain the optimal solution with the Hungarian method. If a connected graph is obtained, go to Step 5. Otherwise, go to Step 3.

Step 3. Randomly select two parts from the N bipartite graph. Choose a point pair from the two selected parts (u_i, u_j) and (v_k, v_l). Find other point pairs and calculate their corresponding values Δt.

Step 4. Compare all Δt values of point pairs. Select the minimum pairs and exchange their heads and tails. Then, $N := N - 1$. If $N = 1$, go to Step 5. If $N > 1$, go to Step 3.

Step 5. Output optimal order of wagons' placing-in to goods operation sites. The computation ends.

Wagons' taking-out separately operation has some same characters with wagons' placing-in separately. Therefore, the model and algorithm of wagons' placing-in separately are also suitable for the problem of wagons' taking-out separately.

3.2.4. Complexity Analysis. If there are n vertices in the Hamilton graph, there are $n/2$ subcycles in the worst case. We should execute a broken circle $(n/2)$-1 times to form only one circle. We should make comparisons $n/2$ times and similarly each circle has $n/2$ vertices at most. Therefore, the complexity of the algorithm in this paper is $o(n^2)$.

4. Complex Problems

4.1. Placing-In and Transferring Combined and Taking-Out and Transferring Combined. Placing-in and transferring combined means that the shunting locomotive transfers wagons during the process of placing-in wagons and then returns to the railway station. The subprocedures are choosing wagon-groups, running to goods sites, aligning wagon location, gathering wagons, transferring wagons, and turning back.

One transferring operation can be regarded as a combination of one taking-out operation and one placing-in operation. Transferring operations can reduce the computational complexity of the algorithm to some extent. Placing-in and transferring combined and taking-out and transferring combined could be considered as the shortest circle problems in a Hamilton graph with visiting priority.

For instance, if we transfer wagons from unloading site v_k to loading site v_l, the shunting locomotive should visit unloading site v_k beforehand, and then it can visit site v_l at some other time. That is to say, site v_k has visiting priority over site v_l, that is, $v_k \succ v_l$ ($\succ$ signifies priority).

The placing-in and transferring combined operation is composed of placing-in operations and transferring operations. Assuming that there is a point set which comprises all placing-in sites, the station, and placing-in sites of all transferring operations, with the broken circle and connection method above we can obtain a satisfactory wagons' placing-in circle for the point set.

Then insert taking-out site of transferring operation into the segments which is behind its placing-in operation site in the circle. For example, we need to transfer wagons from unloading site v_k to loading site v_l. It is assumed that we have already obtained the placing-in order $v_0 v_1 \cdots v_i v_j \cdots v_l \cdots v_n v_0$. If we insert taking-out site v_k of the transferring operation in segment (v_i, v_j), its corresponding route length is increased by $\Delta t' = t_{ik} + t_{kj} - t_{ij}$. We select the circle with minimum increment. Then, a satisfactory placing-in and transferring combined order is obtained.

The taking-out and transferring combined operation consists of taking-out operation and transferring operation. Similarly, we form a point set including all taking-out operation sites, the station, and all taking-out operation sites of transferring operations. We can obtain a satisfactory taking-out circle with the broken circle and connection method for the point set. Then we insert the placing-in site of the transferring operation between the end-points of each segment in front of its taking-out operation site in the circle. We select the circle with minimum increment in terms of route length. Then we can obtain a satisfactory taking-out and transferring combined operation order.

4.2. Placing-In and Taking-Out Combined and Placing-In, Transferring, and Taking-Out Combined. For placing-in and taking-out combined operations, we first formulate a placing-in circle and a taking-out circle, respectively, and transform them into a circle with the broken circle and connection method above. Then a satisfactory operation order of placing-in and taking-out combined operation can be obtained. The placing-in and taking-out operation order gained is relatively centralized, but it can reduce extra shunting work. Although it may prolong running time for the locomotive, it fits with the actual work rather better.

For the placing-in, transferring, and taking-out combined operation, we form a set which includes all taking-out sites and all taking-out sites of transferring operations. On the other hand, we formulate another set which comprises all placing-in sites and all placing-in sites of transferring operations. Then we can obtain a satisfactory placing-in circle and a satisfactory taking-out circle with the broken circle and connection method above. Then we use the method once again. As a result, only one circle is gained in the end. During the process of the last broken circle and connection, we only need to exchange the segments, which are behind the taking-out sites in the taking-out circle, with the segments in the placing-in circle. This ensures a visiting priority relation in which the transferring operation requires and reduces the computational effort. For instance, there is an assignment of transferring some wagons from unloading operation site v_k to loading site v_l. On condition that a taking-out circle $v_0 v_1 \cdots v_k v_i v_j \cdots v_n v_0$ has been gained, we only need to exchange segments of section $v_k v_i v_j \cdots v_n v_0$ for segments of the placing-in circle.

5. Numerical Example

The layout of a railway station is shown in Figure 3. The station is equipped with a shunting locomotive to deliver wagons. There are several tasks in terms of wagons' placing-in and taking-out to do. The tasks are listed in Table 2. The corresponding running times in every line are marked in Figure 3. Now it is necessary to gain the operation order of the shunting locomotive within the least possible time.

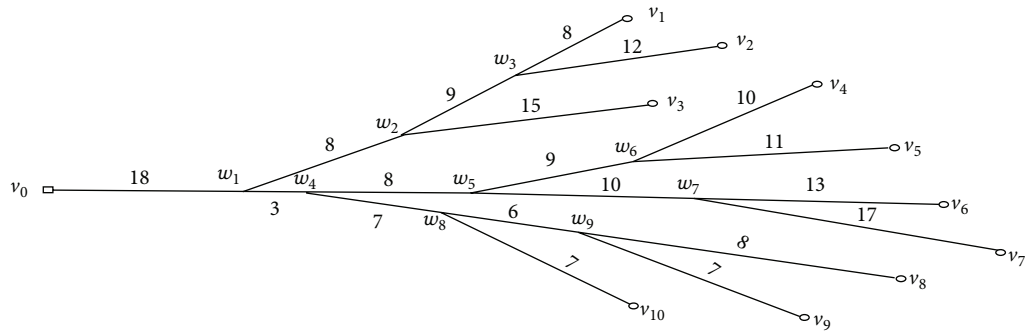


Figure 3: Layout of a railway station.

Table 2: Tasks of the shunting locomotive.

Serial number	Operation sites	Operation contents
1	v_1	Placing-in two wagons
2	v_2	Taking-out two empty wagons and transferring them to site v_4
3	v_3	Taking-out three wagons
4	v_4	Receiving two empty wagons from site v_2
5	v_5	Placing-in one wagon
6	v_6	Taking-out two wagons
7	v_7	Taking-out three empty wagons and transferring them to site v_{10}
8	v_8	Placing-in two wagons
9	v_9	Taking-out three wagons
10	v_{10}	Receiving three empty wagons from site v_7

All operation sites (including the station) are divided into two groups: placing-in operation site set $\{v_1, v_4, v_5, v_8, v_{10}\}$ and taking-out operation site set $\{v_0, v_2, v_3, v_6, v_7, v_9\}$. The priority relations required are $v_2 \succ v_4$ and $v_7 \succ v_{10}$.

Time coefficients for all sites in the railway station are calculated and shown in Table 3, where we set $M = 80$. According to the computational procedure of the Hungarian algorithm, we can obtain the optimal solution to the assignment problem (relaxation problem of the shortest circle in the Hamilton graph): $x_{110} = 1, x_{45} = 1, x_{54} = 1, x_{81} = 1, x_{108} = 1, x_{09} = 1, x_{23} = 1, x_{32} = 1, x_{67} = 1, x_{76} = 1$, and $x_{90} = 1$. The total weight is 316. The corresponding graph of the result is shown in Figure 4.

After the broken circle and connection process has been executed three times, we can obtain a placing-in circle and a taking-out circle and they are shown in Figure 5. Then we execute the last broken circle and connection procedure and compare the total weights of all circles. Exchanging edges (v_2, v_3) and (v_8, v_1), we can obtain a circle with minimum weights, as shown in Figure 6. Hence, the satisfactory wagons' placing-in and taking-out order is $v_0 \to v_9 \to v_6 \to v_7 \to v_2 \to v_1 \to v_{10} \to v_4 \to v_5 \to v_8 \to v_3 \to v_0$, which has total weights of 450.

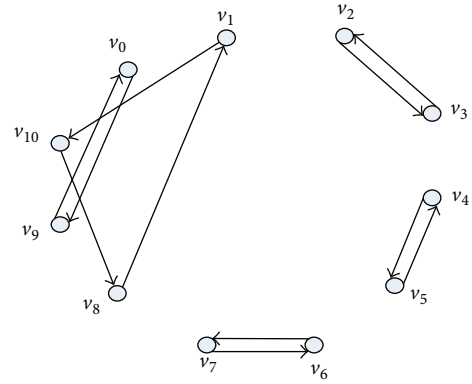


Figure 4: Optimal result of assignment problem.

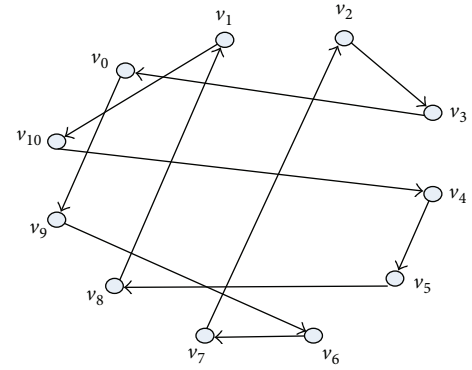


Figure 5: Placing-in circle and taking-out circle (bold lines denote placing-in circle and thin lines denote taking-out circle).

6. Conclusion

Taking the locomotive running times as weights, the problem of wagons' placing-in or taking-out separately can be regarded as a single machine scheduling problem, $1|p_{ij}|C_{\max}$, which can be transformed into the shortest circle problem in a Hamilton graph. The Hungarian algorithm was applied to obtain the optimal solution to the assignment problem. The broken circle and connection method was applied to find a satisfactory order of wagons' placing-in and taking-out. We

Table 3: Time coefficients for all sites in railway station.

Origins destinations	v_0	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{10}
v_0	80	43	47	41	48	49	52	56	42	41	35
v_1	43	80	20	32	55	56	59	63	49	48	42
v_2	47	20	80	36	59	60	63	67	53	52	46
v_3	41	32	36	80	53	54	57	61	47	46	40
v_4	48	55	59	53	80	21	42	46	48	47	41
v_5	49	56	60	54	21	80	43	47	49	48	42
v_6	52	59	63	57	42	43	80	30	52	51	45
v_7	56	63	67	61	46	47	30	80	56	55	49
v_8	42	49	53	47	48	49	52	56	80	15	21
v_9	41	48	52	46	47	48	51	55	15	80	20
v_{10}	35	42	46	40	41	42	45	49	21	20	80

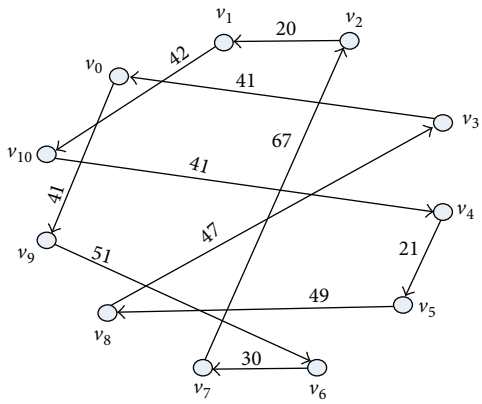


Figure 6: Satisfactory order of wagons' placing-in and taking-out.

estimated the computational complexity of the algorithm to be $o(n^2)$.

We extended the model to other operation patterns: placing-in and transferring combined, taking-out and transferring combined, placing-in and taking-out combined, and placing-in, transferring, and taking-out combined. All these were regarded as the shortest circle problems in a Hamilton graph with visiting priority and could be solved with the extended algorithm provided.

Many experiments also proved that the model and algorithm were feasible on condition that other layouts (radial and mixed goods operation sites) had been converted into branch-shaped form.

The research also provides a new method for solving the problem of the shortest circle in a Hamilton graph. We gained a satisfactory solution with the designed algorithm.

Conflict of Interests

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

References

[1] Y. C. Lei, Z. Y. Tu, W. H. Gui, Z. Wu, and F. Yan, "Optimization of placing-in and taking-out wagons on branch-shaped railway lines based on genetic and ant colony algorithm," *Journal of Central South University (Science and Technology)*, vol. 48, no. 8, pp. 2356–2361, 2011.

[2] W. Q. Li and W. Du, "Mathematical models and algorithms for the problem of fetching and delivering wagons on tree-like railway individual line," *Journal of Henan University (Natural Science)*, vol. 27, pp. 1–8, 1997.

[3] H. Shi, Q. Peng, and H. Guo, "Algorithm by using Hamilton graph to resolve wagon placing-in and taking-out on branch-shaped sidings," *China Railway Science*, vol. 26, no. 2, pp. 132–135, 2005.

[4] S. H. Yu and X. M. Li, "The model of placing-in and taking-out for working wagons and the calculation method," *Railway Transport and Economy*, vol. 24, pp. 46–48, 2002.

[5] W. Du and W. Q. Li, "An algorithm for the problem of fetching and delivering vehicles of through-running train on radial individual line," *Journal of Southwest Jiaotong University*, vol. 30, pp. 503–508, 1995.

[6] N. Mladenović, D. Urošević, S. Hanafi, and A. Ilić, "A general variable neighborhood search for the one-commodity pickup-and-delivery travelling salesman problem," *European Journal of Operational Research*, vol. 220, no. 1, pp. 270–285, 2012.

[7] M. Renaud, R. Stefan, L. Fabien, and P. Olivier, "A branch-and-cut-and-price approach for the pickup and delivery problem with shuttle routes," *European Journal of Operational Research*, vol. 236, pp. 849–862, 2014.

[8] A. Rais, F. Alvelos, and M. S. Carvalho, "New mixed integer-programming model for the pickup-and-delivery problem with transshipment," *European Journal of Operational Research*, vol. 235, no. 3, pp. 530–539, 2014.

[9] H. Hernández-Pérez and J. Salazar-González, "Heuristics for the one-commodity pickup-and-delivery traveling salesman problem," *Transportation Science*, vol. 38, no. 2, pp. 245–255, 2004.

[10] J. Renaud, F. F. Boctor, and J. Ouenniche, "A heuristic for the pickup and delivery traveling salesman problem," *Computers & Operations Research*, vol. 27, no. 9, pp. 905–916, 2000.

[11] M. Held and R. M. Karp, "A dynamic programming approach to sequencing problems," *SIAM Journal on Applied Mathematics*, vol. 10, pp. 196–210, 1962.

[12] J. D. C. Little, K. G. Murty, D. W. Sweeney, and C. Karel, "An algorithm for the traveling salesman problem," *Operations Research*, vol. 11, pp. 972–989, 1963.

[13] M. Padberg and G. Rinaldi, "A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems," *SIAM Review*, vol. 33, no. 1, pp. 60–100, 1991.

[14] M. Grötchel and O. Holland, "Solution of large-scale symmetric travelling salesman problems," *Mathematical Programming*, vol. 51, no. 2, pp. 141–202, 1991.

[15] G. Laporte, "The traveling salesman problem: an overview of exact and approximate algorithms," *European Journal of Operational Research*, vol. 59, no. 2, pp. 231–247, 1992.

[16] D. Gamboa, C. Rego, and F. Glover, "Implementation analysis of efficient heuristic algorithms for the traveling salesman problem," *Computers and Operations Research*, vol. 33, no. 4, pp. 1154–1172, 2006.

[17] C. Rego, D. Gamboa, F. Glover, and C. Osterman, "Traveling salesman problem heuristics: leading methods, implementations and latest advances," *European Journal of Operational Research*, vol. 211, no. 3, pp. 427–441, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 107102, 14 pages
http://dx.doi.org/10.1155/2014/107102

Research Article

Developing a Collaborative Planning Framework for Sustainable Transportation

Okan Örsan Özener

Department of Industrial Engineering, Ozyegin University, 34794 Istanbul, Turkey

Correspondence should be addressed to Okan Örsan Özener; orsan.ozener@ozyegin.edu.tr

Received 31 March 2014; Revised 30 June 2014; Accepted 1 July 2014; Published 22 July 2014

Academic Editor: X. Zhang

Currently, as being the highest petroleum consuming sector in the world, transportation significantly contributes to the total greenhouse gas emissions in the world. Road transportation not only is responsible for approximately 20% of the total emissions of carbon dioxide in the EU and in the US but also has a steadily increasing trend in contributing to global warming. Initiatives undertaken by authorities, such as *Emission cap and trade in the EU,* limit the emissions resulted from the actions of the companies and also give economic incentives to companies to reduce their emissions. However, in logistics systems with multiple entities, it is difficult to assess the responsibilities of the companies both in terms of costs and emissions. In this study, we consider a delivery network with multiple customers served by a single carrier, which executes a delivery plan with the minimum transportation cost, and allocate the resulting costs and the emissions among the customers in a fair manner. We develop allocation mechanisms for both costs and emissions. In order to develop a mechanism that provides further reduction of the emissions, we study a setting where the carrier takes the responsibility of the emissions and reflects the resulting inefficiencies while charging the customers.

1. Introduction and Literature Review

Transportation sector is facing formidable challenges and recently environmental issues lie in the heart of these challenges. Currently, as being the highest petroleum consuming sector in the world, transportation significantly contributes to the total greenhouse gas emissions and hence to the global warming. In the EU, road transportation is responsible for approximately 20% of the total emissions of carbon dioxide. According to European Commission Directorate-General for Energy and Transport [1], in the EU in no other sector has the growth rate of greenhouse gas (GHG) emissions been as high as in transport, which amounted to 23% in 2010 [2] compared to 1990 levels. In the US, transportation is responsible for approximately 28% of the GHG emissions, second to the electricity sector, and compared to 1990 levels it has been increased by 18% [3]. This increasing trend in both the EU and the US is mainly due to the increased demand for travel of individuals and increased freight transportation activities.

On the other hand, road transportation can be considered as the most important mode for passenger and freight transportation. In 2010, total goods transport activities in the EU27 (only intra-EU) have amounted to 3,831 billion tonne kilometers. Road freight transport accounts for 45.8% of the total freight volume in the EU [4]. In the US, road transportation is responsible for hauling 70% of the total freight in 2012. Even though road transportation plays such a crucial role in passenger and freight transportation, it is also the largest contributor to global warming. Over the next 50 years, the GHG emissions of the transportation sector could grow 80% compared to current levels due to increased transportation activities [5].

Recently, there have been several initiatives to reduce the environmental damage caused by transportation activities. According to European Commission Directorate-General for Energy and Transport [1], the environmental suitability of transport activities is a major concern and immediate action should be taken to mitigate the negative effects of transportation on the environment. According to European

Commission Directorate-General for Energy and Transport [1], the EU has recently adopted a climate and energy package that sets a target of reducing GHG emission in the EU by 20% with respect to 1990. The US transportation sector also considers three potential approaches that could reduce GHG emissions as well as fuel consumption: (i) improvements in vehicle technology, (ii) switching to environmentally friendly fuels, and (iii) transportation/travel demand management (TDM), which corresponds to better utilization of the transportation resources [5].

Based on these facts and figures, it is clear that governments and international organizations should play an active role in achieving sustainable transportation. Several governments have already passed legislations to support sustainable transportation including the EU's Transport Protocol of the Alpine Convention [6]. An initiative for reducing GHG emissions is to limit the amount of pollutant emitted by the direct and the indirect actions of a company. This is referred to as "emission cap and trade," which allows controling pollution using economic means. European Union Emission Trading Scheme [7] is an example to these cap and trade initiatives. In this context, companies have certain limits either allocated or sold by the central authority. Companies must limit the emissions resulted by their actions to their thresholds, which are set by the permits they have. If the thresholds are not sufficient, then companies must acquire additional permits from other companies whose emissions are below their limit. The allowance of acquiring additional permits is referred to as "trade," which gives economic incentives to the companies to limit their GHG emissions. However, it is quite difficult to assess the companies' emission responsibilities resulted from their indirect actions.

Having this in mind, we develop mechanisms for the fair assessment of the emission responsibilities of the customers in a logistics system. In this context, a single logistics provider (carrier) may serve several customers along a delivery route in order to minimize the transportation costs. Allocating cost among the customers is required to cover the expenses plus the mark-up of the carrier and allocating emissions is required for emission cap and trade purposes discussed above. The GHG emissions are allocated among the customers in units of grams of CO_2 (gCO_2). Even though the transportation cost and the emissions are correlated as they both depend on the fuel consumption of vehicle, the relationship is not trivial as there are several other factors contributing to each of these values. For instance, the operating cost of a truck is estimated as 49 cents per mile on average; however, only 24.2 cents of this cost is resulted from the fuel consumption and the rest is due to maintenance, repairs, and depreciation [8]. On the other hand, the emissions are largely affected by the payload of the vehicle, as it varies considerably along a delivery route. The cost of emissions depends on the final emission responsibilities of the customers, which determine the total number of permits required by the company as well as the trading cost of a permit on the market. The cost of the permit on the market varies considerably mostly due to the available number of permits. The standard reporting unit is tones of carbon dioxide equivalent (tCO_2-e) and there exist several studies to forecast the price per ton of CO_2 [9]. In this study, we assume the price as \$42 per ton of CO_2 even though this value might fluctuate over time. Note that both per mile cost of transportation and per ton cost emission represent just two parameter values and do not affect the solution methodology at all. On the other hand, the allocation of responsibility is important and affects the cost/emission allocation mechanisms. In the first setting we consider the customers are responsible for both the transportation cost and the emissions directly and the carrier minimizes the transportation cost to serve the customers and allocates both the cost of transportation and the total emissions to the customers individually. In the second setting, the emission responsibility is diverted to the carrier and the carrier decides the minimum total cost plan to serve the customers and allocates a single cost figure, which includes the transportation cost and the emission cost, to the customers. Our motivation for considering the second setting is that it might favor social welfare as the total overall emission might be reduced when the carrier's objective includes this as well.

Unfortunately, neither allocating costs nor allocating emissions are a simple task. The main reason is even though the total cost and the total emission can be easily calculated, it is not trivial to determine which customer is responsible for how much of these totals. We illustrate this on a simple example. Consider a delivery route of a carrier (Figure 1) serving three customer locations. Suppose that the carrier uses a truck with capacity Q. Suppose that Customers B and C are requesting deliveries with weights equal to $0.1Q$ and Customer A is requesting a delivery with a weight equal to $0.8Q$. Suppose that the cost of traveling between any two locations except from A and B is 1 and negligible between A and B. The optimal delivery route is trivial as all demand can be consolidated in one truck (Figure 1). Note that this route also results in the minimum GHG emission, since the largest bulk delivery spends as less time as possible on the vehicle. If we were to allocate both the costs and the emissions proportional to the customers' distance to the depot, then all customers will have equal allocations in both costs and emissions. It is clear that this allocation is not acceptable since Customer A should be allocated a higher emission (due to its load weight) and Customer C should be allocated a higher cost value (due to its location). If we were to allocate both the costs and the emissions proportional to their individual solution values, respectively, then Customers B and C will be allocated the same cost and emission. Again, this is not an acceptable allocation as the cost and the emission contribution of Customer B on this delivery route is less than Customer C and hence should be allocated less in each category (as Customer B has a synergy with Customer A). Finally, if we were to reverse the route, then the contributions of Customer C seem relatively lower as the load delivered to Customer C now spends less time on the vehicle. However, delivery route determination decision is an exogenous decision of the carrier. Hence, the allocation decisions should be independent of the routing decisions. These issues remain even when the carrier takes the responsibility of the emissions and only allocates costs to the customers.

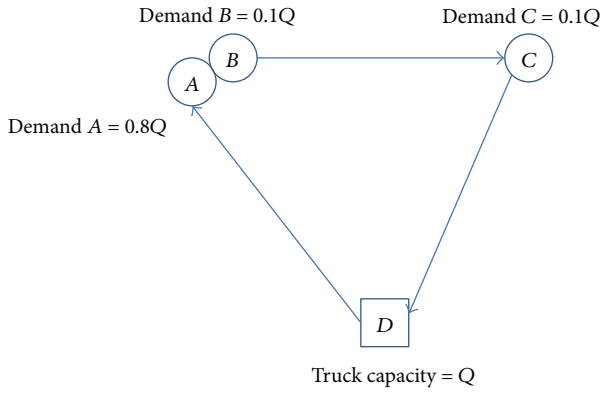


Figure 1: A delivery route.

Before discussing the related works in the literature, we present a list of desirable properties for an allocation method. First, an allocation method should allocate the entire costs/emissions among the participants, which is referred to as "budget balancedness." Second, each individual should be allocated at most its individual solution value, which is called "individual rationality." Third, if we take this property one step further, the same principle should apply to each subset (coalition) of the members, which is referred to as "stability." Fourth, the carrier's routing decisions should not have an effect on the cost/emission allocations. Otherwise, it would not be proper to charge a customer a higher cost/emission value just because the routing decision did not favor that particular customer. Finally, an allocation mechanism should treat each participant "fairly". In our context, a fair allocation method should satisfy budget balanced and stability conditions (referred to as "core allocation" in the literature) and should be independent of the routing decision of the carrier. Faigle et al. [10], Pál and Tardos [11], Engevall et al. [12], and Geomans and Skutella [13] are some examples from the literature which adopt the same fairness definition. In most cases, a core allocation may not exist; hence, our objective is to identify a good approximation to the core.

Allocating the costs/benefits in a logistics system have been studied in the literature under various settings. Tamir [14], Potters et al. [15], and Kuipers [16] consider a traveling salesperson problem (TSP) and discuss the existence of core allocations. Faigle et al. [10] propose a moat-packing based allocation method for TSP to obtain fair allocations. On the other hand, Göthe-Lundgren et al. [17] and Engevall et al. [12] analyze a vehicle routing problem (VRP) setting from a more practical perspective and propose a computational approach to compute nucleolus. Finally, Özener and Ergun [18] consider a vendor managed inventory setting and propose methods to compute cost-to-serve values under this setting.

Recent papers study the carbon emission and transportation relationship from different perspectives. Bektaş and Laporte [19] study the pollution-routing problem, an extension to VRP that takes other factors besides travel distance into account such as greenhouse emissions, fuel, and travel times. Pitera et al. [20] consider a case study on an urban transportation system and propose local search algorithm for the underlying emissions minimization problem. Suzuki [21] proposes a solution approach for a truck-routing problem where the objective is to minimize the fuel consumption and pollutants emission. Xiao et al. [22] study the capacitated VRP with a load-dependent fuel consumption rate and propose a solution approach to this problem using a simulated annealing algorithm with a hybrid exchange rule. Gajanand and Narendran [23] formulate a multiple-route VRP problem to minimize the fuel consumption and show that taking alternative routes may offer reduction in fuel consumption. Xue et al. [24] develop a model of the relationship between traffic flow and vehicle emission under road capacity constraints. Wang et al. [25] focus on the carbon emission rights and propose an auction mechanism which converges to a unique equilibrium.

Finally, Sichwardt [26], Leenders [27], and Naber [28] consider CO_2 allocation among the customers under a VRP setting. Sichwardt [26] analyzes the problem based on the empirical data of a Swedish medium-sized transportation company. The author assumes that the transportation decisions are given based on a combination of the sweep and savings algorithm. After the CO_2 emission calculations on a specific route, the emissions are allocated to the customers based on their distance to depot. Leenders [27] assumes a set of properties as the fairness criteria: (i) individual rationality, (ii) marginality, (iii) efficiency, and (iv) no negative allocation. Based on a comparative analysis, the selected allocation method is the one based on the individual rationality of the shipments. Naber [28] also discusses the vagueness of the term fairness and comes up with a long list of fairness concepts including stability and budget balancedness. The author compares methods such as Shapley Value and nucleolus as well as a proportional-based method. However, the analysis is on a per-route basis and hence the carrier's routing decision may have a huge impact on the final allocation of the emissions. More importantly, as the computations are on an individual route level, the number of customers, which significantly affects the computational complexity of the generic allocation methods, is less than 15. Hence, other than the proportional based method, these approaches are not scalable for real-life sized problems.

Contrary to the studies reviewed above, we propose methods to simultaneously allocate both the transportation costs and the emissions among the customers. We develop an allocation mechanism based on duality and propose an approximation for the Shapley Value. We refer the reader to Owen [29], Kalai and Zemel [30], Pál and Tardos [11], and Geomans and Skutella [13] as examples of the application of the duality-based allocation methods. Our fairness criteria are a more restrictive one compared to the studies discussed above. We computationally show that our methods outperform the proportional methods proposed in those studies. Finally, to the best of our knowledge, this is the first study that considers the setting where the carrier takes over the emission responsibility and allocates a joint cost to the customers.

The remainder of the paper is organized as follows. In Section 2, we provide a formal definition of the problem and list our assumptions. In Section 3, we present solution

approaches for the carrier's routing problem under two different settings. In Section 4, we present our proposed solution approaches, dual LP based and Shapley Value, respectively. In Section 5, we computationally demonstrate how our proposed mechanisms perform in comparison with the other allocation schemes. Concluding remarks are provided in Section 6.

2. Problem Definition

In this section, we provide a formal definition of the problem, explain our emission calculation approach in detail, and finally list our assumptions.

We consider a set of customers $N = \{1, \ldots, n\}$ that are served by a common carrier from a single depot location. Including the depot location, the vertex set is $V = \{0, 1, \ldots, n\}$, where 0 represents the depot and edge set is $E = \{(i, j) : i, j \in V\}$. The distance along an edge (i, j) is denoted by d_{ij} and these figures are assumed to satisfy the triangle inequality. Note that the proposed solution approaches are valid under an asymmetric distance setting and can be used with minor modifications. The carrier serves customers via a set of delivery routes and each route is assigned to one vehicle from a finite number, K, of homogenous vehicles with capacity Q. Customer i is to be delivered a shipment with a weight denoted by q_i, $q_i \leq Q$ for all i. Note that a feasible delivery route corresponds to an ordered customer list with total shipment weight less than equal to the truck capacity.

In this context, delivery routes executed by the carrier determine the total transportation costs and the total emissions. It is important to note that the delivery routes determined by the carrier are not necessarily the optimal set of delivery routes to serve the customers. This is the case in many real life delivery systems; nevertheless, the cost to be allocated among the customers corresponds to the total cost of the selected delivery routes. We assume that the carrier is operating a commercial delivery truck with an operating cost of 49 cents per mile or equivalently 30 cents per kilometer [8]. Note that these values are just parameters and do not actually affect neither the proposed methodology nor the comparative results. We denote the transportation cost along an edge (i, j) by c_{ij}. Finally, there are two distinct approaches for calculating CO_2 emissions of transportation: energy-based approach and activity-based approach. Similar to the studies discussed above, we employ an activity-based approach. However, the difference of our approach is that we use a piecewise linear function with several break points rather than a linear function in order to better approximate the CO_2 emission values. The general formula is as follows:

$$\begin{aligned} CO_2 \text{ emissions} = {} & \text{Transport volume} \times \text{Distance} \times \text{Ave.} \\ & CO_2\text{-emission factor per ton-km.} \end{aligned} \tag{1}$$

Note that the transport volume is the payload of the truck covering that particular distance. The calculation of the average CO_2-emission factor per ton-km is done assuming a 40–44 ton truck and emission factors presented in Table 1, which is acquired from McKinnon and Piecyk [31].

Based on the data above, the emission value along an edge (i, j) with a payload of t, which is denoted by $g_{ij}(t)$, is calculated by identifying the average CO_2-emission factor per ton-km with respect to the payload along the edge and multiplying this factor by the payload volume and the distance. We assume that the truck itself weighs around 14-15 tons and hence the payload volume varies between 0 and 30 excluding the truck's own weight.

In this paper, we consider two different settings based on the emission responsibility (hence the objective function) of the carrier. Our motivation for analyzing two different settings is to assess the performance of these distinct policies from the perspective of fairness in allocating costs and emissions. The settings are as follows.

(i) First setting: the emission responsibility is on the customers and thus the objective of the carrier is to minimize the total transportation costs.

(ii) Second setting: the emission responsibility is on the carrier and thus the objective of the carrier is to minimize the total cost including the transportation costs and the emission costs.

In the first setting, the carrier does not have an incentive to execute low-emission routes and hence treats the emission caused by the delivery routes as a byproduct of the transportation activities. As a result, after the execution of the routes both the total transportation costs and the total emissions are to be allocated among all the customers.

In the second setting, the carrier wants to execute cost-effective and low-emission routes. As a result, after the execution of the routes the carrier has to allocate a joint cost among the customers in a fair manner. For this setting, we assume a unit cost of \$42 per ton of CO_2 emission and one permit is equivalent to one ton of carbon dioxide. Hence, the emission cost is a step function in terms of tons of emission. However, as the carrier might probably have other activities besides serving these customers and will pay an overall emission cost annually for all its activities, it will be an overestimate to use a step function with intervals of tons. Instead, we assume a linear relationship while estimating the emission cost of the delivery routes of the carrier.

3. Solution Method for the Carrier's Routing Problem

In this section, we present a solution method that determines the best routing of shipments from the perspective of the carrier. Note that, in the first setting, the carrier's objective is to minimize the cost of transportation. Hence, the routing decision of the carrier does not take into account the emission caused by the resulting delivery routes. The underlying routing problem is a well-studied problem in the literature, VRP.

In order to solve this VRP, we first iteratively generate clusters using a modified K-means clustering approach, and then based on these clusters we generate a number of feasible delivery patterns. Next, we compute the transportation cost and CO_2 emission of the generated patterns, and finally we

Table 1: Carbon emission factors (gCO_2/ton-km) for 40–44 ton trucks with varying payloads and levels of empty running.

Payload in tonnes	% of truck-kms run empty										
	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
10	81	84.7	88.8	93.4	98.5	104.4	111.1	118.8	127.8	138.4	151.1
11	74.8	78.2	81.9	86.1	90.8	96.1	102.1	109.1	117.3	127	138.6
12	69.7	72.8	76.2	80	84.3	89.2	94.7	101.1	108.6	117.5	128.1
13	65.4	68.2	71.4	74.9	78.9	83.4	88.5	94.4	101.3	109.5	119.3
14	61.7	64.4	67.3	70.6	74.2	78.4	83.2	88.7	95.1	102.7	111.8
15	58.6	61	63.8	66.8	70.3	74.2	78.6	83.7	89.7	96.8	105.3
16	55.9	58.2	60.7	63.6	66.8	70.5	74.6	79.5	85.1	91.7	99.7
17	53.5	55.7	58.1	60.8	63.8	67.2	71.2	75.7	81	87.2	94.7
18	51.4	53.5	55.8	58.3	61.2	64.4	68.1	72.4	77.4	83.3	90.4
19	49.6	51.5	53.7	56.1	58.8	61.9	65.4	69.5	74.2	79.8	86.5
20	48	49.8	51.9	54.2	56.8	59.7	63	66.9	71.4	76.7	83
21	46.6	48.3	50.3	52.5	54.9	57.7	60.9	64.5	68.8	73.9	80
22	45.3	47	48.8	50.9	53.3	55.9	59	62.5	66.5	71.4	77.2
23	44.2	45.8	47.6	49.6	51.8	54.3	57.2	60.6	64.5	69.1	74.7
24	43.2	44.7	46.4	48.3	50.5	52.9	55.7	58.9	62.7	67.1	72.4
25	42.3	43.8	45.4	47.3	49.3	51.7	54.3	57.4	61	65.2	70.3
26	41.5	42.9	44.5	46.3	48.3	50.5	53.1	56	59.5	63.6	68.5
27	40.8	42.2	43.7	45.4	47.3	49.5	52	54.8	58.1	62.1	66.8
28	40.2	41.5	43	44.6	46.5	48.6	51	53.7	56.9	60.7	65.3
29	39.7	41	42.4	44	45.7	47.8	50.1	52.7	55.8	59.5	63.9

solve a set partition problem (SPP) to select the delivery patterns to serve the customers with the minimum cost possible. As the number of customers along a route is limited due to the truck capacity restrictions, we are able to solve these TSPs using exact methods and commercial solvers. Our motivation for choosing this particular solution method among all possible VRP solution methodologies is that we are able to generate a large set of delivery routes, all of which will be used in our allocation methodology.

We use a variant of the K-means clustering algorithm, which is discussed in [32, 33]. In this approach, we partition the customers into a certain number of clusters. In our problem, this number would be equal to the number of available trucks, which is denoted by K. After initiating the first K clusters randomly, we implement a variant of the set partition problem (VSPP) to assign customers to these cluster bases. VSPP ensures that the total weight of the shipments assigned to a cluster does not exceed the truck capacity. The details of the overall solution algorithm are as Procedure 1. Note that H represents the iteration number of the outer loop and F represents the iteration number in the inner loop in generating the delivery patterns. Before presenting the formulation of PS, we introduce the notation used in the formulation. Let x_p be the binary variable indicating whether a pattern is selected in the final solution or not. Let P_i be the set of delivery patterns that visit customer i. The formulation of PS is as follows:

$$PS\colon z_\alpha = \min \sum_{p\in P} \alpha_p x_p, \tag{2}$$

$$\text{s.t.} \quad \sum_{p\in P_i} x_p = 1, \quad \forall i \in N, \tag{3}$$

$$\sum_{p\in P} x_p \le K, \tag{4}$$

$$x_p \in \{0,1\}, \quad \forall p \in P. \tag{5}$$

The objective is to minimize the total transportation costs. Constraints (3) ensure that each customer is served by exactly one delivery route. Constraint (4) ensures that the number of delivery routes executed should be less than or equal to the total number of vehicles. The last constraints are the binary restrictions of the variables.

Solving the carrier's routing problem would determine the minimum cost solution of delivering to all the customers. Note that, in applying Procedure 1, we calculate not only the transportation costs but also the CO_2 emissions of the delivery routes. Here, z_α^* corresponds to the optimal objective function value of PS, which may not correspond the optimal solution to the problem as Procedure 1 does not guarantee optimality. We assume that z_α^* denotes the total cost to be allocated among the customers. We also need to compute the total emissions to be allocated among the customers. Let z_β^* denote that value, which is equal to $\sum_{p\in P} \beta_p x_p^*$. Finally, we are able to generate a large set of feasible delivery patterns, not only the ones executed in the final solution. This is very crucial because, for our proposed allocation methods, we require a range of feasible delivery routes, especially

Initiate a set of delivery patterns $P = \{\}$
For $h = 1, \ldots, H$
  Initiate the cluster centers by randomly choosing K locations out of all customer locations
  For $f = 1, \ldots, F$
    Assign customers to cluster centers by solving VSPP
    For $k = 1, \ldots, K$
      Generate a delivery pattern p for cluster k and add this pattern to set P
      Calculate the centroid for all locations assigned to this cluster center
      Relocate the cluster center to the location closest to center of gravity of the cluster
    End For
  End For
End For
Remove the duplicated patterns in P
For $p = 1, \ldots, |P|$
  Solve a TSP to calculate the cost and the emission of the route α_p and β_p, respectively
End For
Solve PS to select the final K delivery patterns

Procedure 1: Solving the carrier's routing problem.

to determine a fair allocation method independent of the routing solution preferred and executed by the carrier.

In the second setting, the carrier's objective is to minimize the total cost of the delivery routes, which includes the transportation costs and the emission costs. Under this setting, we modify Procedure 1, specifically the step "solving the TSPs to calculate the cost and the emission of the route." Unfortunately, this modification is a major one. In the original procedure, we solve a mathematical model for TSP with regular assignment constraints as well as subtour elimination constraints. However, when the emission costs are also included in the objective function and since the emission costs depend on the payload of the vehicle on a certain link, we need to keep track of the route traveled and hence the payload on the vehicle at any given time. Therefore, we need an entirely different model to handle this situation. The problem resembles the delivery man problem (DMP) studied by Fischetti et al. [34]. In addition to this modification, we calculate the emission values using a piecewise linear approximation with several break points. Hence, we need a further modification of the flow-based model proposed by Fischetti et al. [34]. Unfortunately incorporating the piecewise approximation into this flow formulation makes the model too time consuming to be used for large number of delivery patterns. Therefore, we use a step function for assessing the emission cost in this flow-based model's objective function and calculate the emission cost by using our original piecewise approximation after solving the model. The formulation of our flow-based TSP model, FTSP, is as follows:

$$\text{FTSP}: \gamma_P = \min \sum_{i \in \widetilde{N}} \sum_{j \in \widetilde{N}} c_{ij} r_{ij} + \sum_{i \in \widetilde{N}} \sum_{j \in \widetilde{N}} o_{ij}, \tag{6}$$

$$\text{s.t.} \quad \sum_{i \in \widetilde{N}} r_{ij} = 1, \quad \forall j \in \widetilde{N}, \tag{7}$$

$$\sum_{j \in \widetilde{N}} r_{ij} = 1, \quad \forall i \in \widetilde{N}, \tag{8}$$

$$\sum_{i \in \widetilde{N} \setminus \{0\},} s_{i0} = 1, \tag{9}$$

$$\sum_{i \in \widetilde{N}} s_{ik} - \sum_{j \in \widetilde{N}} s_{kj} = q_k, \quad \forall k \in \widetilde{N}, \tag{10}$$

$$\sum_{i \in \widetilde{N}} s_{i0} - \sum_{j \in \widetilde{N}} s_{0j} = -\sum_{k \in \widetilde{N}} q_k, \tag{11}$$

$$s_{ij} \le M r_{ij}, \quad \forall i, j \in \widetilde{N}, \tag{12}$$

$$\sum_{l=0}^{30} t_{ijl} = 1, \quad \forall i, j \in \widetilde{N}, \tag{13}$$

$$\sum_{l=0}^{30} t_{ijl} (l-1) \le \frac{s_{ij}}{1000}, \quad \forall i, j \in \widetilde{N}, \tag{14}$$

$$\sum_{l=0}^{30} t_{ijl} (l) \ge \frac{s_{ij}}{1000}, \quad \forall i, j \in \widetilde{N}, \tag{15}$$

$$42 \left[\sum_{l=1}^{30} t_{ijl} g_{ij} (l - 0.5) \right] = o_{ij}, \quad \forall i, j \in \widetilde{N}, \tag{16}$$

$$r_{ij} \in \{0, 1\}, \quad \forall i, j \in \widetilde{N}, \tag{17}$$

$$s_{ij} \in \mathbb{Z}^+, \quad \forall i, j \in \widetilde{N}, \tag{18}$$

$$o_{ij} \ge 0, \quad \forall i, j \in \widetilde{N}, \tag{19}$$

$$t_{ijl} \in \{0, 1\}, \quad \forall i, j \in \widetilde{N}, \ \forall l \in \{0, 1, \ldots, 30\}. \tag{20}$$

In the formulation above $\widetilde{N}$ represents the set of customers to be visited by the given TSP pattern p plus the

carrier's depot location. The binary variable r_{ij} denotes whether edge (i, j) is selected in the final TSP tour or not. The integer variable s_{ij} represents the flow volume (weight) along edge (i, j). Note that as q_k values are in kilograms, the flow values along the edges are represented in integer multiples of kilograms and also the weight of the truck is not included in this flow volume. The variable o_{ij} represents the cost of emission along the edge (i, j). Finally, the binary variable t_{ijl} denotes whether on edge (i, j) the flow volume falls within an interval between $l - 1$ and l or not.

The objective function is to minimize the total transportation costs and the emission costs. Constraints (7) and (8) ensure that each customer is visited only once by the delivery pattern. Constraints (9), (10), and (11) ensure that the flow conservation within the network is achieved while satisfying the volume delivery requirements of the customers. In order to represent a positive flow on the vehicle along the selected edges of the pattern, we increase the outgoing flow value from the depot by one kilogram, which is negligible in emission calculations. Constraints (10) also guarantee that there will be no subtours in a feasible solution. Suppose that there is a subtour that does not include the depot (otherwise there is only one grand tour, and hence no subtours). For all the nodes in that particular subtour, we sum up the corresponding constraints of type (10) and this would yield a contradiction. Hence, any feasible solution to FTSP does not have subtours. Constraints (12) guarantee that if an edge has a positive flow then that edge has to be covered by the pattern. Constraints (13), (14), (15), and (16) make sure that the flow on an edge falls within a certain interval between 0 and 30 and the emission cost along an edge is computed according to that specific interval. In Constraints (16), the constant, 42, represents unit cost of emission per ton of CO_2 and $g_{ij}(l - 0.5)$ represents the total amount of emission in tons along edge (i, j) if this edge is covered with a payload of $l - 0.5$ tons. Note that $g_{ij}(l - 0.5)$ values are parameters that can be computed in a preprocessing step. Also, $g_{ij}(l - 0.5)$ is an approximation as even though the real flow value, $(s_{ij} - 1)/1000$, falls within the interval $[l - 1, l]$, it may not be exactly equal to $l - 0.5$. The last four constraints are the binary, integer, and nonnegativity restrictions of the variables. Recall that, after solving the model above, we calculate the total emission cost using our original piecewise approximation and update the total cost of the TSP pattern accordingly. The rest follows by replacing α_p values with γ_p values and solving a modified *PS* to obtain $z^*_\gamma = \sum_{p \in P} \gamma_p x^*_p$.

4. Cost/Emission Allocation Methods

In this section, we describe our proposed allocation methods both for cost and emission allocation. Before describing each method, note that the cooperative game under a VRP setting may have an empty core, which follows directly from the proof on the existence of the core allocation under a TSP setting, and hence both the cost and the emission allocation problems are not guaranteed to have a budget balanced and stable cost allocation. Having this in mind, our objective is to find the most fair allocation method, which we define as being exactly budget balanced and as stable as possible (also note that stability implies individual rationality). Hence, for all the allocation methods, we calculate the percent instability values for all possible groups of customers and decide which allocation method yields better solutions based on these instability values. Finally, ease of computation is another criterion in choosing the best allocation method as the applicability of the proposed methods directly depends on their computational times and the related works in the literature mostly focus on the theoretical contribution and ignore this aspect of the problem.

4.1. An Allocation Method Based on the Distance to Depot. First, we describe an allocation method, which allocates the costs and the emissions proportional to the customers' distances to depot. It is natural to assume that the contribution of a customer to both transportation cost and CO_2 emission of a deliver route is correlated to the customer's distance to the depot, even though this correlation might be stronger in cost contribution than that of emission contribution as the weight of the shipment is ignored. The obvious advantages of such a cost/emission allocation method are as follows: (i) it is a simple/fast allocation method that requires no complex calculations. (ii) it provides individually rational allocations, and (iii) obtained allocations do not depend on the routing solution of the carrier. The disadvantage of this method is that it completely ignores the synergies among the customers and it is very unlikely to obtain approximate core, hence fair, allocations using this proportional method. A particular disadvantage of this method in allocating CO_2 emission is that this method completely ignores the shipment weights of the customer, which is a very important factor in determining the CO_2 emissions. We compute θ_i, the cost allocated to customer i, and σ_i, the emission allocated to customer i, as Procedure 2.

Note that for our second setting, we replace z^*_α value with z^*_γ value in Procedure 2 to obtain cost allocations.

4.2. An Allocation Method Based on the Individual Solutions. Second, we describe an allocation method, which allocates the costs and the emissions proportional to the solutions where the customers are served individually by the carrier. It is clear that the cost allocations will be exactly the same as the proportional allocation based on the distance to the depot. However, the emission allocation will be different as the weight of the shipments becomes a factor in this allocation method. Hence, we expect this allocation method to perform similar to the one discussed above except that the emission allocations might be better in terms of fairness criteria. The details of the procedure are as Procedure 3.

Again, for our second setting, we replace z^*_α value with z^*_γ value in Procedure 3 to obtain cost allocations.

4.3. An Allocation Method Based on Duality. Third, we introduce our first proposed method for allocating the costs and the emissions, which is based on duality. The relationship between core allocation and LP duality is a well-established one since Owen [29]. In this allocation method, we use the

For $i = 1, \ldots, N$

Compute $\theta_i = \dfrac{d_{i0}}{\sum_{j=1}^{N} d_{j0}} z_\alpha^*$

Compute $\sigma_i = \dfrac{d_{i0}}{\sum_{j=1}^{N} d_{j0}} z_\beta^*$

End For

Procedure 2: Allocation method based on the distance to depot.

For $i = 1, \ldots, N$

Compute $\theta_i = \dfrac{2(d_{i0})}{\sum_{j=1}^{N} 2(d_{j0})} z_\alpha^*$

Compute $\sigma_i = \dfrac{g_{0i}(q_i) + g_{i0}(0)}{\sum_{j=1}^{N} (g_{0j}(q_j) + g_{j0}(0))} z_\beta^*$

End For

Procedure 3: Allocation method based on the individual solutions.

dual of the LP relaxation of PS to obtain cost allocations. Note that, from the optimal values of the corresponding dual variables, we obtain weights and by multiplying with z_α^* and z_β^* we obtain the actual allocations.

There are several advantages of the duality-based allocation method: (i) it is a simple/fast allocation method that only requires an LP solution, (ii) obtained allocations do not depend on the routing solution of the carrier, (iii) and it assesses the synergies among the customers better than other allocation methods discussed and hence is more likely to yield fair allocations.

The LP relaxation of PS is as follows:

$$PS_{\text{LP}}: \bar{z}_\alpha = \min \sum_{p \in P} \alpha_p x_p, \tag{21}$$

$$\text{s.t.} \quad \sum_{p \in P_i} x_p = 1, \quad \forall i \in N, \tag{22}$$

$$\sum_{p \in P} x_p \leq K, \tag{23}$$

$$x_p \leq 1, \quad \forall p \in P, \tag{24}$$

$$x_p \geq 0, \quad \forall p \in P. \tag{25}$$

Let y_i be the dual variables associated with constraints (22), let v be the dual variable associated with constraint (23), let u_p be the dual variables associated with constraints (24), and finally let I_p be the set of customers visited by pattern p, and then the dual of the PS_{LP} is as follows:

$$\text{DPS}_{\text{LP}}: \hat{z}_\alpha = \max \sum_{i \in I} y_i + \sum_{p \in P} u_p + vK, \tag{26}$$

$$\text{s.t.} \quad \sum_{i \in I_p} y_i + u_p + v \leq \alpha_p, \quad \forall p \in P, \tag{27}$$

$$y_i \; u.r.s, \quad \forall i \in I, \tag{28}$$

$$v \leq 0, \tag{29}$$

$$u_p \leq 0, \quad \forall p \in P. \tag{30}$$

Solving the dual LP above in polynomial time yields the optimal values for the dual variables y_i, which are the variables to be used in determining the cost allocations. Note that we need to solve another, modified dual LP in order to obtain dual variables for allocating the emissions. In that case, we just replace α_p values with β_p values to obtain $\bar{y}_i^*$. The details of the procedure are as Procedure 4.

To obtain duality-based cost allocations for our second setting, we replace z_α^* value with z_γ^* value and α_p values with γ_p values.

4.4. An Approximation for the Shapley Value. Finally, we introduce an allocation method based on an approximation for the Shapley Value. The Shapley Value is a generic allocation method introduced by Shapley [35]. In a nutshell, Shapley Value is the weighted average of the marginal contribution of each member to each subset of the collaboration. Let $m^i(S)$ be the marginal cost of adding customer i to the subset S and let $l^i(S)$ be the marginal emission of adding customer i to the subset S, and then the Shapley Value calculations are as follows:

$$\theta_i = \sum_{S \subseteq N \setminus i} \frac{|S|!\,|N \setminus (S \cup i)|!}{|N|!} m^i(S), \tag{31}$$

$$\sigma_i = \sum_{S \subseteq N \setminus i} \frac{|S|!\,|N \setminus (S \cup i)|!}{|N|!} l^i(S). \tag{32}$$

The advantages of the Shapley Value allocation method are as follows: (i) obtained allocations do not depend on the

Solve PS_{LP} to obtain $y_i^* \quad \forall i$
Solve $\widetilde{PS}_{\mathrm{LP}}$ to obtain $\tilde{y}_i^* \quad \forall i$
For $i = 1, \ldots, N$

$$\text{Compute } \theta_i = \frac{y_i^*}{\sum_{j=1}^{N} y_i^*} z_\alpha^*$$

$$\text{Compute } \sigma_i = \frac{\tilde{y}_i^*}{\sum_{j=1}^{N} \tilde{y}_i^*} z_\beta^*$$

End For

Procedure 4: Allocation method based on duality.

For $i = 1, \ldots, N$
Let N_i be the set of the customers including the nearest 5 customers and customer i

$$\text{Compute } \theta_i = \sum_{S \subseteq N_i \setminus i} \frac{|S|!|N_i \setminus (S \cup i)|!}{|N_i|!} m^i(S)$$

$$\text{Compute } \sigma_i = \sum_{S \subseteq N_i \setminus i} \frac{|S|!|N_i \setminus (S \cup i)|!}{|N_i|!} l^i(S)$$

End For

Procedure 5: Approximation for the Shapley Value.

routing solution of the carrier, (ii) it assesses the synergies among the customers, and (iii) it satisfies additional properties such as equal treatment of equals, additivity, and dummy properties. A major disadvantage of the Shapley Value is that it requires an exponential effort to compute the allocation unless an implicit method is developed (as the number of subsets grows exponentially). Hence, in its current form Shapley Value is not a suitable method for the large instances of the problem. Therefore, we propose an approximation for the Shapley Value, in which we calculate the weighted average of the marginal contribution of each customer to the subsets that include only the nearest 5 customers. By doing that, we limit the number of subsets and hence limit the computational effort that is required to calculate the Shapley Value. The details of the procedure are as Procedure 5.

To obtain the Shapley Value cost allocations for our second setting, we update $m^i(S)$ values and repeat Procedure 5.

5. Computational Study

We carried out a computational study to demonstrate the performance of our proposed allocation methods. We perform our experiments on randomly generated instances with varying characteristics. These instances are all generated on a region of 1,000 × 1,000 unit square and the Euclidian distances between two locations are divided by 10 to obtain the distances in kilometers. Note that, for calculating the transportation cost, we multiply the distances in kilometers by the operating cost per kilometer multiplier, which is taken as \$0.30.

We generate a total of 20 instances using different parameter configurations. Half of these instances have 25 customers and the rest has 50 customers. We generate a certain portion of the customer locations in clustered regions, representing metropolitan areas. There are 3 clusters in 25-customer instances and 4 clusters in 50-customer instances. Among all the customers, 40% of them are generated within the cluster regions and the remaining ones are randomly placed over the entire map. The load-carrying capacity of the trucks is assumed to be equal to 30 tons. In half of the instances, low volume instances, the shipment weight is uniformly distributed between 5% and 25% of the truck capacity. In the rest of the instances, high volume instances, the shipment weight is uniformly distributed between 10% and 30% of the truck capacity. Low volume and high volume instances are equally distributed among 25- and 50-customer instances.

A major decision in our computational analysis is the number of iterations in our iterative VRP solution procedure. Recall that H represents the iteration number of the outer loop and F represents the iteration number in the inner loop in generating the delivery patterns. Our objective is to generate as many patterns as possible; however, the computational time of the procedure is also a concern. In our computations, we perform two different sets of runs with $H = 50$ and $H = 100$, whereas F is always set to 20. In solving PS, we set the time limit to 3600 seconds (1 hour) and the optimality gap to 1% for the termination of the run. Finally, all the computational experiments are carried out on a 64-bit Windows Server with two 2.4 Ghz Intel Xeon CPU's and 24 GB RAM. The algorithms are implemented using C++ and CPLEX Concert Technology.

Table 2: Instability values of the proposed methods on 25- and 50-customer instances when $H = 50$.

		25-customer				50-customer			
		PR	IS	DU	SV	PR	IS	DU	SV
Cost	Subset	608.9	608.9	178	286.1	1063.1	1063.1	169.2	636.5
	Ave	7.26	7.26	3.59	5.64	10.14	10.14	3.48	6.36
	Max	39.83	39.83	17.43	27.56	51.34	51.34	18.61	30.2
CO_2	Subset	883.5	894.4	304.2	272.7	1020.9	1036.5	279	433.4
	Ave	9.43	9.3	5.27	6.63	11.74	11.89	4.38	6.16
	Max	52.02	51.36	25.81	40.03	57.64	58.83	30.39	32.81

We test the performance of the proposed allocation methods with respect to both solution quality and computational time. Recall that the quality of an allocation method is assessed using our fairness criteria, budget balanced, and stability. As we cannot allow any budget deficiency or surplus, all the allocation methods have to satisfy the budget balanced property. Hence, we need to compute the maximum percent deviation from stability of each allocation method over all the subsets of the customers. Then, we call this maximum percentage deviation as "percent instability" of allocation method and assume that the higher this value is, the less fair the allocation is. Unfortunately, the problem comes with the fact that there are 2^N subsets and this number is huge when $N = 25$ or $N = 50$. Therefore, we use the "approximate stability assessment" procedure developed by Özener et al. [18] to assess the stability of the allocation methods. The purpose of this method is to generate only a fraction of all the possible subsets and, however, at the same time make sure that these subsets are the important ones. The definition of the "importance of a subset" is having the highest possibility of being instable as well as having a high percentage instability value. In that procedure, the inherent assumption is that the customers that are closer to each other have the highest synergy potential and hence constitute an important subset. In their procedure, they pick a random point on the map and assign probabilities to the customers based on their distance to that base point. Then, they generate a subset based on these probabilities and check the percent instability of that subset. We modify this procedure for our application. In our modified procedure, we discard the generated subsets if the customers cannot be served by a single delivery route (total shipment volume exceeds the truck capacity) or their total shipment volumes are less than 75% of the truck capacity. In that sense, we assume that the highest synergy customers subsets are the ones that correspond to the efficient delivery routes. We generate 25,000 subsets for each of the 25-customer instances and to 50,000 for each of the 50-customer instances. Note that the procedure does not allow duplicate subsets.

Table 2 summarizes the performance of the allocation methods when $H = 50$. The first four columns correspond to 25-customer instances and the next four columns correspond to 50-customer instances. Within each four-column group, each column represents an allocation method: proportional allocation method PR, individual solutions based allocation method IS, duality based method DU, and Shapley Value SV, respectively. The first three rows summarize the performance of the methods on allocating costs and the next three rows summarize the performance of the methods on allocating CO_2 emissions. The first row, "Subset," presents the average number of instable subsets (over all the generated subsets) with each cost allocation method. The row "Ave" presents the average of the average percent instability of each cost allocation method over all the generated subsets whereas the next row "Max" presents the average of the maximum percent instability of cost allocation methods over all the generated subsets. The next three rows present the same performance measures for the CO_2 emission allocations.

The computational experiments reveal that the duality-based method provides significantly better allocations especially compared to the proportional and individual solution based allocation methods. The average instability of duality-based cost allocation method over all instances is equal to 3.59% on 25-customer instances and 3.48% on 50-customer instances. The same values for proportional, individual solution based and Shapley Value allocation methods are 7.26%, 7.26%, and 5.64%, respectively, for 25-customer instances and 10.14%, 10.14%, and 6.36%, respectively, for 50-customer instances. Hence, we conclude that the superior performance of the duality-based method is even more significant in larger instances. When we compare the number of instable subsets, again duality-based method outperforms the other methods. When the CO_2 emission allocations are considered, we observe the same pattern, the superiority of the duality-based method, except that proportional and individual solution based allocation methods do not yield the same values.

Table 3 presents the same results except that the classification of the 20 instances is now based on the load volumes. Hence, the first four columns present the results for the low volume instances and the next four columns present the results for the high volume instances. Even though the ranking of the allocation methods does not change with respect to the instance characteristics based on the volume of the loads, we observe that the overall performances of all the allocation methods are slightly worse in low volume instances. The reason for such a difference in performance is due to the fact that, in the low volume instances, the synergies among customers are higher, which lead to more challenging instances in terms of stability.

Tables 4 and 5 present the results when $H = 100$, whereas in the former table the classification is based on the number of customers; in the latter one, the classification

Table 3: Instability values of the proposed methods on low and high volume instances when $H = 50$.

		Low volume				High volume			
		PR	IS	DU	SV	PR	IS	DU	SV
Cost	Subset	863.9	863.9	218	309.2	808.1	808.1	129.2	613.4
	Ave	9.29	9.29	4.04	5.75	8.1	8.1	3.03	6.25
	Max	46.72	46.72	21.52	24.32	44.45	44.45	14.52	33.44
CO_2	Subset	1033.5	1045.8	321.1	307.5	870.9	885.1	262.1	398.6
	Ave	11.32	11.35	5.35	6.42	9.85	9.84	4.3	6.37
	Max	57.16	57.44	33.21	32.49	52.5	52.75	22.99	40.35

Table 4: Instability values of the proposed methods on 25- and 50-customer instances when $H = 100$.

		25-customer				50-customer			
		PR	IS	DU	SV	PR	IS	DU	SV
Cost	Subset	598.9	598.9	98	282.6	1044.3	1044.3	127.5	584.7
	Ave	7.3	7.3	2.92	5.48	10	10	2.93	6.08
	Max	39.7	39.7	13.65	27.26	48.33	48.33	14.87	30.74
CO_2	Subset	914.8	925.2	270.8	280.8	1103.2	1125.5	281.3	510.7
	Ave	9.61	9.52	4.94	6.68	11.85	11.92	4.05	6.16
	Max	54.4	53.68	26.37	40.89	56.65	57.41	20.19	32.32

Table 5: Instability values of the proposed methods on low and high volume instances when $H = 100$.

		Low volume				High volume			
		PR	IS	DU	SV	PR	IS	DU	SV
Cost	Subset	867.4	867.4	134.6	312	775.8	775.8	90.9	555.3
	Ave	9.22	9.22	3.36	5.45	8.08	8.08	2.48	6.1
	Max	44.05	44.05	16.11	23.95	43.97	43.97	12.4	34.04
CO_2	Subset	1059.7	1077.4	259.6	304.8	958.3	973.3	292.5	486.7
	Ave	11.38	11.36	4.55	6.49	10.07	10.07	4.44	6.35
	Max	56.24	56.3	23.26	31.75	54.81	54.8	23.3	41.46

is based on load volumes. From both tables, we observe that the performance of the duality-based method improves whereas the performances of the other allocation methods do not change significantly. Here, one might expect the exact same performance from all the allocation methods except the duality-based method as their procedures are not affected. The differences are due to selecting a different seed in random variable generation in each of these computations.

Table 6 summarizes the computational times of the proposed allocation methods. The proportional method and individual solution based method clearly do not take even a second to compute and hence their computational times are not presented. The first row presents the computational times of the methods on 25-customer instances when $H = 50$, the second row presents the same values for 50-customer instances when $H = 50$, the third row for 25-customer instances when $H = 100$, and finally the last row for 50-customer instances when $H = 100$. We observe that duality-based method does not take longer than a second to compute even on the 50-customer instances when $H = 100$. Even though the Shapley Value method has the highest computational time of all the allocation methods, it is still within the acceptable limits. Finally, the total computational time values also include the stability assessment procedure run times, which are significant as expected. However, we should stress that the assessment computations should not be included in the running time of the allocation methods. We conclude that our proposed methods are not only effective but also computationally efficient.

Under our second setting, we provide additional incentives to decrease the emissions by switching the emission responsibility to the carrier. From Table 7, we observe that the second setting in fact decreases the total emission values. The average percentage improvement is 5.39% and the maximum percentage improvement is 14.60%.

Tables 8 and 9 present the results for the joint cost allocation methods under both classification rules, whereas in the former table $H = 50$ and in the latter $H = 100$. From both tables, we observe that the relative performances of the methods are similar with respect to each other; however, the instability values of all the methods are worse compared to their performance in allocating cost and emissions individually. This is inline with our expectations as in the joint cost setting opportunities for cost saving for the subsets might be higher, which leads to more restrict stability conditions.

Table 6: Computational times of the proposed methods in CPU seconds.

	DU	SV	Total
25 C—50 I	0	10.4	2363.2
50 C—50 I	0.2	27.9	6621.8
25 C—100 I	0	9.9	2278.1
50 C—100 I	0.2	28.5	6841.8

Table 7: Percentage improvement values in the emission values under the second setting.

Instance number	Improvement
1	8.33
2	0.97
3	6.36
4	7.10
5	8.69
6	2.05
7	3.31
8	4.01
9	5.62
10	5.37
11	0.85
12	8.21
13	2.27
14	3.10
15	0.52
16	1.35
17	9.13
18	14.51
19	14.60
20	1.44
Ave	5.39
Max	14.60

Table 8: Instability values of the proposed methods for joint cost allocation when $H = 50$.

		25-customer				50-customer			
		PR	IR	DU	SV	PR	IR	DU	SV
Cost	Subset	983.8	986.2	392.7	511.5	1158.5	1163.7	852.5	608.9
	Ave	8.04	8.04	4.57	5.85	10.32	10.33	5.7	6.05
	Max	45.61	45.63	23.4	32.66	51.6	51.72	32.02	30.92
		Low volume				High volume			
		PR	IR	DU	SV	PR	IR	DU	SV
Cost	Subset	1254.2	1257.2	460.1	551.7	888.1	892.7	785.1	568.7
	Ave	10.09	10.1	4.87	6.14	8.27	8.27	5.4	5.76
	Max	52.75	52.81	27.76	28.98	44.46	44.54	27.65	34.59

Finally, Table 10 summarizes the computational times of the joint cost allocation methods. Compared with the previous performance of the allocation methods, we observe that there is an increase in terms of CPU seconds. This due to the additional complexity of the TSP procedure. Fortunately, the increase is not significant except the stability assessment procedure run time, which is not included in the actual running time of the algorithms.

Based on our detailed computational study, we conclude that the Shapley Value approximation performs better than proportional and individual solution based allocation methods. The duality-based method significantly outperforms all

Table 9: Instability values of the proposed methods for joint cost allocation when $H = 100$.

		25-customer				50-customer			
		PR	IR	DU	SV	PR	IR	DU	SV
Cost	Subset	965.8	968.1	241.1	501.9	1106.1	1110.6	457.2	531
	Ave	8.02	8.03	3.78	5.85	10.54	10.57	4.62	5.74
	Max	45.34	45.36	17.27	32.23	46.3	46.43	46.37	29.49
		Low volume				High volume			
		PR	IR	DU	SV	PR	IR	DU	SV
Cost	Subset	1219.4	1222.9	313.8	534.6	852.5	855.8	384.5	498.3
	Ave	10.12	10.13	4.09	6.03	8.45	8.47	4.31	5.56
	Max	49.31	49.37	19.81	28.14	42.34	42.41	43.83	33.57

Table 10: Computational times of the proposed methods for joint cost allocation in CPU seconds.

	DU	SV	Total
25 C—50 I	0	13.9	8471.4
50 C—50 I	0.1	40.1	20404.7
25 C—100 I	0	14.1	8523.4
50 C—100 I	0	35.8	21301.9

other methods both in allocating costs and CO_2 emissions with respect to the average and the maximum percent instability as well as the number of instable subsets. Finally, the average instability value of the duality-based allocation method is always below 6% regardless of difference instance characteristics.

6. Concluding Remarks

In this paper, we study the cost and emission allocation problem of a logistics system. Road transportation is considered as the largest contributor to the world's GHG emissions and consequently to global warming. In order to mitigate the negative effects of the transportation activities to the environment, certain initiatives are undertaken, such as emission cap and trade in the EU. These initiatives try to give incentives to companies to reduce their emissions resulted from their direct and indirect actions. However, it is quite difficult to assess the companies emission responsibilities resulted from their indirect actions.

Having this in mind, we develop mechanisms for the fair assessment of the emission responsibilities of customers (recipients) in a logistics system. We consider a delivery network with multiple customers served by a single carrier and develop allocation mechanisms for both costs and emissions. Our assessment criterion is the fairness of the allocations. We also analyze the setting where the carrier takes over the emission responsibility and allocates a joint cost to the customers, which provides additional incentives to decrease the emissions. Based on a detailed computational study, we observe that our proposed methods, a duality-based allocation method and an approximation for the Shapley Value, perform significantly better than the allocation methods discussed in the literature. Our computational analysis also establishes that the proposed methods are computationally efficient and can be implemented for real-life sized problems.

Conflict of Interests

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

References

[1] European Commission Directorate-General for Energy and Transport, A sustainable future for transport, 2009, http://ec.europa.eu/transport/media/publications/doc/2009_future_of_transport_en.pdf.

[2] European Commission. Road transport: Reducing CO2 emissions from vehicles, 2014, http://ec.europa.eu/clima/policies/transport/vehicles/indexen.htm.

[3] US Environmental Protection Agency, "Sources of Greenhouse Gas Emissions," 2014, http://www.epa.gov/climatechange/ghgemissions/sources/transportation.html.

[4] European Commission, "EU transport in figures," 2012, http://ec.europa.eu/transport/facts-fundings/statistics/doc/2012/pocketbook2012.pdf/.

[5] S. Mui, J. Alson, B. Ellies, and D. Ganss, "A wedge analysis of the U.S. transportation sector," 2014, http://www.epa.gov/otaq/climate/420r07007.pdf.

[6] Council of the European Union, "The Transport Protocol of the Alpine Convention," 2006, http://europa.eu/legislation_summaries/transport/transport_energy_environment/l24468_en.htm.

[7] European Commission, "The EU emissions trading system (EU ETS)," 2014, http://ec.europa.eu/clima/policies/ets/indexen.htm.

[8] G. Barnes and P. Langworthy, "The per-mile costs of operating automobiles and trucks," State and Local Policy Program, University of Minnesota, 2014, http://conservancy.umn.edu/bitstream/11299/909/1/200319.pdf.

[9] R. Wilson, P. B. Luckow, F. Biewald, E. Ackerman, and E. Hausman, "2012 carbon dioxide price forecast," http://www.synapse-energy.com/Downloads/SynapseReport.2012-10.0.2012-CO2-Forecast.A0035.pdf.

[10] U. Faigle, S. P. Fekete, W. Hochstattler, and W. Kern, "On approximately fair cost allocation in euclidean tsp games," *OR Spectrum*, vol. 20, no. 1, pp. 29–37, 1998.

[11] M. Pál and É. Tardos, "Group strategyproof mechanisms via primal-dual algorithms," in *Proceedings of the 44th Annual IEEE*

Symposium on Foundations of Computer Science (FOCS '03), pp. 584–593, October 2003.

[12] S. Engevall, M. Göthe-Lundgren, and P. Värbrand, "The heterogeneous vehicle-routing game," *Transportation Science*, vol. 38, no. 1, pp. 71–85, 2004.

[13] M. X. Geomans and M. Skutella, "Cooperative facility location games," *Journal of Algorithms*, vol. 50, no. 2, pp. 194–214, 2004.

[14] A. Tamir, "On the core of a traveling salesman cost allocation game," *Operations Research Letters*, vol. 8, no. 1, pp. 31–34, 1989.

[15] J. A. M. Potters, I. J. Curiel, and S. H. Tijs, "Traveling salesman games," *Mathematical Programming*, vol. 53, no. 2, pp. 199–211, 1992.

[16] J. Kuipers, "A note on the 5-person traveling salesman game," *Mathematical Methods of Operations Research*, vol. 38, no. 2, pp. 131–139, 1993.

[17] M. Göthe-Lundgren, K. Jörnsten, and P. Värbrand, "On the nucleolus of the basic vehicle routing game," *Mathematical Programming*, vol. 72, no. 1, pp. 83–100, 1996.

[18] O. O. Özener, Ö. Ergun, and M. Savelsbergh, "Allocating cost of service to customers in inventory routing," *Operations Research*, vol. 61, no. 1, pp. 112–125, 2013.

[19] T. Bektaş and G. Laporte, "The pollution-routing problem," *Transportation Research B: Methodological*, vol. 45, no. 8, pp. 1232–1250, 2011.

[20] K. Pitera, F. Sandoval, and A. Goodchild, "Evaluation of emissions reduction in Urban pickup systems," *Transportation Research Record*, vol. 2224, no. 1, pp. 8–16, 2011.

[21] Y. Suzuki, "A new truck-routing approach for reducing fuel consumption and pollutants emission," *Transportation Research D: Transport and Environment*, vol. 16, no. 1, pp. 73–77, 2011.

[22] Y. Xiao, Q. Zhao, I. Kaku, and Y. Xu, "Development of a fuel consumption optimization model for the capacitated vehicle routing problem," *Computers and Operations Research*, vol. 39, no. 7, pp. 1419–1431, 2012.

[23] M. S. Gajanand and T. T. Narendran, "Green route planning to reduce the environmental impact of distribution," *International Journal of Logistics Research and Applications*, vol. 16, no. 5, pp. 410–432, 2013.

[24] H. Xue, S. Jiang, and B. Liang, "A study on the model of traffic flow and vehicle exhaust emission," *Mathematical Problems in Engineering*, vol. 2013, Article ID 736285, 6 pages, 2013.

[25] M. Wang, C. Dang, and S. Wang, "A Pareto optimal auction mechanism for carbon emission rights," *Mathematical Problems in Engineering*, vol. 2014, Article ID 438104, 7 pages, 2014.

[26] A. Sichwardt, *CO_2 allocation in road transportation for Alwex Transport AB [M.S. thesis]*, University of Linnaeus, 2011.

[27] B. P. J. Leenders, *Carbon allocation in travelling salesman problems: an application in retail distribution [M.S. thesis]*, Eindhoven University of Technology, 2012.

[28] S. K. Naber, *Allocating CO_2 emissions to customers on a distribution route [M.S. thesis]*, Erasmus Universiteit Rotterdam, 2012.

[29] G. Owen, "On the core of linear production games," *Mathematical Programming*, vol. 9, no. 3, pp. 358–370, 1975.

[30] E. Kalai and E. Zemel, "Generalized network problems yielding totally balanced games," *Operations Research*, vol. 30, no. 5, pp. 998–1008, 1982.

[31] A. McKinnon and M. Piecyk, "Measuring and managing CO_2 emission," The European Chemical Industry Council, 2011, http://cefic-staging.amaze.com/Documents/Media%20Center/News/McKinnon-Report-Final-230610.pdf.

[32] K. Ganesh and T. T. Narendran, "CLOVES: a cluster-and-search heuristic to solve the vehicle routing problem with delivery and pick-up," *European Journal of Operational Research*, vol. 178, no. 3, pp. 699–717, 2007.

[33] B. Kim, S. Kim, and S. Sahoo, "Waste collection vehicle routing problem with time windows," *Computers and Operations Research*, vol. 33, no. 12, pp. 3624–3642, 2006.

[34] M. Fischetti, G. Laporte, and S. Martello, "The delivery man problem and cumulative matroids," *Operations Research*, vol. 41, no. 6, pp. 1055–1064, 1993.

[35] L. S. Shapley and L. S. Shapley, "A value for n-person games. Contributions to the Theory of Games, II," *Annals of Mathematical Studies*, vol. 28, pp. 307–317, 1953.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 956536, 10 pages
http://dx.doi.org/10.1155/2014/956536

Research Article

Storage Space Allocation of Inbound Container in Railway Container Terminal

Li Wang, Xiaoning Zhu, and Zhengyu Xie

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Xiaoning Zhu; xnzhu@bjtu.edu.cn

Received 21 February 2014; Revised 14 June 2014; Accepted 29 June 2014; Published 22 July 2014

Academic Editor: Hu Shao

Efficient storage strategy of railway container terminals is important in balancing resource utilization, reducing waiting time, and improving handling efficiency. In this paper, we consider the formulation and solution algorithm for storage space allocation problem of inbound containers in railway container terminal. The problem is formulated as two-stage optimization models, whose objectives are balancing the workload of inbound containers and reducing the overlapping amounts. An algorithm implement process based on rolling horizon approach is designed to solve the proposed models. Computational experiments on an actual railway container terminal show that the proposed approach is effective to solve space allocation problem of inbound container and is significant for the operation and organization of railway container terminals.

1. Introduction

With rapid development of passenger-dedicated railway in China, more transportation capacities of railway freight corridor are released, which puts forward huge opportunities and challenges for railway container transportation. At present, operation and organization of railway container terminals cannot meet developing demands of container transportation. To enable container to rapidly transfer between rail and truck, modern transshipment technologies are required to improve the resource utilization in railway container terminals.

As key resources of railway container terminals, storage spaces are responsible for temporarily storing containers which are moved over long distances by container trains and short distances by trucks. Storage space allocation is important in balancing space utilization, improving the efficiency of container handling, and reducing the turnaround time of containers. The storage space allocation is a vital basis and constraint for other resources utilization in railway container terminal.

Inbound and outbound container operations in railway container terminal are different. Inbound containers arrive predictably in large quantities, are temporarily stored in container yard, and depart one by one in an unpredictable sequence. On the contrary, outbound containers arrive in a random sequence and depart predictably. This paper focuses on the inbound containers in railway container terminals.

The rest of this paper is organized as follows: the relevant literature is reviewed in the next section. The storage space allocation problem of inbound containers in railway container terminals is described in Section 3 and formulated in Section 4. An algorithm implement process based on rolling horizon approach is developed in Section 5. Computational results are reported in Section 6 and finally Section 7 covers the conclusion.

2. Literature Review

The storage space allocation problem of inbound containers in railway container terminal belongs to the storage space allocation problem (SSAP) which is defined as the temporary allocation of the inbound/outbound containers to the storage blocks at each time period with aim of balancing the workload between blocks in order to minimize the storage/retrieval times of containers [1].

The SSAP was firstly formulated for a container terminal in Hong Kong [2]. The problem was solved by a rolling horizon approach. In each planning horizon, the problem

is decomposed into two levels and formulated each level as a mathematical programming model. In order to determine the storage location of arriving export containers by considering its weight, a dynamic programming model is formulated to minimize the number of relocation movements expected for the loading operation [3]. The process of determining the storage locations for outbound containers was divided into two stages: space allocation stage and stage of locating individual containers [4]. A storage location assignment problem for outbound containers of maritime terminal was decomposed into two stages. The problem in the first stage is solved by a mixed integer programming model, while a hybrid sequence stacking algorithm is applied to solve the problem in the second stage [5].

An approach for allocating storage space to groups of outbound containers in port container terminals was proposed with considering the impacts of various space-reservation strategies on the productivity of the loading operation [6]. For finding the best allocation of containers in a yard bay in order to minimize the number of reshuffles, a domain-dependent heuristically guided planner for obtaining the optimized reshuffling plan was proposed with a stacking state and a container demand known [7]. In order to improve the operations efficiency of retrieving inbound containers in a modern automatic container terminal, inbound container space allocation models were proposed to optimally allocate the arrival inbound containers so as to minimize the expected container retrieval time [8]. A novel approach using ant-based control was proposed for allocating containers to storage blocks in a marine container terminal with the competing objectives of balancing the workload among yard blocks and minimizing the distance traveled of trucks between yard blocks and berths [9].

In order to solve the storage space problem of outbound containers for utilizing space efficiently and make loading operations more efficient, two heuristic algorithms are suggested based on the duration-of-stay of containers and the subgradient optimization technique, respectively [4]. An extended version of SSAP was proposed and an efficient genetic algorithm was developed to solve the extended problem in a container terminal [1]. Construction algorithms and a tabu search heuristic are presented for dynamic space allocation problem to optimize the space/resource assignments during the implementation of project activities [10]. Three tabu search heuristics are presented for dynamic space allocation problem. The first heuristic is a simple basic tabu search heuristic. The second heuristic adds diversification and intensification strategies to the first and the third heuristic is a probabilistic tabu search heuristic [11].

A hybrid insertion algorithm is designed for solving the problem which integrates the yard truck scheduling and the storage allocation to minimize the weighted sum of total delay of requests and the total travel time of yard trucks [12]. For the problem of determining the stacking positions for incoming containers in automated container terminals, an online search algorithm was proposed to dynamically adjust and optimize a stacking policy by continuously generating and evaluating variants of stacking policies [13]. A decision support system was present to manage container stacking problem, berth allocation problem, and the quay crane assignment problem in a coordinated way. A domain-oriented heuristic planner for calculating the number of reshuffles needed to allocate the containers in the appropriate place [14]. A construction and a hybrid algorithm (HGT) based on the GRASP and tabu search metaheuristics were proposed to solve the dynamic space allocation problem, where project duration is divided into a number of consecutive periods, each of them associated with a number of activities [15].

According to the literature review above, most studies focused on the storage space allocation problem in maritime container terminals. Specific literature on railway container terminal is scarce. Because storage strategy of inbound and outbound containers in railway container terminals is obviously different from maritime container terminals, existing studies are hardly applied in railway container terminals. In this paper, we consider the storage space allocation problem of inbound container in railway container terminals. Given that container arrival-departure time and operation sequences are known, two-stage storage space allocation models are proposed, whose objectives are to balance workloads among container blocks and assign containers to the optimum positions.

3. Problem Description

The Chinese railway container terminals have advanced arrival-departure lines, storage spaces, and handling equipment. Figure 1 gives a schematic representation of a typical railway container terminal in China. Our study is based on the configuration and layout of the representation in Figure 1.

As observed in Figure 1, container yard of railway container terminal is composed by inbound container yard, outbound container yard, and auxiliary container yard. Since most of inbound container allocation operations occur in inbound container yard, we set inbound container yard as the study scope of this paper.

According to the inbound container status in different handling stages, containers to be handled in inbound container yard can be classified into the following three types.

(i) Inbound containers on container train wait for unloading and allocating to the inbound container yard, abbreviated as ICT.

(ii) ICT containers temporarily stored in inbound container yard wait for loading to trucks for customers, abbreviated as ICTY.

(iii) ICT containers are unloaded and directly loaded to trucks for customers, abbreviated as ICTT.

According to optimization objectives of SSAP, most literatures decomposed the storage space allocation problem into two stages. The first stage is to balance containers workloads among blocks and evenly allocate containers to each block. The second stage is the slot allocation for containers which are allocated to blocks based on the first stage optimization results. In this paper, our study decomposed the storage space allocation problem of inbound container in railway container

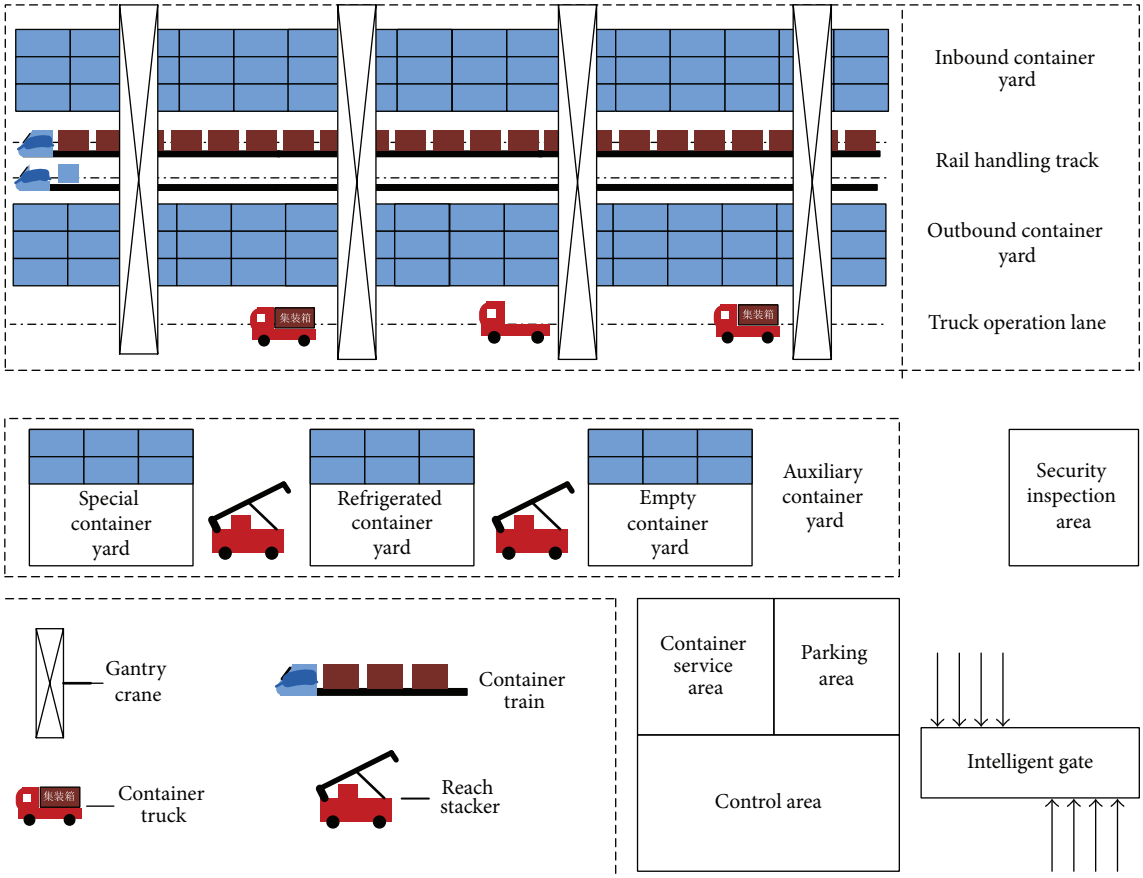


Figure 1: Schematic representation of a railway container terminal.

terminal into two stages based on the optimization objectives of SSAP.

(i) The first stage is workload balance similar to the exciting study, whose objective is to balance workloads of inbound containers among blocks and evenly allocate inbound containers to each block.

(ii) The second stage is container slot allocation based on the first stage optimization results, whose objective is to minimize overlapping amounts of ICT.

4. Problem Formulation

In this section, according to the problem description above, the storage space allocation problem of inbound container in railway container terminal is formulated as two-stage optimization models based on rolling horizon approach. At each planning epoch, we plan for a fixed horizon in immediate future and execute the plan accordingly up to the next planning epoch; then we formulate a new plan based on the latest information; this pattern goes on continually [2]. The rolling of planning horizon is shown in Figure 2. The workload balance is implemented in each planning epoch, and the container slot allocation is implemented in each period of planning epoch.

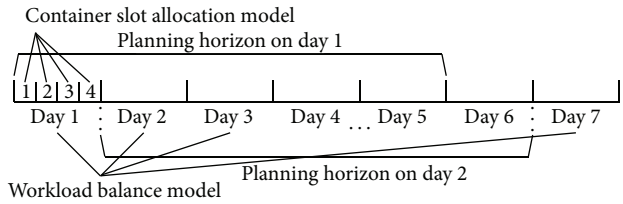


Figure 2: Rolling of planning horizon.

4.1. Assumptions. The following four assumptions are introduced for the formulation of the problem.

(1) There is enough resource, that is, gantry crane and container yard space, to handle the allocation workload at the considered block.

(2) The arriving and departure time of containers are known in advance and there is no time delay in scheduling period.

(3) Handling sequence of containers is assumed to be known.

(4) The containers in the model are assumed to be of one size.

4.2. Workload Balance Model

4.2.1. Notations and Variables. The notations and variables of workload balance model are defined as follows:

IB: total number of blocks in inbound container yard;

T: total number of planning periods in a planning epoch;

C_i: storage capacity of block i, $1 \le i \le IB$;

V_{i0}: initial inventory of block i, $1 \le i \le IB$;

V_{it}: total number of containers in block i at beginning of period t, $1 \le i \le IB$, $1 \le t \le T$;

ICTY_{it}^{0}: number of ICTY in block i that are picked up at beginning of period t, $1 \le i \le IB$, $1 \le t \le T$;

ICT_{tk}: number of ICT that are unloaded from trains in period t and to be picked up in period $t+k$, $1 \le t \le T$, $0 \le k \le T - t$;

ICT_{it}: number of ICT in block i that are unloaded from trains in period t, $1 \le i \le IB$, $1 \le t \le T$;

ICT_{itk}: number of ICT in block i that are unloaded from trains in period t and to be picked up in period $t+k$, $1 \le i \le IB$, $1 \le t \le T$, $0 \le k \le T - t$;

ICTY_{it}: number of ICTY in block i that are picked up in period t, $1 \le i \le IB$, $1 \le t \le T$;

ICTT_t: number of ICTT that are unloaded and directly loaded to trucks in period t, $1 \le t \le T$.

4.2.2. Objective Function. By attention to the problem description in Section 3, the objective function of workload balance model is written as follows:

$$\mathrm{Min} \sum_{t=1}^{T} \left[\max_{\{i\}} \left(\mathrm{ICT}_{it} + \mathrm{ICTY}_{it} \right) - \min_{\{i\}} \left(\mathrm{ICT}_{it} + \mathrm{ICTY}_{it} \right) \right]. \tag{1}$$

Objective function is to balance workloads of inbound containers among blocks and evenly allocate ICT to blocks in each planning epoch.

4.2.3. Constraints. The constraints of workload balance model are introduced as follows to ensure the practical feasibility of the solution.

(1) Constraints on ICT:

$$\mathrm{ICT}_{tk} = \sum_{i=1}^{IB} \mathrm{ICT}_{itk}, \quad t = 1, 2, \ldots, T; \ k = 0, 1, \ldots, T - t, \tag{2}$$

$$\mathrm{ICT}_{it} = \sum_{k=0}^{T-t} \mathrm{ICT}_{itk}, \quad i = 1, 2, \ldots, IB; \ t = 1, 2, \ldots, T. \tag{3}$$

Constraint (2) ensures that the total number of ICT that are unloaded from trucks in period t and to be loaded onto rail vehicles in period $t + k$ is the sum of these containers assigned to all the blocks. Constraint (3) ensures that the total number of ICT in block i that are unloaded from trucks in period t is the sum of these containers loaded onto rail vehicles in period $t + k$ in this block.

(2) Constraints on ICTY:

$$\mathrm{ICTY}_{it} = \mathrm{ICTY}_{it}^{0} + \sum_{k=0}^{t-1} \mathrm{ICT}_{i(t-k)k}, \quad i = 1, 2, \ldots, IB; \quad t = 1, 2, \ldots, T. \tag{4}$$

Constraint (4) indicates that the number of ICTY in block i during period t is the sum of initial number of ICTY in block i and the number of ICTY transferred from the ICT that unloaded in the block in planning epoch.

(3) Constraints on ICTT:

$$\mathrm{ICTT}_t = \sum_{i=1}^{IB} \left(\mathrm{ICT}_{it} - \mathrm{ICT}_{itk} \right), \quad t = 1, 2, \ldots, T, \quad k = t, t+1, \ldots, T. \tag{5}$$

Constraint (5) indicates that the number of ICTY in block i during period t is the sum of initial number of ICTY in block i and the number of ICTY transferred from the ICT that unloaded in the block in planning epoch.

(4) Capacity constraints:

$$V_{it} = V_{i(t-1)} + \left(\mathrm{ICT}_{it} - \mathrm{OICT}_{it} \right), \quad i = 1, 2, \ldots, IB; \quad t = 1, 2, \ldots T \tag{6}$$

$$V_{it} \le C_i, \quad i = 1, 2, \ldots IB; \ t = 1, 2, \ldots T. \tag{7}$$

Constraint (6) indicates the inventory of containers in block i at beginning of period t. Constraint (7) indicates the storage capacity of block i.

(5) Integer constraint:

All variables take up nonnegative integer values. (8)

4.3. Container Slot Allocation Model

4.3.1. Notations and Variables. The notations and variables of container slot allocation model are defined as follows:

N: total number of ICT which are allocated in considered block at the same planning period;

v: the sequence number of allocated container;

B: total number of bays in the considered block;

R: total number of rows in considered block;

L: maximum layer number of stack;

b: bay identifier of container slot, $1 \le b \le B$;

r: row identifier of container slot, $1 \le r \le R$;

l: layer identifier of container slot, $1 \le l \le L$;

$s(r,b,l)$: container slot of r row, b bay, and l layer;

t_{rbl}: departure time of container in container slot of r row, b bay, and l layer;

K_v^{ICT}: overlapping amounts of the v ICT;

M: an infinitesimal number;

S_{rbl}: if $s(r,b,l)$ has container, $S_{rbl} = 1$. Otherwise, $S_{rbl} = 0$;

S_{vrbl}: if the v container is allocated to $s(r,b,l)$, $S_{vrbl} = 1$. Otherwise, $S_{vrbl} = 0$;

$K_{rbl,rb(l-e)}^{\mathrm{ICT}}$: overlapping of $s(r,b,l-e)$ after ICT was allocated to $s(r,b,l)$. If $t_{rbl} < t_{rb(l-e)}$, $K_{rbl,rb(l-e)}^{\mathrm{ICT}} = 1$. Otherwise, $K_{rbl,rb(l-e)}^{\mathrm{ICT}} = 0$;

H_v: if there are no empty bays while the v ICT is allocated, $H_v = 1$. Otherwise, $H_v = 0$.

4.3.2. Objective Functions. By attention to the problem description in Section 3, the objective functions of container slot allocation model are written as follows:

$$\mathrm{Min} \sum_{v=1}^{D} K_v^{\mathrm{ICT}}. \tag{9}$$

The second stage objective function (9) minimizes overlapping amount which is caused by ICT allocated in considered block at same planning period.

4.3.3. Constraints. The constraints of container slot allocation model are introduced as follows to ensure the practical feasibility of the solution.

(1) Overlapping amounts constrains of ICT:

$$K_v^{\mathrm{ICT}} = (1 - H_v) M + H_v \bullet \min \sum_{e=1}^{l-1} K_{rbl,rb(l-e)}^{\mathrm{ICT}}, \quad v = 1,2,\ldots,N, \quad r = 1,2,\ldots,R, \quad b = 1,2,\ldots,B, \quad l = 2,3,\ldots,L. \tag{10}$$

Constraint (10) represents the calculation of overlapping amount which is caused by ICT allocated in considered block.

(2) Allocation constrain:

$$S_{vrbl} - S_{rb(l-1)} \le 0, \quad r = 1,2,\ldots R, \quad b = 1,2,\ldots,B, \quad l = 2,3,\ldots,L, \quad v = 1,2,\ldots,N. \tag{11}$$

Constraint (11) ensures that each container cannot be allocated upon the empty container slot.

(3) Allocation preferences constraints:

$$\sum_{v=1}^{N}\sum_{r=1}^{m} S_{vrbl} - \sum_{v=1}^{N}\sum_{r=R}^{R-(m-1)} S_{vrbl} \le 0, \quad b = 1,2,\ldots,B, \quad l = 1,2,\ldots,L, \quad m = \frac{R}{2} - 1. \tag{12}$$

Constraint (12) indicates the allocation preferences constraint of ICT. It ensures that the allocation positions of ICT are close to the truck operation lane in order to reduce loading time of ICTY.

5. Solution Algorithm

In order to solve the two-stage models presented above, an algorithm implement process based on rolling horizon approach is proposed in this section. The workload balance model is converted to a linear integer programming model and a heuristic algorithm is designed to solve the container slot allocation model. The implement process is shown in Figure 3.

5.1. Workload Balance Model Conversion. The workload balance model proposed in Section 4 is a nonlinear model, because of the objective function of model. In order to obtain a solution, a conversion method should be used to convert the model to a linear model. According to the conversion mentioned in [2], we define $A_t = \max_{\{i\}}(\mathrm{ICT}_{it} + \mathrm{ICTY}_{it})$ and $B_t = \min_{\{i\}}(\mathrm{ICT}_{it} + \mathrm{ICTY}_{it})$. Then the workload balance model can be converted as the linear integer programming model as follows:

$$\mathrm{Min} \sum_{t=1}^{T} (A_t - B_t). \tag{13}$$

The constraints include (2) to (8) and the constraints on A_t and B_t:

$$\begin{aligned} &\mathrm{ICT}_{it} + \mathrm{ICTY}_{it} \le A_t, \quad i = 1,2,\ldots,IB;\ t = 1,2,\ldots,T \\ &\mathrm{ICT}_{it} + \mathrm{ICTY}_{it} \ge B_t, \quad i = 1,2,\ldots,IB;\ t = 1,2,\ldots,T. \end{aligned} \tag{14}$$

After the conversion of workload balance model, it can be solved by Lingo.

5.2. HA Implementation for Container Slot Allocation Model. The notations of HA are shown in Table 1 and the procedure of HA is shown as follows.

Step 1. According to the initial block information at the beginning of period, get a feasible allocated set of container slots F by removing infeasible container slots. Parameter initialization: let $v = 1$, $r = R$, $b = 1$,$K = \{\phi\}$, $S = \{\phi\}$, $K_v^{\mathrm{ICT}} = I$, and go to Step 2.

Step 2. Allocate the v ICT. Search empty container slots in block row from R to $R - (m - 1)$. If empty container slot exists, go to Step 3. If empty container slot does not exist, go to Step 4.

Step 3. Let $S = S \cup \{S_{rbl}^v\}$ and $v = v + 1$; then if $v \le N$, go to Step 2. Otherwise, let $K_v^{\mathrm{ICT}} = 0$ and go to Step 8.

Step 4. If $s(r,b,l) \notin F$, go to Step 6. Otherwise allocate container to $s(r,b,l)$ and judge the feasibility of solution. If the solution is not feasible, go to Step 8. If the solution is feasible, go to Step 5.

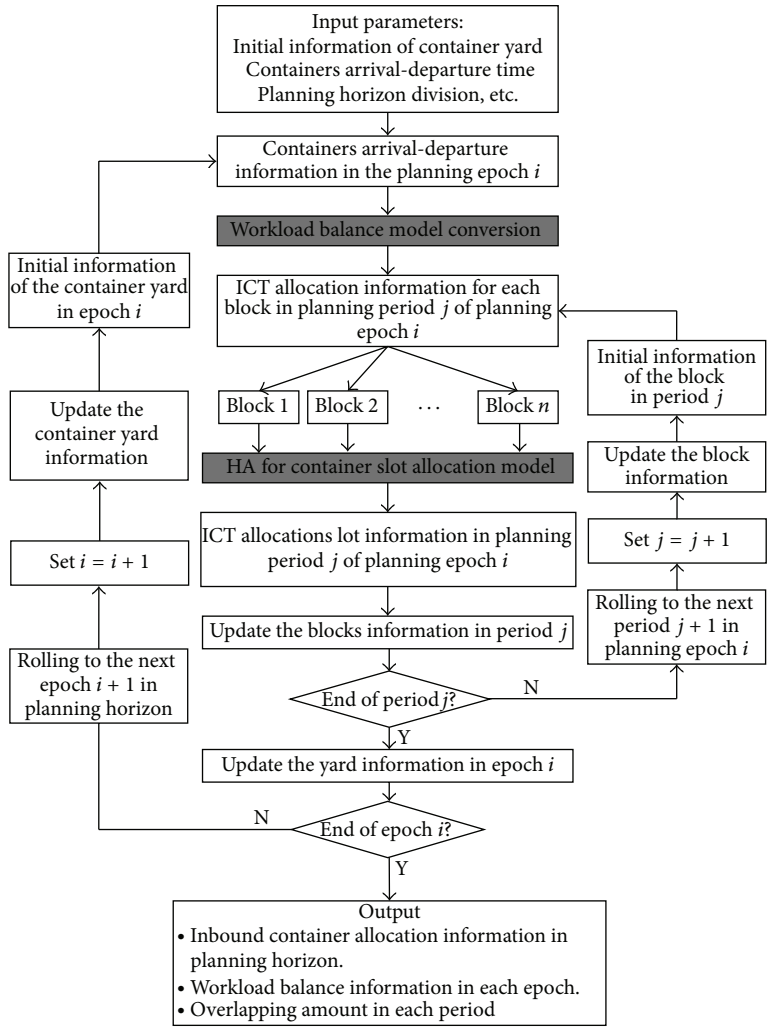


Figure 3: Algorithm implement process.

Table 1: Notations.

Notation	Declaration
K	The set of overlapping amounts
S	The optimal set of slots allocation with minimum overlapping amount
S^v_{rbl}	The container allocated to $s(r,b,l)$
K^{ICT}_n	the overlapping amount increased by the n feasible solution of ICT
$d^{\min}_{vrbl}$	The minimum crane operation distances of the v container allocated to $s(r,b,l)$
A	The alternative set of the v allocated slot with minimum overlapping amount
F	The feasible allocated set of container slots
n	The number of feasible solutions
d_{vrbl}	The operation distance of crane for the v allocated container
I	An arbitrary positive big number

Step 5. Let the solution be K^{ICT}_n and compare $K^{\mathrm{ICT}}_n$and K^{ICT}_v. If $K^{\mathrm{ICT}}_n > K^{\mathrm{ICT}}_v$, go to Step 6. Otherwise, if $K^{\mathrm{ICT}}_n = K^{\mathrm{ICT}}_v$, let $A = A \cup \{S^v_{rbl}\}$ and go to Step 6; otherwise, let $K^{\mathrm{ICT}}_v = K^{\mathrm{ICT}}_n$, $A = \{\phi\}$, $A = A \cup \{S^v_{rbl}\}$, and go to Step 6.

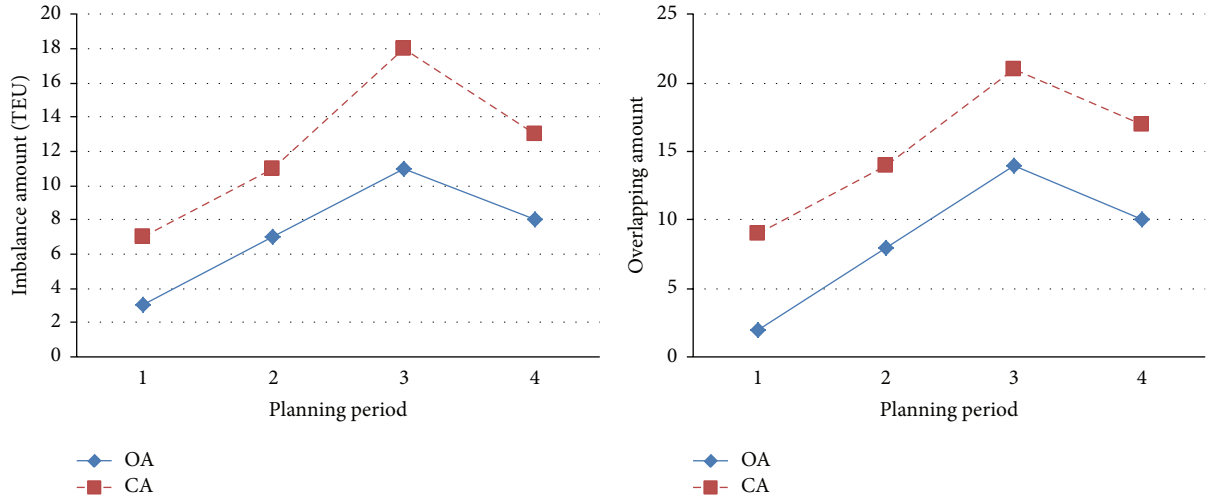
Step 6. Let $b = b + 1$. If $b \le B$, go to Step 4. Otherwise, let $r = r - 1$. If $r \ge 1$, go to Step 4. If $r < 1$, calculate the d_{vrbl} in set A and select $d^{\min}_{vrbl}$, set $K = K \cup \{K^{\mathrm{ICT}}_v\}$, $S = S \cup \{S^v_{rbl}\}$, and go to Step 7.

Step 7. Allocation of the v ICT has finished. Let $r = R$, $b = 1$, and $v = v + 1$; then if $v \le N$, go to Step 2; otherwise go to Step 8.

Step 8. Calculate the overlapping amounts based on K and output the set of container slot allocation S. Procedure terminates.

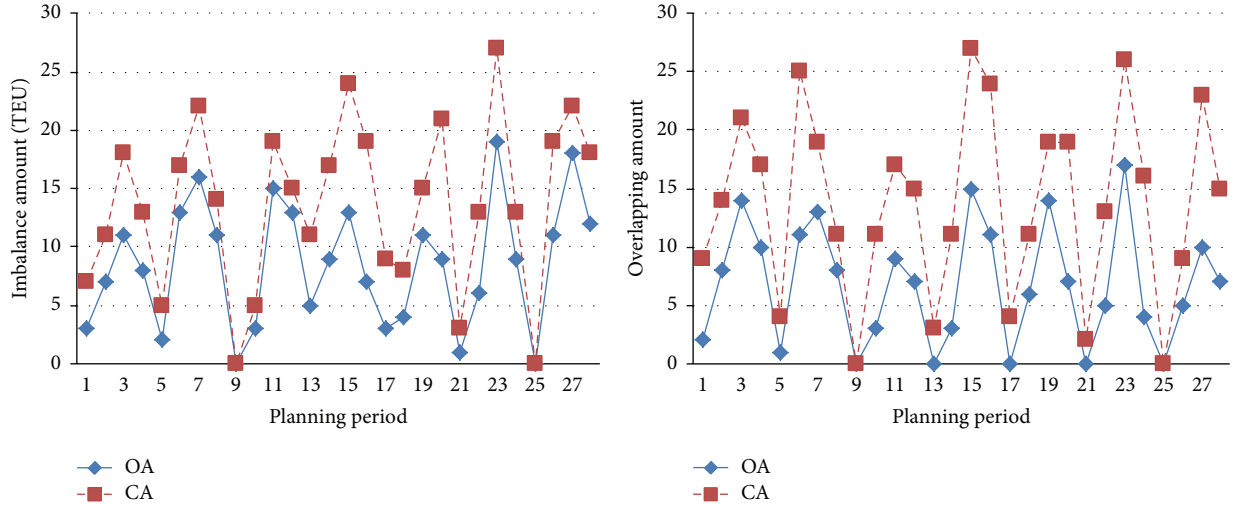
6. Computational Experiment

To illustrate the proposed model and algorithm for space allocation problem of inbound container in railway container terminal, computational experiments are performed by using the actual data from a specific railway container terminal



(a) Imbalance amounts (b) Overlapping amounts

Figure 4: Computational results of 1 day.



(a) Imbalance amounts (b) Overlapping amounts

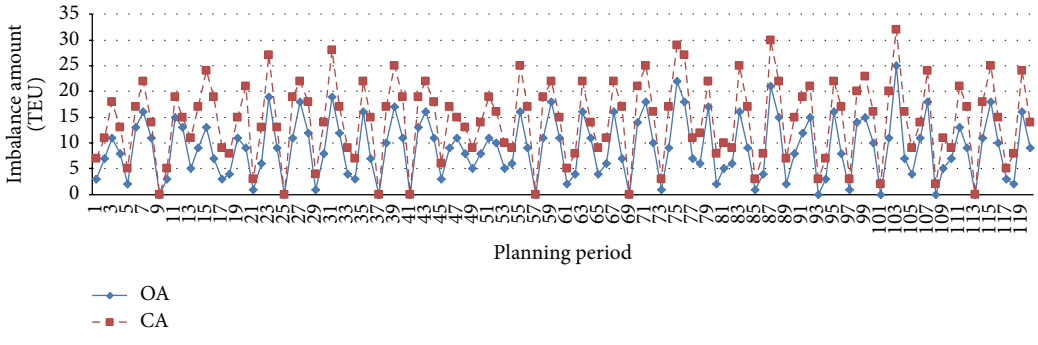
Figure 5: Computational results of 7 days.

in China [16]. In order to evaluate the improvement of our approach, a comparison is made between our approach and random allocation algorithm which is currently used in railway container terminals. Furthermore, to evaluate the effectiveness and practicability of our approach, numerical experiments of 7 days and 30 days are carried out. To implement the proposed algorithm, the parameters related to the specific railway container terminal are needed. There are four blocks in the inbound container yard. Each block is composed by 30 bays, 6 rows, and 2 layers. The numerical experiments are performed based on a personal computer with Intel Core i5-2450M @ 2.50 GHz processors and 4 GB RAM.

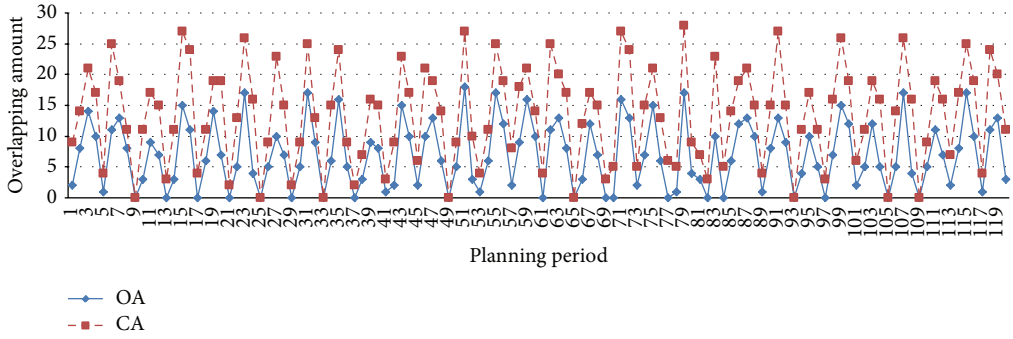
Because most of ICTY are picked up no more than two days after they allocated to blocks, we choose 3 days as a planning horizon, 1 day as a planning epoch, and 6 hours as a planning period. There are 4 planning periods in one planning epoch and 12 planning periods in one planning horizon. A small size sample of 1 day is carried out firstly. The inbound container information of 4 planning periods in 1 day is shown in Table 2. The ICTY loading plan of each block in planning period is shown in Table 3.

Table 2: Inbound container information of 4 planning periods in 1 day (TEU).

Container type	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ICT	20	25	135	40
ICTY	15	65	79	67
ICTT	2	3	7	3



(a) Imbalance amounts



(b) Overlapping amounts

Figure 6: Computational results of 30 days.

Table 3: ICTY loading plan of each block in planning period (TEU).

Container type	$i = 1$				$i = 2$				$i = 3$				$i = 4$			
	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ICTY	4	0	6	5	18	27	13	7	13	22	16	28	27	19	12	9

Table 4: Comparison between OA and RAA in 1 day.

Planning period	Our approach (OA)		Random allocation algorithm (RAA)		GAP_1	GAP_2
	Imbalance amounts	Overlapping amounts	Imbalance amounts	Overlapping amounts		
1	3	2	7	9	57.1%	77.8%
2	7	8	11	14	36.3%	42.9%
3	11	14	18	21	38.8%	33.3%
4	8	10	13	17	38.4%	41.2%

Notes: GAP_1 = (imbalance amounts obtained from RAA − imbalance amounts obtained from OA) ∗ 100/imbalance amounts obtained from RAA.
GAP_2 = (overlapping amounts obtained from RAA − overlapping amounts obtained from OA) ∗ 100/overlapping amounts obtained from RAA.

Based on the computational example above, experiment is conducted, and a comparison between our approach (OA) and random allocation algorithm (RAA) is made to evaluate the performance of OA on space allocation problem of inbound container in railway container terminal. The computational results are shown in Table 4 and Figure 4.

As observed in Table 4 and Figure 4, the imbalance amounts and overlapping amounts obtained by OA are both fewer than the amounts obtained by CA, the average GAP of imbalance amounts is 42.7%, and the average GAP of overlapping amounts is 48.8%. In order to evaluate the effectiveness and practicability of our approach, numerical experiments of 7 days and 30 days are conducted. The computational results of 7 days are shown in Table 5 and Figure 5 and of 30 days are shown in Figure 6.

As observed in Table 5 and Figures 5 and 6, the performance of our approach is satisfactory in solving different size instances. The results of computational experiments indicate that our approach is efficient to solve space allocation problem of inbound container in railway container terminals.

Table 5: Comparison between OA and RAA in 7 days.

Planning period	Our approach (OA)		Random allocation algorithm (RAA)		GAP_1	GAP_2
	Imbalance amounts	Overlapping amounts	Imbalance amounts	Overlapping amounts		
1	3	2	7	9	57.1%	77.8%
2	7	8	11	14	36.4%	42.9%
3	11	14	18	21	38.9%	33.3%
4	8	10	13	17	38.5%	41.2%
5	2	1	5	4	60.0%	75.0%
6	13	11	17	25	23.5%	56.0%
7	16	13	22	19	27.3%	31.6%
8	11	8	14	11	21.4%	27.3%
9	0	0	0	0	0.0%	0.0%
10	3	3	5	11	40.0%	72.7%
11	15	9	19	17	21.1%	47.1%
12	13	7	15	15	13.3%	53.3%
13	5	0	11	3	54.5%	100.0%
14	9	3	17	11	47.1%	72.7%
15	13	15	24	27	45.8%	44.4%
16	7	11	19	24	63.2%	54.2%
17	3	0	9	4	66.7%	100.0%
18	4	6	8	11	50.0%	45.5%
19	11	14	15	19	26.7%	26.3%
20	9	7	21	19	57.1%	63.2%
21	1	0	3	2	66.7%	100.0%
22	6	5	13	13	53.8%	61.5%
23	19	17	27	26	29.6%	34.6%
24	9	4	13	16	30.8%	75.0%
25	0	0	0	0	0.0%	0.0%
26	11	5	19	9	42.1%	44.4%
27	18	10	22	23	18.2%	56.5%
28	12	7	18	15	33.3%	53.3%

Notes: GAP_1 = (imbalance amounts obtained from RAA − imbalance amounts obtained from OA) ∗ 100/imbalance amounts obtained from RAA.
GAP_2 = (overlapping amounts obtained from RAA − overlapping amounts obtained from OA) ∗ 100/overlapping amounts obtained from RAA.

7. Conclusion

In this paper, we considered the space allocation problem of inbound container in railway container terminals with the arrival-departure time and operation sequence of containers known. Two-stage optimization models were proposed; the first stage is workload balance model, whose objective is to balance workloads of inbound containers among blocks and evenly allocate inbound containers to each block. The second stage is container slot allocation model, whose objective is to minimize overlapping amounts of ICT. An algorithm implement process based on rolling horizon approach is designed to solve the proposed models. Computational experiments on an actual railway container terminal show that the models and algorithm proposed in this paper are effective to balance workloads of inbound containers and reduce the overlapping amounts.

In future, proposing a stochastic programming model for container slot allocating problem by considering random factors is a possibility for further research.

Conflict of Interests

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

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant no. 60870014), the Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20130009110001), and the Major Research plan of the National Natural Science Foundation of China (Grant no. 71390332).

References

[1] M. Bazzazi, N. Safaei, and N. Javadian, "A genetic algorithm to solve the storage space allocation problem in a container terminal," *Computers and Industrial Engineering*, vol. 56, no. 1, pp. 44–52, 2009.

[2] C. Zhang, J. Liu, Y. Wan, K. G. Murty, and R. J. Linn, "Storage space allocation in container terminals," *Transportation Research B: Methodological*, vol. 37, no. 10, pp. 883–903, 2003.

[3] K. H. Kim, Y. M. Park, and K. Ryu, "Deriving decision rules to locate export containers in container yards," *European Journal of Operational Research*, vol. 124, no. 1, pp. 89–101, 2000.

[4] K. H. Kim and K. T. Park, "A note on a dynamic space-allocation method for outbound containers," *European Journal of Operational Research*, vol. 148, no. 1, pp. 92–101, 2003.

[5] L. Chen and Z. Lu, "The storage location assignment problem for outbound containers in a maritime terminal," *International Journal of Production Economics*, vol. 135, no. 1, pp. 73–80, 2012.

[6] Y. J. Woo and K. H. Kim, "Estimating the space requirement for outbound container inventories in port container terminals," *International Journal of Production Economics*, vol. 133, no. 1, pp. 293–301, 2011.

[7] M. Rodriguez-Molins, M. A. Salido, and F. Barber, "Intelligent planning for allocating containers in maritime terminals," *Expert Systems with Applications*, vol. 39, no. 1, pp. 978–989, 2012.

[8] M. Yu and X. Qi, "Storage space allocation models for inbound containers in an automatic container terminal," *European Journal of Operational Research*, vol. 226, no. 1, pp. 32–45, 2013.

[9] O. Sharif and N. Huynh, "Storage space allocation at marine container terminals using ant-based control," *Expert Systems with Applications*, vol. 40, no. 6, pp. 2323–2330, 2013.

[10] A. R. McKendall Jr. and J. R. Jaramillo, "A tabu search heuristic for the dynamic space allocation problem," *Computers and Operations Research*, vol. 33, no. 3, pp. 768–789, 2006.

[11] A. R. McKendall Jr., "Improved Tabu search heuristics for the dynamic space allocation problem," *Computers and Operations Research*, vol. 35, no. 10, pp. 3347–3359, 2008.

[12] D. Lee, J. X. Cao, Q. Shi, and J. H. Chen, "A heuristic algorithm for yard truck scheduling and storage allocation problems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 5, pp. 810–820, 2009.

[13] T. Park, R. Choe, Y. Hun Kim, and K. Ryel Ryu, "Dynamic adjustment of container stacking policy in an automated container terminal," *International Journal of Production Economics*, vol. 133, no. 1, pp. 385–392, 2011.

[14] M. A. Salido, M. Rodriguez-Molins, and F. Barber, "A decision support system for managing combinatorial problems in container terminals," *Knowledge-Based Systems*, vol. 29, pp. 63–74, 2012.

[15] G. C. da Silva, L. Bahiense, L. Satoru Ochi, and P. O. Boaventura-Netto, "The dynamic space allocation problem: Applying hybrid GRASP and Tabu search metaheuristics," *Computers and Operations Research*, vol. 39, no. 3, pp. 671–677, 2012.

[16] L. Wang, *Key resources scheduling optimization theory and method of railway container terminal [Ph.D. thesis]*, Beijing Jiaotong University, Beijing, China, 2014, (Chinese).

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 607195, 8 pages
http://dx.doi.org/10.1155/2014/607195

Research Article

A Bayesian Combined Model for Time-Dependent Turning Movement Proportions Estimation at Intersections

Pengpeng Jiao, Tuo Sun, and Lin Du

School of Civil and Transportation Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Correspondence should be addressed to Pengpeng Jiao; jiaopengpeng@bucea.edu.cn

Received 1 April 2014; Revised 23 June 2014; Accepted 1 July 2014; Published 17 July 2014

Academic Editor: X. Zhang

Time-dependent turning movement flows are very important input data for intelligent transportation systems but are impossible to be detected directly through current traffic surveillance systems. Existing estimation models have proved to be not accurate and reliable enough during all intervals. An improved way to address this problem is to develop a combined model framework that can integrate multiple submodels running simultaneously. This paper first presents a back propagation neural network model to estimate dynamic turning movements, as well as the self-adaptive learning rate approach and the gradient descent with momentum method for solving. Second, this paper develops an efficient Kalman filtering model and designs a revised sequential Kalman filtering algorithm. Based on the Bayesian method using both historical data and currently estimated results for error calibration, this paper further integrates above two submodels into a Bayesian combined model framework and proposes a corresponding algorithm. A field survey is implemented at an intersection in Beijing city to collect both time series of link counts and actual time-dependent turning movement flows, including historical and present data. The reported estimation results show that the Bayesian combined model is much more accurate and stable than other models.

1. Introduction

Time-dependent turning movement flows are very important input data for traffic signal control system, route guidance system, and other ITS systems. However, using the existing traffic detection devices, it is impossible to achieve the turning movement flows at intersections directly. Since the real-time link flows of entering and exiting legs can be detected conveniently, the dynamic turning movements estimation methods based on detected time-series of link counts, that is, dynamic origin-destination flows estimation (DODE), have been studied extensively in the literature.

Using different methods, existing researches have formulated many DODE models, which can be generally classified into five categories according to the modeling techniques: parameter optimization method, entropy maximization method, maximum likelihood method, Kalman filtering (KF) method, and variational inequality (VI) method. The previous three categories are extended from the static origin-destination (O-D) estimation problems, the KF model is a kind of state-space method, and the VI model is a rather new technique.

Along a different line, existing DODE models have focused on different objective networks, including intersection, freeway segment, and general road network. Intersection models were the earliest to be developed, and most of them were extended from static O-D estimation models. Cremer and Keller (1981) [1], Cremer (1983) [2], and Cremer and Keller (1984) [3] have constructed relations between time series of link traffic flows and dynamic turning movement flows, that is, dynamic O-D flows, and formulated a series of least square models. Nihan and Davis (1987) [4] further proposed a revised least square model using recursive method and designed the truncation and normalization processes to satisfy the constraints requirement of dynamic turning proportions. Bell [5] also presented a revised optimization model to estimate intersection turning movement flows, which considered the platoon dispersion between entrance and exit legs. All above researches only employed the inequality constraints in the optimization models. To further incorporate the equality constraints, Li and De Moor (1999) [6] put forward a new least square model. Considering that the occasional outliers of detected flows may greatly influence

the results of least square models, Jiao et al. [7] proposed a least absolute deviation model and designed a genetic algorithm to obtain the optimal solution. All these models fall within the scope of parameter optimization method, they can achieve rather accurate estimation results; however, they are not efficient enough.

Introduction of dynamic travel time and dynamic route choice extended intersection models to freeway segment and general network. Most of these models employed state-space formulations, for example, Okutani (1987) [8], Ashok and Ben-Akiva (2000, 2002) [9, 10], Lin and Chang (2007) [11], Li et al. (2009) [12], and so forth. These models were all formulated using KF techniques and were applicable for freeway segment or general road network. Since this paper mainly deals with intersection problems, we do not put too much emphasis on the review of these models, but the efficiency of KF method is rather valuable for time-dependent turning movement proportions estimation at intersections.

More recently, Nie and Zhang (2008) [13] employed the variational inequality technique in the DODE problem and formulated several new VI models. Lou and Yin (2010) [14] inferred turning movements with incomplete information at intersections and further incorporated them into a decomposition scheme for estimating dynamic O-D flows on actuation-controlled signalized arterials. Lu et al. (2013) [15] considered the influences of congestion and presented a single-level nonlinear optimization model to estimate the dynamic O-D flows, integrating a dynamic user equilibrium constraint. All these models have proved to be rather accurate.

Further analyses of existing researches show that the estimated dynamic O-D flows often fluctuate within a range around the actual values, that is, in some specific intervals, some models tend to overestimate the O-D flows, while other models tend to underestimate them. One possible way to approach this problem is to develop a combined model framework that can automatically integrate multiple submodels that run simultaneously.

To the best of our knowledge, there is no existing research which estimates the intersection turning movements in a combined model framework. Fortunately, similar methods have been used in short-term traffic forecast through combining multiple prediction models using the errors the models made in the previous time intervals, for example, Zheng et al. (2006) [16] and Dong et al. (2010) [17]. However, the nature of dynamic turning movement estimation is different to that of short-term traffic forecast. Therefore, the key feature of this paper is to integrate several different estimation models and to present a combined model to estimate the turning movement proportions at intersections.

The rest of this paper is organized as follows. Section 2 illustrates the basic problem statement and variable definitions. Section 3 presents a revised back propagation (BP) neural network model to estimate intersection turning movement proportions and further develops a self-adaptive learning rate approach and the gradient descent with momentum method to solve the BP neural network model. Section 4 formulates an efficient KF method and designs a revised sequential KF algorithm. Based on above two submodels,

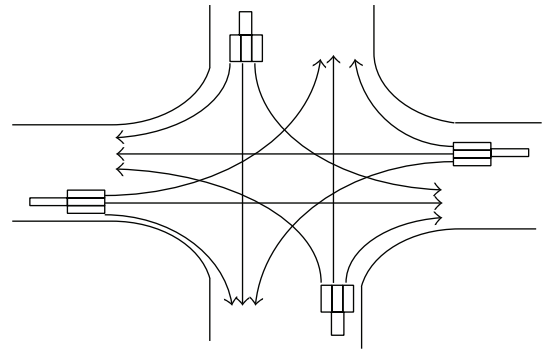


Figure 1: A typical intersection.

Section 5 further proposes a Bayesian combined model, which is calibrated using both historical data and estimated results in past several intervals. Section 6 reports the evaluation results based on practical traffic survey data. Section 7 concludes the paper and recommends some future researches.

2. Problem Statement and Variable Definitions

This paper deals with a typical intersection with r entrance legs and s exiting legs, as described in Figure 1.

For convenience of illustration, we first define following variables used in this paper:

$Q_i(k)$ is detected link flows at entering leg i during interval k, $i = 1, 2, \ldots, r$;

$Y_j(k)$ is detected link flows at exiting leg j during interval k, $j = 1, 2, \ldots, s$;

$T_{ij}(k)$ is turning flows entering from leg i during interval k with destination at leg j, that is, the time-varying turning movement flows;

$B_{ij}(k)$ is turning movement proportions, that is, $B_{ij}(k) = T_{ij}(k)/Q_i(k)$.

Obviously,

$$0 \le B_{ij}(k) \le 1 \tag{1}$$

$$\sum_j B_{ij}(k) = 1 \tag{2}$$

$$Y_j(k) = \sum_{i \ne j} Q_i(k) B_{ij}(k). \tag{3}$$

Additionally, we use the superscript "N" to denote the specific variables in BP neural network model, "KF" to indicate the specific variables in KF model, and "H" to show the historical data.

Based on the $Q_i(k)$ and $Y_j(k)$ obtained from traffic detection devices, the key issue of this paper is to estimate the time-varying turning movements $T_{ij}(k)$ or turning proportions $B_{ij}(k)$ from the detected time series of link counts.

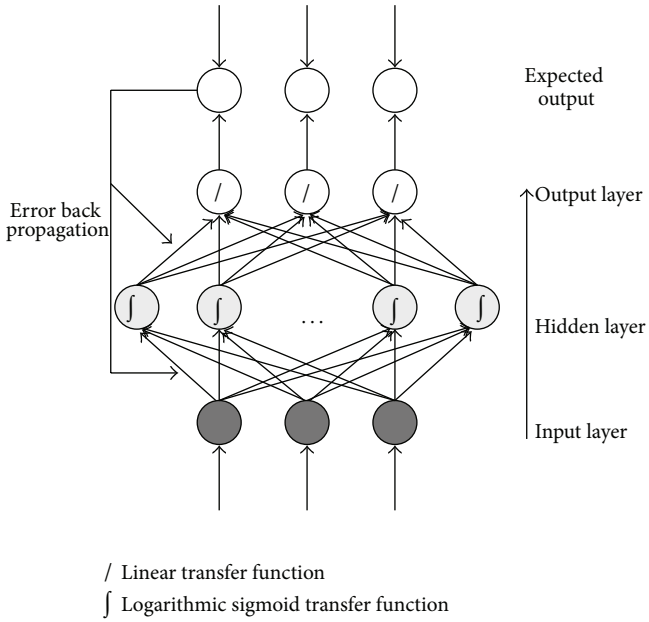


Figure 2: Architecture of back propagation neural network model.

3. Back Propagation Neural Network Model

3.1. Architecture of the BP Neural Network Model. The architecture of the BP neural network model is shown in Figure 2. It mainly has the following advantages.

(1) It has very strong self-learning abilities and is capable of addressing very complicated nonlinear estimation problems, especially for short-term cases.

(2) The large amount of training data can be achieved from traffic surveillance systems directly, which will greatly improve the performance of the estimation.

(3) The back propagation of estimation errors enables the dynamic adjustment of weights, which speeds up the training process and improves the accuracy of the model.

There are three layers in the BP neural network model.

(1) Input layer: three neurons in the input layer, corresponding to the link counts of the upstream approach. Here the number of neurons is subject to the number of upstream lanes and detectors.

(2) Hidden layer: 15 neurons in the hidden layer, which is decided through many experiments. The logarithmic sigmoid transfer function is used in the hidden layer, which outputs results between 0 and 1, corresponding to the actual range of turning movement proportions. The formulation of the transfer function is

$$f(x) = \frac{1}{(1 + e^{-x})}. \tag{4}$$

(3) Output layer: three neurons in the output layer, corresponding to the turning movement flows of left-turn, go straight and right-turn directions, that is, the estimation results. Linear transfer function is used in the output layer.

In this paper, historically detected link flows and surveyed turning proportions are used as the training set to train the BP neural network model, and currently detected link flows are used as the test set to estimate the turning fractions.

3.2. Algorithm of Self-Adaptive Learning Rate and Gradient Descent with Momentum. Since common BP neural network method is rather slow in the training process and often arrives at locally optimal solution [18], this paper adopts the integrated algorithm of self-adaptive learning rate and gradient descent with momentum, which accelerates the training process dramatically and guarantees the reliability of the algorithm.

The self-adaptive learning rate approach is formulated as follows:

$$\begin{aligned} \eta(N) &= 2^{\lambda}\eta(N-1) \\ \lambda &= \operatorname{sign}\left[\frac{\Delta\omega(N)\,\Delta\omega(N-1)}{\eta(N)\,\eta(N-1)}\right], \end{aligned} \tag{5}$$

where $\eta(N)$ is the learning rate of step N, λ is the growth factor of learning rate, $\omega(N)$ is the weight of step N, $\Delta\omega(N)$ is the difference between current weight and the previous weight, and sign[ ] is the function returning the sign of the variable in the square bracket.

Here the learning rate is adjusted automatically according to the gradient in the previous step, with the initial value within [0.01, 0.8].

The gradient descent with momentum method is presented as follows:

$$\begin{aligned} &\omega_{mn}(N+1) \\ &= \omega_{mn}(N) \\ &\quad + \eta(N)\left[(1-mc)\frac{\Delta\omega_{mn}(N)}{\eta(N)} + mc\frac{\Delta\omega_{mn}(N-1)}{\eta(N-1)}\right], \\ &\omega_{lm}(N+1) \\ &= \omega_{lm}(N) + \eta(N)\left[(1-mc)\frac{\Delta\omega_{lm}(N)}{\eta(N)} + mc\frac{\Delta\omega_{lm}(N-1)}{\eta(N-1)}\right], \\ &\theta_{m}(N+1) \\ &= \theta_{m}(N) + \eta(N)\left[(1-mc)\frac{\Delta\theta_{m}(N)}{\eta(N)} + mc\frac{\Delta\theta_{m}(N-1)}{\eta(N-1)}\right], \\ &\alpha_{n}(N+1) \\ &= \alpha_{n}(N) + \eta(N)\left[(1-mc)\frac{\Delta\alpha_{n}(N)}{\eta(N)} + mc\frac{\Delta\alpha_{n}(N-1)}{\eta(N-1)}\right], \end{aligned} \tag{6}$$

where l denotes neurons in the input layer, m indicates neurons in the hidden layer, n shows neurons in the output

layer, ω_{lm} is the weight from neuron m in the hidden layer to neuron l in the input layer, ω_{mn} is the weight from neuron n in the output layer to neuron m in the hidden layer; mc is the momentum factor with the constraint $0 < mc < 1$; θ_m is the threshold of node m in the hidden layer, and α_n is the threshold of node n in the output layer. Both θ_m and α_n keep updating with the training process.

The above algorithm is coded using M language of Matlab software.

4. Kalman Filtering Model

4.1. KF Model. To estimate the dynamic turning movement flows efficiently, we further propose a revised Kalman filtering model using the turning movement proportions $B_{ij}(k)$ as the state variables to reflect the interrelations between the link flows at entering and exiting legs.

We formulate the state transition equation as follows:

$$\mathbf{B}(k) = \mathbf{B}(k-1) + \mathbf{W}(k), \tag{7}$$

where $\mathbf{B}(k)$ is a column vector form of $B_{ij}(k)$ and $\mathbf{W}(k)$ is a column vector of random errors. $\mathbf{W}(k)$ is actually a Gaussian white noise vector, its mean value is $\mathbf{0}$, and its covariance matrix is $\mathbf{D}\delta_{yz}$, where $\mathbf{D}$ is a constant semipositive matrix and δ_{yz} is the Kronecker delta; that is, $\delta_{yz} = 1$ while $y = z$; otherwise $\delta_{yz} = 0$.

According to (3), we formulate the measurement equation as follows:

$$\mathbf{Y}(k) = \mathbf{Q}(k) * \mathbf{B}(k) + \mathbf{e}(k), \tag{8}$$

where $\mathbf{Y}(k)$ is a column vector of detected link flows at exiting legs, $\mathbf{Q}(k)$ is the corresponding measurement matrix, and $\mathbf{e}(k)$ is the column vector of link flows detection errors; its mean value is $\mathbf{0}$, and its covariance matrix is $\mathbf{R}\delta_{yz}$, where $\mathbf{R}$ is also a constant semipositive matrix, just like $\mathbf{D}$.

Equations (7) and (8) constitute the Kalman filtering model to estimate the time-dependent turning movement proportions in a state-space formulation.

4.2. Revised Sequential KF Algorithm. To improve the estimation efficiency, we design a revised sequential KF algorithm [19] to avoid the inverse matrix computation. The truncation and normalization processes [4] are also integrated into the algorithm to modify the results to satisfy the inherent constraints of $B_{ij}(k)$ in (1) and (2).

Furthermore, the initial values of $B_{ij}(k)$ are preset according to the number of lanes turning to different directions instead of conventional average number 0.33, as shown in (9). This revision of initial values of $B_{ij}(k)$ will accelerate the convergence of the revised KF algorithm:

$$B_{ij}(0) = \frac{L_{ij}}{\sum_j L_{ij}}, \tag{9}$$

where L_{ij} is the number of lanes from entering leg i turning to exiting leg j. In case of the presence of mixed lane, it is divided into the involved directions averagely.

Except for the initial values of $B_{ij}(k)$ and the truncation and normalization processes, the other steps in the algorithm are rather similar to existing sequential KF algorithms [19].

This algorithm is also coded using M language of Matlab software.

5. Bayesian Combined Model

Through the above two models, one can obtain the time-dependent turning movement proportions $B_{ij}^{\mathrm{N}}(k)$ and $B_{ij}^{\mathrm{KF}}(k)$, respectively, where $B_{ij}^{\mathrm{N}}(k)$ is the result from BP neural network model and $B_{ij}^{\mathrm{KF}}(k)$ is the result from KF model. To get more accurate and stable estimation results, we integrate these two models into a combined model framework, as shown in Figure 3.

From Figure 3 one can find out that the Bayesian combined model is actually a weighted average result of BP neural network and KF models:

$$\widetilde{B}_{ij}(k) = W^{\mathrm{N}}(k) \times B_{ij}^{\mathrm{N}}(k) + W^{\mathrm{KF}}(k) \times B_{ij}^{\mathrm{KF}}(k), \tag{10}$$

where $\widetilde{B}_{ij}(k)$ is the estimation result of the combined model, $W^{\mathrm{N}}(k)$ is the weight of BP neural network model, and $W^{\mathrm{KF}}(k)$ is the weight of KF model.

By adjusting the above two weights, each submodel may be strengthened or weakened. If we set the weight of a submodel to zero, this model will be neglected. According to a comparison of estimation errors, these weights are decided logically, and the details are shown below.

According to the historical estimated results of two submodels and the historical actual turning proportions, we can get the mean absolute percentage error (MAPE) of two submodels, respectively. Here we use EH^{N} to denote the historic MAPE of BP neural network model and $\mathrm{EH}^{\mathrm{KF}}$ to denote the historic MAPE of KF model. The prior probabilities to choose BP neural network model and KF model are presented as follows:

$$\begin{aligned} \Pr\left(H^{\mathrm{N}}\right) &= \begin{cases} 1 - \mathrm{EH}^{\mathrm{N}}, & \left(\mathrm{EH}^{\mathrm{N}} < 1\right) \\ 0, & \left(\mathrm{EH}^{\mathrm{N}} \geq 1\right) \end{cases} \\ \Pr\left(H^{\mathrm{KF}}\right) &= \begin{cases} 1 - \mathrm{EH}^{\mathrm{KF}}, & \left(\mathrm{EH}^{\mathrm{KF}} < 1\right) \\ 0, & \left(\mathrm{EH}^{\mathrm{KF}} \geq 1\right), \end{cases} \end{aligned} \tag{11}$$

where the superscript "N" means BP neural network model, the superscript "KF" means KF model, the function Pr() denotes a choice probability, $\Pr(H^{\mathrm{N}})$ is the prior probability of model H^{N} (BP neural network model), and $\Pr(H^{\mathrm{KF}})$ is the prior probability of model H^{KF} (KF model). Equations in (11) illustrate the influences of historical estimation errors.

To further reflect the influences of current estimation errors, we define E^{N} and E^{KF} as the MAPEs of currently running BP neural network and KF models, respectively.

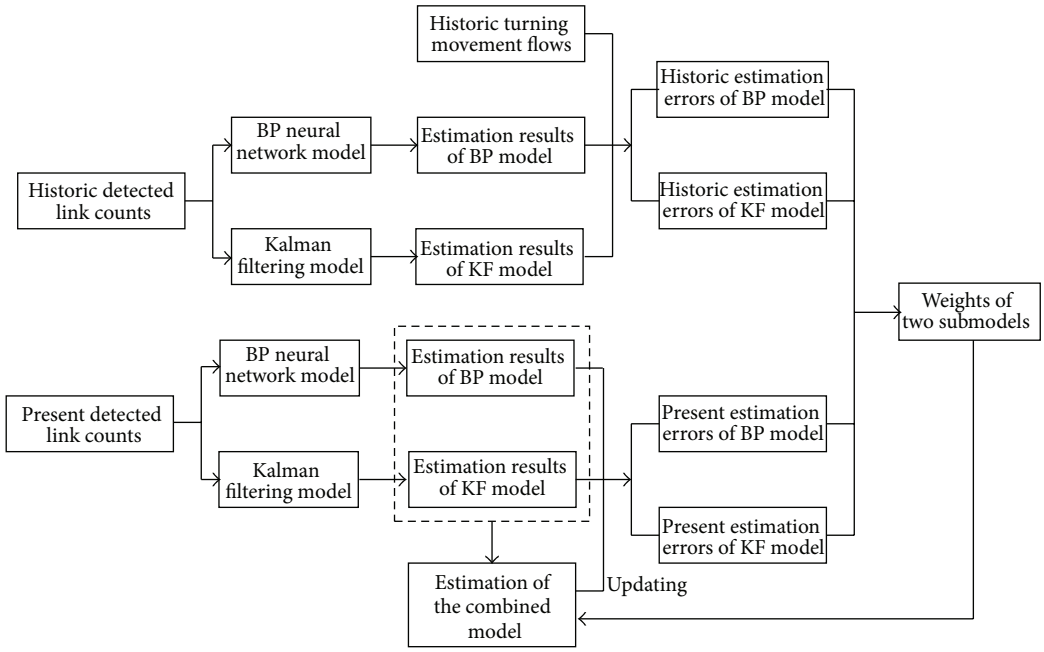


Figure 3: Updating flow of the Bayesian combined model.

Here the MAPEs are from previous 5 intervals; that is, they are updating momentarily:

$$\Pr\left(D \mid H^{\mathrm{N}}\right) = \begin{cases} 1 - E^{\mathrm{N}}, & \left(E^{\mathrm{N}} < 1\right) \\ 0, & \left(E^{\mathrm{N}} \geq 1\right) \end{cases}$$
$$\Pr\left(D \mid H^{\mathrm{KF}}\right) = \begin{cases} 1 - E^{\mathrm{KF}}, & \left(E^{\mathrm{KF}} < 1\right) \\ 0, & \left(E^{\mathrm{KF}} \geq 1\right), \end{cases} \tag{12}$$

where $\Pr(D \mid H^{\mathrm{N}})$ is the probability of generating estimation D using model H^{N} and $\Pr(D \mid H^{\mathrm{KF}})$ is the probability of generating estimation D using model H^{KF}.

In (12), the most important issue is that the present actual turning proportions are unavailable. To address this problem, we use the Bayesian weighted results instead of actual values. For the first five intervals, the corresponding historical actual values are used alternatively. Since the previous five intervals remain rolling, there is actually an updating mechanism in the Bayesian combined model stated here, as shown in Figure 3.

According to the Bayesian theorem, the posterior probability [20] can be formulated as follows:

$$\Pr\left(H^{\mathrm{N}} \mid D\right) = \frac{\Pr\left(D \mid H^{\mathrm{N}}\right)\Pr\left(H^{\mathrm{N}}\right)}{\Pr\left(D\right)}$$
$$\Pr\left(H^{\mathrm{KF}} \mid D\right) = \frac{\Pr\left(D \mid H^{\mathrm{KF}}\right)\Pr\left(H^{\mathrm{KF}}\right)}{\Pr\left(D\right)}$$
$$\Pr\left(D\right) = \Pr\left(D \mid H^{\mathrm{N}}\right)\Pr\left(H^{\mathrm{N}}\right) + \Pr\left(D \mid H^{\mathrm{KF}}\right)\Pr\left(H^{\mathrm{KF}}\right), \tag{13}$$

where $\Pr(H^{\mathrm{N}} \mid D)$ is the posterior probability of BP neural network model and $\Pr(H^{\mathrm{KF}} \mid D)$ is the posterior probability of KF model.

Finally we obtain the weights of two submodels:

$$W^{\mathrm{N}} = \frac{\Pr\left(D \mid H^{\mathrm{N}}\right)\Pr\left(H^{\mathrm{N}}\right)}{\Pr\left(D \mid H^{\mathrm{N}}\right)\Pr\left(H^{\mathrm{N}}\right) + \Pr\left(D \mid H^{\mathrm{KF}}\right)\Pr\left(H^{\mathrm{KF}}\right)}$$
$$W^{\mathrm{KF}} = \frac{\Pr\left(D \mid H^{\mathrm{KF}}\right)\Pr\left(H^{\mathrm{KF}}\right)}{\Pr\left(D \mid H^{\mathrm{N}}\right)\Pr\left(H^{\mathrm{N}}\right) + \Pr\left(D \mid H^{\mathrm{KF}}\right)\Pr\left(H^{\mathrm{KF}}\right)}. \tag{14}$$

Using (10) and (14), we can obtain the final estimation results of the Bayesian combined model.

6. Case Study

To testify the accuracy and efficiency of two submodels and the Bayesian combined model, we collected the real-world data of 2 hours through a field survey during morning peak hours at the intersection of Zhaodengyu road and Pinganli west road, which is located in Xicheng district, Beijing city, China. We achieved enough data for the case study, including time-varying link flows on all entering and exiting legs and dynamic turning movement flows from all entrance legs. The survey was conducted twice during the same morning peak hour periods in two days, respectively. Data of the first day is used as the historical information, and data of the second day is used as currently detected information.

Furthermore, we implemented three algorithms and achieved all estimated turning movement proportions using three models, respectively. A time interval of 3 minutes is used

Table 1: Statistical results of evaluation indices.

Index	B_{14}			B_{31}			B_{41}		
	BP	KF	Bayesian	BP	KF	Bayesian	BP	KF	Bayesian
MAPE	43.41%	37.97%	35.45%	14.06%	11.76%	12.92%	36.82%	28.01%	27.82%
MPE	13.28%	15.88%	15.38%	12.29%	11.71%	12.07%	10.35%	10.51%	10.42%
RMSE	0.064	0.058	0.056	0.142	0.119	0.125	0.114	0.095	0.093
NRMS	43.32%	39.50%	37.65%	17.03%	14.77%	15.21%	42.72%	34.65%	34.71%

in the paper; therefore there are totally 40 intervals in each day.

For the historical estimation, the previous 30 intervals are used to train the BP neural network model, and the results of last 10 intervals are estimated. Furthermore, all these 40 historical intervals are used to train the BP neural network model for present estimation.

The right-turn at east entrance, go straight at west entrance, and left-turn at north entrance, that is, $B_{14}(k)$, $B_{31}(k)$, and $B_{41}(k)$, are taken, for instance, to show the accuracy of the proposed models. The mean average percentage error (MAPE), mean percentage error (MPE), root mean square error (RMSE), and normalized root mean square error (NRMS) between actual and estimated values are selected as the evaluation criteria. The reduction in these indices thus represents the potential improvement of the proposed models and algorithms:

$$\begin{aligned}
\text{MAPE} &= \frac{\sum_{k=1}^{n}\left(\left|\widetilde{B}_{ij}(k) - B_{ij}(k)\right| / B_{ij}(k)\right)}{n} \times 100\% \\
\text{MPE} &= \frac{\sum_{k=1}^{n}\left(\left(\widetilde{B}_{ij}(k) - B_{ij}(k)\right) / B_{ij}(k)\right)}{n} \times 100\% \\
\text{RMSE} &= \sqrt{\frac{\sum_{k=1}^{n}\left[\widetilde{B}_{ij}(k) - B_{ij}(k)\right]^2}{n}} \\
\text{NRMS} &= \frac{\sqrt{n\sum_{k=1}^{n}\left[\widetilde{B}_{ij}(k) - B_{ij}(k)\right]^2}}{\sum_{k=1}^{n} B_{ij}(k)} \times 100\%,
\end{aligned} \tag{15}$$

where n is the total number of time intervals.

The evaluation indices of two submodels and the Bayesian combined model are all reported in Table 1. Because the estimation results of KF model during its initialization period are usually unreliable, the results of the first 20 intervals are excluded from the indices statistics.

Graphical illustrations of these estimation results during last 35 intervals are further illustrated in Figures 4, 5, and 6.

As expected, one can find out the following results from the case study.

(1) All three models are rather accurate. Relatively, the Bayesian combined model yields better estimation results than those from two submodels, which is obvious from the magnitudes of all four evaluation indices in Table 1. It clearly indicates the advantage of the Bayesian combined model framework. If historical data are available, the proposed method can provide accurate estimation of dynamic turning movement proportions.

(2) Concerning the capability to track the dynamic characteristics of turning fractions, the Bayesian combined model still achieves the best results. It is mainly because of the increased system observability from historical data.

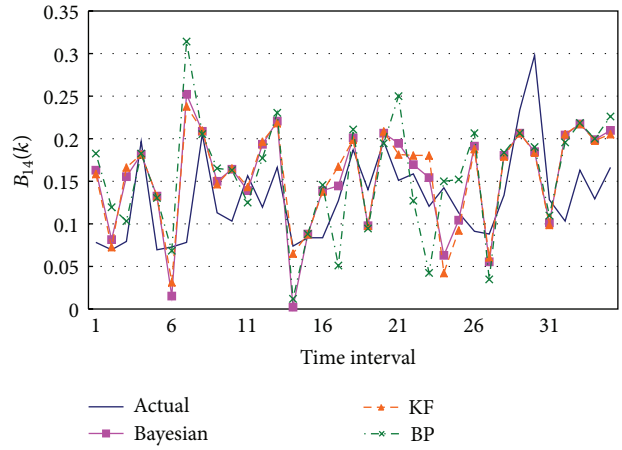


Figure 4: Comparison of estimated turning proportion $B_{14}(k)$ from three models.

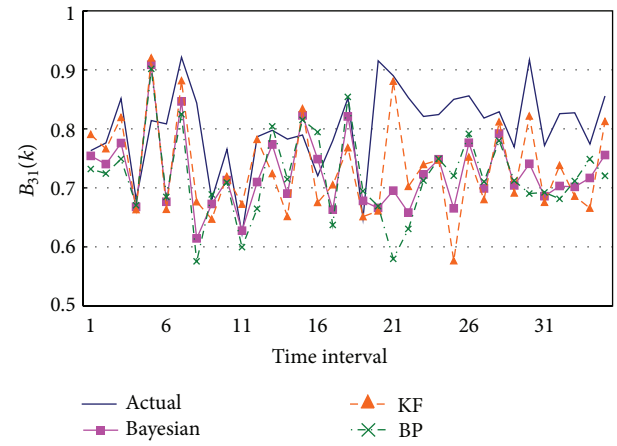


Figure 5: Comparison of estimated turning proportion $B_{31}(k)$ from three models.

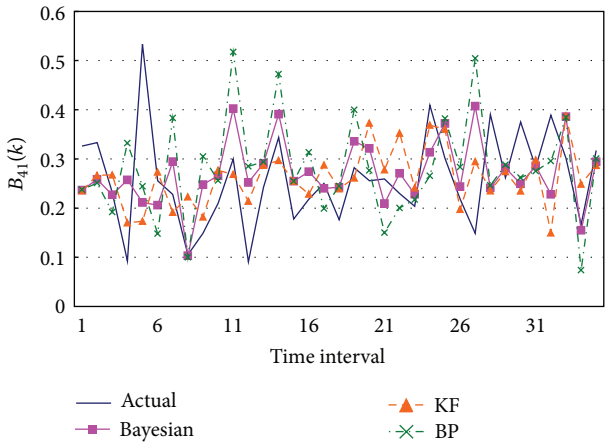


FIGURE 6: Comparison of estimated turning proportion $B_{41}(k)$ from three models.

(3) Results of all three models fluctuate within a range around the actual values; that is, during some intervals, the BP neural network model tends to overestimate the turning fractions, while the KF model tends to underestimate them, and vice versa during some other intervals. However, the turbulence of estimation results from Bayesian combined model is the minimum. It indicates that the proposed Bayesian combined model is the most stable and robust.

(4) The results of go through direction are better than those of left-turn and right-turn directions. It is mainly due to the rather bigger values of the go through proportions.

(5) Estimation results of all turning fractions show that the KF model is rather more accurate than the BP neural network model. The possible reason is that only data of 40 intervals are used for training, which reduces the accuracy of the BP neural network model.

7. Conclusions

This paper addresses three models concerning time-dependent turning movement proportions estimation at intersections: BP neural network model, KF model, and Bayesian combined model. We first present a revised BP neural network model and develop a self-adaptive learning rate approach and the gradient descent with momentum method for solution. For more efficient estimation, we propose a revised KF model and design a modified sequential KF algorithm. Taking into account historical information and estimation errors, we further integrate above two submodels into a Bayesian combined model framework, which is calibrated using both historical and present estimation errors. The reported examples based on practical traffic survey data have demonstrated that the Bayesian combined model is much more accurate and stable than other two submodels.

Further researches are directed towards two aspects. The first is to take into account the travel time of intersection turning vehicles under traffic congestion. The second is to extend the combined model to freeway segment and general road network and to estimate O-D flows for large-scale network.

Conflict of Interests

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

Acknowledgments

The research reported here has been partially supported by National Natural Science Foundation of China Project (51208024), Science and Technology Project of Ministry of Housing and Urban-Rural Development of China (2013-K5-6), and the Importation and Development of High-Caliber Talents Project of Beijing Municipal Institutions (CIT&TCD201404071). The authors thank the anonymous reviewers for their constructive comments and suggestions.

References

[1] M. Cremer and H. Keller, "Dynamic identification of O-D flows from traffic counts at complex intersections," in *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*, pp. 121–142, University of Toronto Press, Toronto, Canada, 1981.

[2] M. Cremer, "Determining the time dependent trip distribution in a complex intersection for traffic responsive control," in *Proceedings of the 4th International IFAC/IFIP/IFORS Conference on Control in Transportation Systems*, Baden-Baden, Germany, 1983.

[3] M. Cremer and H. Keller, "A systems dynamics approach to the estimation of entry and exit O-D flows," in *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, pp. 431–450, Vnu Science Press, Utrecht, The Netherlands, 1984.

[4] N. L. Nihan and G. A. Davis, "Recursive estimation of origin-destination matrices from input/output counts," *Transportation Research B*, vol. 21, no. 2, pp. 149–163, 1987.

[5] M. G. H. Bell, "The estimation of origin-destination matrices by constrained generalised least squares," *Transportation Research B: Methodological*, vol. 25, no. 1, pp. 13–22, 1991.

[6] B. Li and B. De Moor, "Recursive estimation based on the equality-constrained optimization for intersection origin-destination matrices," *Transportation Research B: Methodological*, vol. 33, no. 3, pp. 203–214, 1999.

[7] P. P. Jiao, H. P. Lu, and L. Yang, "Real-time estimation of turning movement proportions based on genetic algorithm," in *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems*, pp. 484–489, Vienna, Austria, September 2005.

[8] I. Okutani, "The kalman filtering approaches in some transportation and traffic problems," in *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, pp. 397–416, Elsevier Science, New York, NY, USA, 1987.

[9] K. Ashok and M. E. Ben-Akiva, "Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows," *Transportation Science*, vol. 34, no. 1, pp. 21–36, 2000.

[10] K. Ashok and M. E. Ben-Akiva, "Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows," *Transportation Science*, vol. 36, no. 2, pp. 184–198, 2002.

[11] P. W. Lin and G. L. Chang, "A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix," *Transportation Research B: Methodological*, vol. 41, no. 5, pp. 554–572, 2007.

[12] J. W. Li, B. L. Lin, Z. H. Sun, and X. F. Geng, "An estimation model of time-varying origin-destination flows in expressway corridors based on unscented Kalman filter," *Science in China E: Technological Sciences*, vol. 52, no. 7, pp. 2069–2078, 2009.

[13] Y. Nie and H. M. Zhang, "A variational inequality formulation for inferring dynamic origin-destination travel demands," *Transportation Research B: Methodological*, vol. 42, no. 7-8, pp. 635–662, 2008.

[14] Y. Lou and Y. Yin, "A decomposition scheme for estimating dynamic origin-destination flows on actuation-controlled signalized arterials," *Transportation Research C*, vol. 18, no. 5, pp. 643–655, 2010.

[15] C. C. Lu, X. Zhou, and K. Zhang, "Dynamic origin-destination demand flow estimation under congested traffic conditions," *Transportation Research C: Emerging Technologies*, vol. 34, pp. 16–37, 2013.

[16] W. Zheng, D. H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.

[17] S. Dong, R. M. Li, L. G. Sun, T. H. Chang, and H. P. Lu, "Short-term traffic forecast system of Beijing," *Transportation Research Record*, no. 2193, pp. 116–123, 2010.

[18] X. Wen, L. Zhou, X. Li, and B. W. Zhang, *Neural Network Simulation and Application Using Matlab [Monograph]*, Science Press, 2003.

[19] Y. Lin, Y. L. Cai, and Y. X. Huang, "Kalman filtering based dynamic OD matrix estimation and prediction for traffic systems," in *Proceedings of the 6th International IEEE Conference on Intelligent Transportation Systems*, vol. 2, pp. 1515–1520, Shanghai, China, 2003.

[20] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, Cambridge, UK, 2003.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 104349, 13 pages
http://dx.doi.org/10.1155/2014/104349

Research Article

Multiagent Based Decentralized Traffic Light Control for Large Urban Transportation System

Yang Xu, Yulin Zhang, and Ming Liu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Correspondence should be addressed to Yang Xu; xuyang@uestc.edu.cn

Received 2 February 2014; Revised 27 May 2014; Accepted 19 June 2014; Published 15 July 2014

Academic Editor: X. Zhang

Intelligent traffic control is an important issue of the modern transportation system. However, in large-scale urban transportation systems, traditional centralized coordination methods suffer bottlenecks in both communication and computation. Decentralized control is hard if there is very limited observation to the whole network as evidences to support joint traffic coordination decisions. In this paper, we proposed a novel decentralized, multiagent based approach for massive traffic lights coordination to promote the large-scale green transportation. Considering that only the traffic from the adjacent intersections may affect the state of a given intersection one time ahead, the key of our approach is using the observations of a local intersection and its neighbors as evidences to support the traffic light coordination decisions. Therefore, we can model the interactions as decentralized agents coordinating with a decision theoretical model. Within a local intersection, constraint optimizing agents are designed to efficiently search for joint activities of the lights. Since this approach involves only local intersection cooperation, it is well scalable and easily implemented with small communication overhead. In the last section, we present our software design on this approach and based on our simulation, this approach is feasible to a large urban transportation system.

1. Introduction

Building intelligent transportation techniques in large urban transportation systems is appealing to reduce fuel consumptions and transportation delays [1]. A key is to enable joint coordination between intersections so that continuous traffic can flow over several intersections in main directions with the least delays [2].

To achieve this, extensive researches have been carried out. Traditional centralized coordination schemas, such as auction [3] and resource optimization [4], although have proven their performance to coordinate a number of intersections, they are infeasible to the large urban transportation system due to their computation and communication bottleneck shown in related literatures [5]. Existing works in decentralized traffic control suffer many limitations due to the nature of partial observability to the whole transportation network as well as its high dynamic traffic patterns. Historical data based approaches, such as reinforcement learning [6] and pattern discovery models [7], rely on the statistic pattern inferred to predict incoming traffics. However, they are incapable of managing dynamic traffic patterns dissimilar to the historical data, such as traffic accident or road maintenance. Some other works use dynamic programming to react to the dynamic real-time traffic according to the sensor readings, but the models are based on some impractical assumptions. For example, Robertson used a decentralized control schema but assumed that the sensor information within the whole area was easily obtained from centralized severs [8]. This may suffer a communication bottleneck in a huge urban area. Shenoda assumed that the coming vehicles follow a Poisson distribution to build a decentralized coordination algorithm [9], which is not always the case, for example, domain of emergency response in disasters. Xie et al. used fixed traffic signal phases to solve the conflicted traffic flows [10], which is not flexible to optimize concurrent traffic flows within the intersections.

In this paper, we proposed a decentralized traffic control approach to enable green transportation in a large urban area.

This approach is built based on a two-level hierarchical agent-based architecture toward robust and flexible coordination. In the top level, decentralized agents are modeled to coordinate with the neighbor intersections. In the bottom level, local agents within an intersection work cooperatively with a constraint optimization model. To enable the intelligent transportation, the key is that agents should make decisions based on the prospection of their local traffic state. Although it is infeasible for these agents to get the global state of the whole network to infer its next local state, we observed that only the traffic in their locally adjacent intersections can affect their coordination decision. Therefore, by closely interacting with their neighbors and sharing their local traffic states, each intersection may be able to gain a complete view of the states necessary to achieve decentralized control. Since this approach requires only the local state of neighbor intersections to cooperate, it is well scalable and easily implemented with small communication overhead.

In our algorithm design, between intersections, we setup a decision theoretical model for decentralized coordination. Since the next-time traffic is solely determined by the current state of the local intersection and its actions, it is a Markov decision process (MDP) and each agent takes the states of adjacent intersections into consideration for their decision model. To solve the uncertainties in state transition functions, we built heuristics either from statistic data or from reinforcement learning so that middle agents are able to jointly choose their best actions to minimize traffic delays. In addition, local agents within an intersection handle the local traffic. The decision process within a local intersection is modeled as a constraint optimization problem (COP) where maximum traffic flow should get through but conflicted traffics should be avoided.

In order to implement our software design, the key is to build each agent and implement the coordination by the interactions of the agents. In the top level, information agent is modeled for each intersection to maintain the local state that it needs to make decisions. In the bottom level, control agents within an intersection work cooperatively to solve the conflicts of different traffic flows. Middle agents are also built to coordinate all these control agents with the COP model. By building these two-level agents, the coordination between intersections is mainly achieved by the information agents, and the conflicts within each intersection are able to be solved by the middle agent and control agents. The system is implemented by RETSINA platform. In addition, it is simulated and the illustrated results proved the feasibility of our approach.

2. State of the Art

Many algorithms are designed to optimize the traffic of large urban transportation system in a decentralized manner. In order to achieve decentralized urban traffic control, the most straightforward approach is to generate optimal coordinated plans for fixed-time operation, such as TRANSYT [11]. However, due to the high dynamics in large urban transportation, this approach is hard to be adaptive to its real-time traffic.

Historical data are widely used to generate adaptive algorithms. Pattern discovery models are developed by categorizing the historical traffic into different patterns and assigning an optimal traffic light control plan for each pattern. PCA and SVM methods are applied for feature extraction, training, and classification of network-level traffic patterns [12] so that sensors within the intersection can detect real-time traffic pattern and choose the predefined plan. But the traffic patterns known as a priori may not cover all the patterns in real domains. An alternative is to incrementally build new patterns by assuming that the traffic is relatively infrequently changing [13]. Reinforcement learning is a set of techniques that is always applied in this domain. Q-learning is applied to learn the control policy for single intersection from historical data [14]. Multiagent reinforcement learning is applied for all the agents to learn the control concurrently, but this approach is not scalable since the reward function is too large to be enumerated [6]. In addition, all these approaches are incapable of managing dynamic traffic patterns dissimilar to the historical data, such as traffic accident or road maintenance.

In order to effectively respond to real-time traffic, close interactions between intersections are carried out to obtain the network-level traffic information and synchronize all the interactions over the network. Distributed constraint optimization problem (DCOP) is applied to formalize the synchronization of different intersections [15]. However, this approach is based on a centralized mediator to refer for all intersections in a given mediation session, which suffers communication and computation bottleneck when the network scales up. ADOPT and DPOP models are also carried out with either huge communication overload or retarded system response time [5].

Some other approaches relax these interactions and synchronize only within local intersections to obtain a good policy. For example, Phase-by-phase system is developed to optimize traffic for local intersections by predicting the traffic flow merely from their neighbors, but this research is based on the assumption that the vehicles to an intersection follow a Poisson process, which is not always the case in the real domain [9]. Xie et al. model the urban traffic control as a synchronously operating scheduling problem [10]. Each intersection agent estimates its future traffic from its upstream neighbors to obtain a myopic projection if the traffic flow is characterized as a cluster sequence, which is inaccurate in real domain with unpredictable vehicle behaviors.

3. Problem Description

A given traffic network can be modeled as an undirected graph $G(V, E)$ shown in Figure 1, where V is the set of intersections and E is the set of roads between intersections. For any two intersections v_i and v_j, $\langle v_i, v_j \rangle \in E$ represents that there is a road that vehicles can get through between v_i and v_j, and v_i and v_j are neighbors. Specifically, $n(v_i)$ is defined as all the neighbors of the intersection v_i, and $|n(v_i)|$ is the number of roads connected to intersection v_i. For example,

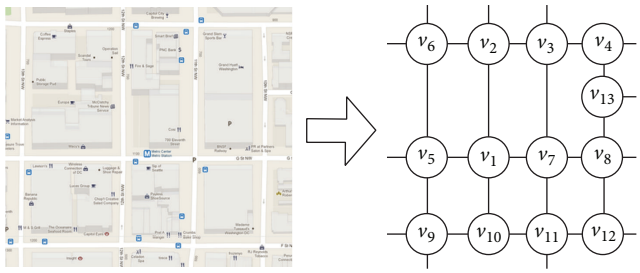


Figure 1: Models of intersections in a given urban area.

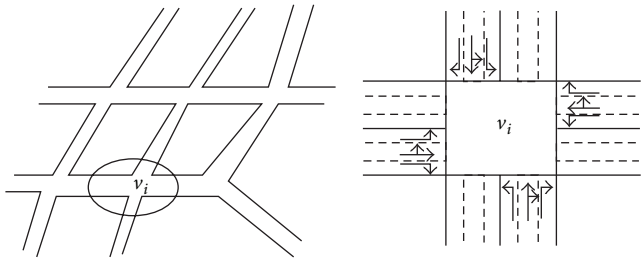


Figure 2: Model of intersections in a transportation network.

in Figure 1, $|n(v_1)| = 4$ and $|n(v_4)| = 3$. For each road $\langle v_i, v_j \rangle$, there are $|n(v_i)| - 1$ turning directions toward v_i, and each of them is called a lane. In Figure 1, there are 3 lanes of each road connecting to v_1 and 2 lanes each to v_{13}. The total number of lanes in v_i can be calculated as $|n(v_i)| \times (|n(v_i)| - 1)$. To be easily referred, the lane from intersection v_j to intersection v_k going through v_i is written as $Ln_i^j(k)$, $v_k \in \{n(v_i) - v_j\}$.

In each intersection, there are a number of vehicles stopping and waiting to go through. We assume that the number of vehicles waiting at lane $Ln_i^j(k)$ at time t, written as $Tf_i^j(k,t)$, can be detected by video cameras installed at each intersection. Because it is unable to detect the waiting traffic one time ahead, $Tf_i^j(k,t+1)$ is unobservable.

As shown in Figure 2, in an intersection, each traffic light takes an action to control a specific lane. Specifically, at time t the action to control the vehicles on the lane directed from v_j to v_k at v_i is written as $£_i^j(k,t) \in \{green, red\}$. Please note that *yellow* is not assigned to $£_i^j(k,t)$ because *yellow* is a fixed interim status necessary from *green* to *red* and is not an independent action.

At each horizon, we model the transition function under the action of each traffic light $£_i^j(k,t)$ as the waiting vehicles transferred from $Tf_i^j(k,t)$ to $Tf_i^j(k,t+1)$. It denotes that after each lane takes an action, waiting traffic gets through the intersection and new traffic arrives, and the waiting traffic is transferred to $Tf_i^j(k,t+1)$.

Since green intelligent control is an intention that a series of traffic lights are coordinated to allow continuous traffic flow over several intersections in main directions [1], its control should allow more vehicles to go through the intersections with the least delay. To maximize the moving vehicles, the sum of vehicles in the waiting queues should be minimized. Therefore, the key is to minimize waiting vehicles in the next time step other than the myopic optimization in the current time. We define the utility function toward the intelligent control for the global transportation network as

$$\mathrm{EU}(G,t) = \sum_{v_i \in V} \sum_{v_j \in n(v_i)} \sum_{v_k \in \{n(v_i) - v_j\}} Tf_i^j(k, t+1). \quad (1)$$

In this formula, the expected utilities of the cooperative transportation network G at time t (defined as $\mathrm{EU}(G,t)$) are the sum of the waiting vehicles in front of intersections at time step $t+1$.

Based on the expected utility function, the goal of cooperative traffic light control over the network G is to find an optimal joint policy π^* for traffic lights coordinating all the intersections so that the expected utility could be minimized. Consider

$$\pi^*(t) = \underset{Jt\_act(G,t)}{\operatorname{argmin}} \mathrm{EU}(G,t), \quad (2)$$

where $Jt\_act(G,t)$ in the transportation network G at time t consists of all the independent activities of each lane over every intersection at that time. It can be written as

$$Jt_{act(G,t)} = \bigcup_{v_i \in V} \bigcup_{v_j \in n(v_i)} \bigcup_{v_k \in \{n(v_i) - v_j\}} £_i^j(k,t). \quad (3)$$

Since $Tf_i^j(k,t+1)$ is not observable, finding the optimal policy π^* is intrinsically unsolvable in a large-scale transportation system [16].

4. Multiagent Based Traffic Control System

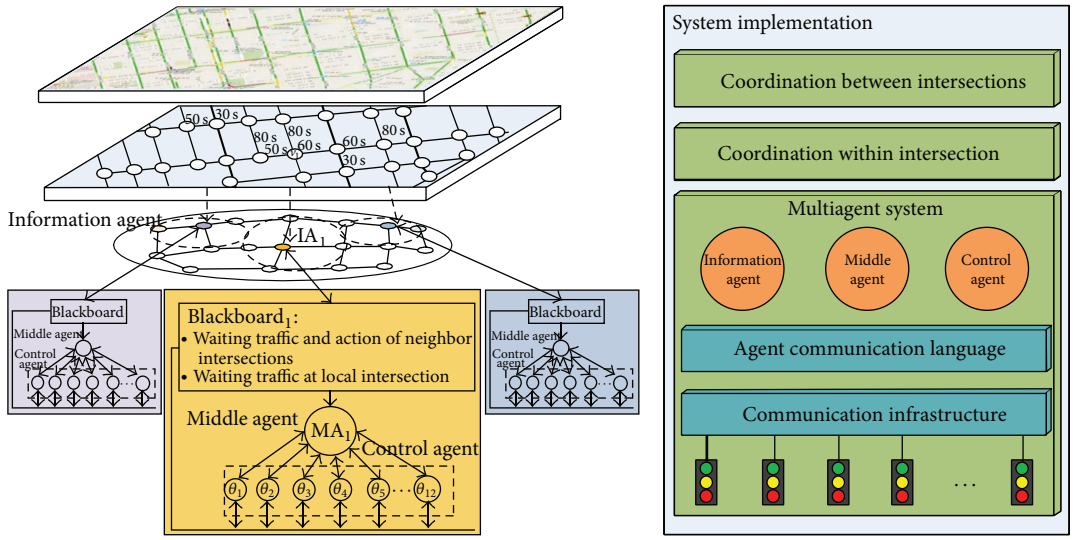
To solve the problem above, the key is to make the decisions to carry out the activities based on the prospection of local intersection state one time ahead. Although neither observation of the local state one time ahead nor accessing the global states of whole transportation network for inference is feasible, the estimation to the next local state of each intersection can still be made according to the following key characters in physical transportation networks.

(i) The state of the network at time step $t+1$ is solely determined by its state at time step t.

(ii) For any intersection v_i, its traffic flow is solely determined by the traffic flow arriving from its adjacent intersections.

Therefore, only the traffic in local intersection and the neighbors can affect its next coordination decision. If close interaction between adjacent intersections can be made to share their local traffic states and inform neighbors in advance of the actions that they are going to carry out, we can build a decentralized traffic coordination to solve this urban traffic optimization problem in a myopic view.

4.1. Decentralized Multiagent System Design. According to this, our multiagent system design is illustrated as Figure 3(a). Based on the analysis in Section 3, the urban traffic network



(a) Multiagent traffic light coordination architecture design (b) System implementation

Figure 3: Decentralized multiagent based traffic light coordination architecture and implementation.

is modeled as a graph. In the graph, each intersection is represented as a node and the road between them has been abstracted as a link. In the next level, since only the traffic of the local intersection and its adjacent intersections can affect its next coordination decision, each intersection has to share their states with their neighbors. For each intersection, we setup an information agent to share the states of the traffic of their lanes with the neighbors as well as receiving and publishing the adjacent intersection states to be used for the intersection's local decisions.

In the bottom level, control agents are responsible to solve their joint optimized decisions within the intersection. To achieve this, the multiagent system consists of three parts. As the first part, a blackboard is used to gather the information provided by its information agent, which are (1) current states of waiting traffic and traffic control actions from each adjacent intersection and (2) the states of waiting traffic of its own intersection. As the second part, each lane within the intersection is represented by a control agent to make decisions for its own lane and negotiate with the other control agent to solve the conflicts of different traffic flows. As the third part, a middle agent is built to coordinate local control agents to reach optimized joint decisions.

The system implementation of our approach is illustrated by Figure 3(b). We use the RETSINA platform to implement the multiagent system with the three basic agents as well as their agent communication language. In addition, we build algorithms for the agents to achieve green intelligent traffic coordination. Between intersections, since an intersection's decision is only determined by the states of its local neighbors, we model the decision as a Markov decision process. Within each intersection, the decision process of the control agents is modeled as a constraint optimization problem (COP). In the rest of this section, we will introduce in detail the agent and algorithm designs.

4.2. Information Agent. In order to coordinate with neighbor intersections and gain the information required for decision, information agent IA_i is built for intersection v_i and we have $IA = \{IA_1, IA_2, \ldots IA_i, \ldots\}$, $|IA| = |V|$. Inherited from the transportation network, the logical network of the information agents follows the same connection of G and agent IA_i has a set of neighbor information agents $n(IA_i) = \{IA_j \mid v_j \in n(v_i)\}$, which represents the information agents of v_i's adjacent intersections.

At time t, agent IA_i is able to gain the local state $Tf(v_i, t)$ of intersection v_i:

$$Tf(v_i, t) = \bigcup_{v_j \in n(v_i)} \bigcup_{v_k \in \{n(v_i) - v_j\}} Tf_i^j(k, t). \quad (4)$$

To make a rational decision toward decentralized traffic optimization, the intersection should know the traffic flows released from neighbor intersections before they arrive. In this case, the agent IA_i has to gain a complete view of the local state of v_i and the actions of their neighbors' incoming action in advance. Hence the state of v_i is defined as follows:

$$S_{v_i}(t) = Tf(v_i, t) \bigcup_{v_j \in n(v_i)} Tf(v_j, t) \bigcup_{v_j \in n(v_i)} £_j(t), \quad (5)$$

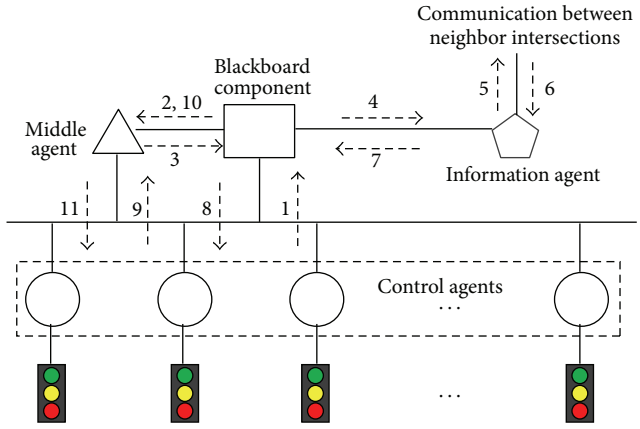
where $£_j(t)$ is the joint action of all the traffic lights, $£_j(t) = \bigcup_{v_l \in n(v_j)} \bigcup_{v_m \in \{n(v_j) - v_l\}} £_j^l(m, t)$, and $£_j^l(m, t)$ is the action for the traffic light controlling lane $Ln_j^l(m)$ at time t. As Figure 4 shows, the local state obtained by the information agent is published on the blackboard for the agents within the intersection.

In formula (5), the state of intersection v_i is composed of the waiting traffic at the local intersection, those of its adjacent intersections as well as their actions to be carried

out at t. Since the actions of the neighbors to be carried out at t should be shared in advance to gain the complete view of $S_{v_i}(t)$, IA_i as a neighbor of the other agents should have to share its action in advance as well. To break this deadlock, a practical protocol has been designed to estimate its action with $£_i^+(t)$. The details of the decentralized coordination model of v_i will be presented in Section 5.1. In addition, another key of this model is how to infer v_i's next-time local state $Tf(v_i, t+1)$ from its current local state, which will be addressed in Section 5.3.

4.3. Middle Agent and Control Agent. Within each intersection, there are conflicts between different traffic lights. Although it is possible for v_i to make a decision for all the traffic lights within the intersection based on the fixed conflict-free traffic signal phases, it is not flexible to deal with unpredictable traffic patterns. Therefore, we build a middle agent and a set of control agents for each intersection. As Figure 4 illustrates, a control agent $\theta_i^j(k)$ is built for each traffic light to monitor and control traffic on lane $Ln_i^j(k)$. Before each round of decision, it monitors the waiting traffic on the lane and publishes it on the blackboard. After the local state is obtained by the information agent and published on the blackboard, all the control agents in the intersection work together to eliminate the conflict actions and provide local optimized activities $£oc_i(t)$ for the middle agent. The middle agent MA_i for intersection v_i is responsible to initialize and coordinate all these control agents within the intersection and also choose the best joint action from the local optimized activities $£oc_i(t)$.

4.4. Interactions between Agents. With the two-level multiagent system, their decentralized coordination is able to be achieved through the information processing process shown in Figure 4. Firstly, the control agent monitors the waiting traffic at the lane and publishes it on the blackboard. Then the middle agent is able to estimate its current action $£_i^+(t)$. In order to gain the local state, the information agent shares its waiting traffic and expected action with neighbors. After the local state is obtained and published on the blackboard, the control agents are able to cooperatively generate the conflict-free local optimized joint actions $£oc_i(t)$ and the middle agent can choose the best action from $£oc_i(t)$.



(1) Current waiting traffic at each lane $Tf_i^j(k,t)$
(2) Current waiting traffic at the intersection $Tf(v_i,t)$
(3) Estimated action $£_i^+(t)$
(4, 5) Waiting traffic and estimated action $\langle Tf(v_i,t), £_j^+(t)\rangle$
(6) Information $\langle Tf(v_j,t), £_j^+(t)\rangle$ from neighbor v_j
(7, 8, 10) Local state $S_{v_i}(t)$
(9) Local optimized action $£oc_i(t)$
(11) Optimal action $£oc_i^*(t)$

Figure 4: Information processing between agents.

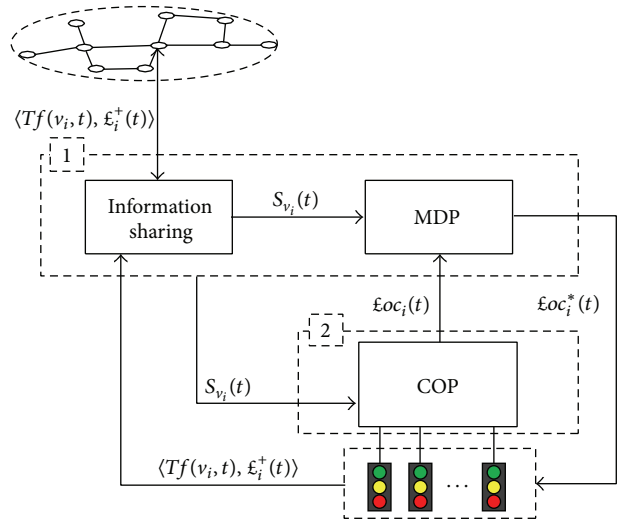


Figure 5: Coordination model of local intersection.

5. Intelligent Traffic Control Algorithm

Based on our multiagent system design, we build algorithms for these agents to generate the optimal action $£oc_i^*(t)$ according to state $S_{v_i}(t)$ and solve the constraints within the intersection. An overview of our traffic control algorithm is illustrated in Figure 5. For each intersection, before making the decision, it has to share its state of local waiting traffic and expected actions. Since the intersection's next time state can be maintained with its information agent, a Markov decision process can be built for the decentralized traffic control as a conflict-free localized optimal action set. Secondly, we model this coordination within the intersection as a constraint optimization problem (COP). In order to solve these models, the key is to figure out the transition from the current local state $Tf_i^j(k,t)$ to its next local traffic state $Tf_i^j(k,t+1)$ one time ahead. Finally, we propose a heuristic algorithm for the local state transition function.

5.1. Coordination between Intersections. In this section, we present the details on how agents in different intersections

coordinate with each other, and what information is shared, so as to make their decentralized actions toward optimized decentralized traffic control. The key is that agents build their own local states to infer the states of their local intersections one time ahead. For a specific middle agent MA_i, its decentralized control process is built as a decision theoretical model $\langle S_{v_i}, £oc_i, T, Tf(v_i)\rangle$.

(i) State: $S_{v_i}(t)$ is the intersection v_i's local state at t. It is built from the information received from their neighbors as well as the observation to the local traffic state.

(ii) Action: as explained in Section 3, agent MA_i is only able to choose the action from the available joint activities set $£oc_i(t)$ worked out by its local agents Θ_i.

(iii) Transition function $T : S \rightarrow Tf(v_i)$ defines the transition from state $S_{v_i}(t)$ to the local traffic state of v_i at $t+1$ by the action.

(iv) Utility function is defined as $Tf(v_i, t+1)$.

MA_i is to find its optimal policy π^{Δ}:

$$\pi^{\Delta} = \underset{£oc_i(t)}{\operatorname{argmin}} Tf\left(v_i, t+1\right). \tag{6}$$

A key challenge of this model is that information agent IA_i has to share its current joint action of all the traffic lights in the interaction with its neighbors one time ahead. And these joint actions of the neighbors are critical to build $S_{v_i}(t)$, which in turn influences their own decision. Therefore, the agent has to make its decision based on sharing its decision result one time ahead and this looping process produces a deadlock.

To break this deadlock, from formula (5), we observe that a heuristic protocol can be designed because adjacent intersections' joint actions only contribute a small portion to the local state. Therefore, an estimated joint action $£_i^+(t)$ to $£_i(t)$ can be defined to solve this deadlock, and formula (5) can be represented as

$$S_{v_i}(t) = Tf\left(v_i, t\right) \bigcup_{v_j \in n(v_i)} Tf\left(v_i, t\right) \bigcup_{v_j \in n(v_i)} £_j^+(t). \tag{7}$$

$£_i^+(t)$ can be estimated in many ways; a practical way is to estimate $£_i^+(t)$ one time ahead according to the local state $Tf(v_i, t)$. Therefore, we design a practical algorithm as

$$Tf\left(v_i, t\right) \longrightarrow £_i^+(t), \tag{8}$$

where the action is determined by the number of waiting traffics of each lane in the intersection. Although it may be imprecise to estimate $£_i^+(t)$ based on historical states, considering that the traffic moves continuously, $£_i^+(t)$ cannot be significantly varied from $£_i(t)$.

5.2. Coordination within Intersection. In this section, we present how the control agents Θ_i in an intersection v_i are coordinated by middle agent MA_i to build the local conflict-free joint activities set $£oc_i(t)$. In order to solve the constraints

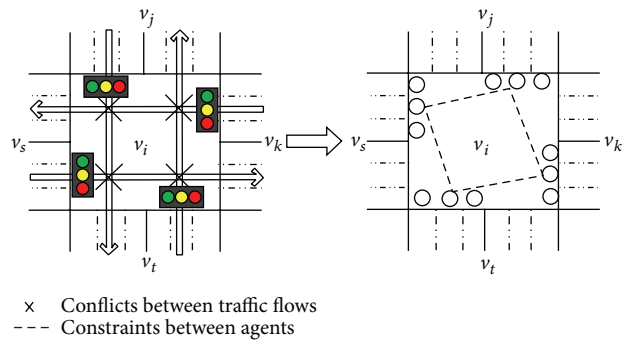


Figure 6: Traffic conflictions in a four-way intersection.

within the intersection, the agents have two policies: fixed policy and dynamic policy. In the fixed policy, $£oc_i(t)$ contains all the conflict-free joint actions, which are predefined. In the dynamic policy, the middle agent gives an order to the control agents and each agent proposes its preferred action according to the order.

In this paper, we focus on the dynamic policy design, and the middle agent has to work out a set of orders for the control agents as their social conventions and start the negotiation from one control agent at each round to get a local joint action. There are also two ways to generate each of the order: random ordering and heavy traffic lane first. Random ordering initializes the order of control agents randomly. On the other way, the control agent with more waiting traffic is given a higher priority.

Following a given order, all the control agents have to coordinate to work out the local optimized conflict-free joint policy. The coordination process of all the agents in Θ_i is built as a constraint optimization model $\langle £_i, \{green, red\}, C_i\rangle$.

(i) $£_i$ defines the variables of COP, where each variable is the traffic light action of $£_i^j(k,t)$ worked out by $\theta_i^j(k)$.

(ii) Each $£_i^j(k,t)$ is only chosen from a binary set $\{green, red\}$.

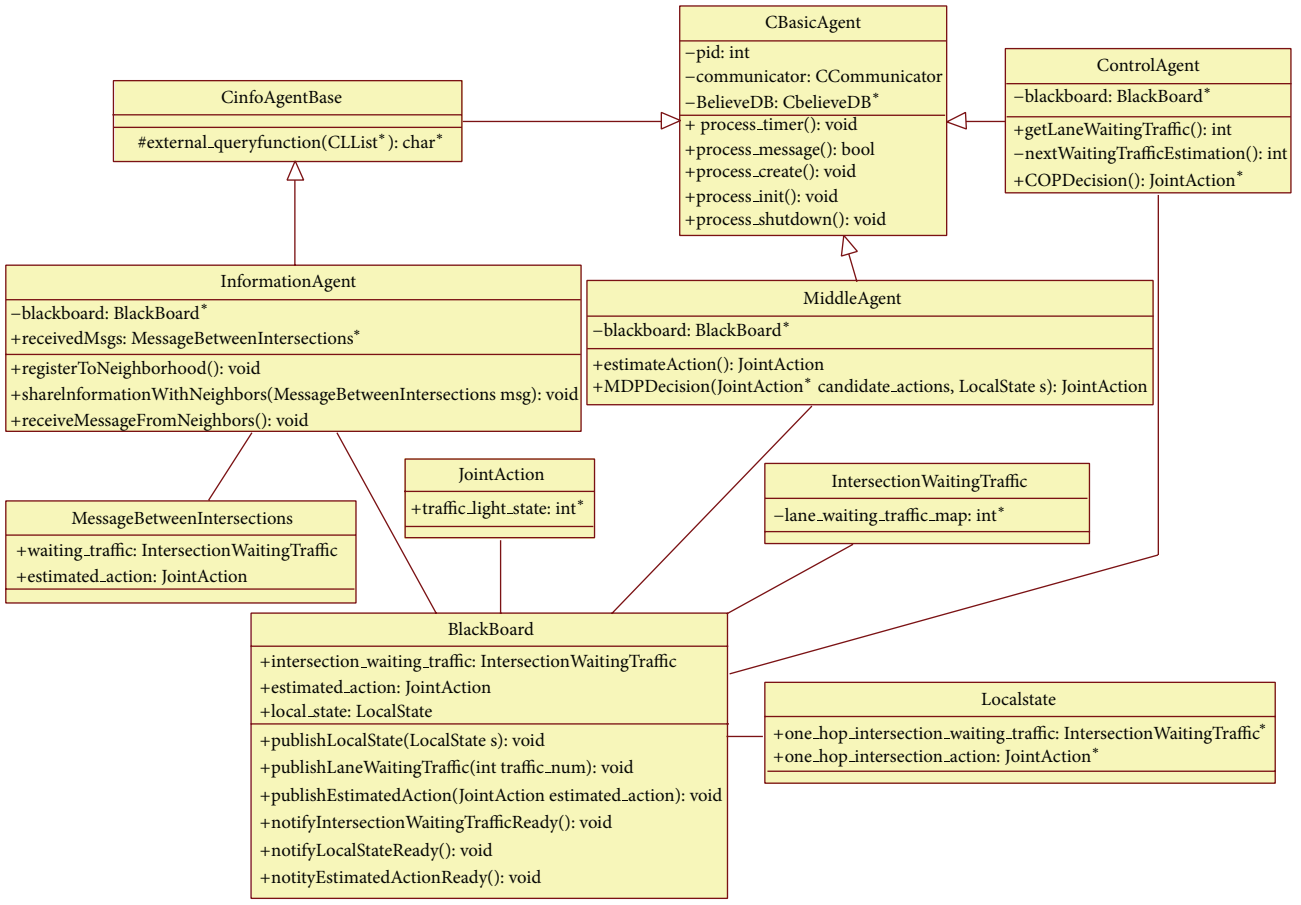
(iii) C_i is the binary constraint set predefined for v_i by domain. For example, typical traffic conflictions in a four-way intersection can be illustrated in Figure 6, where any variables connected with a dash line cannot be *green* at the same time.

The utility function for each assignment of variable $\theta_i^j(k)$ can be formulated to help agent $\theta_i^j(k)$ to locally minimize the waiting traffic of its lane:

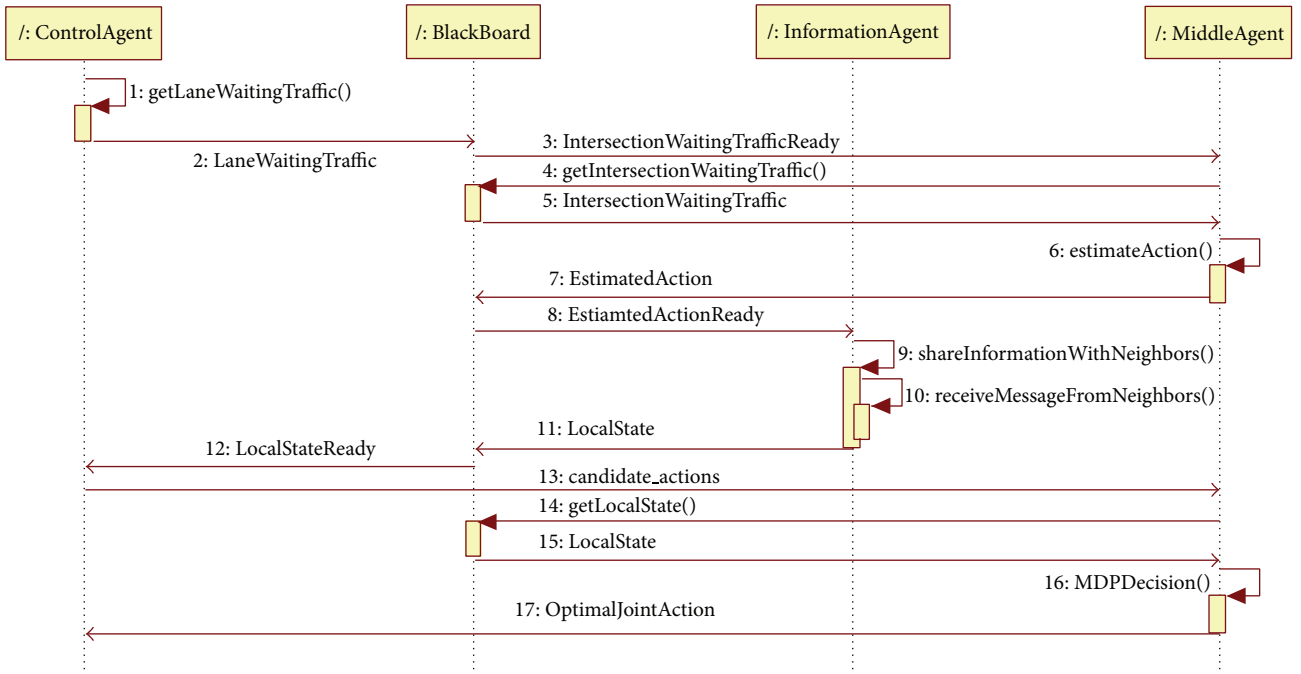
$$\mathrm{EU}\left(Tf_i^j(k), £_i^j(k,t)\right) = Tf_i^j(k, t+1). \tag{9}$$

For each agent $\theta_i^j(k)$, its optimization policy is

$$\pi^{\Delta\Delta} = \underset{£_i^j(k,t)}{\operatorname{argmin}} Tf_i^j(k, t+1). \tag{10}$$



(a) Class diagram



(b) Sequence diagram

Figure 7: Software engineering illustration of the multiagent system design.

(1) $order_i = M_i.generateOrder(\Theta_i)$;
(2) $M_i.SetupSearches(£oc_i)$;
(3) **for all** $£oc_i[m] \in £oc_i$ **do**
(4) $\quad order_i.starter = RandomChoose(\Theta_i)$;
(5) $\quad$ **for all** $\theta_i^j(k) \leftarrow order_i.next()$ **do**
(6) $\quad\quad £oc_i^j(k,t) \leftarrow Lottery(\mathrm{Tf}_i^j(k,t+1), \mathrm{Tf}(v_k,t))$;
(7) $\quad$ **end for**;
(8) **end for**;
(9) **return** $£oc_i$;

Algorithm 1: Decision process of Θ_i within intersection v_i.

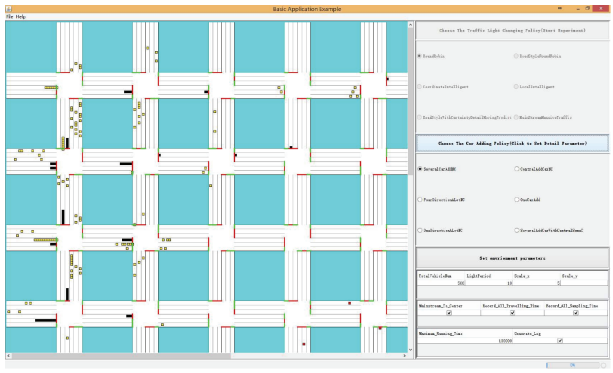


Figure 8: The screen shot of the urban traffic simulator.

According to this policy, we build a myopic heuristic mapping function as

$$£_i^j(k,t) \times Tf_i^j(k+1) \times \sum_{v_l \in n(v_k)-v_i} Tf_k^i(l,t) \longrightarrow [0,1]. \quad (11)$$

In this equation, if $\theta_i^j(k)$ is more likely to help to reduce $Tf_i^j(k+1)$, the agent $\theta_i^j(k)$ is more likely to set $£_i^j(k,t)$ as *green*. The pending downstream traffic $Tf_k^i(l,t)$ to the adjacent intersections should also be considered. Because if there is heavy traffic in the next intersection to go through, discharging more traffic there is not helpful and the lane is less likely to be set as *green*. In addition, to avoid any lane with very little waiting traffic being set as continuous *red*, we define that the continuous red phrase should not exceed $R_{\max}$.

The decision process of local agents Θ_i is described in Algorithm 1. In this algorithm, MA_i initiates an order for the control agents in Θ_i (line 1) and sets up a set of searches for $£oc_i$ (line 2). For each search $£oc_i[m]$, it randomly starts from an agent (line 4), and each agent $\theta_i^j(k)$ sequentially chooses its action $£oc_i^j(k,t+1)$ according to formula (11) (line 6). Since agents' decisions are based on the probabilistic model, the optimal joint activities are not guaranteed in the searches. Therefore, we set the size of $£oc_i$ bigger than the number of lanes in v_i to increase the chance of optimization. However, it also increases the computation complexity for Θ_i.

5.3. Heuristic Transition Function. Although solving the scalable MDP for massive traffic lights control is mathematically feasible, the uncertainties on the state transition function resulted by unpredictable traffic in heavy traffic network will make the computation hard. There are three key factors:

(i) the unpredictable amount of traffic going through under a given green light;

(ii) the uncertainty of line choosing on the adjacent intersection when vehicles passed through a given intersection;

(iii) the unpredictable arriving time of given traffic arrived at the next intersection which depends on the congestion and road conditions as well as their distances between intersections.

These factors may vary significantly under different traffic conditions. For simplicity and clarity of our model, we make the following two assumptions. Firstly, we assume that the traffic flow getting through the intersection follows the exponential queue discharge flow rate model [17]. In this case, during a green-light cycle Δ_i, the maximum number of vehicles getting through an intersection is denoted as $h(\Delta_i)$. Therefore, at each time t, the number of vehicles getting through lane $Ln_i^j(k)$ is denoted as $Td_i^j(k,t)$, which could be estimated as If $£_i^j(k,t) = green$, during the green-light cycle Δ_i

$$Td_i^j(k,t) = \begin{cases} Tf_i^j(k,t), & \text{if } Tf_i^j(k,t) < h(\Delta_i) \\ h(\Delta_i), & \text{otherwise,} \end{cases} \quad (12)$$

otherwise, $£_i^j(k,t) = red$ and $Td_i^j(k,t) = 0$.

Secondly, the probability for each vehicle to choose the lane could be estimated from historical statistics. In this paper, we assume that vehicles will evenly choose the lanes after it gets through an intersection. Thus, the probability $P_i^j(k)$ is

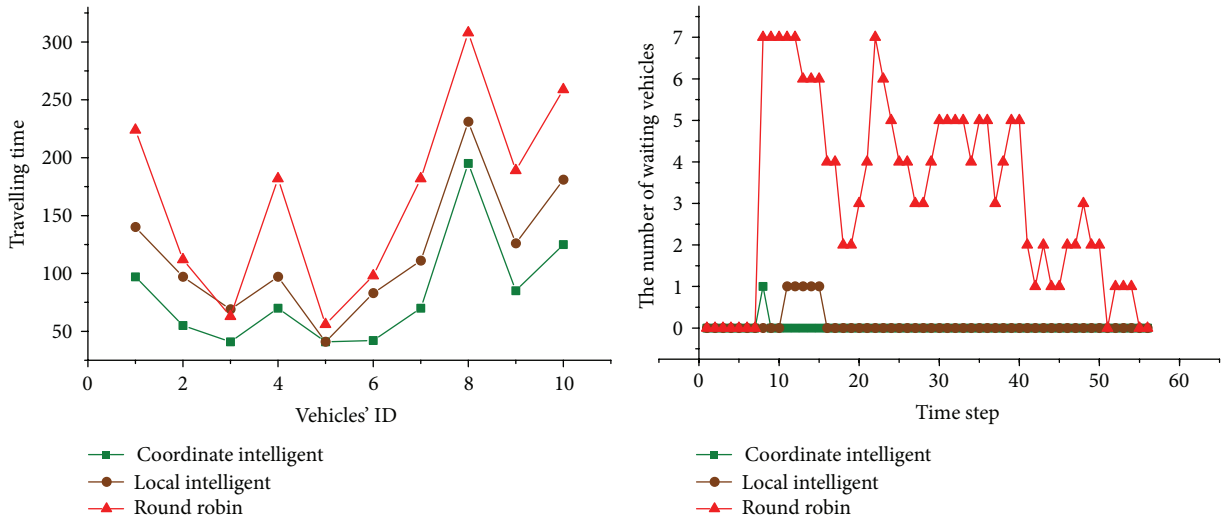
$$\forall v_k \in n(v_i), v_j \in n(v_i) - v_k, \quad P_i^j(k) = \frac{1}{|n(v_i)| - 1}. \quad (13)$$

With the assumptions above, we can establish the transition from $Tf_i^j(k,t)$ to $Tf_i^j(k,t+1)$. Observing that the number of vehicles on a given lane is determined by vehicles' choice of lanes and the number of vehicles released by adjacent intersections in the last cycle, we will have

$$P_i^j(k) \times \sum_{v_l \in n(v_j)-v_i} Td_j^l(i,t). \quad (14)$$

According to the real-time traffic condition, not all vehicles released from its adjacent intersections can arrive at the intersection at the end of the green phase. The number of vehicles arriving at intersection v_i on lane $Ln_i^j(k)$ is denoted as $Tc_i^j(k,t)$. If the weight of next road is $w_{i,j}$, a function to predict the arriving ones within the traffic light period Δ_i can be proposed as

$$Tc_i^j(k,t) = g\left(P_i^j(k) \times \sum_{v_l \in n(v_j)-v_i} Td_j^l(i,t), w_{i,j}\right). \quad (15)$$



(a) The travelling time of all the vehicles

(b) The number of waiting vehicles at each time

FIGURE 9: Results of green-wave effect.

(1) **for** $\mathrm{Ln}_i^j(k) \in v_i$ **do**
(2) **if** $\mathrm{Tf}_i^j(k,t) < h(\Delta_i)$ **then**
(3) $\mathrm{Td}_i^j(k,t) = \mathrm{Tf}_i^j(k,t)$;
(4) **else**
(5) $\mathrm{Td}_i^j(k,t) = h(\Delta_i)$;
(6) **end if**
(7) $P_i^j(k) = \dfrac{1}{|n(v_i)| - 1}$
(8) *calculate* $\mathrm{Tc}_i^j(k,t)$
(9) *calculate* $\mathrm{Tf}_i^j(k,t+1)$;
(10) **end for**
(11) **return** $\mathrm{Tf}(v_i, t+1)$;

ALGORITHM 2: Intersection v_i's transition process from current state to the local state one time ahead.

With the number of vehicles to be released and the ones to arrive at the intersection, the transition function to update the waiting vehicles at next time $Tf_i^j(k,t+1)$ can be computed as

$$Tf_i^j(k,t+1) = Tf_i^j(k,t) + Tc_i^j(k,t) - Td_i^j(k,t). \quad (16)$$

Algorithm 2 presents the process of state transition of intersection v_i from state t to the local state one time ahead $t+1$. For each lane $Ln_i^j(k)$, it firstly estimates the number of vehicles that can get through this intersection when the traffic light is set *green* according to formula (12) (line 2–6). Next, it calculates the probability of each line that the vehicles will choose (line 7). After the number of vehicles released from adjacent intersections and their lane choosing probability are figured out, the number of vehicles arriving at this lane at next time is able to be calculated according to formula (15) (line 8). Finally, according to the transition function (16), $Tf_i^j(k,t+1)$ could be solved (line 9). When all lanes' one time ahead states are estimated, the local state at $t+1$ is worked out (line 11).

6. Software Design

In this section, we present the multiagent system design as well as the information processing process. Our two-level agents are built based on RETSINA [18] for its advantage of the programming platform and multiagent coordination mechanism. RETSINA is developed by Robotics Institute of Carnegie Mellon University. It implements all the basic types of agents and the agent communication language (ACL) as well as the agent management service. In addition, RETSINA also provides a peer to peer interaction mechanism for the multiagent systems in a distributed infrastructure. It is implemented with C and C++, which could be easily encoded in embedded traffic control devices. In our multiagent system design, there are four key components.

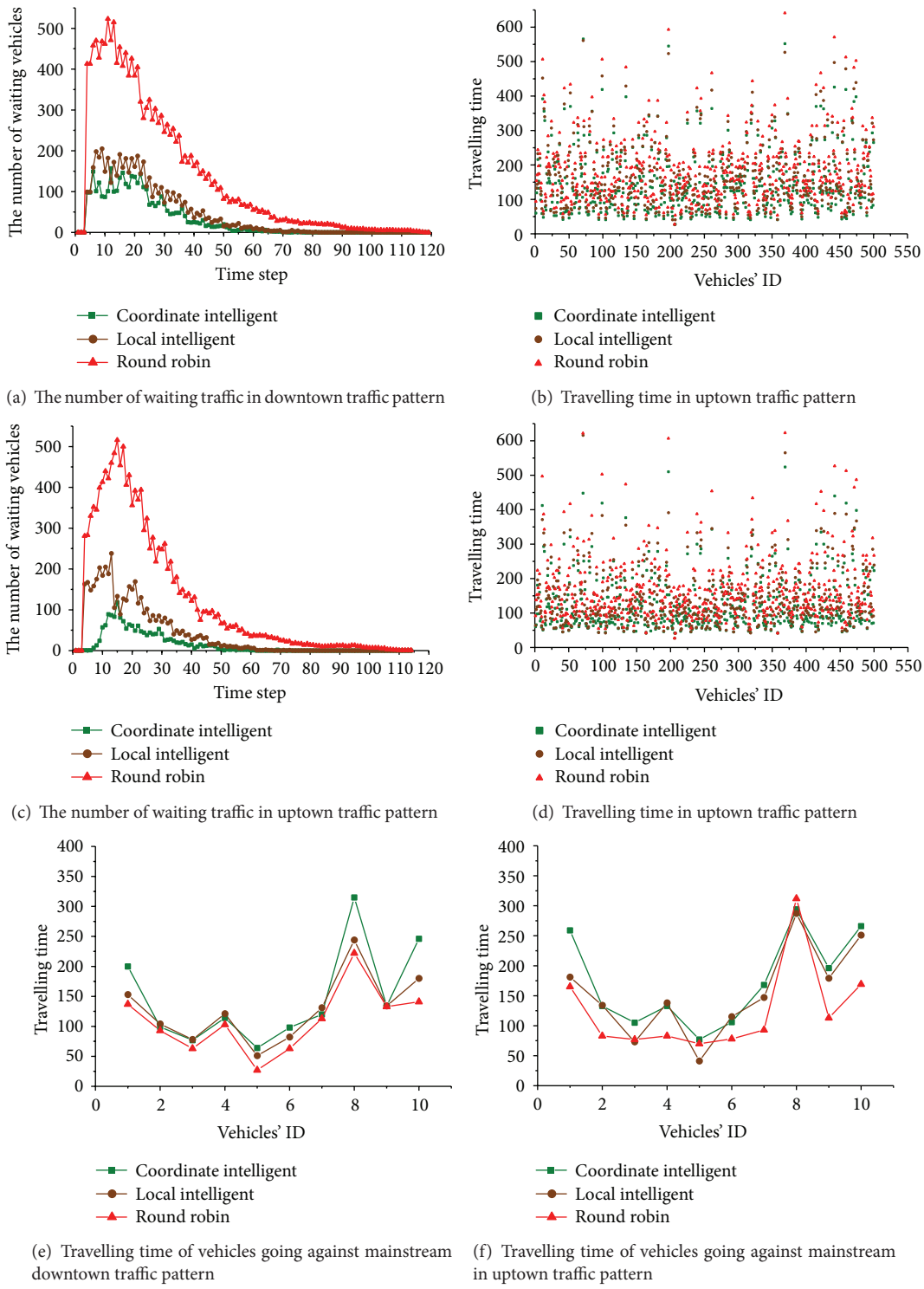
Information agent responds to share the local states with the other information agents of the adjacent intersections. It is implemented based on the RETSINA information agent, which carries out the specific task to communicate with other information agents by using ACL.

Control agent is customized to make decisions for each lane that it represents. It is based on the RETSINA task agent to carry out the information process described in Algorithm 2.

Middle agent is also built based on RETSINA basic task agent. It responds to generate the order for the control agents within the intersection and choose the best policy to achieve joint decentralized control.

Blackboard is a single instance for each intersection in the multiagent system. It is used to provide information publish service for all agents within the intersection.

The multiagent traffic control system designed is illustrated in Figure 7. In Figure 7(a), information agent inherits



(a) The number of waiting traffic in downtown traffic pattern

(b) Travelling time in uptown traffic pattern

(c) The number of waiting traffic in uptown traffic pattern

(d) Travelling time in uptown traffic pattern

(e) Travelling time of vehicles going against mainstream downtown traffic pattern

(f) Travelling time of vehicles going against mainstream in uptown traffic pattern

Figure 10: Results in two typical traffic patterns.

from RETSINA information agent. Both middle agent and control agent inherit from RETSINA task agent. All these agents have access to the blackboard component and the methods related to the cooperative decision. The interaction process in Figure 4 is presented as the sequence diagram illustrated in Figure 7(b). All the agents have their own threads and lifetime, and they interact with each other asynchronously. With the help of blackboard component, all these agents are able to publish and get the information critical to joint intelligent traffic coordination.

7. Simulation and Results

In order to manifest the feasibility of our approach, we build an abstract traffic simulator for evaluation. The screen shot of the urban traffic simulator is shown in Figure 8. In this simulator, we build a grid network with a number of intersections, where each intersection connects four adjacent intersections. Vehicles are simulated to go through the network but they have to wait in front of the intersections until the light of their lanes are green.

In our simulation, we load different control schemas to control the traffic so that we can compare the performance with our design. In our experiment, the route of each vehicle is randomly produced and we choose three different schemas to control the traffic: our design labeled as *coordinate intelligent*, local real-time traffic control policy labeled as *local intelligent,* and the traditional round-robin policy labeled as *round robin*. The local real-time traffic control policy (*local intelligent*) is based on the literature [15], whose objective is to release maximum waiting vehicles merely based on the local state of the interaction. We hypothesized that without the prospection one time ahead to build the intelligent control, *local intelligent* should perform worse than our approach. *Round robin* control policy assigns uniform time slices for all the conflict-free traffic signal phases, which are initialized in deployment. It schedules all these conflict-free traffic signal phases in periodic sequential order [19]. Since *round robin* schedules the traffic in a fixed control manner, ignoring the real-time traffic demand, it should perform the worst in unbalanced traffic. The results are compared with two criterions:

(i) the average traveling time for each vehicle to go through the transportation network;

(ii) the average number of waiting vehicles in front of intersections.

7.1. Traffic Control with Green-Wave Effect. When the traffic is sparse, the green-waved effect is the most straightforward way to test intelligent traffic control performance [20]. Therefore, we initialized a 5×5 grid transportation network and 10 vehicles are randomly generated from the margin of this network to get through under the three control schemas. The experimental results are shown in Figure 9. When there are only few vehicles in the network, in *Round Robin* schema, the vehicle not catching up green light has to wait although no vehicle goes the "green" lane. Because of one time ahead intelligent control, *coordinate intelligent* is easier to create the green wave effect that allows the coming vehicles to get through the intersection without stops. Therefore, both the waiting time and the number of waiting vehicles in the *coordinate intelligent* schema are the least in both Figures 9(a) and 9(b).

7.2. Traffic Control in Different Traffic Patterns. In urban transportation system, there are two typical traffic patterns [21]. The downtown rush hour emerges in the morning when a lot of vehicles are driven towards the town for work, while the uptown rush hour emerges in the afternoon when massive vehicles go the other way.

We test our approach under these two typical traffic patterns. In this experiment, we initialized a 5×5 grid network. In the downtown traffic pattern, we simulated 500 vehicles, which are evenly generated from the margin of the network, driving toward downtown within the initial 10 time steps, and as shown in Figure 10(a), the number of waiting vehicles reaches a peak between time steps 10 and 20 as they overload the intersections in the grid. Each traffic control schema is applied to route traffic to their destination and decrease the waiting traffics. Traditional *round robin* way, which only works well in handling balanced traffic, has a poor performance in responding to the unbalanced traffic. As expected, with the one time ahead intelligent control, our approach performs best to have the mainstream traffic flows get through the intersections quickly. Both the average traveling time of those vehicles and the number of waiting vehicles in this schema stay the least as shown in Figure 10(b). In the next section, we also simulated 500 vehicles in an uptown traffic pattern, evenly generated from downtown, driving to spread out of the network. Similar to Figures 10(a) and 10(b), our schema works best in Figures 10(c) and 10(d).

Since the intelligent traffic control should always try to have the mainstream traffic flows go through the intersections with high priorities, we test whether this is the case in our design. In the two traffic patterns, we put 10 vehicles to be driven against the mainstream. Figure 10(e) shows the average traveling time of the 10 vehicles to go uptown while there are 500 vehicles going downtown. Figure 10(f) shows the result of the 10 vehicles to go downtown while the 500 vehicles are going uptown. As expected, to evacuate the heavy traffic, both *local intelligent* and *coordinate intelligent* schemas have to give higher priorities to the mainstream and sacrifice the minority from the other direction. Therefore, as Figures 10(e) and 10(f) indicate, it costs more traveling time for the 10 vehicles to get through the network. Besides, the results also show that the evacuation ability of *coordinate intelligent* is higher than *local intelligent*.

7.3. Traffic Control in Different Network Scales. In order to test the scalability of our intelligent control, we perform experiments in different scales of transportation networks. We initialized 500 vehicles in the network with two traffic patterns described in Section 7.2. As shown in Table 1, when the scale of the network increases, it takes the vehicles more time to travel through this network. Due to the intelligent traffic control of our approach, it outperforms the other schemas.

7.4. Traffic Control in Different Traffics. In this section, we test our approach with different number of traffics. We initialized a 5×5 grid network with 500 to 2500 vehicles to go through. As shown in Table 2, in two typical traffic patterns, heavy traffic is more likely to cause congestions with longer average traveling time for each vehicle. However, our approach performed best.

Table 1: The average traveling time of different traffic controls in different network scales.

	Scale	Coordinate intelligent	Local intelligent	Round robin
Downtown traffic pattern	5 × 5	150.3	155.7	177.1
	6 × 6	165.5	182.1	210.1
	7 × 7	183.9	195.0	237.9
	8 × 8	216.5	219.0	255.0
	9 × 9	219.5	233.4	295.1
	10 × 10	241.8	252.8	327.0
Uptown traffic pattern	5 × 5	121.7	146.9	166.6
	6 × 6	177.0	206.4	236.0
	7 × 7	195.5	237.7	269.6
	8 × 8	257.7	298.9	336.7
	9 × 9	288.1	334.2	394.2
	10 × 10	359.6	401.2	478.2

Table 2: The average traveling time of different traffic controls with different number of vehicles.

	Vehicle number	Coordinate intelligent	Local intelligent	Round robin
Downtown traffic pattern	500	149.7	160.3	175.7
	1000	179.7	186.6	208.3
	1500	222.6	230.2	240
	2000	243.1	257.1	270.1
	2500	241.4	281.1	293.9
Uptown traffic pattern	500	128.7	142.5	167.3
	1000	138.8	181.9	221.2
	1500	164.1	219.9	277.9
	2000	180.4	240.9	328.7
	2500	201.8	280.9	382.6

8. Conclusion and Future Work

In this paper, we presented a multiagent based decentralized traffic light coordination approach for large urban transportation system. In order to improve the control efficiency, we use the prospection of local state one time ahead to make rational decision and build a two-level multiagent architecture and intelligent traffic control algorithms to coordinate these agents. Experiments manifest that our approach is feasible and scalable to improve the decentralized traffic control efficiency.

Although we are capable of dealing with some of the challenges, we leave many of the others in the future. Firstly, in our model we primarily considered video cameras as input sensors; however, more sensors should be considered as valuable inputs. Although those sensors are helpful to refine the model, as a challenge, they may also bring heavy computation. Secondly, traffic flow estimation methods should be polished to improve the efficiency. Moreover, deployment in real domain is the key to evaluate our approach.

Conflict of Interests

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

Acknowledgments

This research has been sponsored in part by National Natural Science Foundation of China nos. 61370151 and 61202211, National Science and Technology Support Program of China 2012BAI22B05, and Huawei Research Foundation YB2013120141.

References

[1] M. Ángel Sotelo, J. W. C. van Lint, U. Nunes et al., "Introduction to the special issue on emergent cooperative technologies in intelligent transportation systems," *IEEE Transactions on Intelligent Transporation Systems*, vol. 13, no. 1, pp. 1–5, 2012.

[2] A. Warberg, J. Larsen, and R. Jorgensen, "Green wave traffic optimization—a survey," DTU Technical Report, 2008.

[3] S. Heiko and B. Klemens, "Agent-based traffic control using auctions," in *Cooperative Information Agents XI*, pp. 119–133, 2007.

[4] C. Diakaki, M. Papageorgiou, and K. Aboudolas, "A multivariable regulator approach to traffic-responsive network- Wide signal control," *Control Engineering Practice*, vol. 10, no. 2, pp. 183–195, 2002.

[5] R. Junges and A. L. C. Bazzan, "Evaluating the performance of DCOP algorithms in a real world, dynamic problem," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 599–606, ACM, 2008.

[6] 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.

[7] B. C. Silva, D. Oliveira, A. L. C. Bazzan, and E. W. Basso, "Adaptive traffic control with reinforcement learning," in *Proceedings of the 4th Workshop on Agents in Traffic and Transportation*, pp. 80–86, ACM, 2006.

[8] D. I. Robertson and R. D. Bretherton, "Optimizing networks of traffic signals in real time—the SCOOT method," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 1, pp. 11–15, 1991.

[9] M. Shenoda and R. Machemehl, "Development of a phaseby-phase, arrival-based," Tech. Rep., SWUTC, 2006.

[10] X.- F. Xie, S. F. Smith, and G. J. Barlow, "Schedule-driven coordination for real-time traffic network control," in *Proceedings of the 22nd International Conference on Automated Planning and Scheduling (ICAPS '12)*, pp. 323–331, June 2012.

[11] TRANSYT-7F, *TRANSYT-7F User's Manual*, Transportation Research Center, University of Florida, 1988.

[12] J. Ren, X. Ou, Y. Zhang, and D. Hu, "Research on networklevel traffic pattern recognition," in *Proceedings of the IEEE Intelligent Transportation Systems*, pp. 500–504, 2002.

[13] B. C. da Silva, E. W. Basso, A. L. C. Bazzan, and P. M. Engel, "Dealing with non-stationary environments using context detection," in *Proceedings of the 23rd International Conference on Machine Learning (ICML '06)*, pp. 217–224, June 2006.

[14] B. Abdulhai, R. Pringle, and G. J. Karakoulas, "Reinforcement learning for true adaptive traffic signal control," *Journal of Transportation Engineering*, vol. 129, no. 3, pp. 278–285, 2003.

[15] D. de Oliveira, A. L. C. Bazzan, and V. Lesser, "Using cooperative mediation to coordinate traffic lights: a case study," in *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 463–470, ACM, Utrecht, The Netherlands, July 2005.

[16] D. Greenwood, B. Burdiliak, I. Trencansky, and H. Armbruster, "GreenWave distributed traffic intersection control," in *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 1413–1414, ACM, 2009.

[17] R. Akçelik and M. Besley, "Queue discharge flow and speed models for signalised intersections," in *Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, pp. 99–118, 2002.

[18] K. Sycara, M. Paolucci, M. Van Velsen, and J. Giampapa, "The retsina mas infrastructure," in *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 29–48, ACM, 2003.

[19] Y. Zhang, Y. Xu, T. Sun, and P. Liu, "Green-waved cooperative coordination algorithm for decentralized traffic control," in *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT '12)*, pp. 75–79, Macau, China, December 2012.

[20] A. Warberg, J. Larsen, and R. M. Jorgensen, "Green wave traffic optimization—a survey," Tech. Rep., Technical University of Denmark, 2008.

[21] S. Guo, J. Tan, J. Duan et al., "Characteristics of atmospheric non-methane hydrocarbons during haze episode in Beijing, China," *Environmental Monitoring and Assessment*, vol. 184, no. 12, pp. 7235–7246, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 230537, 7 pages
http://dx.doi.org/10.1155/2014/230537

Research Article

Modeling Complex System Correlation Using Detrended Cross-Correlation Coefficient

Keqiang Dong,[1] Hong Zhang,[2] and You Gao[1]

[1] *College of Science, Civil Aviation University of China, Tianjin 300300, China*
[2] *Renai College, Tianjin University, Tianjin 300300, China*

Correspondence should be addressed to Keqiang Dong; hongzhangdong@163.com

Received 6 April 2014; Revised 19 June 2014; Accepted 1 July 2014; Published 15 July 2014

Academic Editor: X. Zhang

The understanding of complex systems has become an area of active research for physicists because such systems exhibit interesting dynamical properties such as scale invariance, volatility correlation, heavy tails, and fractality. We here focus on traffic dynamic as an example of a complex system. By applying the detrended cross-correlation coefficient method to traffic time series, we find that the traffic fluctuation time series may exhibit cross-correlation characteristic. Further, we show that two traffic speed time series derived from adjacent sections exhibit much stronger cross-correlations than the two speed series derived from adjacent lanes. Similarly, we also demonstrate that the cross-correlation property between the traffic volume variables from two adjacent sections is stronger than the cross-correlation property between the volume variables of adjacent lanes.

1. Introduction

Many diversified complex systems are composed of constituents that mutually interact in a complex fashion. The complexity of the mutual interaction, such as the output of each constituent which depends not only on its own past but also on the past values of other constituent outputs, can be additionally studied if memory is included. Such complex systems are characterized by both long-range correlations and long-range cross-correlations. A number of studies suggest the existence of these properties in diverse systems. Applying the random matrix theory, Stanley et al. demonstrated the cross-correlation properties between individual stocks traded in the Korean stock market [1]. By analyzing 48 world financial indices, Wang et al. found the long-range power-law cross-correlations in the absolute values of returns [2]. Podobnik et al. studied the cross-correlation in successive differences of air humidity and air temperature [3]. Du et al. provided cross-correlation time delay model to improve earthquake relocation forecasts [4].

These studies provide strong empirical evidences for the existence of cross-correlations between the dynamics of natural systems. Pearson's correlation coefficient (PCC), which is used to represent the linear correlation between two time series which are both assumed to be stationary [5, 6], is commonly used to gain insight into the dynamics of cross-correlations in time series. Nevertheless, in natural systems, the nonlinear and nonstationary characteristics are usually present [7, 8]. Therefore, PCC may not be suitable to describe the cross-correlations between time series that are nonlinear or nonstationary. To address the drawbacks of PCC, the detrended cross-correlation analysis (DCCA) method is employed in this paper.

The DCCA method, which is a modification of standard covariance analysis in which the global average is replaced by local trends [9, 10], was proposed by Podobnik and Stanleys. The performance of detrended cross-correlation analysis method was systematically tested for the effect of nonstationarities [9–11]. After that, numerous issues referring to a broad range of applications [12–15] were established to investigate cross-correlational signal in the presence of nonstationarities.

In analogy with the cross-correlation coefficient, Zebende recently introduced the detrended cross-correlation (DCCA) coefficient [6]. One of the outstanding advantages of the nonlinear cross-correlation coefficient is that it can investigate the cross-correlations at different time scales [16, 17]. After

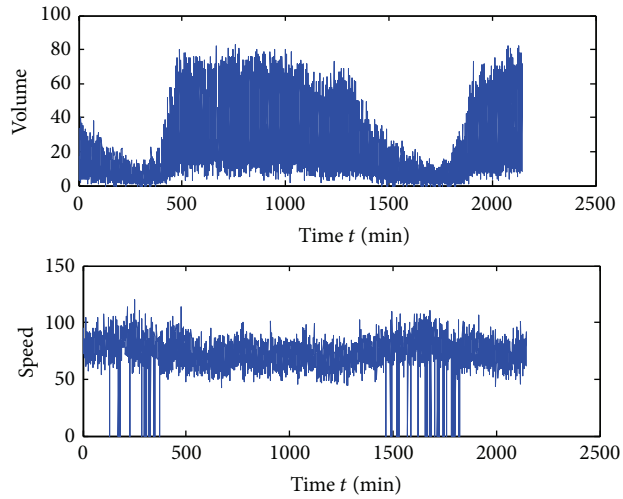


FIGURE 1: The time series plot of the speed data and volume data observed at the Beijing Third Ring Road.

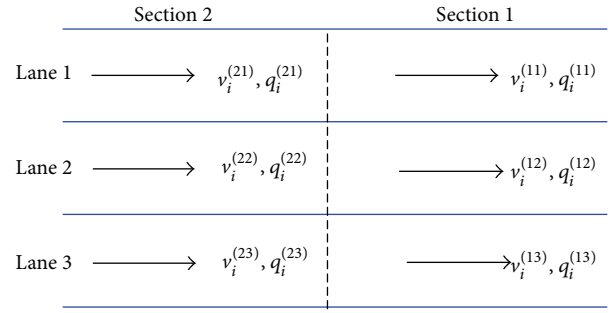


FIGURE 2: The twelve datasets of traffic time series.

that, Cao et al. adopted the DCCA coefficient to analyze and quantify cross-correlations between the Chinese exchange market and stock market [16]. Vassoler et al. quantified the cross-correlations between time series of air temperature and relative humidity by DCCA coefficient [18]. Podobnik et al. showed that the tendency of the Chinese stock market to follow the US stock market is extremely weak by using the DCCA coefficient [19]. Wang et al. studied the statistical properties of the foreign exchange network at different time scales applying the DCCA coefficient [20].

Here, using the DCCA coefficient method, we model the traffic data collected on the Beijing Third Ring Road as the input data which can be readily observed from conventional point detectors. The preliminary test results demonstrate that the cross-correlation property between the traffic series from two adjacent sections is stronger than the cross-correlation property between the series of adjacent lanes and disjoint lanes. The scaling results suggest the feasibility of estimating cross-correlations in traffic variables using point detector data via the proposed approach.

The organization of this paper is as follows. In the next section, we present the dataset and DCCA coefficient method. In Section 3, we show the main empirical results and discussion. Finally, we draw some conclusions in Section 4.

2. Data and Methodology

2.1. The Dataset. Traffic systems have a number of parameters that can be measured. The speed and volume are employed in collecting and studying traffic data here. The data was observed on the Beijing Third Ring Road (BTRR) over a period of about 7 days, from 0:00 AM on March 21, 2011, to 23:30 PM on March 27, 2011. Figure 1 shows the time series plot of the speed data and volume data observed at the Beijing Third Ring Road.

The BTRR is a closed road system without any traffic-signal control. There are three main lanes as well as one or two auxiliary lanes related to on-and-off ramps for each direction. The data were downloaded from the Highway Performance Measurement Project (FPMP). The periodic time of detecting is 2 min and the distance between two adjacent detectors is about 500 m. For investigating the cross-correlations in traffic time series, we will analyze twelve datasets as follows (see Figure 2):

(1) $\{v_i^{mn} : m = 1, 2;\ n = 1, 2, 3\}$: the speed series of Lane n in Section m;

(2) $\{q_i^{mn} : m = 1, 2;\ n = 1, 2, 3\}$: the volume series of Lane n in Section m.

2.2. DCCA Coefficient Method. DCCA coefficient method is an extension of detrended cross-correlation analysis (DCCA) and detrended fluctuation analysis (DFA) method, and both methods are based on random walk theory [6, 21, 22]. For two nonstationary time series $\{x_k\}$ and $\{y_k\}$, $k = 1, 2, \ldots, T$, where T is the length of data, the DCCA coefficient is given as follows.

Step 1. Compute the profiles of underlying time series using

$$X(i) = \sum_{k=1}^{i} \left(x_k - \langle x \rangle\right),$$
$$Y(i) = \sum_{k=1}^{i} \left(y_k - \langle y \rangle\right), \tag{1}$$

where $\langle x \rangle = (1/k)\sum_{j=1}^{k} x_j$ and $\langle y \rangle = (1/k)\sum_{j=1}^{k} y_j$ are the mean.

Step 2. Cut the profiles X and Y into $N_s = [N/s]$ nonoverlapping segments of equal length s, respectively. In each segment v, we calculate the local trend by a least-square fit of the data and obtain the difference between the original time series and the fits.

Step 3. Calculate the covariance of the residuals in each segment:

$$f^2_{\mathrm{DCCA}}(s,v) = \frac{1}{s}\sum_{i=1}^{s}\left(X_{(v-1)s+i} - \widetilde{X}_{i,v}\right)\left(Y_{(v-1)s+i} - \widetilde{Y}_{i,v}\right),$$
$$f^2_{\mathrm{DFA},\{x_i\}}(s,v) = \frac{1}{s}\sum_{i=1}^{s}\left(X_{(v-1)s+i} - \widetilde{X}_{i,v}\right)^2,$$
$$f^2_{\mathrm{DFA},\{y_i\}}(s,v) = \frac{1}{s}\sum_{i=1}^{s}\left(Y_{(v-1)s+i} - \widetilde{Y}_{i,v}\right)^2, \tag{2}$$

for each segment $v = 1, 2, \ldots, N_s$, and

$$f^2_{\mathrm{DCCA}}(s,v) = \frac{1}{s}\sum_{i=1}^{s}\left(X_{N-(v-N_S)s+i} - \widetilde{X}_{i,v}\right) \times \left(Y_{N-(v-N_S)s+i} - \widetilde{Y}_{i,v}\right),$$
$$f^2_{\mathrm{DFA},\{x_i\}}(s,v) = \frac{1}{s}\sum_{i=1}^{s}\left(X_{N-(v-N_S)s+i} - \widetilde{X}_{i,v}\right)^2, \tag{3}$$
$$f^2_{\mathrm{DFA},\{y_i\}}(s,v) = \frac{1}{s}\sum_{i=1}^{s}\left(Y_{N-(v-N_S)s+i} - \widetilde{Y}_{i,v}\right)^2,$$

for each segment $v = N_s + 1, N_s + 2, \ldots, 2N_s$. Here $\widetilde{X}_{i,v}$ and $\widetilde{Y}_{i,v}$ are the fitting polynomials in segment v, respectively. Then the averages over all segments to obtain the fluctuation function are as follows:

$$f^2_{\mathrm{DCCA}}(s) = \frac{1}{2N_s}\sum_{v=1}^{2N_s} f^2_{\mathrm{DCCA}}(s,v), \tag{4}$$

$$f_{\mathrm{DFA},\{x_i\}}(s) = \left\{\frac{1}{2N_s}\sum_{v=1}^{2N_s} f^2_{\mathrm{DFA},\{x_i\}}(s,v)\right\}^{1/2}, \tag{5}$$

$$f_{\mathrm{DFA},\{y_i\}}(s) = \left\{\frac{1}{2N_s}\sum_{v=1}^{2N_s} f^2_{\mathrm{DFA},\{y_i\}}(s,v)\right\}^{1/2}. \tag{6}$$

Step 4. For the two nonstationary time series $\{x_i\}$ and $\{y_i\}$, the DCCA coefficient is defined as the ratio between the detrended covariance function $f^2_{\mathrm{DCCA}}(s)$ of (4) and two detrended variance functions $f_{\mathrm{DFA}}(s)$ of (5) and (6):

$$\rho_{\mathrm{DCCA}}(s) = \frac{f^2_{\mathrm{DCCA}}(s)}{f_{\mathrm{DFA},\{x_i\}}(s)\, f_{\mathrm{DFA},\{y_i\}}(s)}, \tag{7}$$

where $\rho_{\mathrm{DCCA}}(s)$ ranges from -1 to 1 [6, 20]. A value of $\rho_{\mathrm{DCCA}}(s) = 1$ or $\rho_{\mathrm{DCCA}}(s) = -1$ implies that the two nonstationary time series $\{x_i\}$ and $\{y_i\}$ are completely cross-correlated or anti-cross-correlated, at the time scale s, whereas a value of $\rho_{\mathrm{DCCA}}(s) = 0$ indicates that there is no cross-correlation between the two time series $\{x_i\}$ and $\{y_i\}$ [6, 19]. Obviously, the DCCA coefficient $\rho_{\mathrm{DCCA}}(s)$ is a function of the different window size s of data, which means that it can investigate the cross-correlations between two time series $\{x_i\}$ and $\{y_i\}$ at different window scales.

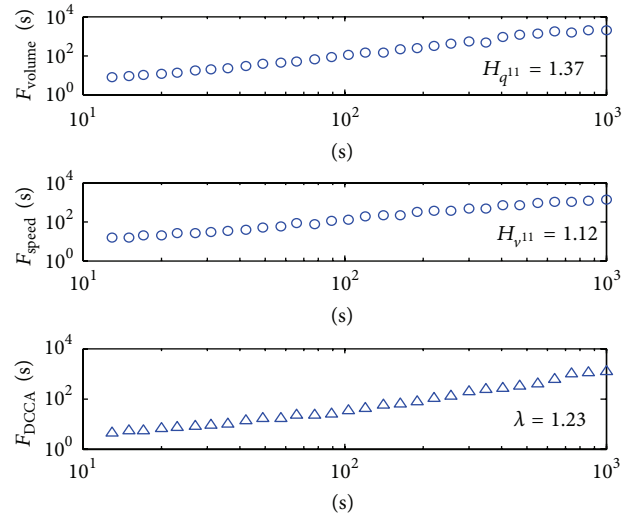


Figure 3: The DFA and DCCA fluctuate function for traffic speed fluctuation series $\{v_i^{(11)}\}$ and the traffic volume fluctuation series $\{q_i^{(11)}\}$.

3. Empirical Results and Analysis

3.1. The Cross-Correlation of the Speed and Volume Series. For two nonstationary cross-correlated time series $\{x_i\}$ and $\{y_i\}$, the power-law relationship $f^2_{\mathrm{DCCA}}(s) \sim s^{2\lambda}$ exists. The scaling exponent λ represents the degrees of the cross-correlation between the two time series $\{x_i\}$ and $\{y_i\}$. For time series $x_i = y_i$, the DCCA fluctuate function reduces to the DFA fluctuate function $f_{\mathrm{DFA}}(s)$.

In order to study the dynamics of the traffic time series over time, we first consider two time series, both of which can be considered as two outputs of traffic system: the traffic speed fluctuation series $\{v_i^{(11)}\}$ and the traffic volume fluctuation series $\{q_i^{(11)}\}$. Here $\{v_i^{(11)}\}$ are the speeds of Lane 1 in Section 1 and $\{q_i^{(11)}\}$ are the volumes of Lane 1 in Section 1.

Figure 3 displays the DFA and DCCA curve obtained between traffic speed fluctuation series $\{v_i^{(11)}\}$ and the traffic volume fluctuation series $\{q_i^{(11)}\}$. The curves exhibit obvious power-law behavior with DFA exponent $H_{q^{(11)}} = 1.37$, $H_{v^{(11)}} = 1.12$ and the DCCA exponent $\lambda = 1.23$, implying long-range autocorrelation and cross-correlations in traffic dynamics.

It is apparent that the traffic flow series can be characterized by a local variability of the DCCA coefficient as shown in Figure 4. The small fluctuations exhibited by the $\rho_{\mathrm{DCCA}}(s)$ provide evidence that a more complex evolution dynamics characterizes the traffic flow.

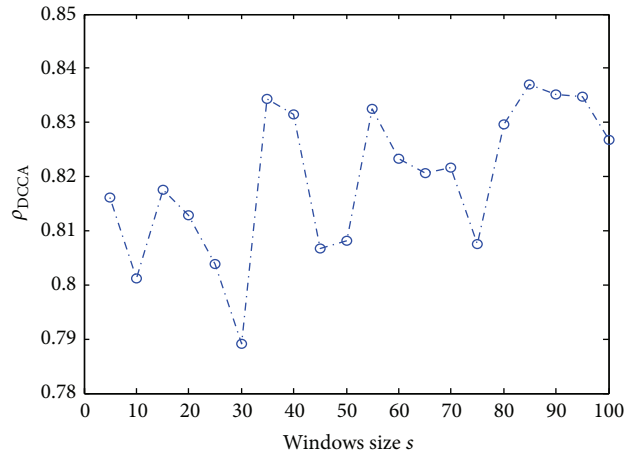


FIGURE 4: The DCCA coefficient between traffic time series.

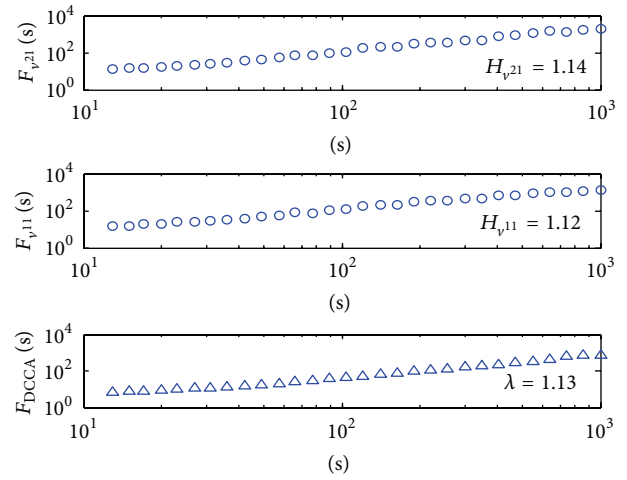


FIGURE 5: The DFA and DCCA curve of traffic speed variables $\{v_i^{(11)}\}$ and $\{v_i^{(21)}\}$.

3.2. The Cross-Correlation of the Speed Series. It is worth noticing the fact that, according to the definition of cross-correlation [9], each of the two variables at any time depends not only on its own past values but also on past values of the other variable.

Here, we firstly investigate the cross-correlations between two traffic speed fluctuation variables $\{v_i^{(11)}\}$ and $\{v_i^{(21)}\}$, which are derived from two adjacent sections of a highway and simultaneously recorded every two minutes (see Figure 2). Figure 5 displays the DFA and DCCA curve for traffic speed fluctuation variables $\{v_i^{(11)}\}$ and $\{v_i^{(21)}\}$. The curves also exhibit obvious power-law behavior with DFA exponent $H_{v^{(21)}} = 1.14$, $H_{v^{(11)}} = 1.12$ and the DCCA exponent $\lambda = 1.13$, implying long-range autocorrelation and cross-correlations in traffic speed time series.

The DCCA coefficient curve is given in Figure 6. We find that $\rho_{DCCA}(s)$ fluctuate around the value $\rho_{DCCA} = 0.97$ and show that the cross-correlated behavior between the time series $\{v_i^{(11)}\}$ and $\{v_i^{(21)}\}$ is very strong.

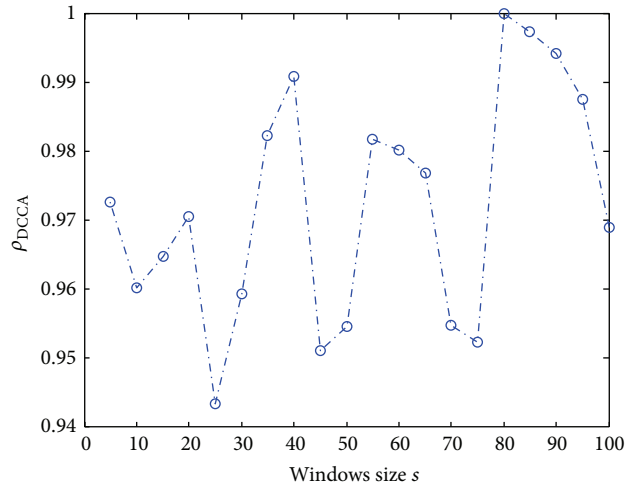


FIGURE 6: The cross-correlations between two traffic speed fluctuation variables from two adjacent sections.

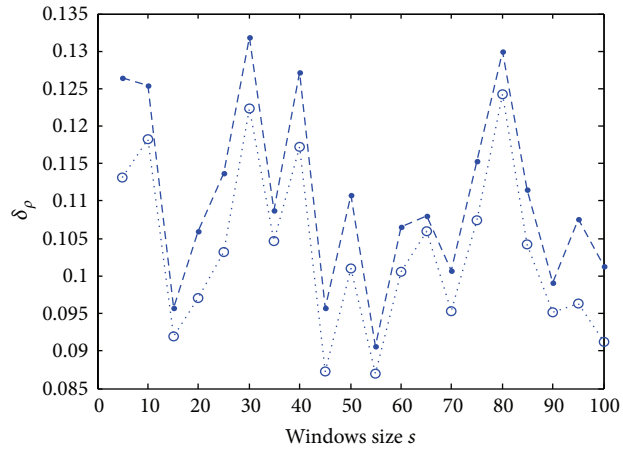


FIGURE 7: The error function $\delta(s)$ of traffic speed series for the data from two adjacent sections and the data from two adjacent lanes (circles), and $\delta'(s)$ for the data from two adjacent sections and the data from two disjoint lanes (filled dots).

And then, we consider the case when two time series of variables $\{v_i^{(11)}\}$ and $\{v_i^{(12)}\}$ are derived from two adjacent lanes (see Figure 2). For convenience, we study the difference between the DCCA coefficient of the data from two adjacent sections of one lane and the data from two adjacent lanes by using the error function.

The error function is defined as $\delta(s) = \rho_{DCCA}(s) - \rho^{(1)}_{DCCA}(s)$, where $\rho_{DCCA}(s)$ is the DCCA coefficient of traffic speed fluctuation variables $\{v_i^{(11)}\}$ and $\{v_i^{(21)}\}$ and $\rho^{(1)}_{DCCA}(s)$ is the DCCA coefficient of traffic speed fluctuation variables $\{v_i^{(11)}\}$ and $\{v_i^{(12)}\}$.

From Figure 7, we can see that the error function $\delta(s) > 0$ (circles) indicates that the cross-correlation of speed series between two adjacent lanes is weaker than the time series of two adjacent sections.

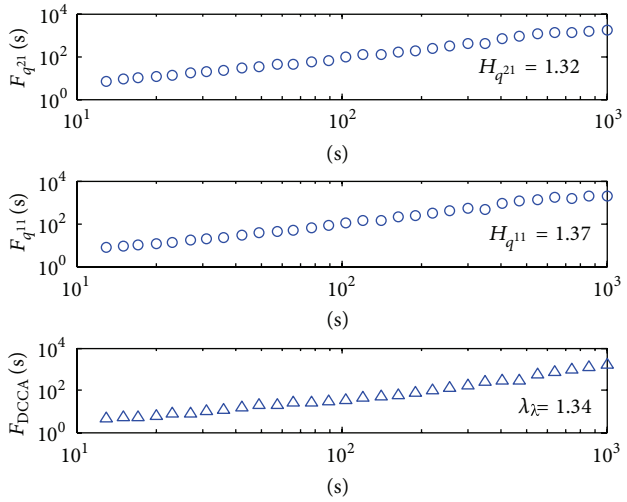


FIGURE 8: The DFA and DCCA curve of traffic volume variables $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$.

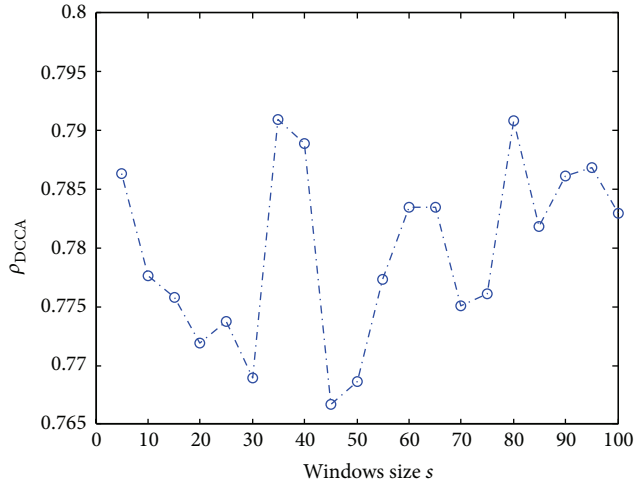


FIGURE 9: The cross-correlations between two traffic volume fluctuation variables from two adjacent sections.

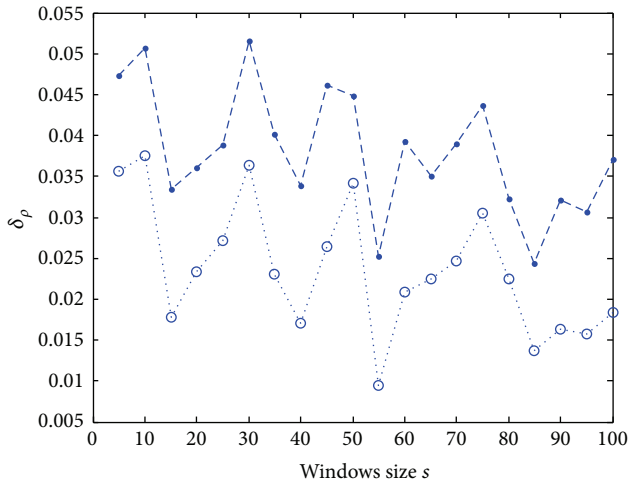


FIGURE 10: The error function $\delta(s)$ of traffic volume series for the data from two adjacent sections and the data from two adjacent lanes (circles) and $\delta'(s)$ for the data from two adjacent sections and the data from two disjoint lanes (filled dots).

In addition, we also find that cross-correlation exists between the two time series of variables $\{v_i^{(11)}\}$ and $\{v_i^{(13)}\}$, which are derived from Lane 1 and Lane 3 (see Figure 2). We employ the error function $\delta'(s) = \rho_{DCCA}(s) - \rho_{DCCA}^{(2)}(s)$ once again, where $\rho_{DCCA}^{(2)}(s)$ is the DCCA coefficient of traffic speed fluctuation variables $\{v_i^{(11)}\}$ and $\{v_i^{(13)}\}$. For comparison, the error function $\delta'(s) = \rho_{DCCA}(s) - \rho_{DCCA}^{(2)}(s)$ is also plot in Figure 7 (filled dots). Obviously, the error function $\delta'(s) > 0$ (filled dots) indicates that the cross-correlation between speed series from two disjoint lanes is weaker than the cross-correlation between the time series of two adjacent sections.

To analyze the statistical properties of the speed time series, we compute the P value for $\rho_{DCCA}(s)$ and $\rho_{DCCA}^{(1)}(s)$. The result $P = 6.38 \times 10^{-16}$ indicates that the difference between two quantities is statistically significant. Similarly, the P value of $\rho_{DCCA}(s)$ and $\rho_{DCCA}^{(2)}(s)$ also shows significant difference ($P = 2.51 \times 10^{-22}$).

3.3. The Cross-Correlation of the Traffic Volume Series. Next, we investigate the cross-correlations between two traffic volume time series $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$ (see Figure 2). The DFA curves in Figure 8 show that each of two volume time series $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$ exhibits autocorrelated behavior by DFA exponent $H_{q^{(21)}} = 1.32$, $H_{q^{(11)}} = 1.37$. Figure 8 also illuminates that the cross-correlated behavior between $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$ exists by DCCA exponent.

Figure 9 shows the DCCA coefficient of traffic volume fluctuation variables $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$. The DCCA coefficient $\rho_{DCCA}(s)$ fluctuates around the value $\rho_{DCCA} = 0.78$ and shows that the cross-correlations between $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$ exists.

Further, we investigate the case when two time series of variables $\{q_i^{(11)}\}$ and $\{q_i^{(12)}\}$ are derived from two adjacent lanes (see Figure 2). The error function is employed once again. In Figure 10, we give the error function $\delta(s) = \rho_{DCCA}(s) - \rho_{DCCA}^{(1)}(s)$, where $\rho_{DCCA}(s)$ is the DCCA coefficient of traffic volume variables $\{q_i^{(11)}\}$ and $\{q_i^{(21)}\}$ and $\rho_{DCCA}^{(1)}(s)$ is the DCCA coefficient of traffic volume fluctuation variables $\{q_i^{(11)}\}$ and $\{q_i^{(12)}\}$. The error function $\delta(s) > 0$ (circles) demonstrates that the cross-correlation of volume fluctuation series between two adjacent lanes is weaker than the time series of two adjacent sections.

For convenience, Figure 10 also shows the error function $\delta'(s) = \rho_{DCCA}(s) - \rho_{DCCA}^{(2)}(s)$, where $\rho_{DCCA}^{(2)}(s)$ is the DCCA coefficient of traffic volume series $\{q_i^{(11)}\}$ and $\{q_i^{(13)}\}$. Similarly, it is apparent that the cross-correlation of volume series between two disjoint lanes is weaker than the time series of two adjacent sections by direct observation of the error function $\delta'(s) > 0$ (filled dots).

In the statistical analysis, the P value for $\rho_{\mathrm{DCCA}}(s)$ and $\rho^{(1)}_{\mathrm{DCCA}}(s)$ is 2.22×10^{-8} which indicates that the difference between two quantities is statistically significant. $\rho_{\mathrm{DCCA}}(s)$ and $\rho^{(2)}_{\mathrm{DCCA}}(s)$ of volume time series are also statistically significant based on permutation testing ($P = 6.61 \times 10^{-14}$).

4. Conclusion

In the paper, we consider DCCA coefficients method to understand the complexity of traffic dynamic. The technique has been implemented on the time series of the original traffic variables from adjacent lanes and adjacent sections. For the traffic speed time series and volume time series, the DCCA coefficients fluctuate around the value $\rho_{\mathrm{DCCA}} = 0.82$ and provide evidence that cross-correlation characteristic exists in traffic dynamic. Then, we apply DCCA coefficients method to study the cross-correlation between traffic speed series. We find that two traffic speed fluctuation parameters derived from adjacent sections exhibit much stronger correlation than the traffic parameters derived from adjacent lanes and disjoint lanes. Similarly, by applying DCCA coefficients method to traffic volume series, the cross-correlation property between the volume variables from two adjacent sections is stronger than the cross-correlation property between the volume variables of adjacent lanes and disjoint lanes.

The relationship of traffic series between two adjacent sections or lanes in China is investigated with the data from BTRR. The results that the traffic series between two adjacent sections or lanes exhibit cross-correlation are attributable to each of the two variables at any time depending not only on its own past values but also on past values of the other variable. Therefore, the findings presented here encourage us to think that this method reveals the relation in anomalous traffic conditions.

Conflict of Interests

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

Acknowledgments

The financial support from the funds of the Fundamental Research Funds for the Central Universities under Grant no. 3122013C005 and the National Natural Science Foundation of China under Grant no. U1233201 is gratefully acknowledged.

References

[1] G. Oh, C. Eom, F. Wang, W.-S. Jung, H. E. Stanley, and S. Kim, "Statistical properties of cross-correlation in the korean stock market," *European Physical Journal B*, vol. 79, no. 1, pp. 55–60, 2011.

[2] D. Wang, B. Podobnik, D. Horvatić, and H. E. Stanley, "Quantifying and modeling long-range cross correlations in multiple time series with applications to world stock indices," *Physical Review E*, vol. 83, Article ID 046121, 2011.

[3] B. Podobnik, I. Grosse, D. Horvatić, S. Ilic, P. C. Ivanov, and H. E. Stanley, "Quantifying cross-correlations using local and global detrending approaches," *European Physical Journal B*, vol. 71, no. 2, pp. 243–250, 2009.

[4] W. Du, C. H. Thurber, and D. Eberhart-Phillips, "Earthquake relocation using cross-correlation time delay estimates verified with the bispectrum method," *Bulletin of the Seismological Society of America*, vol. 94, no. 3, pp. 856–866, 2004.

[5] R. Gu, Y. Shao, and Q. Wang, "Is the efficiency of stock market correlated with multifractality? An evidence from the Shanghai stock market," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 2, pp. 361–370, 2013.

[6] G. F. Zebende, "DCCA cross-correlation coefficient: quantifying level of cross-correlation," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 4, pp. 614–618, 2011.

[7] Z. Wu, N. E. Huang, S. R. Long, and C. Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 38, pp. 14889–14894, 2007.

[8] X. Li and P. Shang, "Multifractal classification of road traffic flows," *Chaos, Solitons and Fractals*, vol. 31, no. 5, pp. 1089–1094, 2007.

[9] B. Podobnik and H. E. Stanley, "Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series," *Physical Review Letters*, vol. 100, no. 8, Article ID 084102, 2008.

[10] B. Podobnik, D. Horvatic, A. M. Petersen, and H. E. Stanley, "Cross-correlations between volume change and price change," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 52, pp. 22079–22084, 2009.

[11] W. Zhou, "Multifractal detrended cross-correlation analysis for two nonstationary signals," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 6, Article ID 066211, 2008.

[12] S. Hajian and M. S. Movahed, "Multifractal detrended cross-correlation analysis of sunspot numbers and river flow fluctuations," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, no. 21, pp. 4942–4957, 2010.

[13] B. Podobnik, D. Wang, D. Horvatic, I. Grosse, and H. E. Stanley, "Time-lag cross-correlations in collective phenomena," *EPL*, vol. 90, no. 6, Article ID 68001, 2010.

[14] K. Dong, Y. Gao, and N. Wang, "EMD method for minimizing the effect of seasonal trends in detrended cross-correlation analysis," *Mathematical Problems in Engineering*, vol. 2013, Article ID 493893, 7 pages, 2013.

[15] X. Zhao, P. Shang, and J. Huang, "Continuous detrended cross-correlation analysis on generalized Weierstrass function," *The European Physical Journal B*, vol. 86, no. 2, pp. 58–63, 2013.

[16] G. Cao, L. Xu, and J. Cao, "Multifractal detrended cross-correlations between the Chinese exchange market and stock market," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 20, pp. 4855–4866, 2012.

[17] G. J. Wang and C. Xie, "Cross-correlations between Renminbi and four major currencies in the Renminbi currency basket," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 6, pp. 1418–1428, 2013.

[18] R. T. Vassoler and G. F. Zebende, "DCCA cross-correlation coefficient apply in time series of air temperature and air relative humidity," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 7, pp. 2438–2443, 2012.

[19] B. Podobnik, Z. Q. Jiang, W. X. Zhou, and H. E. Stanley, "Statistical tests for power-law cross-correlated processes," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 84, no. 6, Article ID 066118, 2011.

[20] G. Wang, C. Xie, Y. Chen, and S. Chen, "Statistical properties of the foreign exchange network at different time scales: evidence from detrended cross-correlation coefficient and minimum spanning tree," *Entropy*, vol. 15, no. 5, pp. 1643–1662, 2013.

[21] D. ben-Avraham and S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems*, Cambridge University Press, Cambridge, UK, 2000.

[22] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis," *Physica A*, vol. 295, no. 3-4, pp. 441–454, 2001.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 914515, 8 pages
http://dx.doi.org/10.1155/2014/914515

Research Article

A Two-Stage Model for Project Optimization in Transportation Infrastructure Management System

Zhang Chen,[1] Liyuan Liu,[1] Li Li,[1] and Hui Li[2]

[1] *Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China*
[2] *Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA*

Correspondence should be addressed to Li Li; 622lilian@tongji.edu.cn

Received 11 April 2014; Revised 3 June 2014; Accepted 18 June 2014; Published 13 July 2014

Academic Editor: X. Zhang

Mathematical optimization is very important for project decision in the Transportation Infrastructure Management System (TIMS). However, it has not been widely employed in TIMS due to poor performance of conventional optimization models in calculation speed and practical application. Therefore, it is necessary to improve the performance of optimization models. According to the process of decision-making in transportation management, a novel two-stage project optimization model, including budget allocation and project distribution, was proposed in this paper. Moreover, the methods of dynamic programming (DP) and genetic algorithm (GA) were applied to obtain an effective solution. The findings indicate that the new optimization method can provide a satisfactory and reasonable maintenance schedule for transportation infrastructure maintenance agencies whose routine management will benefit from the newly proposed model.

1. Introduction

Transportation infrastructure includes roads, bridges, tunnels, airports, railways, and seaports. As a result, Transportation Infrastructure Management System (TIMS) correspondingly covers many subsystems, among which Pavement Management System (PMS) and Bridge Management System (BMS) are the most important ones. Whatever kind of subsystem it is, project optimization is a key element in the process of decision-making for the infrastructure management. Specifically, project optimization [1] refers to finding an optimal maintenance strategy with maximized benefit through arranging pavement maintenance reasonably in terms of time and space in the planning period. Most of existing research efforts on project optimization of TIMS focused on PMS.

Since decision supporting system was introduced into the second generation of PMS, project optimization has been paid much attention by pavement researchers and management agencies. Currently, there are mainly two categories of project optimization methods for network-level pavement management system, namely, prioritization method and mathematical optimization [2, 3]. Prioritization method is to carry out project selection based on some principles prescreened and then to determine the maintenance strategy for each year in the planning period. In 1994, Hass et al. [4] summarized the characteristics of different prioritization methods, among which the two most popular methods are based on infrastructure performance parameters and economic analysis parameters. The prioritization method based on performance parameters is adopted to arrange road maintenance projects by California, US, where road roughness, damaged condition, and average daily traffic volume are considered as influencing factors to develop the prioritization principles. Hudson uses rainfall, number of freeze-thaw cycles, and damage severity as parameters to establish an expression for prioritization through regression or variance analysis so as to determine the priority index of every section requiring reconstruction in the road network. While prioritization based on performance parameters is convenient for calculation and of high pertinence, the results may be far from the economic optimum. Thus the prioritization method based on economic analysis parameters is relatively advanced regarding economic optimum. In

1980s, the prioritization method based on economic analysis parameters was very popular [5–8]. The State of Washington, US, used the total cost, including initial project construction cost, maintenance cost, user operating expense, delay cost, and pavement salvage, as the prioritization indicator in their PMS while the performance index was adopted in UK. At the same time, Hass et al. used equivalent annual cost in Canada, which refers to the ratio of the project construction cost and the corresponding expected lifetime to sequence. Among the PMSs adopting the prioritization method, the most typical one is PAVER in US. PAVER uses the method of benefit-cost-increment to determine the priorities of maintenance projects according to budget optimization. Another representative one is the Pavement Maintenance Decision Support System of Shanghai (PMDSSS), China. According to different managerial preferences of road administrators and agencies, this PMDSSS has developed eight different prioritization principles taking into account either performance parameters or economic analysis parameters or both, with different combinations of traffic volume, pavement damage condition, riding quality, structural capability, and economic indicators. The principles greatly help policy makers to formulate large or medium maintenance plans and long-term rehabilitation strategies in the planning years. However, the tradeoff between the maintenance strategy and the time has not been included in the prioritization method, which leads to huge disparity between the calculated result and the actual optimal solution. Therefore, many researchers have been, instead, focusing on the mathematical optimization, which refers to considering each project in the planning period, the possible maintenance plan, and the implementation schedule through mathematical calculations.

The idea of optimization using mathematical method was proposed in 1970. The mathematical optimization for PMS falls into two categories: static and dynamic. Integer programming is mostly used in the former. For instance, PMS in Denmark [9], HDM-III in World Bank [10], RAMS of Texas, US, and PARS of Ontario, Canada, all adopt 0-1 integer programming [11, 12]. PMS in the State of Indiana, US [13, 14], uses the number of workday needed for maintenance as the decision variable, and constraints from budget, manpower, machine, and material are also included to build up the integral programming model. In the dynamic programming, PMS of Arizona, US [15], was the first to successfully introduce Markov decision process to the network-level pavement management system. When improving the PAVER system, Feighan et al. [16] managed to develop the strategy for maintenance and reconstruction that minimizes the cost of road network through dynamic programming. In 1994, Liu and Yao [17] used Markov decision process to minimize maintenance cost. In 1995, Zou [18] formulated a dynamic model to predict the pavement performance and the adaptive iteration algorithm. In addition, to solve the problems of pavement maintenance strategy, he adopted the analytical hierarchy process model and the heuristic optimization technique. In 2001, Nunoo [19] optimized the maintenance plan of integrated pavement using shuffled complex evolution algorithm. In 2003, Chan et al. [20, 21] attempted to use genetic algorithm to develop a multiobjective programming model to optimize the strategy for highway pavement maintenance. Meanwhile, they adopted heuristic algorithm to yield the result. Ferreira et al. [22] also developed a probabilistic segment-linked optimization model together with a genetic algorithm heuristic with the objective of minimizing total discounted cost of M&R actions. While mathematical optimization can produce the optimal calculation results, the large number of factors to be considered, huge data processing, and barely satisfactory calculation speed greatly limit its practical application. In order to apply mathematic optimization to project optimization, the key point is to find a model or algorithm which can save computing resources as well as meeting the practical needs.

Compared to the existing research studies, the main contributions of this paper are listed as follows. Firstly, the project optimization is separated into two independent and interrelated processes, namely, budgets allocation and project distribution, and a new two-stage model is developed. Secondly, the technique of dynamic programming (DP) and genetic algorithm (GA) are applied to solve the model and yield an effective solution. Finally, the new programming method is verified to be effective through the case study in Shanghai and the poor calculation speed and the practical application limitations of conventional methods are improved.

2. Integer Programming in Conventional Project Optimization

Integer programming (IP) in the conventional project optimization model can be stated as follows:

$$\text{Objective} \quad \max Z = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{m} X_{ijt} * B_{ijt} \tag{1}$$

$$\text{Subject to} \quad \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{m} X_{ijt} * C_{ijt} \le A \tag{2}$$

$$\sum_{j=1}^{m} X_{ijt} = 1 \quad (i = 1, 2, \ldots, N; t = 1, 2, \ldots, T) \tag{3}$$

$$X_{ijt} = \begin{cases} 1 & \text{if treatment } j \text{ is applied in segment } i \text{ in year } t \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$B_{ijt}, C_{ijt}, A, S_t, V_t \ > 0, \tag{5}$$

where Z is total maintenance benefit; B_{ijt}, C_{ijt} are, respectively, the benefit and cost caused by implementing treatment j in project i in year t; A is the total budget in planning period; T is the length of planning period, normally 5 or 10 years; N is the amount of total road units; m is the total number of treatments for each project.

The model above is a large integer programming problem in which the set of feasible solutions is very huge. Therefore, it is necessary to simplify the model in order to calculate

conveniently as well as meeting practical needs. Equation (1) is transformed to the following form:

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{m}X_{ijt} * B_{ijt} = \sum_{i=1}^{N}\sum_{j=1}^{m}X_{ij1} * B_{ij1} + \cdots + \sum_{i=1}^{N}\sum_{j=1}^{m}X_{ijk} * B_{ijk} + \cdots + \sum_{i=1}^{N}\sum_{j=1}^{m}X_{ijT} * B_{ijT}. \quad (6)$$

If the expression $f_t(Y_t) = \sum_{i=1}^{N}\sum_{j=1}^{m} X_{ijt} * B_{ijt}$ (Y_t is the budget of year t) is considered as the maintenance benefit of year t, then the total maintenance benefit in planning years should be equal to the summation of maintenance benefit of each year. Therefore,

$$\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{m}X_{ijt} * B_{ijt} = f_1(Y_1) + \cdots + f_t(Y_t) + \cdots + f_T(Y_T) = \sum_{t=1}^{T} f_t(Y_t), \quad (7)$$

where Y_t is budget of year t and $f_t(Y_t)$ is maintenance benefit of year t.

It is shown in (1) and (7) that solution of project optimization can be approached with the following steps. The first step is budget optimizing to get the budget allocation in each year in the planning period. The second step is project portfolio (including optimization of project scheme and schedule) to get the optimal benefit maintenance strategy in each year based on the budget allocation of first step. After several iterations of optimization, selection, and comparison through the repetition of above two steps, the optimal maintenance strategy under the optimal budget allocation can be achieved finally. Actually, this method reflects well the actual iterative decision-making process of the government agencies, which is "budget allocation, project arrangement, budget adjustment, and project adjustment." Through this iterative method, project optimization is divided into two relatively simple processes, namely, budget optimization and project distribution.

3. Two-Stage Optimization Approach

According to the above analysis, project optimization can be divided into two stages. The first one is about how to allocate budgets. The second one determines maintenance projects in each year based on the budget allocation in the first stage. The first stage is defined as budget allocation model and the second as project distribution model, which build up the two-stage optimization approach.

3.1. Budget Allocation Model. Budget allocation model is, under certain amount of fund, to find a reasonable budget allocation for each year in planning period in order to maximize the benefit of budget. The model focuses on the reasonable way of budget allocation.

In the process of allocating budgets, pavement management agencies need to consider not only the performance of road network but also the budget limits (maximal and minimal budgets for each year are expressed with S_t and V_t, resp.). Therefore, the budget allocation model can be designed as follows:

$$\text{Objective} \quad \max \sum_{t=1}^{T} f_t(Y_t) = f_1(Y_1) + f_2(Y_2) + \cdots + f_t(Y_t) + \cdots + f_T(Y_T) \quad (8)$$

$$\text{Subject to} \quad \sum_{t=1}^{T} Y_t \le A \quad (9)$$

$$Y_t \le S_t \quad (t = 1, 2, \ldots, T) \quad (10)$$

$$Y_t \ge V_t \quad (t = 1, 2, \ldots, T). \quad (11)$$

In fact, there are a large number of maintenance projects in the road network for the planning period, and each project has several treatments. Different treatments bring different maintenance benefits. Therefore, it is necessary to pose some limitations to budget allocation considering that pavement management agencies may utilize fuzzy decisions unconsciously in the process of decision-making which makes their decisions rather reasonable. Firstly, budget change is discontinuous when management agencies are allocating or adjusting budgets; therefore the budget change is set as integer times of the smallest unit. Secondly, maximal maintenance benefit in the planning period under certain budget allocation schedule equals the summation of maximal benefits in each year:

$$\sum_{t=1}^{T} f_t(Y_t) = \max(f_1(Y_1)) + \cdots + \max(f_t(Y_t)) + \cdots + \max(f_T(Y_T)) = \sum_{t=1}^{T} \max(f_t(Y_t)). \quad (12)$$

3.2. Project Distribution. Project distribution model is about how to arrange maintenance projects under given budget allocation schedule $(Y_1, Y_2, \ldots, Y_T)$ in each year in order to obtain the maximal maintenance benefit. It can be designed as follows:

$$\text{Objective} \quad \sum_{t=1}^{T}\left(\max \sum_{i=1}^{n}\sum_{j=1}^{m}\left(X_{ijt}B_{ijt}\right)\right) \quad (13)$$

$$\text{Subject to} \quad \sum_{i=1}^{n}\sum_{j=1}^{m}\left(X_{ijt}C_{ijt}\right) \leqq Y_t, \quad (t = 1, 2, \ldots, T) \quad (14)$$

$$\sum_{j=1}^{m} X_{ijt} \le 1 \quad (i = 1, 2, \ldots, n; t = 1, 2, \ldots, T), \quad X_{ijt} = 0 \text{ or } 1, \quad (15)$$

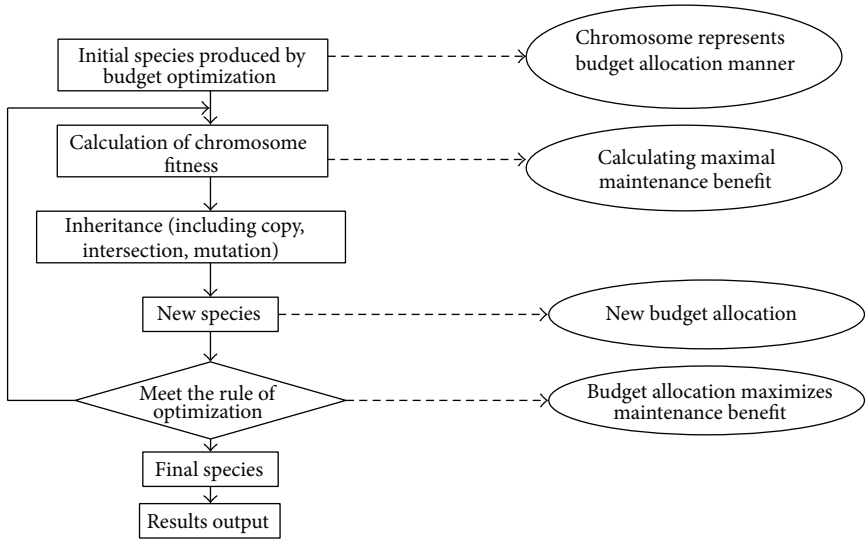


Figure 1: Calculation process of budget allocation with GA.

where n is the total amount of projects in road network; m is the total amount of treatments for each project; Boolean variable $X_{ijt} = 1$ (or 0) means treatment j is applied (or not) to project i in year t.

3.3. Relation of Budget Allocation Model and Project Distribution Model. The results of the budget allocation model and the project distribution model are interactional. At first, annual budget will be provided through the budget allocation model. Then, the maximal maintenance benefit and the project schedule will be constructed through the project distribution model according to the result of the first model and then a feedback will be offered to the first model to judge whether the solution is the best or not. If not, the budget allocation model will optimize again and produce a new array of budgets in each year, then the corresponding total maintenance benefit and maintenance schedule will be obtained through the project distribution model. Through iterations of the two models, the optimal strategy will be obtained in the planning years.

4. Solving the Two-Stage Model

The key to applying the mathematic optimization method into practice is to find a reasonable solving method which ensures certain accuracy and meets practical requirements. In the objective function of (8), it is difficult to be expressed for $f_i(Y_i)$ by explicit functions, which causes difficulty in effectively solving it with normal algorithm. Given that genetic algorithm (GA) is able to search the global optimal solution in a complicated space while its objective functions are not necessary to be explicit functions, GA is chosen to solve the budget allocation model (8).

In the budget allocation model, decision (maintenance strategy) of each stage (planning year) is the function of pavement network condition given by the decision of last stage. That is, each stage will influence the next stage through the network condition resulting from the different treatment strategy. Therefore, under a given budget allocation $(Y_1, Y_2, \ldots, Y_T)$, the formulation of pavement network maintenance strategy in planning years is a multistage problem in which all stages have mutual connections, namely, a classic dynamic programming problem. Therefore, the method of dynamic programming is employed to solve the projects distribution model (13).

4.1. GA for Budget Allocation Model. Before the calculation with GA, some parameters including length of chromosome, species group scale, intersection rate, and mutation rate have to be set in advance according to the scale of pavement network, length of planning period, and budget. Then, the budget allocation model can be solved in the following process, as shown in Figure 1.

(1) Initial species: each chromosome in the species represents a budget allocation mode.

(2) Fitness function: the fitness function of chromosome is derived from the maximal maintenance benefit under a budget allocation mode. The function is computed using the project allocation model.

(3) Genetic manipulation of chromosome: through the genetic manipulation of chromosome, the budget allocation with superior maintenance benefit can be inherited and those inferior ones will be eliminated.

(4) Termination of iteration: certain iteration times can be set as the termination condition of GA.

In the above calculation process, the fitness function of chromosomes, which are not subject to the budget constraints, can be limited by penalty function or be given a very small fitness value. The chromosome with the biggest fitness value gets to be obtained by GA and the corresponding budget allocation manner is the final solution of the budget allocation model.

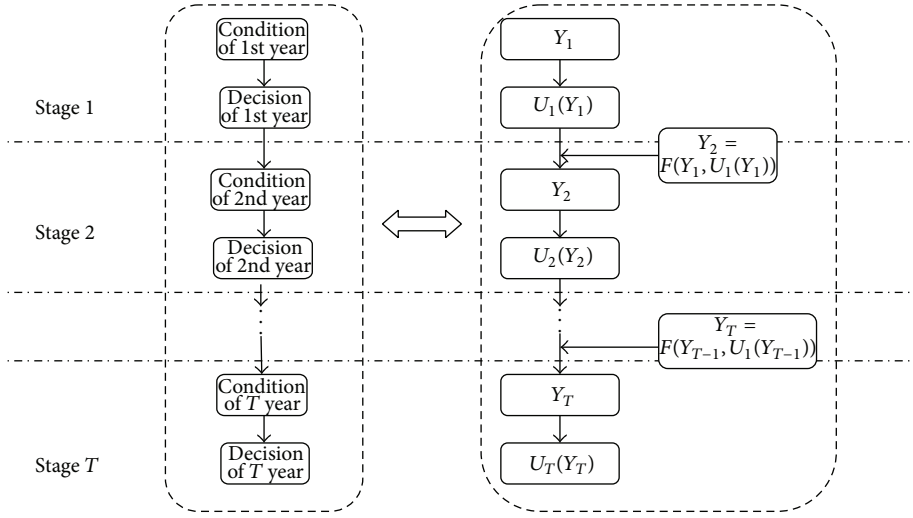


Figure 2: DP process of project distribution model.

4.2. DP for Project Distribution Model. Project distribution model is mainly used to arrange project implementation in each year under a given optimization result of budget allocation. The output results should include a list of maintenance projects in each planning year. Meanwhile, through the project distribution model, the fitness value can be achieved for the above-mentioned chromosome which can serve as the basis of calculation for further optimization.

As discussed above, the determination of maintenance strategy of pavement network in each stage is a multistage decision-making problem in which each stage is interactional. The process has the following characters: at the beginning of each stage, optimal decision of each stage is only related to road network condition, but not related to the decision of the previous stage in which the road network condition is known. The combination of optimal decision in each stage is the best strategy for the planning period. On the other hand, the process of decision-making is a multistage chain in which each stage does not influence the following stages. Therefore, the dynamic programming is a powerful tool to solve this kind of problem.

4.2.1. Calculation Process of DP. According to the characteristics of DP and project distribution model, the process of DP in project distribution can be described as shown in Figure 2.

In Figure 2, the whole planning period of T years can be divided into T stages. Vector Y_t (pavement network condition in year t, budget of year t) represents initial condition of each stage. The decision can be made according to $U_t(Y_t) = U_t$ (pavement network condition in year t, budget of year t). Finally the maximal maintenance benefit, benefit in this stage, can be calculated based on the $U_t(Y_t)$ and the condition transition function $Y_t + 1 = F(Y_t, U_t(Y_t))$.

4.2.2. The Calculation of Maximal Maintenance Benefit (Determination of $U_t(Y_t)$). Calculation model of maximal maintenance benefit $U_t(Y_t)$ in each year can be devised as follows:

$$\max \quad \sum_{i=1}^{n}\sum_{j=1}^{m}\left(X_{ij}B_{ij}\right) \tag{16}$$

$$\text{s.t} \quad \sum_{i=1}^{n}\sum_{j=1}^{m}\left(X_{ij}C_{ij}\right) \le y \tag{17}$$

$$\sum_{j=1}^{m} X_{ij} \le 1, \quad X_{ij} = 0 \text{ or } 1, \tag{18}$$

where n is the number of projects; m is the number of treatments for each project; y is the total maintenance budget; $X_{ij} = 1$ (or 0) means treatment j is selected (or not) in project i.

The following is a recursion equation built up by the DP method, in which there are 3 treatments for each project. The cost of each treatment is defined as integer and described from small to large as w_{i1}, w_{i2}, and w_{i3}. At the same time, the corresponding benefit is v_{i1}, v_{i2}, and v_{i3}, respectively:

$$f(i,m) = \begin{cases} f(i+1,m) & 0 < m < w_{i1} \\ \max\left(f(i+1,m),\ f(i+1,m-w_{i1}) + v_{i1}\right) & w_{i1} < m < w_{i2} \\ \max\left(f(i+1,m),\ f(i+1,m-w_{i1}) + v_{i1},\ f(i+1,m-w_{i2}) + v_{i2}\right) & w_{i2} < m < w_{i3} \\ \max\left(f(i+1,m),\ f(i+1,m-w_{i1}) + v_{i1},\ f(i+1,m-w_{i2}) + v_{i2},\ f(i+1,m-w_{i3}) + v_{i3}\right) & X_i > m > w_{i3}. \end{cases} \tag{19}$$

Table 1: Results of new method.

Year	No. of segments to be treated	Length of road to be treated (m)	Area of pavement to be treated (m^2)	Total maintenance cost (¥10,000)	Average PCI of pavement network	Average PII of pavement network
2005	60	17859	187210	1320	89.6	483649.2
2006	56	17234	223318	1470	88.3	361548
2007	63	17298	221853	1540	87.1	571428.7
2008	81	23026	303050	1450	86.8 (minimal)	888703.5
2009	89	23103	307059	1700	87	782662.4
2010	112	28812	334866	1670	88.3	780185.2
2011	85	22393	310335	1670	88.8	755266.7
2012	80	22567	328590	1660	89.3	675452.1
2013	98	22582	272024	1700	90.6	520031.3
2014	75	22688	319542	1580	91.1 (minimal)	504012.8
Total	**799**	**217562**	**2807847**	**15760**	**88.69**	**6322940**

Then constraints can be formulated as follows:

$$f(n,m) = \begin{cases} 0 & m < w_{n1} \\ v_{n1} & w_{n2} < m < w_{n1} \\ \max(v_{n1}, v_{n2}) & w_{n2} < m < w_{n3} \\ \max(v_{n1}, v_{n2}, v_{n3}) & m > w_{n3}. \end{cases} \quad (20)$$

The function of $f(i,m)$ can be solved through the backward deduction method and the optimal solution for each stage can be achieved through the backtracking method.

5. Model Verification through Case Study

The model and algorithm are tested and verified through a case study with the data collected from asphalt pavements of 12 districts in Shanghai in 2004. The planning period lasts 10 years, the total number of sections surveyed is 867, the total road length is 238,332 meters, and the total area is 3,093,293 square meters.

(1) Parameters of funds are provided as follows:

(a) total budget: 160 million RMB,
(b) minimum investment for each year: 13 million RMB,
(c) maximum investment for each year: 17 million RMB,
(d) the step change of investment for each year: 100 thousand RMB.

(2) Parameters of GA are provided as follows:

(a) reproduction rate: 0.7,
(b) mutation rate: 0.05,
(c) crossover rate: 0.3.

The calculation software is programmed with Microsoft Visual Basic 6.0. The model proposed and developed in this paper (hereafter referred to as "new method") is compared with the current pavement maintenance method (hereafter referred to as "conventional method"), in which maintenance would be conducted once the PCI is lower than 75, currently adopted by pavement management agencies in Shanghai. The calculation results are listed in Tables 1 and 2.

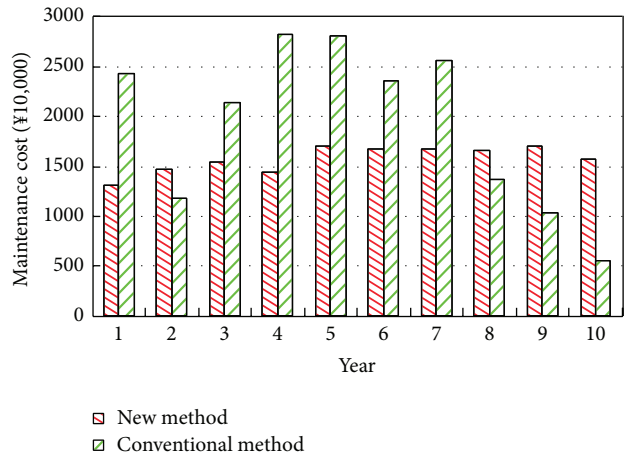


Figure 3: Maintenance cost comparison of road network.

In the tables, the maintenance benefit of the pavement network is indicated with PII (Pavement Improvement Index). PII mainly considers restoration of pavement indicators, interval time between two maintenance services, and social impact. The comparison of the data in Tables 1 and 2 is shown in Figures 3 and 4.

Based on the case study results, the following observations can be drawn.

(1) The total maintenance benefit increases (the PII improves from 6246469 to 6322940 after optimization with the new method), while the average PCI of road network in the planning period decreases after adopting the new method. At the same time, the total cost in the planning period is significantly reduced (the total cost decreases from ¥19286 to ¥15760, a drop

Table 2: Results of conventional method.

Year	No. of segments to be treated	Length of road to be treated (m)	Area of pavement to be treated (m^2)	Total maintenance cost (¥10,000)	Average PCI of pavement network	Average PII of pavement network
2005	78	24152	269404	2440	90.27	572719
2006	43	13433	173309	1188	88.58 (minimal)	302004
2007	73	20923	257218	2145	87.88	621756
2008	133	37972	461683	2820 (maximal)	89.26	1131875
2009	121	33224	446845	2810	90.77	962347
2010	118	29992	366621	2356	92.51	749037
2011	115	31383	435860	2559	94.38	864322
2012	79	19932	279639	1370	95.09	551289
2013	63	15851	224198	1039	95.13 (maximal)	340663
2014	34	8620	130474	559 (minimal)	94	150450
Total	**857**	**235482**	**3045251**	**19286**	**91.78 (average)**	**6246469**

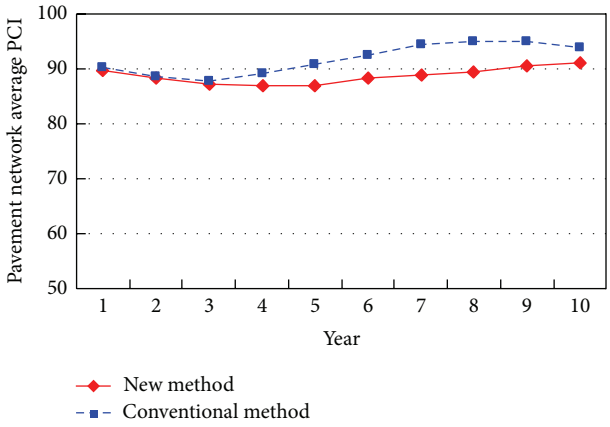


Figure 4: Average PCI comparison of road network.

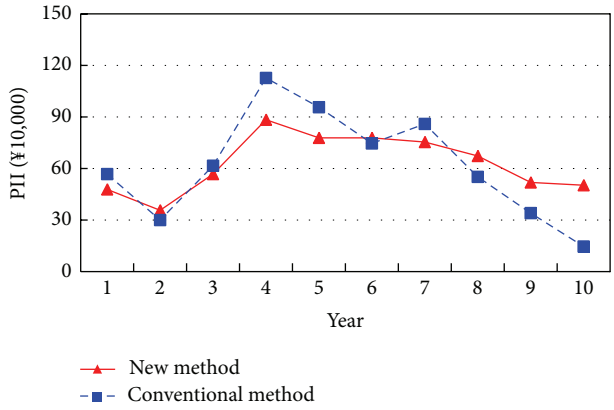


Figure 5: PII comparison.

of 18%). That is, pavement management agencies will achieve better maintenance benefit through spending fewer resources with the new method.

(2) As shown in Figure 3, through the conventional method, budget demand in each planning year is significantly different (i.e., the maximal budget demand is ¥28,200,000 in the fourth year, and the minimal demand is ¥5,590,000 in the tenth year. The variance of the investment in each year is up to 2447). That seriously violates the rule of a stable financial plan of local finance department. Through the new method, the budget of each planning year is well-controlled. The difference of each year's budget effectively decreases (i.e. the maximal budget demand is ¥17,000,000 in the fifth and ninth year, and the minimal demand is ¥13,200,000 in the first year. The variance of the investment in each year is lowered down to 387). It well reflects the rule of a stable financial plan of finance department.

(3) Figures 4 and 5 indicate that the PCI and the PII of road network in each year, similar to the maintenance budget, change remarkably when adopting the conventional method. On the contrary, the new method can significantly mitigate the variation of PCI and PII. The variance of PCI decreases from 8.3 to 4.5 and the variance of PII decreases from 928403 to 505346.

In general, pavement management agencies will get more maintenance benefit with fewer funds through the new model. Meanwhile, the amount of investment and the variability of pavement network PCI in each planning year can be effectively controlled to meet the rule of a stable project plan for pavement management agencies. Consequently, the new model and the solving method proposed in this paper are effective and practical for the infrastructure management.

6. Conclusion

Choosing a proper model and algorithm is very critical for methods of mathematic optimization. To solve the problems of poor calculation speed and practical application limitation of the conventional mathematic method, a two-stage model consisting of the budget allocation model and the project distribution model is developed in this paper. In addition,

the model is solved through dynamic programming and genetic algorithm. Furthermore, it is verified to be effective by practical data in a case study. The findings indicate that the new optimization method can provide a satisfactory and reasonable maintenance schedule for transportation infrastructure maintenance agencies whose routine management will benefit from the newly proposed model.

Conflict of Interests

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

References

[1] T. Hegazy, "Optimization of resource allocation and leveling using genetic algorithms," *Journal of Construction Engineering and Management*, vol. 125, no. 3, pp. 167–175, 1999.

[2] N. Bandara and M. Gunaratne, "Current and future pavement maintenance prioritization based on rapid visual condition evaluation," *Journal of Transportation Engineering*, vol. 127, no. 2, pp. 116–123, 2001.

[3] T. F. Fwa, K. C. Sinha, and J. D. N. Riverson, "Highway routine maintenance programming at network level," *Journal of Transportation Engineering*, vol. 114, no. 5, pp. 539–554, 1988.

[4] R. Hass, W. R. Hudson, and J. Zaniewski, *Modern Pavement Management*, Krieger, Malabar, Fla, USA, 1994.

[5] J. J. Hajek and W. A. Phang, "Prioritization and optimization pavement of preservation treatments," Transportation Research Record 1216, National Research Council, Washington, DC, USA, 1989.

[6] P. B. Still, "A Pavement management system for UK roads," *Highways and Transportation*, vol. 37, no. 2, pp. 6–9, 1990.

[7] R. Haas, "A method for integrated priority programming and budget level analysis for pavement maintenance and rehabilitation," in *Proceedings of the 5th International Conference on the Structure Design of Asphalt Pavements*, vol. 1, pp. 626–635, Delft, The Netherlands, 1982.

[8] R. Dekker, R. E. Wildeman, and F. A. van der Duyn Schouten, "A review of multi-component maintenance models with economic dependence," *Mathematical Methods of Operations Research*, vol. 45, no. 3, pp. 411–435, 1997.

[9] P. Ullidtz, "A danish pavement management system," in *Proceedings of the NAPMC*, vol. 3, pp. 6.84–6.95, Toronto, Canada, 1985.

[10] D. Mrawira, W. J. Welch, M. Schonlau, and R. Haas, "Sensitivity analysis of computer models: World bank HDM-III model," *Journal of Transportation Engineering*, vol. 125, no. 5, pp. 421–428, 1999.

[11] S. Chi, J. Hwang, M. Arellano, Z. Zhang, and M. Murphy, "Development of network-level project screening methods supporting the 4-year pavement management plan in Texas," *Journal of Management in Engineering*, vol. 29, no. 4, pp. 482–494, 2013.

[12] R. Kher and W. D. Cook, "PARS: the MTC mode for programming and financial planning in pavement rehabilitation," in *Proceedingsof the National Arts Marketing Project Conference (NAMPC '85)*, vol. 3, pp. 6.23–6.40, Toronto, Canada, 1985.

[13] B. Colucci-Rios and K. C. Sinha, "An optimal pavement management system at the network level," in *Proceedings of the 1st North American Pavement Management Conference (NAPMC '85)*, vol. 3, pp. 6.68–6.83, Toronto, Canada, 1985.

[14] R. F. Carmicheal, "Implementing pavement management system at the Indiana Department of Highways," in *Proceedings of the 6th International Conference on Structure Design of Asphalt Pavements*, vol. 1, pp. 834–857, Ann Arbor, Mich, USA, 1987.

[15] K. Golabi, R. B. Kulkarni, and G. B. Way, "A statewide pavement management system," *Interfaces*, vol. 12, no. 6, pp. 5–21, 1982.

[16] K. J. Feighan, M. Y. Shahin, and K. C. Sinha, "A dynamic programming approach to optimization for pavement management systems," in *Proceedings of Second North American Conference on Managing Pavements (SNACMP '87)*, vol. 2, pp. 195–206, Toronto, Canada, 1987.

[17] B. Liu and Z. Yao, "Optimization method in network pavement management system," *China Journal of Highway and Transport*, vol. 7, no. 3, pp. 1–9, 1994.

[18] P. Zou, "On optimization method of pavement maintenance and rehabilitation programming," *China Journal of Highway and Transport*, vol. 8, no. 2, pp. 1–8, 1995.

[19] C. Nunoo, *Optimization of Pavement Maintenance and Rehabilitation programming Using Shuffled Complex Evolution Algorithm*, Florida International Univesrity, Miami, Fla, USA, 2001.

[20] W. T. Chan, T. F. Fwa, and J. Y. Tan, "Optimal fund-allocation analysis for multidistrict highway agencies," *Journal of Infrastructure Systems*, vol. 9, no. 4, pp. 167–175, 2003.

[21] T. F. Fwa, W. T. Chan, and K. Z. Hoque, "Multiobjective optimization for pavement maintenance programming," *Journal of Transportation Engineering*, vol. 126, no. 5, pp. 367–374, 2000.

[22] A. Ferreira, A. Antunes, and L. Picado-Santos, "Probabilistic segment-linked pavement management optimization model," *ASCE Journal of Transportation Engineering*, vol. 128, no. 6, pp. 568–577, 2002.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 708482, 13 pages
http://dx.doi.org/10.1155/2014/708482

Research Article

Exploring Urban Taxi Drivers' Activity Distribution Based on GPS Data

Xiaowei Hu, Shi An, and Jian Wang

School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150090, China

Correspondence should be addressed to Xiaowei Hu; xiaowei_hu@hit.edu.cn

Received 10 April 2014; Revised 9 June 2014; Accepted 19 June 2014; Published 13 July 2014

Academic Editor: X. Zhang

With the rapid development of information communication technology and data mining technology, we can obtain taxi vehicle's real time operation status through the large-scale taxi GPS trajectories data and explore the drivers' activity distribution characteristics. Based on the 204 continuous hours of 3198 taxi vehicles' operation data of Shenzhen, China, this paper analyzed the urban taxi driver's activity distribution characteristics from different temporal and spatial levels. In the time level, we identified the difference with taxi daily operation pattern (weekday versus weekends), continuous time in one day, passengers in vehicle time, and taxi drivers' operation frequency; in the space level, we explored the taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations. This research can be helpful for urban taxi drivers' operation and behavior pattern identification, as well as the contribution to the geographical activity space analysis.

1. Introduction

As an important component of the urban public passenger transportation system, taxi offers an all-weather, convenient, comfortable, and personalized travel services for the urban residents, as well as playing a key role in the urban passenger transportation development [1]. Taking the capital of China, Beijing, as an example, there were 66,600 registered taxi vehicles in the year 2009, and these taxis carried 680 million passengers annually (which means more than 2 million passengers per day), accounting for 9.4 percent of the total passenger volume of urban public transport (including the conventional bus, metro transit, and taxi service) [2].

Due to the 24 hours operation per day, taxis account for a higher proportion in the urban traffic volume and air emission pollution [3]; more researchers attempt to analyze taxi operation status and taxi driver's routing choice based on the GPS traces data, aiming to reduce taxi driver's vacant travel distance and cruising time, and also this can save energy resource. In the taxi operation status analysis aspect, researchers analyze the taxi daily average operation distance, taxi passenger demand distribution, daily average operation frequency, and different driver's operation patterns; the reader can be referred to Hu and Feng [4], Chen [5], Jiang et al. [6], Wen et al. [7], Liu et al. [8], Hu et al. [9], and Zhang and He [10].

In the above studies, researchers have paid more attention to the basic taxi operation status and routing choice analysis, but they paid less attention to the taxi driver's temporal and spatial activity space analysis. Meanwhile, each taxi driver's activity space is quite different from the other drivers, and the analysis of taxi driver's travel activity space is a more interesting research field to explore. When the taxi is with a passenger, the driver will complete the trip in accordance with the passenger's travel will by adopting the shortest path from the origin to the destination [11]; this destination set can be called the taxi driver's drop-off locations; when the taxi is without passenger, the driver always wants to search the next potential passenger during the shortest possible time near or around his last drop-off passenger location [10, 12]; this set can be regarded as the taxi driver's pick-up locations. Although the taxi driver will not pick up his passengers at the same location each day, each taxi driver's pick-up passenger's location and drop-off passenger's location are close to each other; the relationship between taxi driver's pick-up locations and drop-off locations can be explored to describe the spatial distribution of taxi drivers.

Due to the randomness and lack of regularity of taxi operation characteristics, it is improper to adopt visit places' order or arrive at a fixed time to describe taxi driver's travel activity pattern. So how to describe the regularity of taxi driver's operation spaces and time period, is there any difference between weekends versus weekdays? In the urban road network, when a taxi completes a trip to arrive at the passenger's destination, then the taxi driver would be at the without passenger stage and search for the next passenger within a range of regions/areas. So taxi drivers may search and cruise around some region to find their next passenger; there may be a relationship between taxi driver's covering area and operation frequency. We assume that there is relationship between the two activity spaces, drop-off locations and pick-up locations. Also, each taxi driver's activity space dynamics can be explored as time evolves.

As a geographic concept, the activity space means "a composite of the locations where an individual conducts routine activities" [13–15]. A major application of activity theory is empirical measurement and analysis of space time activity (STA) data or records of where and when individuals conducted activities over a daily, weekly, or monthly cycle. Wang and Cheng [15] had summarized the basic components in the activity theory, which includes the activity, activity frequency, activity destination, trip, transport mode(s), and the activity space.

The special work characteristics of the taxi drivers, as a special group, make their daily activity space quite different from the commute workers or students' daily mobility variability, such as the regular arrival at the workplace at a relatively fixed time [16], visiting some shop or restaurant frequently [17], and visiting some places following a certain order [18]. Some researchers focus on taxi driver's cognitive space or travel time difference with the general public, which can reflect taxi driver's spatial-temporal cognitive of the urban space. The pioneer work can be traced to Giraudo and Peruch [19], Peruch et al. [20], Spiers and Maguire [21], and Wakabayashi et al. [22]. Besides the traditional comparison of the cognitive maps method, Wakabayashi et al. [22] also adopted the SDE (standard deviational ellipse) to analyze the difference between the taxi drivers and the university students.

The aim of this paper is to provide measurements to analyze the taxi driver's activity space from the time level and individual level; in the time level, the taxi driver's daily (workdays versus weekends) pick-up locations and drop-off locations will be measured and analyzed, as well as the exploration of relationship between the two activity spaces; in the individual level, each taxi driver's activity space dynamics can be explored as time evolves. We will introduce the geographic concept into this research and attempt to analyze the taxi driver's activity space, which will reveal the relationship between the taxi driver's pick-up locations and drop-off locations and help drivers reduce cruising time and improve operation frequency. This research can also be helpful for saving taxi energy consumption and lowering air pollution emissions, to achieve a more sustainable and environmental development of urban taxi industry.

The paper is organized as follows. Section 2 provides an overview on taxi driver's behavior analysis and activity space measurements comparison. In Section 3, we present the taxi GPS traces data source and taxi driver's activity space measures in detail. Section 4 describes and compares the analysis results. Section 5 is the discussions. Finally, we conclude this paper in Section 6.

2. Literature Review

Around the world, research on taxi drivers' travel behavior analysis has been paid more attention from Pailhous [23] and Michon [24]. There are many influence factors on the taxi driver's travel choice behavior in the actual road network environment, such as the driver's road network familiarity, traffic control devices, road construction, and other urban traffic management and control measures. With the development of ICT (information communication technology), it provides a new way to explore taxi driver's route choice or way-finding behavior, which involves the shortest path optimal and driving destination or direction selection based on the huge number of taxi trajectories. Based on the taxi GPS traces and data mining technology, especially one can obtain the experienced/smart taxi driver's route choice behavior in real time and provide guides for other general public's shortest path optimal choice [11].

Another research field focuses on the taxi driver's practical travel behavior. Murakami and Wagner [25] adopted PDA (personal digital assistant) with a GPS receiver to obtain taxi vehicle's speed and daily travel information, which can be seen as the first trial in analysis of taxi driver's travel behavior by using GPS traces data, and this method can have higher data quality on travel start and end times, total trip time, and destination locations than daily travel diary and telephone retrieval methods. Liu et al. [8] described the searching behavior difference between the top taxi drivers and the ordinary divers in Shenzhen and found that the top drivers pay more attention on the less competitive and more profitable locations, rather than the central business district. Based on the taxi GPS trajectory data in Wuhan, Yue et al. [26] forecasted the taxi driver's pick-up locations on the basis of the selected drop-off locations. These researches also can help us to understand taxi driver's travel behavior, but they have paid less attention to the taxi driver's searching behavior dynamic pattern in their pick-up and drop-off activity space.

The activity theory has been adopted to analyze different kinds of traveler's spatial mobility dynamics. The special work characteristics of taxi drivers, as a special group, make their daily activity space quite different from the commute workers or students' daily mobility variability. Due to their professional relationship, the taxi drivers need to arrive at the destination according to the passenger's travel will or search the next potential passenger according to their individual characteristics (the driving experience and the familiarity of the road network or place); there is not a fixed or constant mobility route for the taxi driver to follow. Based on Wang and Cheng's [15] summary, we can identify taxi driver's basic components in the activity theory, which is shown in Table 1.

Table 1: Taxi driver's basic components of activity theory.

Component	Classification
Activity	Depending on the purpose, taxi driver's activity can be classified into pick-up passenger activity and drop-off passenger activity
Activity frequency	The number of times the pick-up passenger activity or drop-off passenger activity occurs during a given time period, in an hour or a day
Activity destination	The location where a pick-up passenger activity or a drop-off passenger occurs
Trip	Movement between the pick-up passenger activity and drop-off passenger activity destinations
Transport mode (s)	By taking taxi vehicle
Activity space	A composite of the locations where a taxi driver conducts pick-up passenger activity and drop-off passenger activity, so it can be divided into pick-up activity space (location) and drop-off activity space (location)

From Table 1, we can know that some basic components of taxi driver' activity have been determined, such as the activity, trip, and transport mode(s). So we pay more attention to the taxi driver's activity frequency, activity destination, and activity space.

For the taxi driver's activity frequency, we can adopt a taxi's daily number of operation service times to describe this component; this also can be divided into the pick-up passenger activity and the drop-off passenger activity. Due to the different passenger's destination demand, the taxi driver's drop-off locations have the irregularity characteristics and will be in accordance with the passenger's requirement; the taxi driver will search the next passenger near the last drop-off location, so the pick-up locations will have some relationship with the drop-off locations. The question will be focused on the pick-up locations and the drop-off location, which is also the taxi driver's activity space analysis.

In the research of Ge et al. [27], they had clustered the taxi drivers' pick-up locations during a certain time interval to obtain the centroids (mean centre) of these pick-up locations, which can be served as the recommended pick-up locations with a certain probability. This leverage can help a new taxi driver to search the potential passenger in a certain area; meanwhile it provides a reference for us to explore the dynamics in the activity spaces of taxi drivers over time and the relationship between each taxi driver's daily pick-up locations' area and drop-off locations' area. Susilo and Kitamura [16] had extended the action space to the second moment of the activity locations that it contains, and then they examined the day-to-day variation in the second moment. The second moment of activity locations has given us inspirer how to analyze each taxi driver's day-to-day variation in the pick-up and drop-off activity space (location).

Here we hypothesize that each taxi driver's activity space mean centre (centroid) has the relationship with the centroid of the whole taxi driver's space area, just as Susilo and Kitamura [16] analyzed the worker's daily activity locations relationship. We also analyze taxi driver's day-to-day variation on activity space and statistically analyze the variation of the second moments.

3. Taxi Driver's Activity Space Measures

There are various measurements to measure the activity space in the previous researches, such as the SDE (standard deviational ellipse), mean centre, the x/y axis ratio, SDC (standard deviational circle), and kernel density. Based on existing researches, we divided these measurements into two categories, the spatial distribution category and the extended second moments of activity locations measurement category.

3.1. The Spatial Distribution Category. The spatial distribution estimates the basic parameters about the distribution; they include mean centre, standard deviation of the X and Y coordinates, standard deviational ellipse, standard distance deviation, and convex hull.

The mean centre (MC) is the average location of taxi service events in the space (including the pick-up and drop-off event), which can be calculated by [28]

$$\left(\dot{x}_{\mathrm{mc}}, \dot{y}_{\mathrm{mc}}\right) = \left[\frac{\sum_{j=1}^{n} x_j}{n}, \frac{\sum_{j=1}^{n} y_j}{n}\right], \tag{1}$$

where $\dot{x}_{\mathrm{mc}}$, $\dot{y}_{\mathrm{mc}}$ are the coordinates of the mean centre, which can determine the space location of the MC, x_j, y_j are the coordinates of taxi service event j in the two-dimensional, respectively, and n is the total number of taxi service events.

Standard distance deviation (SDD) [28, 29] can describe the absolute dispersion degree for each taxi service event relative to the mean centre (MC); the formula can be expressed by

$$\mathrm{SDD} = \sqrt{\frac{\sum\left(x_j - \dot{x}_{\mathrm{mc}}\right)^2 + \sum\left(y_j - \dot{y}_{\mathrm{mc}}\right)^2}{n}}, \tag{2}$$

where x_j, y_j are the coordinates of taxi service event j and $\dot{x}_{\mathrm{mc}}$, $\dot{y}_{\mathrm{mc}}$ are the coordinates of the mean centre.

Based on the mean centre and standard distance deviation, we can draw the standard deviational circle (SDC), which can express the dispersion of taxi service in all directions of space.

standard deviational ellipse (SDE) [28, 29] can determine the directional factors of the spatial distribution and find the main direction of the taxi service event in space. The calculation of SDE can be expressed by θ (the SDE y-axis in the clockwise rotation angle), where e_1 is the major axis of SDE and e_2 the minor axis of SDE. The detailed formulas are as follows:

$$\begin{aligned}\theta = \arctan\Bigg(\Bigg(&\left[\sum\left(x_j - \dot{x}_{\mathrm{mc}}\right)^2 - \sum\left(y_j - \dot{y}_{\mathrm{mc}}\right)^2\right] \\ &+\Bigg(\left\{\left[\sum\left(x_j - \dot{x}_{\mathrm{mc}}\right)^2 - \sum\left(y_j - \dot{y}_{\mathrm{mc}}\right)^2\right]^2 \\ &+4\left[\sum\left(x_j - \dot{x}_{\mathrm{mc}}\right)\left(y_j - \dot{y}_{\mathrm{mc}}\right)\right]^2\right\}\Bigg)^{1/2}\Bigg) \\ &\times\left(2\sum\left(x_j - \dot{x}_{\mathrm{mc}}\right)\left(y_j - \dot{y}_{\mathrm{mc}}\right)\right)^{-1}\Bigg), \\ e_1 = 2S_x = &\sqrt{\frac{\sum\left[\left(x_j - \dot{x}_{\mathrm{mc}}\right)\cos\theta - \left(y_j - \dot{y}_{\mathrm{mc}}\right)\sin\theta\right]^2}{n-2}}, \\ e_2 = 2S_y = &\sqrt{\frac{\sum\left[\left(x_j - \dot{x}_{\mathrm{mc}}\right)\sin\theta - \left(y_j - \dot{y}_{\mathrm{mc}}\right)\cos\theta\right]^2}{n-2}}, \\ A = &\pi S_x S_y, \end{aligned} \tag{3}$$

where A is area of SDE, S_x is the semimajor axis of SDE, S_y is the semiminor axis of SDE, $\dot{x}_{\mathrm{mc}}$, $\dot{y}_{\mathrm{mc}}$ and x_j, $y_j$are consistent with formulas (1) and (2).

Figure 1 has shown a taxi operation example in Shenzhen from 0 am on April 18, 2011 to 12 am on April 26, 2011 (which consists of 204 continuous hours) in Shenzhen, China. The red points represent the taxi drop-off activity locations, while the cyan points represent the taxi pick-up activity locations. From Figure 1, we can find that the taxi's operation drop-off locations and pick-up locations' mean centre (centriod), standard deviational ellipse distribution, and the two activity spaces' standard deviation of the X and Y coordinates are quite different, and the y-axis is rotated clockwise through an angle. But the mean centres of the drop-off locations and pick-up locations are quite near to each other; we may use the extended second moments of activity locations measurement category to analyze this relationship and distribution.

From Figure 2, we can know that the taxi driver's daily operation has no fixed origin; this is quite different from other workers or students' origin-home location; meanwhile, there is no fixed operation lines or order for taxi driver to follow. But each taxi driver's operation has a mean centre that can be as a fixed location; based on this we can extend to measure taxi driver's daily operation relationship.

Kamruzzaman and Hine [30] analyzed the correlations between the different measures of activity space size; this method can help us to understand different activity space

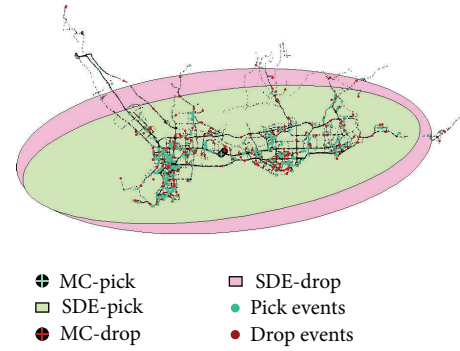


Figure 1: Drop-off locations, pick-up locations, SDE, and mean centre figure of one taxi vehicle operation.

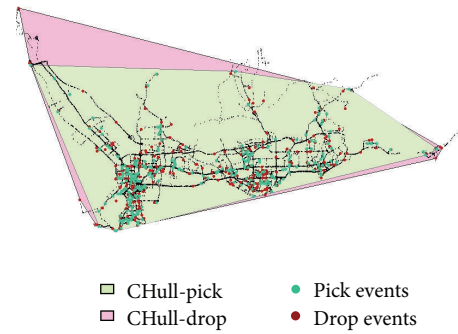


Figure 2: The convex hull figure of one taxi vehicle operation.

measures relationship of each taxi driver's operation. We will analyze these relationships:

(1) area size of SDE and area size of SDD of pick-up locations,

(2) area size of SDE and area size of SDD of drop-off locations,

(3) area size of SDE comparison between the pick-up and drop-off locations,

(4) area size of SDD comparison between the pick-up and drop-off locations.

3.2. The Extended Second Moments of Activity Locations Measurement Category. The researches of Susilo and Kitamura [16] and Ge et al. [27] have given us inspirer how to analyze each taxi driver's day-to-day variation in the pick-up and drop-off activity space. Each taxi driver's daily activity space area mean centre may have the relationship with the centroid of the whole taxi drivers' activity space, just as Susilo and Kitamura [16] analyzed the worker's daily activity locations relationship. We can analyze taxi driver's day-to-day variation on activity space and statistically analyze the second moments of activity locations:

(1) the Great-circle distance between the mean centre of each taxi driver's drop-off location and pick-up location, with calculating the statistics distribution,

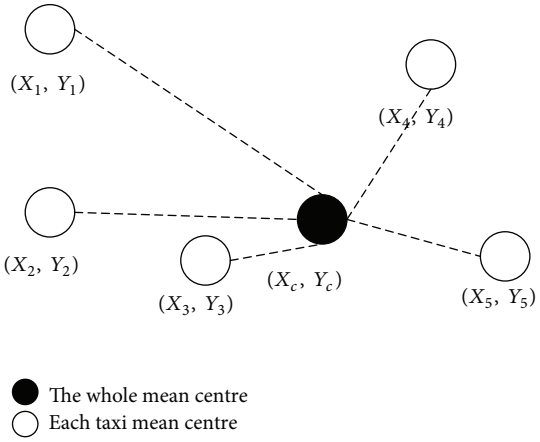


Figure 3: An illustration of the locations mean centre of each taxi driver and the all taxi drivers.

which can be referred to in Figure 1 of the MC pick and MC drop (by using (1)),

(2) the Great-circle distance between the drop-off (pick-up) locations mean centre of each taxi driver and all taxi drivers, with calculating the statistics distribution, which is shown in Figure 3,

(3) the fullness of activity spaces to describe the individual SDE feature class.

The next part is how to calculate the Great-circle distance between two points on the globe surface. Here we adopt the simple spherical law of cosines formula to calculate it [31, 32], assuming the Erath as a spherical earth (ignoring the ellipsoidal effects). The spherical law of cosines formula gives well-conditioned results down to distances as small as around 1 meter. Given two points' latitude and longitude, the formula of the spherical law of cosines is as follows:

$$\begin{aligned} D &= R * \operatorname{acrcos}\left[\sin\left(\mathrm{lat}_1\right) * \sin\left(\mathrm{lat}_2\right) + \cos\left(\mathrm{lat}_1\right)\right. \\ &\qquad \left. * \cos\left(\mathrm{lat}_2\right) * \cos\left(\mathrm{long}_2 - \mathrm{long}_1\right)\right] \\ &= 2R * \arcsin\left(\left(\sin^2\left(\frac{\mathrm{lat}_2 - \mathrm{lat}_1}{2}\right) + \cos\left(\mathrm{lat}_1\right)\right.\right. \\ &\qquad \left.\left. * \cos\left(\mathrm{lat}_2\right) \sin^2\left(\frac{\mathrm{long}_2 - \mathrm{long}_1}{2}\right)\right)^{1/2}\right), \end{aligned} \tag{4}$$

where R is the Earth radius; here we use the average radius of the earth, which is equal to 6371.004 kilometers in average; lat_i, long_i represent the i point's latitude and longitude (in radian), respectively; D is the Great-circle distance between the two points.

Kamruzzaman and Hine [33] had proposed the fullness of activity spaces to describe the individual SDE feature class. The formula can be expressed as

$$F = \begin{cases} \dfrac{l_{Y,\mathrm{axis}}}{l_{X,\mathrm{axis}}}, & \text{if } l_{Y,\mathrm{axis}} > l_{X,\mathrm{axis}}, \\ \dfrac{l_{X,\mathrm{axis}}}{l_{Y,\mathrm{axis}}}, & \text{else if,} \end{cases} \tag{5}$$

where F is the fullness of activity spaces; $l_{Y,\mathrm{axis}}, l_{X,\mathrm{axis}}$ represent the length of the activity spaces locations' SDE y-axis and X-axis, respectively.

4. Results Analysis

4.1. Data Sources and Processing. In this research, we use the Taxi GPS traces data, which contains 13798 taxi records of nine consecutive days, from April 18, 2011 to the noon of April 26, 2011, and the total hours are 204 hours in Shenzhen, China. Among the 13798 taxis equipped with GPS, only 8845 taxi records can be used for analysis, and the other 4953 taxi GPS traces records have wrong or no record. After the data qualification, there are 3198 taxi vehicle's GPS traces covering 204 continuous hours. For each taxi equipped with GPS, the records information including taxi ID, date, time, location (longitude, latitude), velocity, driving direction, and taxi operation status (having passengers or not) is recorded every 5 or 10 seconds for the driving environment effects; under the effect of the traffic environment, the GPS record collection interval time period is not constant.

In the taxi GPS traces data processing, we adopt ArcGIS 9.3, R statistics software and Crime Stat software to do data mining and statistics work. Figure 4 shows the GPS data mining process.

We have divided the GPS data mining into three levels: the time level analysis, personal (each taxi driver) level, and the top 1% and last 1% taxi driver's comparison. From the time level we want to find taxi daily operation pattern (weekdays versus weekends), continuous time in one day, passenger's in vehicle time, and taxi's operation service frequency. This can be used in taxi operation planning, taxi pricing model, and taxi management policy implementation. From the personal level, we want to identify taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations, which can be used in taxi dispatching and setting taxi stop location. From the comparison of the top and last 1% taxi drivers, the difference can be identified, which will be helpful for reducing passenger's waiting time and increasing driver's daily operation service frequency.

4.2. Time Level Analysis. Our first analysis has taken all the 3198 taxi vehicles operations as a whole to explore their operation temporal pattern per hour, which includes the pick-up passengers' locations and drop-off passengers' locations analysis.

Figure 5 shows all the taxi operation service frequency analysis; we can find these useful pieces of information.

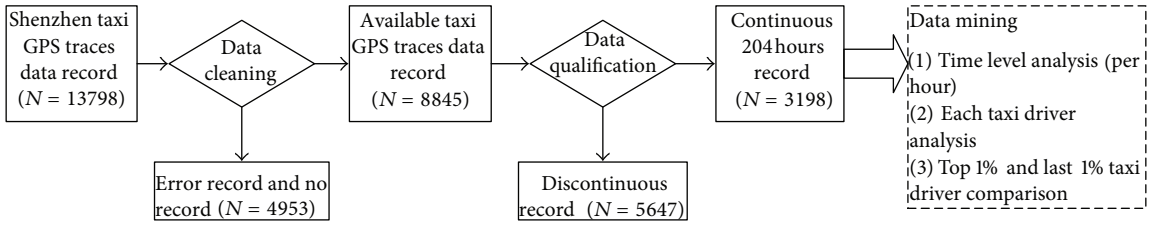


Figure 4: Taxi GPS data mining process.

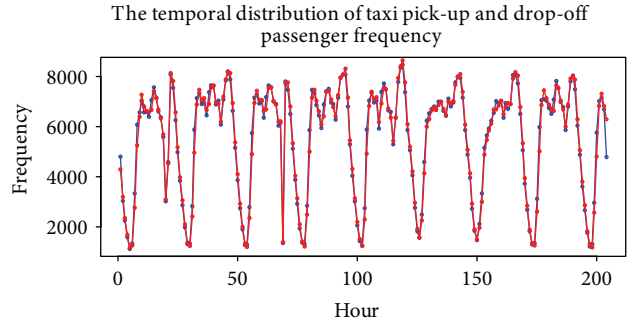


Figure 5: Taxi operation service frequency analysis.

(1) In weekdays, there are three peak times in one day, which are 9 am to 10 am, 3 pm to 4 pm, and 10 pm to 11 pm, and the highest peak time interval is during 10 pm to 11 pm.

(2) At weekends, there are only two peak times, which are 11 am to 12 pm on Saturday and 2 pm to 3 pm on Sunday, as well as the 10 pm to 11 pm at night. Meanwhile, the beginning time of trip increasing at weekdays is an hour later than weekdays.

(3) The trend of pick-up and drop-off changes with the time is almost the same from Monday to Friday, and the lowest service time is from 5 am to 6 am.

(4) When the service frequency increases, the number of pick-up locations will be higher than the drop-off locations; on the contrary, when the service frequency decreases, the number of pick-up locations will be lower than the drop-off locations.

Here we adopt the service frequency divided by the total taxi numbers ($N = 3198$) and obtain each taxi daily operation service frequency situation. The results are shown in Table 2; we can see that the lowest taxi service frequency is less than 0.4 per hour, and the highest is almost 2.7 per hour. The average number of each taxi operating per day is between 39 and 44 times, which is higher than other cities statistics, such as in Harbin where there are 32 taxi service times for each taxi operating per day in 2008 [3], as well as 31 to 33 taxi service times per day in Beijing in 2007 [34]. During 2006 to 2007, Chen [5] also calculated the average operation time for each taxi per day, which is between 35 and 40 times per taxi per day, which is also lower than that in 2011.

According to the researches of Kamruzzaman and Hine [30, 33], we analyze the correlations between the different measures of activity space area, which are shown in Table 3 and Figure 6.

From Figure 6 and Table 3 we can find that the drop-off locations cover area range bigger than the pick-up locations, which can also reflect the disorder and discrete characteristic of the passenger's destination distribution. Meanwhile, the area of location's SDD is bigger than the SDE, regardless of the pick-up or drop-off locations.

4.3. The Personal Level. In this part, we have calculated the distance between the mean centre of each taxi driver's pick-up location and drop-off location and then analyzed the distribution of these distances and the relationship between the pick-up locations and drop-off locations.

Table 4 shows the statistical information of the entire 3198 taxi vehicle's service frequency and the distance between the mean centre of each taxi driver's pick-up locations and drop-off locations. With the combination of Figure 7, more than 80 percent of the taxi vehicles operation service frequency is between 300 and 500 times during the 204 hours.

Figure 8 shows the distance distribution and the frequency distribution curve of taxi pick-up and drop-off passenger locations' mean centre, respectively. The distance between each taxi vehicle's drop-off and pick-up locations' mean centre is below 5000 meters; however, the mean centre of pick-up (or drop-off) locations' distance between each taxi vehicle and the whole operation is changing from 500 meters to 30000 meters, which means each taxi vehicle's operation discrete and randomness. The density line color is blue, and the normal distribution line color is red.

Table 2: Each taxi driver's daily average operation service frequency statistics.

Time periods	April 18 Monday		April 19 Tuesday		April 20 Wednesday		April 21 Thursday		April 22 Friday		April 23 Saturday		April 24 Sunday		April 25 Monday		April 26 Tuesday	
	Drop	Pick	Drop	Pick	Drop	Pick	Drop	Pick	Drop	Pick	Drop	Pick	Drop	Pick	Drop	Pick	Drop	Pick
0:00–0:59	1.35	1.50	1.62	1.56	1.68	1.61	1.67	1.60	1.70	1.65	1.91	1.83	1.90	1.83	1.67	1.59	1.58	1.53
1:00–1:59	1.00	0.95	1.24	1.20	1.28	1.21	1.26	1.21	1.35	1.27	1.63	1.58	1.59	1.53	1.24	1.16	1.19	1.13
2:00–2:59	0.74	0.71	0.96	0.90	0.91	0.86	0.98	0.92	0.99	0.95	1.32	1.27	1.30	1.24	0.89	0.84	0.90	0.83
3:00–3:59	0.53	0.50	0.66	0.62	0.63	0.59	0.66	0.61	0.69	0.64	0.93	0.87	0.91	0.86	0.59	0.57	0.57	0.55
4:00–4:59	***0.36***	***0.35***	0.43	***0.41***	0.42	***0.40***	0.44	0.43	0.47	0.45	0.60	0.57	0.60	0.58	0.42	***0.41***	0.41	***0.38***
5:00–5:59	0.39	0.42	***0.39***	0.42	***0.37***	0.41	***0.38***	***0.42***	***0.39***	***0.40***	***0.49***	***0.49***	***0.47***	***0.46***	***0.39***	0.42	***0.37***	0.41
6:00–6:59	0.86	1.04	0.75	0.89	0.74	0.87	0.78	0.89	0.72	0.86	0.70	0.78	0.62	0.67	0.82	0.97	0.81	0.93
7:00–7:59	1.64	1.90	1.55	1.84	1.54	1.80	1.56	1.83	1.54	1.84	1.29	1.43	0.94	1.05	1.57	1.87	1.50	1.80
8:00–8:59	2.00	2.05	2.15	2.23	2.17	2.24	2.24	2.34	2.14	2.20	1.88	1.95	1.53	1.61	2.15	2.21	2.14	2.19
9:00–9:59	***2.27***	***2.20***	***2.34***	***2.30***	***2.33***	***2.28***	***2.34***	***2.29***	***2.31***	***2.30***	1.97	2.04	1.71	1.77	***2.33***	***2.22***	***2.29***	***2.24***
10:00–10:59	2.13	2.06	2.20	2.16	2.21	2.16	2.20	2.15	2.22	2.18	2.08	2.10	1.83	1.85	2.23	2.20	2.13	2.09
10:00–11:59	2.07	2.06	2.23	2.17	2.21	2.20	2.07	2.01	2.24	2.20	2.11	2.13	1.91	1.95	2.15	2.11	1.97	1.50
12:00–12:59	2.06	2.00	2.08	2.01	2.07	1.99	1.90	1.86	1.93	1.85	2.09	2.09	2.09	2.05	2.10	2.04		
13:00–13:59	2.09	2.21	2.16	2.31	2.09	2.25	2.01	2.15	2.17	2.31	2.14	2.19	2.09	2.19	2.07	2.21		
14:00–14:59	***2.32***	***2.37***	2.36	2.39	2.36	***2.39***	2.31	2.32	***2.36***	***2.42***	***2.19***	***2.19***	2.14	2.17	***2.39***	***2.45***		
15:00–15:59	2.23	2.24	***2.39***	***2.39***	***2.36***	2.37	***2.32***	***2.35***	2.34	2.34	2.09	2.08	***2.20***	***2.19***	2.36	2.35		
16:00–16:59	2.09	2.07	2.18	2.15	2.19	2.19	2.21	2.17	2.11	2.08	2.02	2.01	2.03	1.99	2.20	2.19		
17:00–17:59	1.98	1.99	2.16	2.20	2.15	2.16	2.12	2.16	1.99	2.06	2.19	2.22	2.12	2.17	2.09	2.12		
18:00–18:59	1.78	1.75	1.94	1.90	1.94	1.89	2.00	1.96	1.70	1.65	2.17	2.13	2.17	2.10	1.88	1.84		
19:00–19:59	0.97	0.95	2.25	2.22	1.94	1.92	2.27	2.24	2.00	1.98	2.19	2.18	2.16	2.16	2.16	2.13		
20:00–20:59	1.42	1.44	2.47	2.45	0.44	0.42	2.48	2.49	2.48	2.43	2.40	2.43	2.47	2.52	2.48	2.44		
21:00–21:59	***2.54***	***2.52***	***2.56***	***2.56***	2.41	***2.44***	2.53	2.53	2.63	2.61	2.50	***2.50***	***2.56***	***2.52***	***2.50***	***2.52***		
22:00–22:59	2.45	2.36	2.54	2.46	***2.43***	2.33	***2.60***	***2.53***	***2.70***	***2.65***	***2.53***	2.48	2.51	2.42	2.46	2.34		
23:00–23:59	2.06	1.96	2.17	2.07	2.13	2.03	2.24	2.13	2.43	2.30	2.31	2.24	2.12	2.04	1.97	1.89		
Total	39.33	39.6	43.78	43.81	41	41.01	43.57	43.59	43.6	43.62	43.73	43.78	41.97	41.92	43.11	43.09	15.86	15.58

Table 3: The statistics of the different measures of activity space area.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Pick-up location SDD area (Km^2)	1227	1410	1477	1483	1547	1796
Pick-up location SDE area (Km^2)	1075	1200	1257	1266	1318	1532
Drop-off location SDD area (Km^2)	1344	1481	1544	1580	1655	2727
Drop-off location SDE area (Km^2)	1129	1271	1336	1361	1427	2173

Table 4: The statistics of the different measures of each driver's operation activity space.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Service frequency (times)	150	302	369	356	413	597
Distance between the two mean centres (meters)	0.0	252.1	392.5	432.4	565.5	5121.8
Pick-up location SDC area (Km^2)	34.55	477.32	629.26	623.68	782.41	1947.05
Pick-up location SDE area (Km^2)	20.66	345.98	451.03	452.95	555.66	2806.99
Drop-off location SDC area (Km^2)	35.4	563.6	727.2	727.1	889.5	6279.4
Drop-off location SDE area (Km^2)	18.56	425.63	542.24	549.60	663.40	2921.57

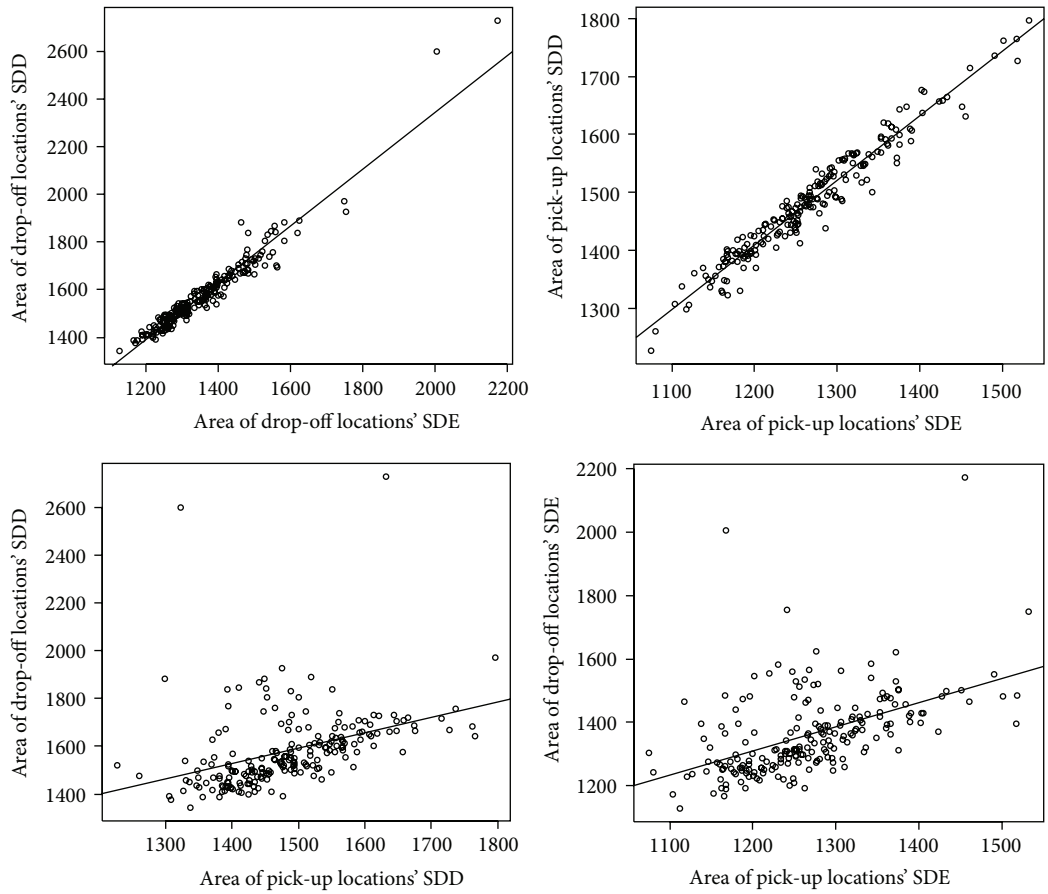


Figure 6: The correlations between the different measures of activity space area.

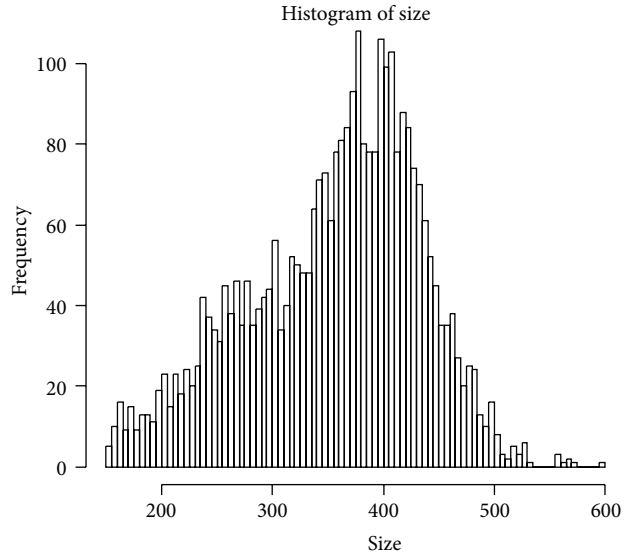


Figure 7: Taxi's service frequency distribution graph.

From Figure 8, we can also find that the frequency distribution of taxi pick-up and drop-off passenger locations' mean centre has some aggregate character, so we adopt nearest neighbor hierarchical clustering method to calculate each taxi vehicles' pick-up and drop-off location, which are shown in Figures 9 and 10, respectively.

The nearest neighbor hierarchical clustering method can identify small geographical environments where there are concentrated pick-up or drop-off activities. And the linkages between several small clusters can be seen through the second- and third-order clusters; there are different scales to the clustering of points' different geographical levels. Also the analysis can show the number of points found per cluster and the density of points found per cluster.

From Figure 11 we can find that the drop-off locations cover area range bigger than the pick-up locations, which can also reflect the disorder of the passenger's destination distribution.

Kamruzzaman and Hine [33] had proposed the fullness of activity spaces to describe the individual SDE feature class. This can be seen in Figure 12.

4.4. The Comparison between the Top and Last 1% Taxi Drivers' Level. From Figure 13 and Table 5 we can find the difference between the top 1% and last 1% taxi searching time/operation time. The top 1% taxi drivers have shorter time both in searching and in operation time than the last 1% taxi drivers. There is more than 22 percent of the last 1% taxi driver's

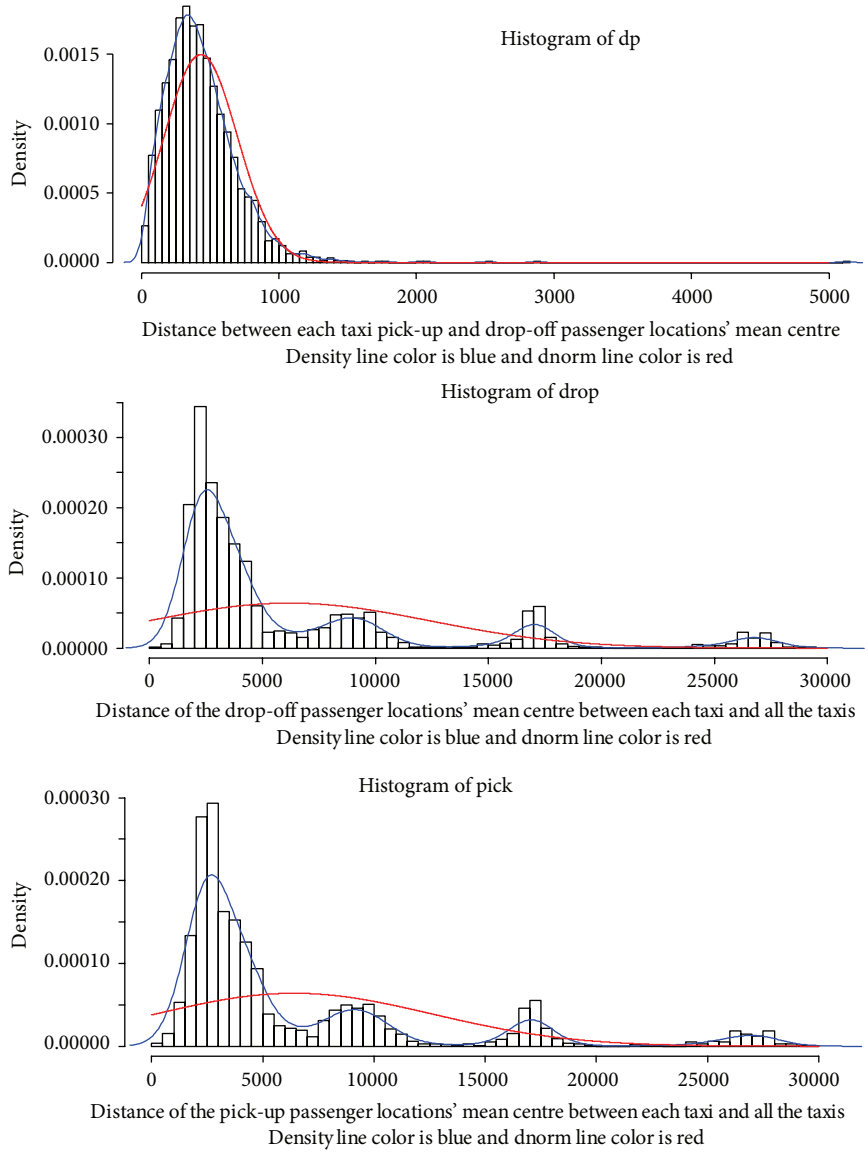


Figure 8: The frequency distribution curve of taxi pick-up and drop-off passenger locations' mean centre.

Table 5: The statistics of the top 1% and last 1% taxi searching time/operation time.

Percent	Searching time (minutes)		Operation time (minutes)	
	Top 1%	Last 1%	Top 1%	Last 1%
25%	3	7	6	7
50%	6	20	9	12
75%	14	42	15	24
85%	23	>60	20	35

searching time which is above 60 minutes; furthermore, the top 1% taxi drivers are only 5 percent above 60 minutes.

Referring to the concept of 85 percent vehicle speed, here we calculate the 85 percent searching time and operation time of the top 1% and last 1% taxi drivers, which are shown in Table 5. The top 1% taxi drivers' 85 percent operation and searching time are at the interval from 20 to 23 minutes; however, the last 1% taxi drivers' 85 percent operation and searching time is 35 and more than 60 minutes, respectively. That is why the top 1% taxi drivers can operate more service time than the last 1% taxi drivers.

5. Discussions

In the time level, we have found that taxi's operations in weekdays and weekends have different operation status; in weekdays there are three peak hours during one day and in weekends there are only two peak hours. Meanwhile, the time level of taxis operation reflects its difference with the normal traffic demand peak hour distribution. There are always two peak hours in the normal traffic demand (as the dual hump curve), which corresponds to going to work and getting off

Figure 9: (a) First-, second-, and third-order pick-up locations' mean centre "hot spots": ellipses. (b) First-, second-, and third-order pick-up locations' mean centre "hot spots": convex hull.

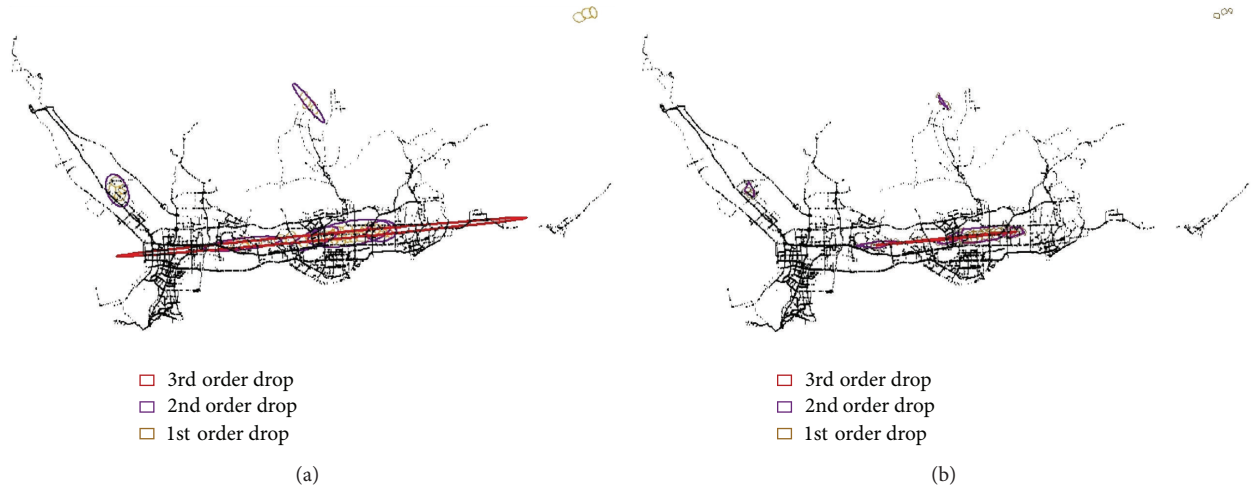


Figure 10: (a) First-, second-, and third-order drop-off locations' mean centre "hot spots": ellipses. (b) First-, second-, and third-order drop-off locations' mean centre "hot spots": convex hull.

work; in addition to this, the taxi operation demand has the third peak hour at night in weekdays, which is also the biggest peak hour. So in the traffic demand forecasting and planning process, we should pay more attention to this new phenomenon and take the taxi demand into account separately.

In the study field of taxi service model, researchers usually adopt virtual customer origin-destination demand pattern to analyze the model, such as Arnott [35], Yang and Wong [36], and Wong et al. [37]. While with the taxi GPS traces data, we can obtain the customers real-time origin and destination demand, which can assist researchers to validate the taxi service model.

On the personal (each taxi driver) level, the distance between each taxi driver's pick-up locations' mean centre and drop-off locations' mean centre is almost around 1000 meters, which reflects the taxi driver's cruise and search for next passenger' pick-up location will be around the previous drop-off location. The cluster analysis also identifies the small geographical environments where there are concentrated pick-up and drop-off activities. This finding can be useful in the taxi dispatching and setting taxi stop location chosen.

Through the comparison between the top and last 1% taxi drivers' operation status, the difference between the searching time and operation time statistics can help the taxi operation company to make strategy for increasing the taxi driver's daily operation service frequency, which will also reduce the passenger's waiting time and increase their satisfaction.

6. Conclusions

This paper first took the taxi driver's pick-up and drop-off passengers' location as the objective, based on large-scale taxi GPS trace data analyzing the urban taxi driver's temporal

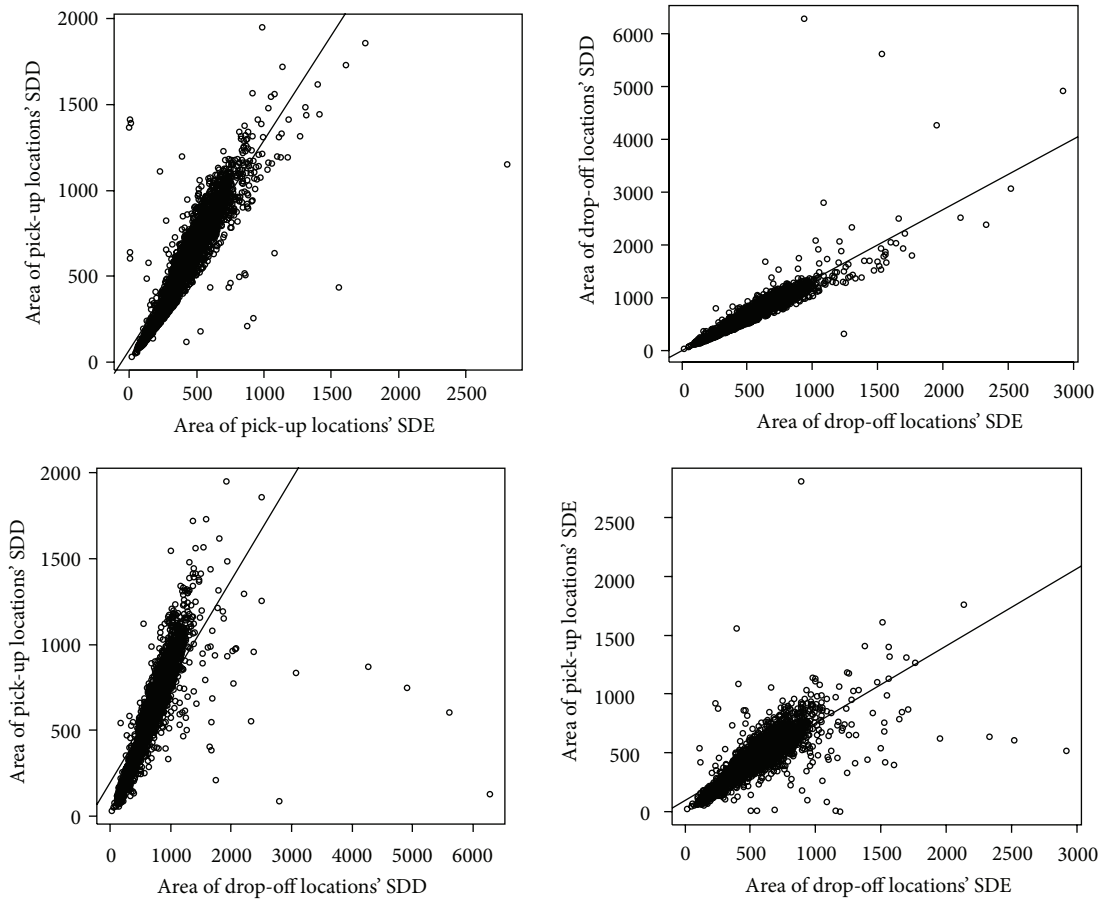


Figure 11: The correlations between the different measures of activity space area.

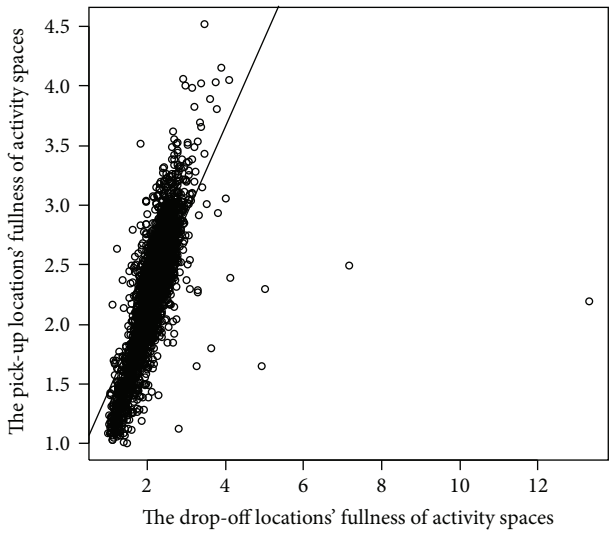


Figure 12: The fullness of activity spaces of each taxi driver's pick-up location and drop-off location.

and spatial distribution. From the time level, we found the difference between taxi daily operation pattern (weekday and weekends), continuous time in one day, passenger's in-vehicle time, and taxi's operation frequency. From the personal level (each taxi vehicle), we have identified the taxi driver's searching pattern, including searching activity space distribution and the relationship between the pick-up locations and the drop-off locations. Through the comparison between the

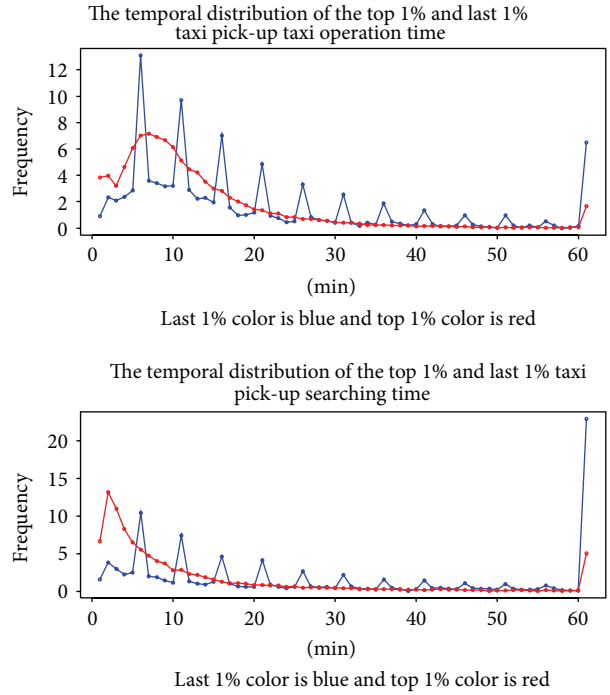


Figure 13: The temporal distribution of the top 1% and last 1% taxi searching time/operation time.

top and last 1% taxi drivers difference on searching time, operation time, and service frequency, which can obtain the pattern of top and last 1% taxi drivers operation pattern. This paper can be helpful for the urban taxi management authority to identify hot spots and develop policies that reduce the taxi drivers' searching time, road congestion, and taxi air pollution emission. The measurements of the taxi driver's activity space analysis also make some contribution to the geographical dynamics analysis.

The next step is to identify the relationship between the taxi driver's pick-up activity and the land-use attribute, as well as each taxi driver's temporal distribution, which will help us to understand taxi drivers' searching behavior pattern.

Conflict of Interests

The authors declare that they have no conflict of interests.

Acknowledgments

This research was supported by the National High-Tech Research and Development Program of China (863 Program, no. 2012AA112310), the China Postdoctoral Science Foundation funded project (no. 2013M540299), the Fundamental Research Funds for the Central Universities (no. HIT.NSRIF.2015075), and the Heilongjiang Postdoctoral Fund (no. LBH-Z13111). The authors would like to express their sincere thanks to Professor Yanfeng Ouyang (University of Illinois at Urbana-Champaign) and the anonymous reviewers for their helpful comments and valuable suggestions on this paper.

References

[1] S. S. Han, "Managing motorization in sustainable transport planning: the Singapore experience," *Journal of Transport Geography*, vol. 18, no. 2, pp. 314–321, 2010.

[2] BTRC, "Beijing Transport," Annual Report, Beijing Transportation Research Center, Beijing, China, 2010, (Chinese).

[3] S. An, X. Hu, and J. Wang, "Urban taxis and air pollution: a case study in Harbin, China," *Journal of Transport Geography*, vol. 19, no. 4, pp. 960–967, 2011.

[4] X. Hu and J. Feng, "Research on characteristics of taxi traffic based on GPS data," *Urban Transport of China*, vol. 5, no. 2, pp. 91–95, 2007 (Chinese).

[5] L.-H. Chen, *Study of taxi operation trait based on GPS float data [M.S. thesis]*, Tongji University, 2008 (Chinese).

[6] B. Jiang, J. Yin, and S. Zhao, "Characterizing the human mobility pattern in a large street network," *Physical Review E*, vol. 80, no. 2, Article ID 021136, 2009.

[7] J. Weng, W. Liu, Z. Chen, and J. Rong, "Research on floating car data based taxi operation and management," *Journal of Beijing University of Technology*, vol. 36, no. 6, pp. 779–784, 2010 (Chinese).

[8] L. Liu, C. Andris, and C. Ratti, "Uncovering cabdrivers' behavior patterns from their digital traces," *Computers, Environment and Urban Systems*, vol. 34, no. 6, pp. 541–548, 2010.

[9] J. Hu, H. Xie, and G. Zhong, "A method for determining taxi reasonable scale based on floating car data," in *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation, Wuhan, China*, pp. 1255–1262, 2011.

[10] Z. Zhang and X. He, "Analysis and application of spatial distribution of taxi service in city subareas based on taxi GPS data," in *Proceedings of the 11th International Conference*

of Chinese Transportation Professionals: Towards Sustainable Transportation Systems (ICCTP '11), pp. 1232–1243, Nanjing, China, August 2011.

[11] Y. Zheng, J. Yuan, W. Xie, X. Xie, and G. Sun, "Drive smartly as a taxi driver," in *Proceedings of the Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing (UIC-ATC '10)*, pp. 484–486, IEEE Computer Society, Washington, Wash, USA, 2010.

[12] D. Luo, *Urban mixed traffic network equilibrium analysis under the influence of taxi services*, Central South University, 2009 (Chinese).

[13] R. G. Golledge and R. J. Stimson, *Spatial Behavior*, The Guilford Press, New York, NY, USA, 1997.

[14] K. W. Axhausen, "Travel diaries: an annotated catalogue," Working Paper, Imperial College, London, UK, 1994.

[15] D. Wang and T. Cheng, "A spatio-temporal data model for activity-based transport demand modelling," *International Journal of Geographical Information Science*, vol. 15, no. 6, pp. 561–585, 2001.

[16] Y. O. Susilo and R. Kitamura, "Analysis of the day-to-day variability in the individual's action space: an exploration of the six-week Mobidrive travel diary data," *Transportation Research Record*, no. 1902, pp. 124–133, 2005.

[17] M. C. González, C. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.

[18] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[19] M. Giraudo and P. Peruch, "Spatio-temporal aspects of the mental representation of urban space," *Journal of Environmental Psychology*, vol. 8, no. 1, pp. 9–17, 1988.

[20] P. Peruch, M.-D. Giraudo, and T. Garling, "Distance cognition by taxi drivers and the general public," *Journal of Environmental Psychology*, vol. 9, no. 3, pp. 233–239, 1989.

[21] H. J. Spiers and E. A. Maguire, "The dynamic nature of cognition during wayfinding," *Journal of Environmental Psychology*, vol. 28, no. 3, pp. 232–249, 2008.

[22] Y. Wakabayashi, S. Itoh, and Y. Nagami, "The use of geospatial information and spatial cognition of taxi drivers in Tokyo," *Procedia Social and Behavioral Sciences*, vol. 21, pp. 353–361, 2011.

[23] J. Pailhous, *La Representation de l'Espace Urbain*, Presses Universitaires de France, Paris, France, 1970.

[24] J. A. Michon, "Processing of temporal information and the cognitive theory of time experience," in *The Study of Time*, J. T. Fraser, F. C. Haber, and C. H. Muller, Eds., pp. 242–257, Springer, Heidelberg, Germany, 1972.

[25] E. Murakami and D. P. Wagner, "Can using global positioning system (GPS) improve trip reporting?" *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 2-3, pp. 149–165, 1999.

[26] Y. Yue, H. Wang, B. Hu, Q. Li, Y. Li, and A. G. O. Yeh, "Exploratory calibration of a spatial interaction model using taxi GPS trajectories," *Computers, Environment and Urban Systems*, vol. 36, no. 2, pp. 140–153, 2012.

[27] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. J. Pazzani, "An energy-efficient mobile recommender system," in *Proceeding of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*, pp. 899–908, New York, NY, USA, July 2010.

[28] J. Lee and D. W. S. Wong, *Statistical Analysis with Arcview GIS*, John Wiley & Sons, New York, NY, USA, 2001.

[29] Ned Levine & Associates, *CrimeStat: A Spatial Statistics Program for the Analysis of Crime Incident Locations (v 3.0)*, Ned Levine & Associates, Houston, Tex, USA; National Institute of Justice, Washington, DC, USA, 2004.

[30] M. Kamruzzaman and J. Hine, "Participation index: a measure to identify rural transport disadvantage?" *Journal of Transport Geography*, vol. 19, no. 4, pp. 882–899, 2011.

[31] R. W. Sinnott, "Virtues of the Haversine," *Sky and Telescope*, vol. 68, no. 2, article 159, 1984.

[32] K. Gade, "A non-singular horizontal position representation," *Journal of Navigation*, vol. 63, no. 3, pp. 395–417, 2010.

[33] M. Kamruzzaman and J. Hine, "Analysis of rural activity spaces and transport disadvantage using a multi-method approach," *Transport Policy*, vol. 19, no. 1, pp. 105–120, 2012.

[34] L. Song, D. Zhang, J. Chen, J. An, and Y. Su, "Analysis of taxi operation characteristics with traffic control," *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 6, pp. 127–121, 2008 (Chinese).

[35] R. Arnott, "Taxi travel should be subsidized," *Journal of Urban Economics*, vol. 40, no. 3, pp. 316–333, 1996.

[36] H. Yang and S. C. Wong, "A network model of urban taxi services," *Transportation Research B: Methodological*, vol. 32, no. 4, pp. 235–246, 1998.

[37] K. I. Wong, S. C. Wong, and H. Yang, "Modeling urban taxi services in congested road networks with elastic demand," *Transportation Research B: Methodological*, vol. 35, no. 9, pp. 819–842, 2001.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 928980, 10 pages
http://dx.doi.org/10.1155/2014/928980

Research Article

Measurement of International Roughness Index by Using Z-Axis Accelerometers and GPS

Yuchuan Du, Chenglong Liu, Difei Wu, and Shengchuan Jiang

Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China

Correspondence should be addressed to Shengchuan Jiang; jiangsc87@gmail.com

Received 3 March 2014; Revised 21 May 2014; Accepted 29 May 2014; Published 30 June 2014

Academic Editor: Andy H. F. Chow

The International Roughness Index (IRI) is a well-recognized standard in the field of pavement management. Many different types of devices can be used to measure the IRI, but these devices are mainly mounted on a full-size automobile and are complicated to operate. In addition, these devices are expensive. The development of methods for IRI measurement is a prerequisite for pavement management systems and other parts of the road management industry. Based on the quarter-car model and the vehicle vibration caused by road roughness, there is a strong correlation between the in-car Z-axis acceleration and the IRI. The variation of speed of the car during the measurement process has a large influence on IRI estimation. A measurement system equipped with Z-axis accelerometers and a GPS device was developed. Using the self-designing measurement system based on the methodology proposed in this study, we performed a small-scale field test. We used a one-wheel linear model and two-wheel model to fit the variation of the Z-axis acceleration. The test results demonstrated that the low-cost measurement system has good accuracy and could enhance the efficiency of IRI measurement.

1. Introduction

The IRI was developed in 1986 using the results of the International Road Roughness Experiment performed in Brazil in 1982 [1]. Since then, the IRI has become a well-recognized standard for the measurement of road roughness. The main advantages of the IRI are that it is stable over time and transferable throughout the world.

The IRI is an index defined by applying the algorithm proposed by Sayers [2] to a measured realization of the longitudinal profile. The measurement of roughness is quite difficult and complex because it depends on the vehicular characteristics in addition to the actual pavement situation [3]. Moreover, the road roughness levels are readily affected by vehicle structures and driving speed. During the course of half a century of development, engineers and scientists have invented several techniques and methods for measuring road roughness. The measurement devices can be divided into four general types [4]: response-type road roughness measuring systems (RTRRMS), direct profile measurements, indirect profile measurements, and subjective rating panels. Type I devices measure the pavement roughness by correlating the RTRRMS measurements with the IRI calculated from a profile, for example, using a bump integrator or NAASRA roughness meter. Type II devices measure the road profile directly, which involves measuring each wheel track separately, for example, using a 3-meter long beam or laser road surface tester (LRST). Type III devices measure the longitudinal profile over the wavelength range of interest, for example, using a General Motors Research (GMR) profilometer. Type IV devices evaluate the pavement quality based on assessment guidance and personal experience. The common measurement methods are compared in Table 1.

Most highway agencies collect IRI data using a laser road surface tester or GMR profilometer. This equipment measures surface profiles at traffic speeds and provides excellent results for use in network analysis for pavement management systems. However, because these devices are mounted on a full-size van, automobile, or trailer, it is difficult to use them on the roadway for short periods of time. In addition, these devices are rather expensive and delicate. For these reasons, they are not effective for providing feedback to contractors' crews. Hajek et al. [5] analyzed the influence of several different factors on the IRI data that was collected

Table 1: Characteristics of common road roughness measurement methods.

Method of measurement	Principal of operation	Measurement device	Advantage	Disadvantage
3-Meter Long Beam	Direct profile measurement	Aluminum beam	Simple, reliable data collection	Inefficient, lower mechanization
Level	Direct profile measurement	Level and rod	Time-stable, straightforward	Time-consuming, inconvenient
Bump integrator	RTRRMS	LVDT, displayer, and a car/trailer	Simple, fast	Time-instable, affected by vehicle vibration, performs at the same speed
Laser road surface tester	Direct profile measurement	Car with laser device and calculators	Straightforward, efficient	High cost of operation and maintenance
GMR profilometer	Indirect profile measurement	LVDT, accelerometer, potentiometer, and a van	Convenient, efficient	Precise instrument required, high cost of operation and maintenance

and made several recommendations for IRI measurement. IRI measurement at the network level has become a routine practice for many road agencies in recent years. On the other hand, IRI measurement at the project level is also required, primarily for accepting or price-adjusting paving contractors' products. The development of methods for IRI measurement is a prerequisite for a pavement management system (PMS) and other parts of the road management industry.

An accelerometer is a device that measures the acceleration in one, two, or three orthogonal axes, and they are used widely in the fields of civil engineering, biology, and industry. Accelerometers can be used to measure the vehicle driving status, where they facilitate evaluations of the overall vehicle performance and response. This information can then be used to make adjustments to various vehicle subsystems, as necessary [6–8]. Accelerometers can also be used to measure seismic activity, inclination, machine vibration, dynamic distance, and speed, with or without the influence of gravity [9, 10].

This study focused on building a model for estimating the IRI, as well as developing an effective and low-cost system for measuring the IRI. The steps of this study, which are reflected by this organization of this paper, are as follows.

(i) Introduction to the principle of using Z-axis accelerometers to measure IRI.

(ii) Modeling the relationship between the variation in the in-car Z-axis acceleration and the IRI.

(iii) Development of a self-designed measurement system with Z-axis accelerometers and a GPS device.

(iv) Field testing results.

2. The Principle of Using Z-Axis Accelerometers to Measure IRI

2.1. International Roughness Index. To study the effects of the road pavement characteristics on the ride quality, we need a valid measurement of the pavement roughness and a comprehensive index to evaluate both road roughness and the ride quality, as perceived by road users. To meet this objective, a fundamental index (IRI) was established by the World Bank in 1986.

The most often employed and most useful model of a vehicle suspension system for developing a low-level controller for a vehicle suspension is the quarter-car model, in which only one quarter of the vehicle is taken into consideration. The model is two-dimensional because only movement in the Z direction is taken into consideration. It consists basically of a single wheel, which is represented in the form of a spring. A general representation of a two-degree-of-freedom quarter-car model is shown in Figure 1.

In this model, the sprung and unsprung masses that correspond to one corner of the vehicle are denoted by m_s and m_u, respectively. The suspension system is represented by a linear spring of stiffness K_s and a linear damper with a damping rate C_s, while the tire is modeled by a linear spring of stiffness K_t. Y is the input. By drawing free body diagrams and applying Newton's Second law, we obtain the following differential equations [11]:

$$m_s\ddot{Z}_s + C_s\left(\dot{Z}_s - \dot{Z}_u\right) + K_s\left(Z_s - Z_u\right) = 0,$$
$$m_u\ddot{Z}_u + C_s\left(\dot{Z}_u - \dot{Z}_s\right) + K_s\left(Z_u - Z_s\right) + K_t\left(Z_u - Y\right) = 0. \quad (1)$$

We can eliminate the masses from the equations, leaving the equations in this form:

$$\ddot{Z}_s + C_s\left(\dot{Z}_s - \dot{Z}_u\right) + K_1\left(Z_s - Z_u\right) = 0,$$
$$u\ddot{Z}_u + C\left(\dot{Z}_u - \dot{Z}_s\right) + K_2\left(Z_u - Z_s\right) + K_1 Z_u = K_1 Y. \quad (2)$$

Using the response of the quarter-car model at a travel speed of 80 km/h, calculated for each point along the distance of travel, the IRI can be defined as follows [1]:

$$\mathrm{IRI} = \frac{1}{L}\int_0^L \left|Z_s - Z_u\right| dx, \quad (3)$$

where L is the distance along the road on which measurement is performed.

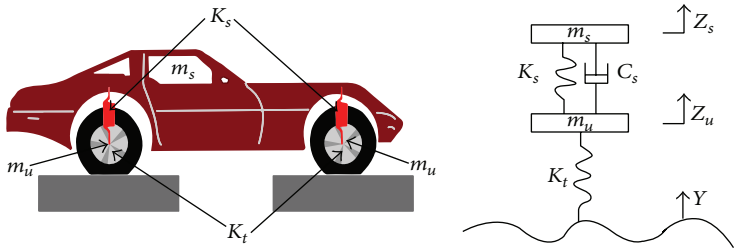


FIGURE 1: Quarter-car vehicle model. (a) Vehicle representation; (b) simplified representation.

2.2. Power Spectral Density. Power spectral density is a probabilistic method, which is a measure of the mean squared value of a random variable. In general, this method is used for random vibration analysis, which describes how the power of a signal or time series is distributed over different frequencies.

If we regard the pavement as a continuous surface, the IRI sequence of road profiles is a random phenomenon that obeys a zero-mean Gaussian distribution. This can be regarded as a stationary stochastic process. Therefore, it is appropriate to describe the pavement characteristics using the power spectral density.

Define $X(t)$ as a stationary stochastic process and $R_x(t)$ as its autocorrelation function, and if the Fourier transform of $R_x(t)$ exists, then

$$S_x(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} R_x(\tau)\, e^{-i\omega\tau} d\tau. \tag{4}$$

$S_x(\omega)$ is the PSD of $X(t)$, ω is angular frequency, and $R_{x(\tau)}$ can be described by the inverse Fourier transform of the power spectral density:

$$R_x(\tau) = \int_{-\infty}^{+\infty} S_x(\omega)\, e^{i\omega\tau} d\omega. \tag{5}$$

These two equations form a Fourier pair called the Wiener-Khintchine formula [12, 13].

2.3. Correlation between IRI and PSD. To evaluate the dynamic actions transmitted from a vehicle's movement on the road surface, we need to develop equations that express the physical state of the systems. According to our definition, a quarter-car model meets the condition of a LTI (linear time invariant) system. Regarding $Y(t)$ as systematic excitation, $Z_s(t)$ and $Z_u(t)$ as systematic response, based on the transmission property of LTI system, the frequency response function can be solved by means of Laplace transform [13]:

$$\begin{aligned} H_{Z_sY}(\omega) &= \frac{k_t\left(jC_s\omega + k_s\right)}{\Delta(\omega)}, \\ H_{Z_uY}(\omega) &= \frac{k_t\left(-m_s\omega^2 jC_s\omega + k_s\right)}{\Delta(\omega)}, \\ \Delta(\omega) &= \left(-m_s\omega^2 jC_s\omega + k_s\right)\left(-m_u\omega^2 jC_s\omega + k_s + k_u\right) \\ &\quad - \left(jC_s\omega + k_s\right)^2. \end{aligned} \tag{6}$$

The pavement roughness is assumed to be a random stationary variable and, using the exciting force of the quarter-car model, the systematic response can be expressed as follows:

$$\begin{aligned} Z_s(t) &= H_{Z_sY}(\omega)\, Y(t), \\ Z_u(t) &= H_{Z_uY}(\omega)\, Y(t). \end{aligned} \tag{7}$$

Define $Z(t) = Z_s(t) - Z_u(t)$, and regard $Z(t)$ as the response of the system. One has

$$\begin{aligned} Z(t) &= H_{Z_sY}(\omega)\, Y(t) - H_{Z_uY}(\omega)\, Y(t) \\ &= \frac{k_t m_s \omega^2}{\Delta(\omega)} Y(t) = H_{ZY}(\omega)\, Y(t). \end{aligned} \tag{8}$$

As mentioned previously, large volumes of measurement data show that the pavement roughness conforms to the fundamental hypothesis of a vibration source random field, which is a zero-mean local ergodic Gaussian random field. According to the principles of a LTI system, the systematic response is also a random stationary variable, so we may also set $S_Y(\omega)$ as the PSD of the pavement roughness, and thus we can determine the PSD and the mean squared value of $Z(t)$. Consider

$$\begin{aligned} S_z(\omega) &= \left|H_{ZY}(\omega)\right|^2 S_y(\omega), \\ \varphi_z^2 = R_z(0) &= \int_{-\infty}^{\infty} \left|H_{ZY}(\omega)\right|^2 S_y(\omega)\, dw. \end{aligned} \tag{9}$$

The mean squared value represents the equivalent amplitude of signals, and thus it can represent the size of the signal amplitude because they are approximately equal and they conform to a linear correlation. Thus, we may derive the expression for the IRI as follows [11]:

$$\mathrm{IRI} = A \times \varphi_z = A \times \sqrt{\int_{-\infty}^{\infty} \left|H_{ZY}(\omega)\right|^2 S_y(\omega)\, dw}. \tag{10}$$

According to the random process theory, $\omega = 2\pi f$, $n = f/v$, and $x = vt$, and thus a different expression for the PSD can be obtained:

$$S_x(n) = 2\pi v S_x(\omega). \tag{11}$$

Acceleration is the second derivative of vertical displacement. Consider

$$S_a(\omega) = \left(\frac{\omega}{v}\right)^4 S_x(\omega),$$
$$\text{IRI} = A \times \varphi_z = A \times \sqrt{\int_{-\infty}^{\infty} \left|H_{ZY}(\omega)\right|^2 \left(\frac{\omega}{v}\right)^4 S_a(\omega)\, dw}. \quad (12)$$

Therefore, it can be seen that there is a linear relationship between the IRI and the square root of pavement power spectral density function. Matlab was used to calculate the measured acceleration values. Then, the measured acceleration values were fitted to the known actual road surface roughness values. Finally, we obtained the conversion formula for the measured acceleration values and the IRI.

3. Correction Model Based on Hybrid Sensors

3.1. Measurement of IRI by Multiple Z-Axis Accelerometers. Because the work described in this paper was aimed at establishing a connection between the Z-axis acceleration (in the direction of gravity) and the IRI, we selected for testing and modeling a section of the road network of Shanghai for which the IRI had already been measured by the highway administration in 2012. A strategy of averaging multiple measurements was applied to reduce the influence of measurement errors. In the tests, the Z-axis accelerometers were placed flat in the car to eliminate the effect of gravity caused by the weight of the components; at the same time, the accelerometers were completely fixed inside the car, so that they would fully reflect the sprung vibration of the car. In accordance with the quarter-car model, accelerometers were fixed separately at the center of the vehicle and over the four wheels. Each link in the road network was tested four times.

3.1.1. Single-Wheel Model. The square root of each value of the power spectral density of the road-induced acceleration was calculated using Matlab, and then values that were obviously too small or large were eliminated. The averages of the square root were calculated for the left and right wheels. The results are shown in Table 2.

The square root values for the left and right wheels were calculated separately and compared with the standard IRI values. It was found that for both the left and the right wheels, there was a linear relationship between the square root value and the standard IRI. The model obtained for the right wheels was

$$\text{IRI} = 2.1204X_r - 2.8401, \quad (13)$$

where X_r is the square root of the power spectral density of the acceleration for the right wheels on the link. R^2 is 0.9173; the fit of the model is very good. Similarly, the model for the left wheels was

$$\text{IRI} = 1.7887X_l - 3.3525, \quad (14)$$

where X_l is the square root of the power spectral density of the acceleration for the left wheels on the link. R^2 is 0.8707; the fit of the model is again very good.

Table 2: Results for the left and right wheels.

Number of links selected	Left-wheel square root value	Right-wheel square root value	Standard IRI
1	5.1634	4.2223	5.704
2	6.3474	5.2826	9.062
3	4.0195	3.2460	3.451
4	3.9012	3.2666	3.874
5	4.3624	3.4117	3.959
6	3.5366	2.2714	2.281
7	5.2493	4.1268	5.915
8	4.3359	3.5055	4.838
9	4.2007	3.0775	4.340
10	5.0837	3.7241	5.403
11	4.4185	3.1880	3.991
12	4.4069	3.5530	4.290
13	5.0762	4.1863	5.614
14	4.9560	4.2084	5.701
15	3.7991	3.2827	3.629
16	4.0984	3.2920	3.963
17	3.9900	3.3378	4.861
18	3.0861	2.5950	2.517
19	3.5595	3.3126	3.537
20	3.2307	2.5859	2.697

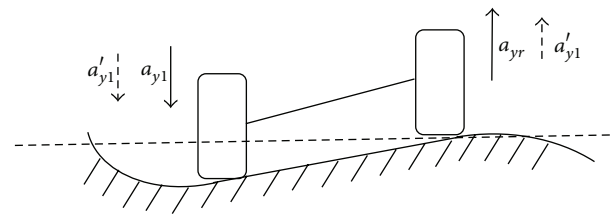


Figure 2: Mutual effects of left and right wheel acceleration.

3.1.2. Two-Wheel Model. During the observation process, we found that there was a relationship between the accelerations of the left and right wheels. The car body is a rigid structure, and thus when driving over the continuous surface of a pavement, if a change is produced on one side, a corresponding change will inevitably appear on the other side, as shown in Figure 2. Thus, although the tires and the suspension system can reduce the effects of the two wheels on each other, the quarter-car model used in the model described above does not provide an accurate simulation. Therefore, we can improve the accuracy of the model by considering the influences of the wheel accelerations on each other.

First, we consider the relationship between the accelerations of the left and right wheels. We used SPSS to analyze the correlations between the accelerations of the left and right wheels, and the results are shown in Table 3.

Although there is a nice relationship between the accelerations of the left and right wheels, the relationship is relatively complex, so we cannot directly establish a fitting model. However, because of the linear relationship between

Table 3: Correlations.

		Right wheel acceleration	Left wheel acceleration
Right wheel acceleration	Correlation coefficient	1.000	.851**
	Significance (2-tailed)	·	.000
	N	31	31
Left wheel acceleration	Correlation coefficient	.851**	1.000
	Significance (2-tailed)	.000	·
	N	31	31

**Correlation is significant at the 0.01 level (2-tailed).

Table 4: Results of multivariate linear fitting using SPSS.

(a) Model Summary

Model	R	R Square	Adjusted R Square	Std. error of the estimate	Change statistics				
					R square change	F change	df1	df2	Sig. F change
1	.971[a]	.942	.936	.3869706	.942	139.067	2	17	.000

[a]Predictors: (constant), right, left.

(b) ANOVA[b]

	Sum of squares	df	Mean square	F	Sig.
1					
Regression	41.650	2	20.825	139.067	.000[a]
Residual	2.546	17	.150		
Total	**44.195**	**19**			

[a]Predictors: (constant), right, left.
[b]Dependent variable: iri.

(c) Coefficients[a]

Model	Unstandardized Coefficients		Standardized coefficients	t	Sig.
	B	Std. error	Beta		
1					
(Constant)	−3.442	.495		−6.948	.000
left	.782	.318	.405	2.456	.025
right	1.300	.369	.581	3.524	.003

[a]Dependent variable: iri.

the square root of the power spectral density of the single-wheel accelerations and the IRI, one might guess that there is also a linear correlation between the power spectral density of the square root of the two-wheel accelerations and the standard IRI. This means that

$$\mathrm{IRI} = a_1 X_l + a_2 X_r + a_3 + \Delta. \quad (15)$$

Therefore, we used multiple linear fitting of the left- and right-wheel square root values with the standard IRI.

We used the package spss17.0 to perform multiple linear regression. The regression results were as follows:

$$\mathrm{IRI} = 0.782 X_l + 1.300 X_r - 3.442. \quad (16)$$

R^2 is 0.942, which satisfies the requirement for precision. Therefore, the goodness of fit for the two-wheel linear model is better than that for the one-wheel linear model, so the two-wheel model can improve the fit (Table 4).

3.2. Velocity Correction. In the actual measurements, because of the limitations imposed by the conditions, the speed of the vehicle was not equal to the specified speed of 80 km/h. At the same time, the relationship between the IRI and the speed is not simply a monotonic increase or decrease, but it is very complex, and it depends on features of the pavement surface such as the waveform. Perera et al. [14] and Xiao-qing and Li-jun [11] measured the IRI values of pavements over a long period at different speeds and compared the measurements with the IRI value at the specified speed of 80 km/h. The results showed that the relationship between the speed and the IRI value is complex and it does not follow any definite proportionality relationship. The experiment showed that the

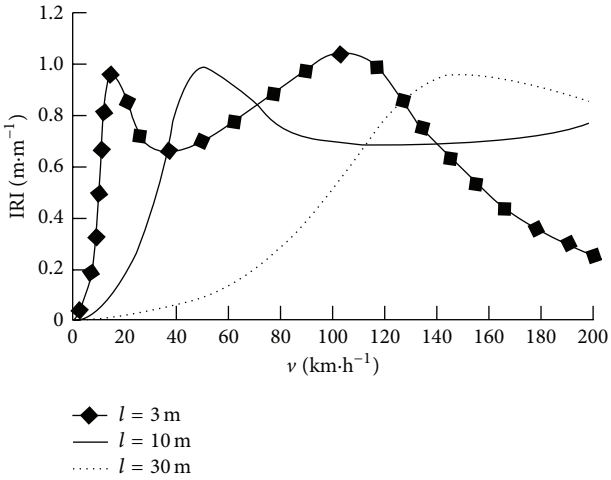


Figure 3: IRI values for different speeds and different waveforms (l is the road wavelength).

IRI varied with the change in velocity (from 0–200 km/h) and wavelength (divided into long wave $l = 30$ m, medium wave $l = 10$ m, and short wave $l = 3$ m), as derived in Figure 3.

Because of the nonlinear relationship between IRI value and speed, the variation of speed during the measurement process has a large influence on the IRI measurement. Therefore, when Z-axis accelerometers are used to measure the IRI, a modification for speed is required. At the same time, it is also important in an IRI measurement to consider the length of road over which measurement is performed and the start and end points. Consequently, we chose a GPS system to allow us to perform modification for speed and so on.

To analyze the influence of speed on the results of the model, we chose an experimental road with light traffic where the traffic speed could reach 80 km/h. It has been found from field tests that when the vehicle speed reaches 60 km/h, certain areas of the vehicle tire are not in contact with the road pavement. This violates our earlier assumption that the distance between the tire and pavement is zero. Therefore, in our experiment, we chose speeds of 20, 30, 40, 50, and 60 km/h on the test road, measured the Z-axis acceleration of the car wheels, and calculated the IRI for the link for each speed separately using the model. If the results were consistent, this would indicate that the model did not depend on the measurement speed. If not, we could modify the model by fitting the calculated results to the vehicle speed. Because the same model was used to calculate the IRI value in each case, the different IRI values for different speeds were related only to the power spectral density values of the acceleration. Therefore, we fitted the power spectral density and speed directly. The results are shown in Table 5.

When the power spectral density value and the speed v are fitted, the model is

$$\text{PSD} = 0.0263v^2 + 0.6027v. \tag{17}$$

R^2 is 0.9991; this model has a very good fit.

Using this model, we can consider the reason why different speeds will lead to different results of the calculation.

Table 5: Values of power spectral density for different speeds.

Speed (km/h)	20	30	40	50	60
Power spectral density	20.9189	41.7322	68.044	115.6324	130.4095

The spatial frequency (in units of m^{-1}) that characterizes the collection frequency of the numerical data is the number of samples per meter. Because the frequency at which the sensor reads data is constant, the spatial frequency of the sequence of acceleration data is related only to the speed. At the same time, because of the small variation of the measuring speed, a quadratic fitting curve is better than a quartic curve. As a result, after modification for the effect of speed, the IRI model is

$$\text{IRI} = 0.782\alpha X_l + 1.300\alpha X_r - 3.442, \tag{18}$$

where α is the correction coefficient for the effect of speed, given by

$$\alpha = \frac{1}{\sqrt{0.0003865v^2 + 0.0009125v}}. \tag{19}$$

4. System Design

According to the fitting model we constructed in Section 3, we developed a coupled system for measuring the IRI. The system comprised two sections: a hardware platform and a software system.

4.1. Description of the Hardware Platform. The hardware platform was used to collect Z-acceleration data for pavements, as well as GPS data, which was transmitted to the software system. The hardware platform comprised the following (Figure 4):

(i) Z-axis accelerometer sensors, type MMA8451Q

(ii) GPS module, type Ublox NEO-6M

(iii) ZigBee modules

(iv) Microcontroller units (MCU), type TC12C5608AD.

The Z-axis accelerometer is a triaxial, intelligent, low-power, mechanical acceleration sensor. This type of sensor can access data from both a low-pass filter and a high-pass filter and therefore greatly reduces the peak data requirements for data analysis and achieves faster data transfer. The sample frequency of the accelerometers was set to 0.1 second, which made the pavement appear more like a continuous surface and this ensured the data processing efficiency. The GPS module contained an integrated audio frequency chip, a baseband chip, and a core CPU. The core controller could connect multiple Z-axis gravity accelerometers and GPS devices. The positional accuracy of GPS is 1 meter, which is adequate for field experiments and road tests.

4.2. Description of the Software Platform. The software system was used to obtain real-time data from hardware devices and it comprised two main components (Figure 5) as follows.

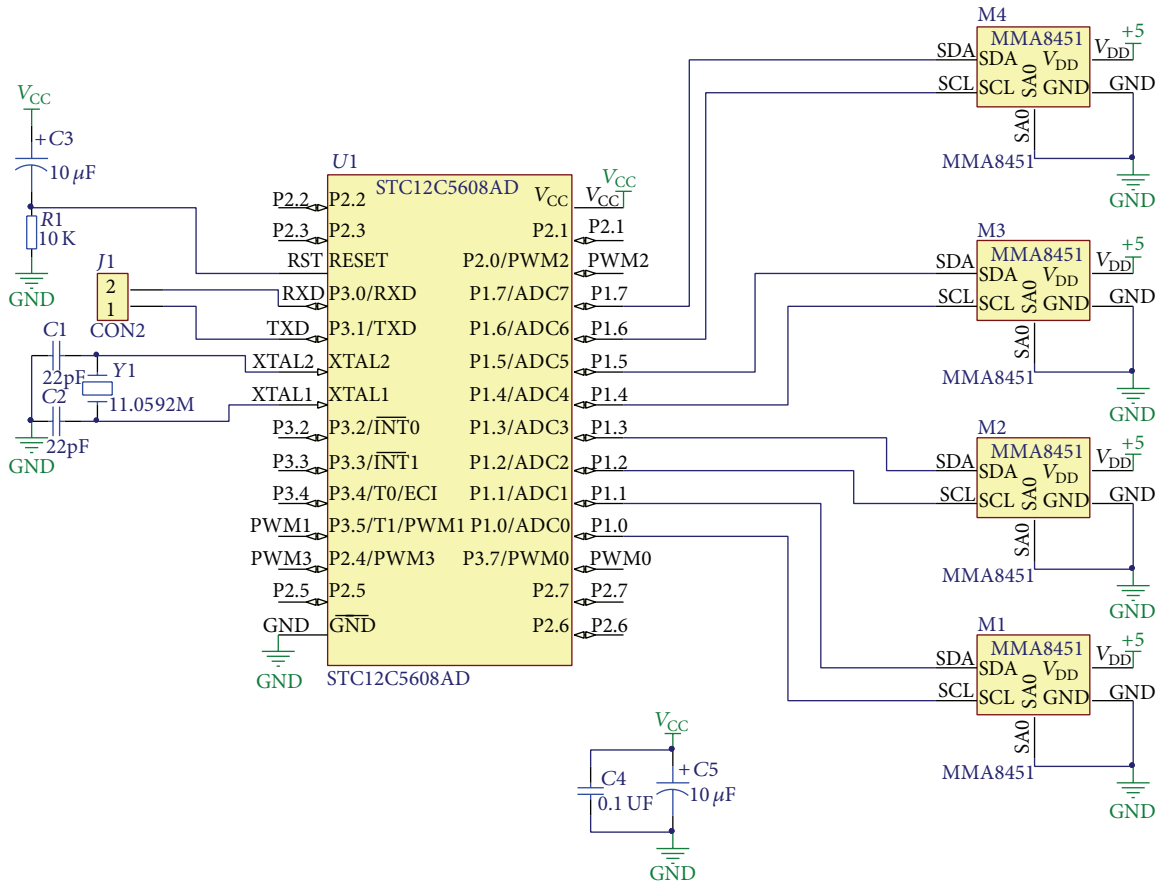


Figure 4: Schematic diagram of the system circuit.

(i) Data acquisition software based on C# for obtaining the acceleration rates from various devices and storing these data.

(ii) A Matlab GUI data processor (based on Matlab GUI) for calculating the IRI (or RQI) using the model constructed above.

4.3. Description of the System Design. The system could record real-time Z-axis accelerations in different pavement conditions at various frequencies, as well as using different parameters, such as those used to describe the driving conditions (Figure 6).

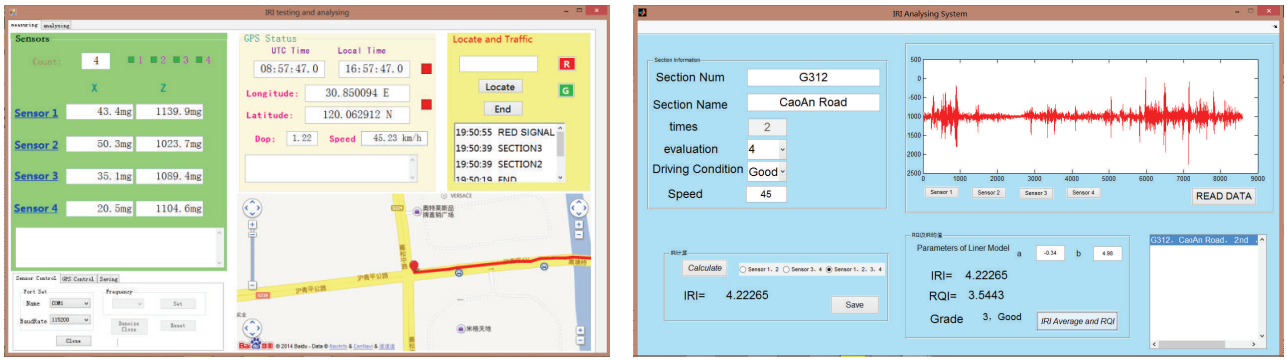
The system design had a two-tier construction. The accelerometers obtained the real-time acceleration and transmitted them to the data acquisition software via a Zigbee module. GPS was also received as geographic information from satellites and transmitted to the software via Bluetooth. When the data flowed into software tier, the acquisition software could match the GPS information and acceleration with the time data, as well as reading the real-time changes based on the wave patterns and locating the position we measured on the e-map. The data obtained from the data acquisition software were used by the Matlab GUI data processor. The processor removed any abnormal data based on residual analysis and we then used the model described above to calculate the PSD of pavements and to evaluate the IRI value.

5. Field Test

To verify the accuracy of the model, we used the above measurement system to perform a field test. The accelerometers were placed flat and fixed completely inside the vehicle, so they fully reflected the sprung vibration of the vehicle. As described in Section 3, the accelerometers were fixed separately over the wheels. Given that the practical tracks on which both the front and back wheels travel were basically the same, it was not necessary to collect acceleration data repeatedly. By contrast, the tracks of the left and right wheels were totally different, so their mutual effects could not be neglected. At the same time, the back wheels were affected less by the engine, so the accelerators were set immediately above the right back wheel and left back wheel in the test car.

Eight typical roads in Shanghai, for which the IRIs were provided by the Shanghai Highway Administration Bureau, were selected for the field test. The specific method employed was as follows.

(i) We confirm the stake mark of the testing origin and destination and recorded the positional data.



(a) (b)

Figure 5: (a) Data acquisition software, (b) Matlab GUI data processor.

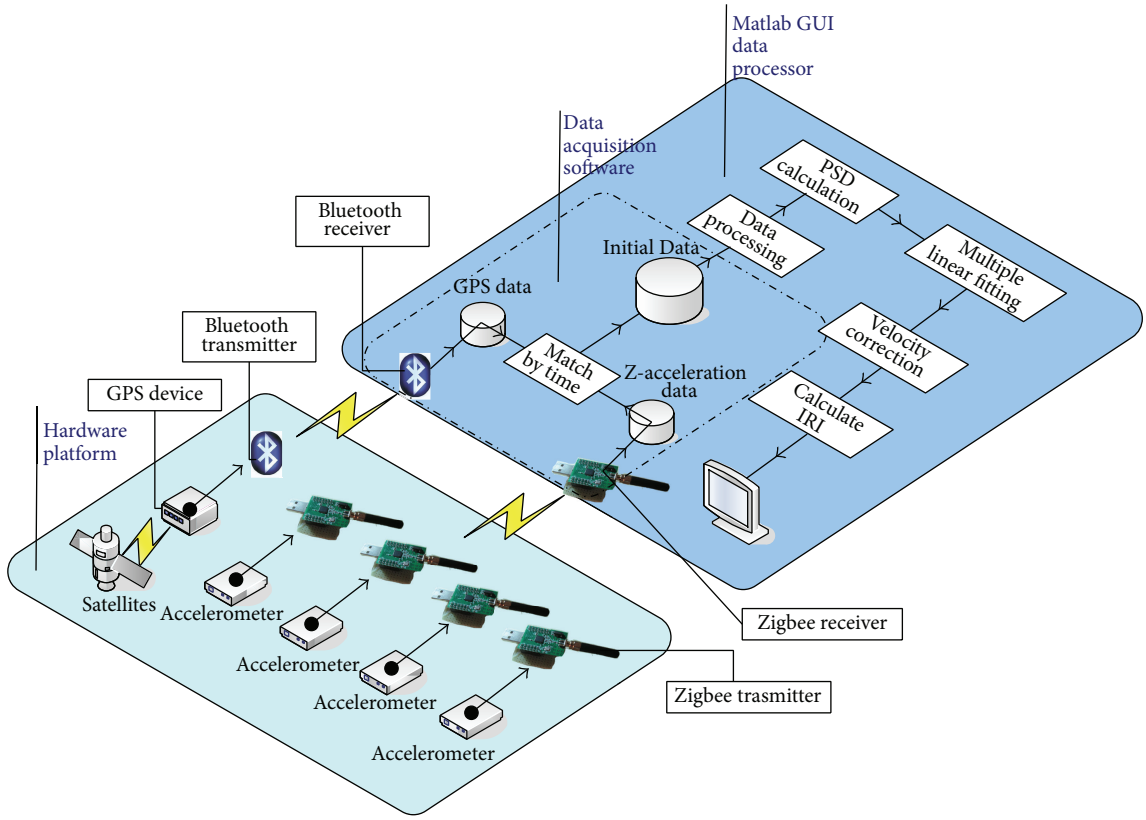


Figure 6: Schematic diagram of the measurement system.

(ii) The experimental car with the measurement system was used to acquire the vertical acceleration (accuracy: 0.01 mg, frequency: 10 Hz) and GPS (accuracy: 1 m, frequency: 1 Hz) at about 60 km/h, where data were recorded for both the left and right wheels.

(iii) The acceleration data were imported into Matlab GUI, before calculating the approximate IRIs of these roads using the linear fitting model.

We selected the measurement data for links 21 to 28 as the input of the model to calculate the IRI. We then calculated the relative error between the approximate IRI obtained from the model and the actual IRI (detected by the vehicle bearing road laser profilometer; Highway Administration of Shanghai, 2012) to test the suitability of the model. The results are shown in Table 6.

When we used the model presented in Section 3 to fit the Z-axis acceleration to the IRI, the results showed that the relative error of the approximate IRI value was lower than 15% and the standardized residual was between −2 and +2. Therefore, this model can meet the needs of the majority of pavement measurements. Thus, this method based on Z-axis accelerometers and GPS devices is feasible for measuring the IRI.

Table 6: Approximate values of the IRIs obtained using the model.

Link number	Right-wheel square root value	Left-wheel square root value	Actual IRI	Approximate IRI	Relative error (%)	Standardized residual
21	3.0182	4.3766	3.9041	4.109	4.98	0.78773
22	3.7904	4.3842	4.9140	4.566	7.61	−0.76443
23	3.4982	4.0119	4.2430	4.292	1.14	0.350071
24	3.1138	3.9016	3.6570	4.086	10.49	1.416846
25	4.4060	5.4115	6.5001	6.137	5.92	−0.80682
26	5.0087	5.5317	6.4629	6.020	7.36	−1.03084
27	5.7322	4.2199	3.9496	4.315	8.46	1.238302
28	4.9631	4.3613	4.4039	3.904	12.82	−1.19086

6. Conclusions

To address the problems of pavement roughness measurement, we established an IRI estimation model based on regression analysis. Based on the multiple linear fitting model and velocity correction model, we developed a coupled system that can record the real-time Z-axis acceleration in different pavement conditions, at different times, and with different values for various other parameters.

The variation in the in-car Z-axis acceleration caused by road roughness can be regarded as a combination of the vibration produced by different mechanical components, and thus the vertical acceleration is strongly correlated with the IRI. The quarter-car model was a LTI system and the mean squared value of the power spectral density could represent the equivalent amplitude of signals, which can represent the size of the signal amplitude, and thus we used a regression method to model the variation in the Z-axis acceleration and the IRI. We used the power spectral density sequence of the Z-axis acceleration to model the IRI. An innovative feature of the measurement process was that multiple local accelerations were considered in order to improve the goodness of fit of the model.

Because the relationship between the IRI value and the speed is nonlinear, variation of the speed during the measurement process has a large influence on the measured IRI value. The length of road along which the measurement is performed and the start and end points are also important in this measurement, so we used a GPS device to allow us to take account of speed. The influence of speed on the results of the model was analyzed, and we then put forward the concept of a speed correction coefficient to improve the reliability of the model.

We used the IRI evaluation model and system to measure the IRI of some typical roads in Shanghai. When our model was used to fit the Z-axis acceleration to the IRI, the results showed that the relative error of the estimated IRI was less than 15%.

Conflict of Interests

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

Acknowledgments

This work was based on the results of a research project, which was supported by a research Grant no. 2012AA112402 from the Ministry of Science and Technology of the People's Republic of China and a research Grant no. 11511501100 from Shanghai Science and Technology Committee. The work of the last author was supported by Program for Changjiang Scholars and Innovative Research Team in University and Shanghai Pujiang Program (11PJD022). The authors take sole responsibility for all views and opinions expressed in this paper.

References

[1] M. W. Sayers, T. D. Gillespie, and W. D. Paterson, "Guidelines for the conduct and calibration of road roughness measurements," World Bank Technical Paper 46, The World Bank, Washington, DC, USA, 1986.

[2] M. W. Sayers, "On the calculation of IRI from longitudinal road profile. TRB Paper No. 95 0842," in *Proceedings of the 74th Annual Meeting on Transportation Research Board (TRB '95)*, Washington, DC, USA, 1995.

[3] J. R. Prasad, S. Kanuganti, P. N. Bhanegaonkar, A. K. Sarkar, and S. Arkatkar, "Development of relationship between roughness (IRI) and visible surface distresses: a study on PMGSY roads," *Procedia—Social and Behavioral Sciences*, vol. 104, pp. 322–331, 2013.

[4] M. W. Sayers, T. D. Gillespie, and A. V. Queiroz, "The international road roughness experiment. Establishing correlation and a calibration standard for measurements," Tech. Rep. WTP45, 1986.

[5] J. J. Hajek, T. J. Kazmierowski, and G. Musgrove, "International roughness index as a measure of customer satisfaction ," in *Proceedings of the Annual Meeting of the Transportation Association of Canada*, Victoria, Canada, 1995.

[6] T. J. Kwon, M. Gerla, V. K. Varma, M. Barton, and T. R. Hsing, "Efficient flooding with passive clustering—an overhead-free selective forward mechanism for ad hoc/sensor networks," *Proceedings of the IEEE*, vol. 91, no. 8, pp. 1210–1220, 2003.

[7] H. Sabbineni and K. Chakrabarty, "Location-aided flooding: an energy-efficient data dissemination protocol for wireless sensor networks," *IEEE Transactions on Computers*, vol. 54, no. 1, pp. 36–46, 2005.

[8] J. E. Jefferies, R. W. DeMay, and G. L. Lachinyan, "Rental/car-share vehicle access and management system and method," U.S. Patent Application 13/830,754, 2013.

[9] T. Dishongh, F. Guilak, and M. Morris, "Apparatus for monitoring physiological, activity, and environmental data," U.S. Patent Application 11/641,973[P], 2006.

[10] A. Prakash, B. N. Sharma, and T. J. Kazmierowski, "Investigation into observational variations in pavement condition survey," in *Proceedings of the 3rd International Conference on Managing Pavements*, vol. 2, pp. 290–301, Washington, DC, USA, 1994.

[11] Z. Xiao-qing and S. Li-jun, "Relationship between international roughness index and velocity of quarter car," *Journal of Tongji University*, vol. 33, no. 10, pp. 1323–1327, 2005.

[12] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 1997.

[13] X. Zhou, L. Yan, and L. Sun, "Study and validation of the relationship between international roughness index and power spectral density," *China Civil Engineering Journal*, vol. 40, no. 1, pp. 99–104, 2007.

[14] R. W. Perera, C. Byrum, and S. D. Kohn, *Investigation of Development of Pavement Roughness*, 1998.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 187275, 11 pages
http://dx.doi.org/10.1155/2014/187275

Research Article

Uncertain Programming for Network Revenue Management

Deyi Mou and Xiaoxin Wang

Institute of Mathematics for Applications, Civil Aviation University of China, Tianjin 300300, China

Correspondence should be addressed to Xiaoxin Wang; 971252012@qq.com

Received 18 March 2014; Revised 14 May 2014; Accepted 19 May 2014; Published 9 June 2014

Academic Editor: Andy H. F. Chow

The mathematical model for airline network seat inventory control problem is usually investigated to maximize the total revenue under some constraints such as capacities and demands. This paper presents a chance-constrained programming model based on the uncertainty theory for network revenue management, in which the fares and the demands are both uncertain variables rather than random variables. The uncertain programming model can be transformed into a deterministic form by taking expected value on objective function and confidence level on the constraint functions. Based on the strategy of nested booking limits, a solution method of booking control is developed to solve the problem. Finally, this paper gives a numerical example to show that the method is practical and efficient.

1. Introduction

After the deregulations in the airline industry, the revenue management techniques have become indispensable for airline seat inventory control. A central problem in airline revenue management is determining optimal decision rules for sequentially accepting or denying itinerary requests. So it is necessary to develop mathematical models to determine complex booking control strategies.

The optimization methods on seat inventory control problem with multiple-fare classes can be separated into the single-leg optimization method and the network optimization method. Based on the optimized frequency, each optimization method can be sorted out the static method and the dynamic method. In brief, the dynamic nature of the arrivals of the requests over time is not explicitly considered in the static method, whereas the mutability in the demand is taken into account in the dynamic method as the end of the reservation period approaches and the seat capacity diminishes.

The single-leg optimization method firstly appeared in the research of Littlewood [1]. He studied a seat inventory control problem with two-fare classes on a single leg and proposed a marginal seat revenue rule applied into a two-price, single-leg model. Belobaba [2, 3] extended this idea to a multiclass problem and introduced the expected marginal seat revenue heuristic for the general approach. Wollmer [4], Brumelle and McGill [5], and Robinson [6] further studied the single-leg problem with multiple-fare classes. They developed algorithms to find the optimal booking control policy under the assumption that the probability distributions of the demands for different fare classes were known. Lee and Hersh [7] developed a discrete-time dynamic programming model to find an optimal booking control policy without requiring any assumptions about the arrival mode for the manifold booking classes. Liang [8] proposed a continuous-time, stochastic, dynamic programming model and showed that a threshold control policy was optimal. Feng and Xiao [9] presented a stochastic control model to dynamically tackle with seat inventory control problem.

For the network optimization method, Glover et al. [10] initially described a minimum cost network flow formulation with deterministic demand without focusing on the stochastic elements. After that, a solution method for the sequential allocation of seats under the assumption of stochastic demand was provided by Wang [11]. Wollmer [12] proposed a linear programming model that considered stochastic demand. Dror et al. [13] proposed a similar deterministic network minimum cost flow formulation that allowed for cancellations as deterministic losses on arcs in the network. Curry [14] developed a combined mathematical model for a multiclass seat inventory control problem. Williamson

[15] studied two network-based mathematical programming models. The first model incorporated probabilistic demand and the second model simplified the problem by substituting stochastic demand by its expectation. Wong et al. [16] applied nesting techniques into a multiclass seat inventory control problem. de Boer et al. [17] proposed stochastic linear programming for network revenue management and developed the nesting technique of Williamsion. Bertsimas and De Boer [18] and van Ryzin and Vulcano [19] used simulation-based optimization methods that also investigated nesting over the network. Cooper and Homem-De-Mello [20] proposed a decomposition method combining mathematical programming methods and Markov decision process. Recently, İlker Birbil et al. [21] proposed a framework for solving airline revenue management problems on large networks.

With the fast development of civil aviation industry, many airlines often create new routes. Due to lacking reliable data and accurate information, the methods in the literature above become invalid for these new routes. On the other hand, when unconventional sudden events such as war, atrocious weather, and earthquake, happen, the cumulative data of computer reservation system is no longer trustworthy. There are some limitations when the traditional stochastic models above deal with such problems at this situation. In the two cases, we have to invite some experts to evaluate their degree of belief that each event will occur. However, humans tend to overweigh unlikely events (Kahnema and Tversky [22]); thus, the degree of belief may have a much larger range than the real frequency. In this situation, if we insist on dealing with the degree of belief using the probability theory, some counterintuitive results will be obtained (Liu [23]). In revenue management of the above two cases, as we stated before, the domain experts invited are likely to overrate the market demand on the new routes and underestimate the market demand under the circumstances of unconventional sudden events. If the belief degree of the market demand is treated as probability, we have no choice but to increase the capacity on the new routes and reduce the capacity under the circumstances of unconventional sudden events. This will cause great losses in revenue for airlines. This conclusion seems unacceptable and then the belief degree cannot be treated as probability.

In order to deal with the experts' degree of belief, the uncertainty theory was founded by Liu [24] and refined by Liu [25] in 2013. Many researchers have contributed to this area. The uncertainty theory has been applied to uncertain programming, uncertain risk analysis, uncertain game, uncertain inference, uncertain logic, uncertain finance, and uncertain optimal control (Liu [25]). Nowadays, the uncertainty theory has become a branch of axiomatic mathematics to model human uncertainty (Liu [26]).

Depending on the analysis as mentioned above, we think that it is necessary to apply the uncertainty theory as a basic approach to model the uncertainty in the revenue management of the two cases above. In this paper, we propose the chance-constrained programming model based on the uncertainty theory to deal with the uncertain factors. The rest of this paper is structured as follows. In Section 2, some basic concepts and properties in uncertainty theory used throughout this paper are introduced. In Section 3, an uncertain programming model is constructed. According to inverse uncertainty distribution, the model can be transformed to its deterministic form. In Section 4, we present a solution method of booking control on the basis of the strategy of nested booking limits. After that, a numerical example is given in Section 5. At last, a brief summary is presented in Section 6.

2. Preliminaries

In this section, some basic definitions and arithmetic operations of uncertainty theory needed throughout this paper are presented.

Definition 1 (Liu [24]). Let Γ be a nonempty set and $\mathscr{L}$ a σ-algebra over Γ. Each element $\Lambda \in \mathscr{L}$ is called an event. The set function $\mathscr{M}$ is called an uncertain measure if it satisfies the following four axioms:

Axiom 1 (Normality). $\mathscr{M}\{\Gamma\} = 1$;

Axiom 2 (Monotonicity). $\mathscr{M}\{\Lambda_1\} \leq \mathscr{M}\{\Lambda_2\}$ whenever $\Lambda_1 \subset \Lambda_2$;

Axiom 3 (Self-Duality). $\mathscr{M}\{\Lambda\} + \mathscr{M}\{\Lambda_c\} = 1$ for any event Λ;

Axiom 4 (Countable Subadditivity). For every countable sequence of events $\{\Lambda_i\}$, we have

$$\mathscr{M}\left\{\bigcup_{i=1}^{\infty}\Lambda_i\right\} \leq \sum_{i=1}^{\infty}\mathscr{M}\left\{\Lambda_i\right\}. \tag{1}$$

Definition 2 (Liu [24]). Let Γ be a nonempty set, $\mathscr{L}$ a σ-algebra over Γ, and $\mathscr{M}$ an uncertain measure. Then the triple $(\Gamma, \mathscr{L}, \mathscr{M})$ is called on uncertainty space.

Definition 3 (Liu [24]). An uncertain variable ξ is a measurable function from an uncertainty space $(\Gamma, \mathscr{L}, \mathscr{M})$ to the set of real numbers; that is, for any Borel set B of real numbers, the set

$$\{\xi \in B\} = \{\gamma \in \Gamma \mid \xi(\gamma) \in B\} \tag{2}$$

is an event.

For a sequence of uncertainty variables $\xi_1, \xi_2, \ldots, \xi_n$ and a measurable function f, Liu [24] proved that $\xi = f(\xi_1, \xi_2, \ldots, \xi_n)$ defined as $\xi(\gamma) = f(\xi_1(\gamma), \xi_2(\gamma), \ldots, \xi_n(\gamma))$, $\forall \gamma \in \Gamma$ is also an uncertain variable. In order to describe an uncertain variable, a concept of uncertainty distribution is introduced as follows.

Definition 4 (Liu [24]). The uncertainty distribution Φ of an uncertain variable ξ is defined by

$$\Phi(x) = \mathscr{M}\{\xi \leq x\}, \tag{3}$$

for any real number x.

Peng and Iwamura [27] proved that a function $\Phi: R \to [0, 1]$ is an uncertainty distribution if and only if it is

a monotone increasing function except for $\Phi(x) \equiv 0$ or $\Phi(x) \equiv 1$. The inverse function Φ^{-1} is called the inverse uncertainty distribution of ξ. Inverse uncertainty distribution is an important tool in the operation of uncertain variables.

Theorem 5 (Liu [24]). *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent uncertain variables with regular uncertainty distributions $\Phi_1, \Phi_2, \ldots, \Phi_n$, respectively. If $f(x_1, x_2, \ldots, x_n)$ is strictly increasing with respect to $x_1, x_2, \ldots, x_m$ and strictly decreasing with respect to $x_{m+1}, x_{m+2}, \ldots, x_n$, then*

$$\xi = f(\xi_1, \xi_2, \ldots, \xi_n) \tag{4}$$

is an uncertain variable with inverse uncertainty distribution

$$\Psi^{-1}(\alpha) = f\left(\Phi_1^{-1}(\alpha), \ldots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1-\alpha), \ldots, \Phi_n^{-1}(1-\alpha)\right). \tag{5}$$

Expected value is the average of an uncertain variable in the sense of uncertain measure. It is an important index to rank uncertain variables.

Definition 6 (Liu [24]). Let ξ be an uncertain variable. Then the expected value of ξ is defined by

$$E[\xi] = \int_0^\infty \mathcal{M}\{\xi \geq r\}\, dr - \int_{-\infty}^0 \mathcal{M}\{\xi \leq r\}\, dr, \tag{6}$$

provided that at least one of the two integrals is finite.

In order to calculate the expected value via inverse uncertainty distribution, Liu and Ha [28] proved that

$$E[\xi] = \int_0^1 f\left(\Phi_1^{-1}(\alpha)\right), \ldots, \Phi_m^{-1}(\alpha), \Phi_{m+1}^{-1}(1-\alpha), \ldots, \Phi_n^{-1}(1-\alpha)\, d\alpha \tag{7}$$

under the condition described in Theorem 5. Generally, the expected value operator E has no linearity property for arbitrary uncertain variables. But, for independent uncertain variables ξ and η with finite expected values, we have

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta], \tag{8}$$

for any real numbers a and b.

3. Uncertain Programming Model for Multiple-Leg Network Seat Inventory Control

3.1. Problem Description. Because an airline wants to maximize revenue from the whole network, the researchers on this field focus on the network-based models now. Airlines usually provide thousands of such combinations of origin, destination and fare class (ODF). Therefore determining a comprehensive booking control strategy for the entire network is crucially important.

The objective of network seat inventory control is to maximize the airline's expected revenue from its supply of ODF combinations. Each ODF in the network is constitutive of one or more flight legs. The limited capacity on each flight leg has to be made full use of in the most profitable way. This can be achieved by limiting the number of seats available to the less lucrative classes. So the problem is to allocate all seats of each flight leg to the related ODF in the most profitable way. Due to its economic importance in the airline, the problem has been extensively studied.

In this paper, the network seat inventory control problem will be modeled by the chance-constrained programming based on the uncertainty theory in which the fare and the demand of each ODF are assumed to be uncertain variables with given uncertainty distributions.

3.2. Model Development. At first, we introduce the following notations to represent the mathematical formulation throughout the remainder of this paper:

x_{ODF} : the number of seats reserved for each separate ODF;

N : the total number of flight legs in the ODF network;

S_l : the set of ODF combinations available on flight leg;

C_l : the seat capacity on leg l;

D_{ODF} : the deterministic aggregated demand for each ODF;

f_{ODF} : the fare required for an ODF.

In order to facilitate the analysis, we make some reasonable assumptions as follows.

(a) The flight market demand exceeds its capacity supply.

(b) Overbooking is not considered by the model discussed here.

Next, based on the analysis of the decision making process, the general problem is formulated as follows [15]:

$$\begin{aligned} \max \quad & \sum_{\text{ODF}} f_{\text{ODF}} x_{\text{ODF}} \\ \text{s.t.} \quad & \\ & \sum_{\text{ODF} \in S_l} x_{\text{ODF}} \leq C_l \quad \forall \text{flight legs } l = 1, \ldots, N, \\ & x_{\text{ODF}} \leq D_{\text{ODF}} \quad \forall \text{ODF}, \\ & x_{\text{ODF}} \geq 0 \text{ integer} \quad \forall \text{ODF}. \end{aligned} \tag{9}$$

In the above model, the quantities f_{ODF} and D_{ODF} are all assumed to be crisp numbers. However, when there are new routes created by the airlines or the emergency takes place sometimes, the quantities generally are not fixed but obtained from experience evaluation or expert knowledge. In this case, we may assume the quantities are uncertain variables. Then the model (9) is only a conceptual model rather than a mathematical model because there does not exist a natural ordership in an uncertain world. Here we take expected value criterion on the objective function and

confidence level on the constraint functions (Liu [25]). Then the model (9) turns into the following mathematical model:

$$\begin{aligned} \max \quad & E\left[\sum_{\mathrm{ODF}} f_{\mathrm{ODF}} x_{\mathrm{ODF}}\right] \\ \text{s.t.} \quad & \sum_{\mathrm{ODF}\in S_l} x_{\mathrm{ODF}} \le C_l \quad \forall \text{flight legs } l = 1,\ldots,N, \\ & M\left\{x_{\mathrm{ODF}} \le D_{\mathrm{ODF}}\right\} \ge \beta_{\mathrm{ODF}} \quad \forall \mathrm{ODF}, \\ & x_{\mathrm{ODF}} \ge 0 \text{ integer} \quad \forall \mathrm{ODF}, \end{aligned} \tag{10}$$

where β_{ODF} are some predetermined confidence levels for all ODF.

In practical applications, the uncertainty distributions of uncertain variables f_{ODF} and D_{ODF} and the confidence levels β_{ODF} are determined by linear interpolation method, the principle of least squares, the method of moments, and the Delphi method from expert's experimental data (Liu [25]). How do we obtain expert's experimental data? Liu [25] proposed a questionnaire survey for collecting expert's experimental data. In this paper, we assume that the uncertainty distributions of uncertain variables f_{ODF} and D_{ODF} and the confidence levels β_{ODF} have been determined.

In order to solve model (10), firstly, we introduce two corollaries which were from the uncertainty theory (Liu [25]).

Corollary 7. *Assume the objective function $f(x,\xi_1,\xi_2,\ldots,\xi_n)$ is strictly increasing with respect to $\xi_1,\xi_2,\ldots,\xi_m$ and strictly decreasing with respect to $\xi_{m+1},\xi_{m+2},\ldots,\xi_n$. If $\xi_1,\xi_2,\ldots,\xi_n$ are independent uncertain variables with uncertainty distributions $\Phi_1,\Phi_2,\ldots,\Phi_n$, respectively, then the expected objective function $E[f(x,\xi_1,\xi_2,\ldots,\xi_n)]$ is equal to*

$$\int_0^1 f\left(x,\Phi_1^{-1},\ldots,\Phi_m^{-1},\Phi_{m+1}^{-1},\ldots,\Phi_n^{-1}\right)d\alpha. \tag{11}$$

Corollary 8. *Assume the constraint function $g(x,\xi_1,\xi_2,\ldots,\xi_n)$ is strictly increasing with respect to $\xi_1,\xi_2,\ldots,\xi_k$ and strictly decreasing with respect to $\xi_{k+1},\xi_{mk2},\ldots,\xi_n$. If $\xi_1,\xi_2,\ldots,\xi_n$ are independent uncertain variables with uncertainty distributions $\Phi_1,\Phi_2,\ldots,\Phi_n$, respectively, then the chance constrain*

$$\mathcal{M}\left\{g\left(x,\xi_1,\xi_2,\ldots,\xi_n\right)\le 0\right\}\ge\alpha \tag{12}$$

holds if and only if

$$\begin{aligned} & g\left(x,\Phi_1^{-1}(\alpha),\ldots,\Phi_k^{-1}(\alpha),\Phi_{k+1}^{-1}(1-\alpha),\ldots,\right. \\ & \left.\Phi_n^{-1}(1-\alpha)\right)\le 0. \end{aligned} \tag{13}$$

Secondly, the next theorem shows that the model (10) is equivalent to a deterministic model, for which many efficient algorithms have been designed.

Theorem 9. *Assume that f_{ODF} and D_{ODF} are independent uncertain variables with uncertainty distributions ϕ_{ODF} and ψ_{ODF}. Then the model* (10) *is equivalent to the following model:*

$$\begin{aligned} \max \quad & \sum_{ODF} x_{ODF}\int_0^1 \phi_{ODF}^{-1}(\alpha)\,d\alpha \\ \textit{s.t.} \quad & \sum_{ODF\in S_l} x_{ODF} \le C_l \quad \forall \textit{flight legs } l = 1,\ldots,N, \\ & x_{ODF} \le \psi_{ODF}^{-1}(1-\beta_{ODF}) \quad \forall ODF, \\ & x_{ODF} \ge 0 \textit{ integer} \quad \forall ODF. \end{aligned} \tag{14}$$

Proof. The function $\sum_{\mathrm{ODF}} f_{\mathrm{ODF}} x_{\mathrm{ODF}}$ is strictly increasing with respect to f_{ODF} and f_{ODF} are independent uncertain variables with uncertainty distributions ϕ_{ODF}, respectively. By using Corollary 7, we obtain

$$E\left[\sum_{\mathrm{ODF}} f_{\mathrm{ODF}} x_{\mathrm{ODF}}\right] = \sum_{\mathrm{ODF}} x_{\mathrm{ODF}}\int_0^1 \phi^{-1}{}_{\mathrm{ODF}}(\alpha)\,d\alpha. \tag{15}$$

Since that

$$M\left\{x_{\mathrm{ODF}} \le D_{\mathrm{ODF}}\right\} \ge \beta_{\mathrm{ODF}} \tag{16}$$

is equivalent to

$$M\left\{-D_{\mathrm{ODF}} + x_{\mathrm{ODF}} \le 0\right\} \ge \beta_{\mathrm{ODF}} \tag{17}$$

and the function $-D_{\mathrm{ODF}} + x_{\mathrm{ODF}}$ is strictly decreasing with respect to D_{ODF} with uncertainty distribution ψ_{ODF}, it follows from Corollary 8 that we have

$$-\psi_{\mathrm{ODF}}^{-1}(1-\beta_{\mathrm{ODF}}) + x_{\mathrm{ODF}} \le 0, \tag{18}$$

that is,

$$x_{\mathrm{ODF}} \le \psi_{\mathrm{ODF}}^{-1}(1-\beta_{\mathrm{ODF}}). \tag{19}$$

The theorem is thus verified. □

4. Solution Method of Booking Control

4.1. The Strategy of Nested Booking Limits. In the model (14), we use $\overline{f}_{\mathrm{ODF}}$ to denote $\int_0^1 \phi_{\mathrm{ODF}}^{-1}(\alpha)d\alpha$ and $\overline{D}_{\mathrm{ODF}}$ to denote $\psi_{\mathrm{ODF}}^{-1}(1-\beta_{\mathrm{ODF}})$. Then the relaxation of the model (14) is

$$\begin{aligned} \max \quad & \sum_{\mathrm{ODF}} \overline{f}_{\mathrm{ODF}} x_{\mathrm{ODF}} \\ \text{s.t.} \quad & \sum_{\mathrm{ODF}\in S_l} x_{\mathrm{ODF}} \le C_l \quad \forall \text{flight legs } l = 1,\ldots,N, \\ & x_{\mathrm{ODF}} \le \overline{D}_{\mathrm{ODF}} \quad \forall \mathrm{ODF}, \\ & x_{\mathrm{ODF}} \ge 0 \quad \forall \mathrm{ODF}. \end{aligned} \tag{20}$$

Table 1: Simulated reservation data.

ODF	$t = 10$	$t = 9$	$t = 8$	$t = 7$	$t = 6$	$t = 5$	$t = 4$	$t = 3$	$t = 2$	$t = 1$	$t = 0$
ABY	3	3	3	5	4	1	3	7	8	11	8
ABT	13	5	8	7	1	2	6	2	5	5	2
BCY	0	2	1	5	4	8	6	4	7	7	4
BCT	15	5	11	9	9	8	2	3	4	2	0
ACY	2	0	0	1	2	9	1	7	5	3	5
ACT	10	2	5	6	2	4	7	4	3	1	1

Table 2: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(42, 2)	(66, 2)	(41, 3)	(71, 4)	(35, 1)	(50, 3)

The whole booking time should be partitioned into a few time periods of reservation; for example, a day is a time period of reservation. In order to facilitate the analysis, we describe the model (20) in the form of the matrix and the vector considering the time period of reservation. For this, we introduce the following notations:

I: the total number of flight legs in the ODF network;

i: index for set of flight legs;

J: the total number of the ODF;

j: index for set of the ODF;

t: index for the time period of reservation;

$C^t = [C_1^t, C_2^t, \ldots, C_i^t, \ldots, C_I^t]^T$: the seat capacity on each flight leg in the reservation time period t;

$F^t = [\overline{f}_1^t, \overline{f}_2^t, \ldots, \overline{f}_j^t, \ldots \overline{f}_J^t]^T$: the corresponding $\overline{f}_{\text{ODF}}$ for each separate ODF in the reservation time period t;

$D^t = [\overline{D}_1^t, \overline{D}_2^t, \ldots, \overline{D}_j^t, \ldots, \overline{D}_J^t]^T$: the corresponding $\overline{D}_{\text{ODF}}$ for each separate ODF in the reservation time period t;

$X^t = [x_1^t, x_2^t, \ldots, x_j^t, \ldots, x_J^t]^T$: the number of seats reserved for each separate ODF in the reservation time period t;

$A = (a_{ij})_{I \times J}$: the matrix denotation of flight legs that each ODF travels, where if the ODF j travels the flight leg i, then $a_{ij} = 1$, otherwise $a_{ij} = 0$;

A^j: the jth column of the matrix A, denoting the flight legs that the ODF j travels.

The model (20) is described as follows in the form of the matrix and the vector:

$$\begin{aligned} \text{LP}\left(C^t, D^t\right) = \max \quad & \left(F^t\right)^T X^t \\ \text{s.t.} \quad & \\ & AX^t \le C^t, \\ & 0 \le X^t \le D^t. \end{aligned} \tag{21}$$

The dual problem of the model above can be described as follows:

$$\begin{aligned} \text{DLP}\left(C^t, D^t\right) = \min \quad & \left[\left(P^t\right)^T C^t + \left(Q^t\right)^T D^t\right] \\ \text{s.t.} \quad & \\ & A^T P^t + Q^t \ge F^t, \\ & P^t, Q^t \ge 0, \end{aligned} \tag{22}$$

where $P^t = [P_1^t, P_2^t, \ldots, P_i^t, \ldots, P_I^t]^T$ denotes the shadow prices corresponding to each flight leg and $Q^t = [Q_1^t, Q_2^t, \ldots, Q_j^t, \ldots, Q_J^t]^T$ denotes the shadow prices corresponding to each $\overline{D}_{\text{ODF}}$.

Bid price control method is one of the prevalent methods of network seat inventory control. The bid price of each ODF is equal to the sum of shadow prices of the flight legs that the ODF crosses. A booking request for a passenger from the ODF is rejected if the bid price of the ODF exceeds the fare for the ODF and is accepted otherwise. Although bid price control method has been used in the actual operations of the airlines, it has a few shortcomings as follows.

(a) Each ODF's contribution to network revenue is not considered in the bid price control method.

(b) When calculating shadow prices using the relevant models, the solution of the model may be degenerate solution. This will cause the multiple bid prices of an ODF.

(c) The fares of most of the passengers on the flight just exceed the bid prices so that airlines suffer losses.

For this, we present a nesting control method based on the network contribution value for the above uncertain programming model.

First, we define an ODF's net contribution value to network revenue in the reservation time period t as the expected fare for the ODF in the reservation time period t minus the opportunity cost of the ODF in the reservation time period t, that is,

$$\text{NCV}_j^t = \overline{f}_j^t - \text{OC}_j^t, \tag{23}$$

where NCV_j^t denotes the net contribution value of the ODF j to network revenue in the reservation time period t and OC_j^t denotes the opportunity cost of the ODF j in the reservation time period t.

Table 3: Parameters of normal distribution of fare.

$N(e_i', \sigma_i')$	1	2	3	4	5	6
	(1000, 80)	(800, 40)	(400, 20)	(320, 10)	(1200, 90)	(960, 50)

Table 4: The seat inventory control for $t = 10$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 10$	ABY	1000	200	42	140	3	3
	ABT	800	0	63	63	13	13
	BCY	400	80	41	105	0	0
	BCT	320	0	64	64	15	15
	ACY	1200	80	35	98	2	2
	ACT	960	−160	10	0	10	0

Table 5: The seat inventory control result.

The combination of the ODF	The total number of accepted seats
ABY	52
ABT	54
BCY	48
BCT	58
ACY	30
ACT	4
The total expected revenue	172800

Table 6: The seat inventory control result of the bid price control method.

The combination of the ODF	The total number of accepted seats
ABY	48
ABT	54
BCY	45
BCT	57
ACY	30
ACT	8
The total expected revenue	171120

Table 7: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(40, 2)	(56, 2)	(39, 3)	(61, 4)	(34, 1)	(42, 3)

The opportunity cost OC_j^t of the ODF j in the reservation time period t is calculated based on the DLP model as follows:

$$\mathrm{OC}_j^t = \left(P^t\right)^T A^j \tag{24}$$

for $1 \le j \le J$.

Now the opportunity cost of the ODF in the reservation time period t is known, so we figure up the net contribution value of the ODF to network revenue in the reservation time period t. However, the solution of the model DLP may be degenerate solution and this phenomenon will cause multiple net contribution values of the ODF to network revenue. In this case, we will use the following method to calculate the net contribution value of the ODF to network revenue in the reservation time period t. If $C^t - A^j \ge 0$, then $\mathrm{NCV}_j^t = \overline{f}_j^t - [\mathrm{LP}(C^t, D^t) - \mathrm{LP}(C^t - A^j, D^t)]$. Otherwise, $\mathrm{NCV}_j^t = -\infty$ where ∞ denotes a large enough number.

Finally, we rank the ODF on the basis of their net contribution value to network revenue to determine the nesting level. If some of the ODFs have the same net contribution value to network revenue, we can rank the ODF on the basis of their expected fare.

4.2. The Algorithm for Nested ODF-Based Booking Control. Every time a booking request arrives for any ODF in the network, a quick decision should be made whether or not to accept the request. We have to specify a booking control strategy for the decision. We propose the algorithm for nested ODF-based booking control.

The notations used in the following algorithm are given as below:

$X^t = [x_t^1, x_t^2, \ldots, x_t^j, \ldots, x_t^J]^T$: the number of seats reserved for each separate ODF in the reservation time period t after ranking the elements of X^t according to the nesting level;

$\mathrm{PL}_t = [\mathrm{PL}_t^1, \mathrm{PL}_t^2, \ldots, \mathrm{PL}_t^j, \ldots, \mathrm{PL}_t^J]^T$: seat protect level for each ranked ODF in the reservation time period t;

M^k $(1 \le k \le j)$: the set of flight legs that the kth ODF travels;

$b_t = [b_t^1, b_t^2, \ldots, b_t^j, \ldots, b_t^J]$: the number of booking requests for the ranked ODF that have already been accepted in the reservation time period t;

BL_t^j: seat booking limit for the jth ODF.

The heuristic algorithm for nested ODF-based booking control is as follows.

Step 1. Calculate each ODF's net contribution value to network revenue in the reservation time period t and determine the nesting level. By solving the model above, we can obtain $X^t = [x_1^t, x_2^t, \ldots, x_j^t, \ldots, x_J^t]^T$.

Table 8: The seat inventory control for $t = 9$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 9$	ABY	1000	200	40	122	3	3
	ABT	800	0	48	48	5	5
	BCY	400	80	39	89	2	2
	BCT	320	0	50	50	5	5
	ACY	1200	80	34	82	0	0
	ACT	960	−160	0	0	2	0

Table 9: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(38, 2)	(47, 2)	(36, 3)	(51, 4)	(32, 1)	(36, 3)

Step 2. Determine seat protect level for each ODF. According to ranking results of Step 1, we rank the elements of X^t and obtain $X^t = [x_t^1, x_t^2, \ldots, x_t^j, \ldots, x_t^J]^T$. Seat protect level is determined as $\mathrm{PL}_t = [\mathrm{PL}_t^1, \mathrm{PL}_t^2, \ldots, \mathrm{PL}_t^j, \ldots, \mathrm{PL}_t^J] = [x_t^1, x_t^2, \ldots, x_t^j, \ldots, x_t^J]^T$.

Step 3. At the beginning of the reservation time period t, it is obvious that $b_t = 0$. Determine seat booking limit for the ranked ODF as follows.
For the first ODF,

$$\mathrm{BL}_t^1 = \min\left\{N_i^t \mid 1 \le i \le I, i \in M^1\right\}; \tag{25}$$

for the jth ODF ($2 \le j \le J$), on the flight leg i ($1 \le i \le I$), the total number of the seats protected for the lower level ODF than the jth ODF is $\Pi_i^t = \sum_{k=1}^{j-1} \mathrm{PR}_k^t$, where the kth ODF travels the flight leg i and the number of the seats available on the flight leg i is $(C_i^t - \Pi_i^t)$. So we have

$$\mathrm{BL}_t^j = \min\left\{C_i^t - \Pi_i^t \mid 1 \le i \le I, i \in M^j\right\}. \tag{26}$$

Step 4. If $\mathrm{BL}_t^j > b_t^j$, then accept the booking request, let $b_t^j = b_t^j + 1$; if $\mathrm{BL}_t^j = b_t^j$, then decline the request.

Step 1 determines the nesting level for each separate ODF; Step 2 determines seat protect level for each ranked ODF; Step 3 determines seat booking limit for the ranked ODF; Step 4 develops the standard of accepting or rejecting the booking requests.

When entering the next reservation time period, let

$$C^t = C^t - b_t^T\left[A^1 \;\; A^2 \;\; \cdots \;\; A^j \;\; \cdots \;\; A^J\right]^T, \qquad t = t - 1, \quad b_t = 0, \tag{27}$$

and then go into the algorithm above.

5. Numerical Experiment

In this section, we apply the model and the algorithm of the former two sections to airline seat inventory control and give an optimal policy.

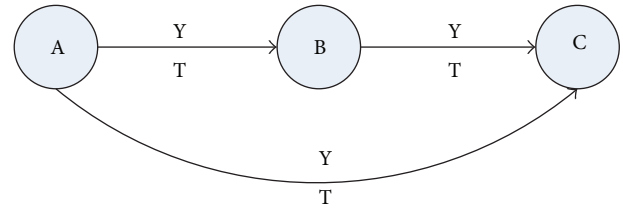


Figure 1: Schematic.

An airline prepares to open a new route, which is from A to C through B. So there are two legs and three segments. Every segment has two fare classes with Y and T, so there are six ODFs. These are shown in Figure 1.

That is,

$$\begin{gathered} N = 2, \\ l = 1, 2, \\ S_1 = \{\mathrm{ABY}, \mathrm{ABT}, \mathrm{ACY}, \mathrm{ACT}\}, \\ S_2 = \{\mathrm{BCY}, \mathrm{BCT}, \mathrm{ACY}, \mathrm{ACT}\}. \end{gathered} \tag{28}$$

Furthermore, there are 140 seats available in the flight and the booking period is partitioned into 11 time intervals.

In order to simulate the booking process, some simulated reservation data about each time interval are given in Table 1.

When $t = 10$, there is $C_1 = C_2 = 140$.

According to experts' experience, the demands for ABY, ABT, BCY, BCT, ACY, and ACT follow a normal uncertainty distribution $N(e_i, \sigma_i)$, $i = 1, 2, 3, 4, 5, 6$, respectively. The fare f_i for ABY, ABT, BCY, BCT, ACY, and ACT follows a normal uncertainty distribution $N(e_i', \sigma_i')$, $i = 1, 2, 3, 4, 5, 6$, respectively. Tables 2 and 3 give the value of the quantities.

Note that the normal uncertain variable $N(e, \sigma)$ has an expected value e and an inverse uncertainty distribution

$$\Phi^{-1}(x) = e + \frac{\sqrt{3}\sigma}{\pi} \ln \frac{1 - x}{x}. \tag{29}$$

Furthermore, $\beta = 0.9$.

According to Section 3 and Section 4, the result of the seat inventory control for $t = 10$ is obtained in Table 4.

When $t = 9, 8, 7, 6, 5, 4, 3, 2, 1, 0$, the results of the seat inventory control are listed in Appendix.

Finally, the result of simulation is listed as Table 5.

The numerical experiment shows that the computation time of the algorithm is 0.4–0.8 seconds.

Table 10: The seat inventory control for $t = 8$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 8$	ABY	1000	200	38	114	3	3
	ABT	800	0	44	44	8	8
	BCY	400	80	36	84	1	1
	BCT	320	0	48	48	11	11
	ACY	1200	80	32	76	0	0
	ACT	960	−160	0	0	5	0

Table 11: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(36, 2)	(39, 2)	(33, 3)	(42, 4)	(30, 1)	(30, 3)

Table 12: The seat inventory control for $t = 7$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 7$	ABY	1000	200	36	103	5	5
	ABT	800	0	37	37	7	7
	BCY	400	80	33	74	5	5
	BCT	320	0	41	41	9	9
	ACY	1200	80	30	67	1	1
	ACT	960	−160	0	0	6	0

Table 13: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(33, 2)	(32, 2)	(30, 3)	(34, 4)	(28, 1)	(24, 3)

Table 14: The seat inventory control for $t = 6$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 6$	ABY	1000	200	23	90	4	4
	ABT	800	0	29	29	1	1
	BCY	400	80	30	61	4	4
	BCT	320	0	31	31	9	9
	ACY	1200	80	28	57	2	2
	ACT	960	−160	0	0	2	0

Table 15: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(30, 2)	(25, 2)	(27, 3)	(26, 4)	(25, 1)	(19, 3)

Table 16: The seat inventory control for $t = 5$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 5$	ABY	1000	200	30	83	1	1
	ABT	800	0	25	28	2	2
	BCY	400	80	27	49	8	8
	BCT	320	0	19	19	8	8
	ACY	1200	80	25	53	9	9
	ACT	960	−160	3	3	4	3

Table 17: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(27, 2)	(20, 2)	(23, 3)	(20, 4)	(22, 1)	(15, 3)

Table 18: The seat inventory control for $t = 4$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 4$	ABY	1000	200	27	68	3	3
	ABT	800	0	19	19	6	6
	BCY	400	80	23	24	6	6
	BCT	320	0	1	1	2	1
	ACY	1200	80	22	41	1	1
	ACT	960	−160	0	0	7	0

Table 19: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(23, 2)	(15, 2)	(19, 3)	(14, 4)	(19, 1)	(11, 3)

Table 20: The seat inventory control for $t = 3$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 3$	ABY	1000	440	23	58	7	7
	ABT	800	240	15	16	2	2
	BCY	400	0	18	18	4	4
	BCT	320	−80	0	0	3	0
	ACY	1200	240	19	35	7	7
	ACT	960	0	1	1	4	1

Table 21: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(19, 2)	(11, 2)	(15, 3)	(10, 4)	(16, 1)	(8, 3)

Table 22: The seat inventory control for $t = 2$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 2$	ABY	1000	200	19	41	8	8
	ABT	800	0	6	6	5	5
	BCY	400	0	10	10	7	7
	BCT	320	−80	0	0	4	0
	ACY	1200	0	16	22	5	5
	ACT	960	−240	0	0	3	0

Table 23: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(14, 2)	(8, 2)	(10, 3)	(6, 4)	(12, 1)	(5, 3)

Table 24: The seat inventory control for $t = 1$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 1$	ABY	1000	200	14	23	11	11
	ABT	800	0	5	5	5	5
	BCY	400	0	10	10	7	7
	BCT	320	−80	0	0	2	0
	ACY	1200	0	4	9	3	3
	ACT	960	−240	0	0	1	0

Table 25: Parameters of normal distribution of demand.

$N(e_i, \sigma_i)$	1	2	3	4	5	6
	(9, 2)	(3, 2)	(5, 3)	(2, 1)	(7, 1)	(2, 1)

Table 26: The seat inventory control for $t = 0$.

	ODF	$\overline{f_j^t}$	NCV_j^t	x_j^t	BL_t^j	Simulated reservation data	The number of accepted seats
$t = 0$	ABY	1000	0	4	4	8	4
	ABT	800	200	0	0	2	0
	BCY	400	0	4	4	4	4
	BCT	320	−80	0	0	0	0
	ACY	1200	−200	0	0	5	0
	ACT	960	−440	0	0	1	0

We use bid price control method for the above simulation data and the result is listed as Table 6.

Comparing Tables 5 and 6, we can see that the total expected revenue using the bid price control method is 171120 RMB, while the total expected revenue using the proposed method in this paper is 172800 RMB. Numerical simulation results show that the proposed method in this paper is effective for improving the airline's revenue.

6. Conclusions

To consider network revenue management problem under conditions of new routes and unconventional sudden events, we established an uncertain programming model. Based on the strategy of nested booking limits, a heuristic algorithm for booking control was developed. Numerical test was performed to evaluate the model and the solution algorithm. The test results show that the model and the solution are all effective. There are some suggestions for future research:

(i) the impact of the nesting heuristics on total revenue;

(ii) more complex hub-spoke network;

(iii) considering dynamic factors such as the arrival order of the requests;

(iv) integrated uncertain and stochastic model.

Appendix

(1) When $t = 9$, there is $C_1 = 122$, $C_2 = 123$, $\beta = 0.8$. (See Tables 7, 3, and 8.)

(2) When $t = 8$, there is $C_1 = 114$, $C_2 = 116$, $\beta = 0.8$. (See Tables 9, 3, and 10.)

(3) When $t = 7$, there is $C_1 = 103$, $C_2 = 104$, $\beta = 0.8$. (See Tables 11, 3, and 12.)

(4) When $t = 6$, there is $C_1 = 90$, $C_2 = 89$, $\beta = 0.8$. (See Tables 13, 3, and 14.)

(5) When $t = 5$, there is $C_1 = 83$, $C_2 = 74$, $\beta = 0.8$. (See Tables 15, 3, and 16.)

(6) When $t = 4$, there is $C_1 = 68$, $C_2 = 46$, $\beta = 0.8$. (See Tables 17, 3, and 18.)

(7) When $t = 3$, there is $C_1 = 58$, $C_2 = 38$, $\beta = 0.8$. (See Tables 19, 3, and 20.)

(8) When $t = 2$, there is $C_1 = 41$, $C_2 = 26$, $\beta = 0.8$. (See Tables 21, 3, and 22.)

(9) When $t = 1$, there is $C_1 = 23$, $C_2 = 14$, $\beta = 0.8$. (See Tables 23, 3, and 24.)

(10) When $t = 0$, there is $C_1 = 4$, $C_2 = 4$, $\beta = 0.8$. (See Tables 25, 3, and 26.)

Conflict of Interests

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

Acknowledgment

This work has been financed by the Fundamental Research Funds for the Central Universities (no. 3122014 K007).

References

[1] K. Littlewood, "Forecasting and control of passengers," *AGIFORS Symposium Proceedings*, vol. 12, pp. 95–117, 1972.

[2] P. P. Belobaba, *Air travel demand and airline seat inventory management [Ph.D. thesis]*, Flight Transportation Laboratory, Massachusetts Institute of Technology, Cambridge, Mass, USA, 1987.

[3] P. P. Belobaba, "Application of a probabilistic decision model to airline seat inventory control," *Operations Research*, vol. 37, no. 2, pp. 183–197, 1989.

[4] R. D. Wollmer, "An airline seat management model for a single leg when lower fare classes book first," *Operations Research*, vol. 40, pp. 26–37, 1992.

[5] S. L. Brumelle and J. I. McGill, "Airline seat allocation with multiple nested fare classes," *Operations Research*, vol. 41, pp. 127–137, 1993.

[6] L. W. Robinson, "Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes," *Operations Research*, vol. 43, pp. 252–263, 1995.

[7] T. C. Lee and M. Hersh, "Model for dynamic airline seat inventory control with multiple seat bookings," *Transportation Science*, vol. 27, no. 3, pp. 252–265, 1993.

[8] Y. Liang, "Solution to the continuous time dynamic yield management model," *Transportation Science*, vol. 33, no. 1, pp. 117–123, 1999.

[9] Y. Feng and B. Xiao, "A dynamic airline seat inventory control model and its optimal policy," *Operations Research*, vol. 49, no. 6, pp. 938–949, 2001.

[10] F. Glover, R. Glover, J. Lorenzo, and C. McMillan, "The passenger mix problem in the scheduled airlines," *Interfaces*, vol. 12, pp. 73–79, 1982.

[11] K. Wang, "Optimal seat allocations for multi-leg flights with multiple fare types," *AGIFORS Symposium Proceedings*, vol. 23, pp. 225–246, 1983.

[12] R. D. Wollmer, "A hub-spoke seat management model," Unpublished Internal Report, McDonnell Douglas Corporation, Long Beach, Calif, USA, 1986.

[13] M. Dror, P. Trudeau, and S. P. Ladany, "Network models for seat allocation on flights," *Transportation Research B*, vol. 22, no. 4, pp. 239–250, 1988.

[14] R. E. Curry, "Optimal airline seat allocation with fare classes nested by origins and destinations," *Transportation Science*, vol. 24, no. 3, pp. 193–204, 1990.

[15] E. L. Williamson, *Airline network seat inventory controlMethodologies and revenue impacts [Ph.D. thesis]*, Massachusetts Institute of Technology, Cambridge, Mass, USA, 1992.

[16] J. T. Wong, F. S. Koppelman, and M. S. Daskin, "Flexible assignment approach to itinerary seat allocation," *Transportation Research B*, vol. 27, no. 1, pp. 33–48, 1993.

[17] S. V. de Boer, R. Freling, and N. Piersma, "Mathematical programming for network revenue management revisited," *European Journal of Operational Research*, vol. 137, no. 1, pp. 72–92, 2002.

[18] D. Bertsimas and S. De Boer, "Simulation-based booking limits for airline revenue management," *Operations Research*, vol. 53, no. 1, pp. 90–106, 2005.

[19] G. van Ryzin and G. Vulcano, "Simulation-based optimization of virtual nesting controls for network revenue management," *Operations Research*, vol. 56, no. 4, pp. 865–880, 2008.

[20] W. L. Cooper and T. Homem-De-mello, "Some decomposition methods for revenue management," *Transportation Science*, vol. 41, no. 3, pp. 332–353, 2007.

[21] Ş. İlker Birbil, J. B. G. Frenk, J. A. S. Gromicho, and S. Zhang, "A network airline revenue management framework based on decomposition by origins and destinations," *Transportation Science*, 2013.

[22] D. Kahnema and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, pp. 263–292, 1979.

[23] B. Liu, "Why is there a need for uncertainty theory?" *Journal of Uncertain Systems*, vol. 6, no. 1, pp. 3–10, 2012.

[24] B. Liu, *Uncertainty Theory*, Springer, Berlin, Germany, 2nd edition, 2007.

[25] B. Liu, *Uncertainty Theory*, Springer, Berlin, Germany, 4th edition, 2013.

[26] B. Liu, *Uncertain Theory: A Branch of Mathematics for Modeling Human uncertainty*, Springer, Berlin, Germany, 2010.

[27] Z. Peng and K. Iwamura, "A sufficient and necessary condition of uncertainty distribution," *Journal of Interdisciplinary Mathematics*, vol. 13, no. 3, pp. 277–285, 2010.

[28] Y. Liu and M. Ha, "Expected value of function of uncertain variables," *Journal of Uncertain Systems*, vol. 4, pp. 181–186, 2010.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 436512, 7 pages
http://dx.doi.org/10.1155/2014/436512

Research Article

Optimal Coordinated Strategy Analysis for the Procurement Logistics of a Steel Group

Lianbo Deng,[1] Zhuqiang Qiu,[2] Pengfei Liu,[2] and Wenzhong Xiao[1]

[1] *School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China*
[2] *School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China*

Correspondence should be addressed to Lianbo Deng; lbdeng@csu.edu.cn

Received 17 February 2014; Accepted 19 May 2014; Published 4 June 2014

Academic Editor: Andy H. F. Chow

This paper focuses on the optimization of an internal coordinated procurement logistics system in a steel group and the decision on the coordinated procurement strategy by minimizing the logistics costs. Considering the coordinated procurement strategy and the procurement logistics costs, the aim of the optimization model was to maximize the degree of quality satisfaction and to minimize the procurement logistics costs. The model was transformed into a single-objective model and solved using a simulated annealing algorithm. In the algorithm, the supplier of each subsidiary was selected according to the evaluation result for independent procurement. Finally, the effect of different parameters on the coordinated procurement strategy was analysed. The results showed that the coordinated strategy can clearly save procurement costs; that the strategy appears to be more cooperative when the quality requirement is not stricter; and that the coordinated costs have a strong effect on the coordinated procurement strategy.

This paper is dedicated to the memory of our best friend, Dr. Zhuqiang Qiu

1. Introduction

According to their scope, coordinated procurement logistics can be divided into two forms: internal coordinated procurement and enterprise alliance coordinated procurement. In recent years, through eliminating outdated production capacity and merging and reorganizing between corporations, China's steel industry has achieved the scale production. However, the advantage of large = scale economies has not been brought into full play, and the internal logistics system between subsidiaries lacks integral coordination.

China's steel industry output accounted for 4% of GDP. The coordinated logistics of this industry shows the following relevant characteristics.

(1) Industry characteristics: the main raw materials for steel enterprises are iron ore, scrap steel, coke, coking coal, and so on. There are obvious homogeneity and substitutability requirements, which provide the operation space for coordinated procurement management.

(2) The competitive environment of the raw material market: the iron ore, coal, and other major upstream industries have a higher industrial concentration than the steel industry, so the whole steel industry faces a relatively unfavourable situation for the negotiation of the prices of raw materials. There is a high correlation between the logistics procurement cost and the degree of synergy of the steel industry.

(3) The resources and industrial layout of China's steel industry: the demand for steel and the distribution of resources in the various regions of China are not balanced, and this leads to a high cost of steel circulation and a low circulation efficiency. Steel production in 2012 was 716.54 million tons, but the total transportation volume of raw materials and steel product was at least 1 billion tons. Collaboration and the integration of procurement logistics are beneficial

for the steel industry, allowing it to reduce circulation costs significantly and improve market competitiveness.

Increasing the level of coordinated logistics procurement and reducing the production costs of the industry have therefore become a key strategy for the survival and development of China's steel industry.

2. Literature Review

Looking at the procurement logistics for a single steel enterprise, Roy and Guin [1] built a conceptual model of just-in-time purchasing for a steel company in India. They considered the identification and classification of raw material, supplier availability, and goods consolidation of distribution outlets. For the raw material procurement of a large steel plant, considering three main factors (the selection of raw material model, supplier, and order quantity), Gao and Tang [2] constructed a multiple objective linear programming (MOLP) model for procurement decisions. Hafeez and colleagues [3] considered factors such as human resources, organization, and technology, using the dynamic structure of the integrated system, and described a two-level steel supply chain that achieves a minimized inventory level under the condition of capacity constraints and limits on raw material procurement lead time.

By analysing the procurement policy for iron ore and coke of Japanese steel enterprises in the late twentieth century, Chang [4] argued that changing technology and institutional structure made Japanese procurement decisions more consistent and that this laid the foundations for coordinated procurement between enterprises. Potter et al. [5] made a thorough study of the development process of the British steel supply chain from the traditional mode to the integrated mode in the twentieth century and analysed in detail the impact of the changes on inventory, ordering lead time, and asset utilization. Faes et al. [6] considered that coordinated procurement can lead to better internal exchange of information, an improved market negotiation strategy, significant cost savings, a greater impact on the monopoly market, and a better understanding of the market and cost structure. Akkermans et al. [7] established a theoretical model for coordination and studied the important effects of nontechnological factors on achieving synergy. Essig [8] found that coordinated procurement can reduce transaction costs, allow a lower purchase price to be obtained, and lead to a more efficient use of procurement staff. Bishop [9] showed that coordinated procurement can lead to the integration of the purchase process, better continuity and coordination, and economies of scale. Tella and Virolainen [10] argued that coordinated procurement members wanted to reduce their procurement costs and achieve lower management costs, lower logistics costs, and higher mobility of the inventory. Helo [11] proved that demand coordination was important for improving the capacity of the supply chain.

Türkay et al. [12] established a model and made a quantitative analysis of the cooperation between businesses in the chemical industry. Kraljic's [13] model briefly described the procurement strategy for different materials from the perspective of the profit impact and supply risk involved in procurement. Fu and Piplani [14] established a model that evaluated supplier coordination based on inventory to simulate and assess distributors' performance before and after coordination. The calculations showed that the coordination of suppliers can improve the performance of the whole supply chain. Keskinocak and Savaşaneril [15] used a game theory method to study the coordinated procurement of two competing purchasers. Goyal and Satir [16] used an indirect group strategy to seek a combination of the optimal basic cycle and order frequency to make the total relevant cost a minimum, to achieve optimization of multispecies coordinated procurement. Federgruen and Zheng [17] adopted a direct group strategy and used a heuristic algorithm to optimize the coordinated procurement. Chakravarty and Goyal [18] adopted a dependent and group strategy and used dynamic programming to optimize coordinated procurement. Gurnani [19] studied the design of a supplier quantity discount programme, which is the coordinated procurement of two heterogeneous buyers with different requirement processes and cost parameters. For multiperiod multiproduct batch procurement, Lu et al. [20] established a mixed integer programming model with a constraint on transport capacity and variable transport price, which determined the optimal procurement quantity, by using the Lagrange relaxation theory. Xiang et al. [21] assumed that a group regularly orders and intensively purchases, under the condition of independent demand, from the subsidiaries in a group company and established an optimal order quantity model.

This paper studies the optimization of a group's internal coordinated procurement logistics, when combined with the characteristics f the raw material procurement logistics of steel enterprises. Comparing with other studies, we take the full logistics cost of coordinated procurement strategy and the quality of demand into account in the coordinated procurement problem. On the basis of the optimization model in [22], this paper improves the solution algorithm and analyzes the effect of different parameters on the coordinated procurement strategy.

The rest of the paper is organized as follows. Section 3 presents a brief description of optimization model. In Section 4, we describe our approach in solution algorithm. The base example and its results are shown in Section 5. Results under the conditions of different parameters are analysed in Section 6. Lastly, the conclusions of our findings are summarized in Section 7.

3. Optimization Model

This paper studies the CPS of a steel group company that is equipped with a coordinated procurement department and has m subsidiaries (or similar procurement entities). The set of subsidiaries is $I = \{i, i = 1, \ldots, m\}$, and the order quantity of raw materials is Q_i during a period of length t. For simplicity, we assume that all subsidiaries in the coordinated procurement alliance have the same purchase frequency.

In the supply market, there are n suppliers providing the raw materials; the set of suppliers is $J = \{j, j = 1, \ldots, n\}$. We introduce $\gamma(i, j)$, the quality satisfaction degree (QSD) of subsidiary i for the raw material provided by supplier j, where $\gamma(i, j) \in [0, 1]$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. That is,

$$\gamma(i, j) \geq \gamma_0(i), \quad i = 1, \ldots, m, j = 1, \ldots, n, \tag{1}$$

where $\gamma_0(i)$ is the basic requirement of subsidiary i for raw materials.

The CPS is to procure raw materials for all or some of the subsidiaries by the coordinated procurement department. It can be expressed as $\pi = \{\pi_k = (I_k \leftrightarrow j_k) \mid I_k \subset I, j_k \in J\}$, where π_k is a sub-CPS of the CPS, namely, the supply relationship between the set of subsidiaries I_k and the supplier j_k. We introduce

$$\delta_k = \begin{cases} 1 & |I_k| \geq 2, \pi_k \in \pi \\ 0 & |I_k| = 1, \pi_k \in \pi, \end{cases} \tag{2}$$

where $|I_k|$ is the number of the subsidiaries in I_k, for sub-CPS π_k. When $\delta_k = 1$, the CPS should be adopted; when $\delta_k = 0$, the independent procurement strategy (IPS) should not be followed.

The sub-CPS should satisfy

$$I_{k_1} \cap I_{k_2} = \emptyset \quad \pi_{k_1}, \pi_{k_2} \in \pi \tag{3}$$

$$\bigcup_{\pi_k \in \pi} I_k = I \tag{4}$$

$$Q(\pi_k) = \bigcup_{i \in I_k} Q_i \quad \pi_k \in \pi, \tag{5}$$

where $Q(\pi_k)$ is the ordering quantity of raw material for the sub-CPS π_k.

Coordinated logistics procurement costs include order preparation costs, storage costs, purchase costs, and transportation costs.

Let c^d be the order preparation cost of one batch, q_k the order quantity of one batch, and c^s the storage cost per unit of raw materials in the coordinated procurement.

The supply price p_k and the unit transportation cost q_k of a sub-CPS π_k are, respectively,

$$p_k = p_{j_k} - rq_k \tag{6}$$

$$C_k^y = C_0^y - C^y q_k, \tag{7}$$

where p_{j_k} and r, respectively, refer to the initial price and the discount coefficient, $p_{j_k} > 0$ and $r \geq 0$; C_0^y and C^y refer to the parameters of the transportation cost, $C_0^y > 0$ and $C^y \geq 0$.

In contrast to independent procurement strategy (IPS), the additional coordination costs $C_0 (C_0 > 0)$ need to be paid in CPS.

Thus, the total logistics costs for sub-CPS π_k are

$$C_k = C_0 \delta_k + Q(\pi_k) l_{j_k} C_k^y + Q(\pi_k) p_k + \frac{c^d Q(\pi_k)}{q_k} + \frac{t}{2} c^s q_k. \tag{8}$$

Then the optimum order quantity and the optimal cost of π_k are, respectively,

$$\begin{aligned} q_k^* &= \sqrt{\frac{c^d Q(\pi_k)}{\left((t/2) c^s - C^y l_{j_k} Q(\pi_k) - rQ(\pi_k)\right)}}, \\ C_k^* &= C_0 \delta_k + Q(\pi_k)\left(C_0^y l_{j_k} + p_{j_k}\right) \\ &\quad + 2\sqrt{\frac{t}{2} c^d c^s Q(\pi_k) - \left(C^y l_{j_k} + r\right)[Q(\pi_k)]^2}. \end{aligned} \tag{9}$$

The above-described situation can, in accordance with Xiao and Qiu [22], be formulated as a multiobjective optimization model as follows:

$$\max \quad F_1 = \sum_{\pi_k \in \pi} \sum_{i \in \pi_k} \gamma(i, j) \tag{10}$$

$$\max \quad F_2 = \frac{1}{\sum_{\pi_k \in \pi} C_k^*} \tag{11}$$

$$\text{s.t.} \quad \text{Formations (1), (3), (4), and (5).} \tag{12}$$

In this model, the objective function equation (10) is to maximize the QSD for the aggregated demand; the objective function equation (11) is to minimize total procurement cost when all sub-CPSs take the most economic order quantity.

4. Solution Algorithm

In order to solve the multiobjective model, we introduce a balancing factor α $(0 \leq \alpha \leq 1)$ of the QSD of total demand, to balance the two objectives. Then the objectives are transformed into the following:

$$\max F = \alpha \sum_{\pi_k \in \pi} \sum_{i \in \pi_k} \gamma(i, j) + \frac{(1 - \alpha)\beta}{\sum_{\pi_k \in \pi} C_k^*}, \tag{13}$$

where β is the cost conversion coefficient of F_2.

In order to obtain the optimal solution of the global situation, we use an intelligent optimization algorithm, the simulated annealing (SA) algorithm.

The annealing schedule of the SA algorithm refers to a set of process parameters used to control the algorithm, including the generation of a neighbourhood solution, the control of temperature, the number of iterations at each temperature, and the termination rule.

Since sub-CPS π_k stands for the supplier relationship between the subsidiary sets I_k and the supplier j, we can adopt a certain rule to select the supplier for each subsidiary; the subsidiaries which select the same supplier and their chosen supplier are composed of the sub-CPS π_k. Automatically the CPS, based on the above initial solution generation method, satisfies the constraints (3) and (4).

According to the objective function, we can select suppliers as follows. For $i \in I$, let its set of alternative suppliers be $J_i = \{j \mid \gamma(i, j) \geq \gamma_0(i)\}$. $\forall j \in J_i$, a sub-CPS which consists

of i and j separately, is defined as $\pi(i, j) = \{(i \leftrightarrow j)\}$. The objective function of this sub-CPS is

$$F(i, j) = \alpha \sum_{\pi(i,j)} \gamma(i, j) + (\alpha - 1) \sum_{\pi(i,j)} \frac{1}{C^*_{\pi(i,j)}}. \quad (14)$$

Then, we choose a supplier j from J_i according to a selection probability $\varphi(i, j)$, which is

$$\varphi(i, j) = \frac{F(i, j)}{\sum_{j' \in J_i} F(i, j')}. \quad (15)$$

According to (15) we select the corresponding supplier for each $i \in I$ and merge the sub-CPSs for the same supplier into a new sub-CPS. The definition of J_i ensures that all sub-CPSs satisfy the constraint (1) and are feasible.

On the basis of the above analysis, the algorithm is as follows.

Step 0. Initialize. Set the initial temperature to T_0, the current temperature $T = T_0$, the current iterations $h = 1$, the optimal solution $\pi_{\text{opt}} = \Phi$, and the objective function $F_{\text{opt}} = \infty$. Use (15) to calculate $\varphi(i, j)$, $\forall i \in I$, $j \in J$.

Step 1. Randomly generate a sub-CPS π_i, $\forall i \in I$.

Step 2. Calculate the objective function value F_i of each sub-CPS π_i using (13).

Step 3. Update the current solution according to the Metropolis criterion: if $F < F_{\text{opt}}$, let $\pi_{\text{opt}} = \pi_0$ and $F_{\text{opt}} = F$; otherwise randomly generate a numerical value ρ in $(0, 1)$ and if $\rho < \exp(-(F - F_{\text{opt}})/T_i)$, then set $\pi_{\text{opt}} = \pi_0$, $F_{\text{opt}} = F$. Otherwise the neighbourhood solution is refused.

Step 4. Judge the number of iterations at the same temperature. The number of iterations at each temperature is restricted by the lower limit of iterations L, the accepting rate, δ, of the neighbourhood solution, and the upper limit H.

If the iterations satisfy the restrictions, then set $h = h + 1$ and move to Step 1; otherwise stop the iterations at the same temperature and move to Step 5.

Step 5. Judge the convergence rule which is a minimum temperature T_f. If it is not satisfied, then update the current temperature $T = \varepsilon T$, where ε is a constant close to 1 and move to Step 1; otherwise terminate the algorithm and output the optimal solutions π_{opt} and F_{opt}.

5. Example Analysis

5.1. The Base Example. A steel group company has four subsidiary companies: i_1, i_2, i_3, and i_4. A certain raw material is offered by five suppliers: j_1, j_2, j_3, j_4, and j_5. The coordinated cost of coordinated procurement $C_0 = 10000$. The transportation cost coefficient $C_0^y = 2$, $C^y = 0.001$. The order price discount coefficient $r = 0.01$. The preparation cost of each batch order $c^d = 50$. The unit storage cost for the raw materials $c^s = 10$.

Table 1: Demand of subsidiaries.

Subsidiary	$\gamma_0(i)$	Demand
i_1	0.60	350
i_2	0.65	410
i_3	0.62	220
i_4	0.70	190

Table 2: Transportation distance and initial price.

Supplier	j_1	j_2	j_3	j_4	j_5
Transportation distance	1200	1000	1500	1800	1300
Initial price	35	36	30	32	38

Table 3: QSD of suppliers.

Subsidiary	Supplier				
	j_1	j_2	j_3	j_4	j_5
i_1	0.95	0.92	0.55	0.95	0.92
i_2	0.90	0.95	0.95	0.62	0.95
i_3	0.50	0.85	0.87	0.60	0.85
i_4	0.94	0.88	0.81	0.90	0.64

Table 4: Sub-CPSs for the base example.

π_i	$Q(\pi_i)$	q_i	$\overline{\gamma}_i$	C_i^*
π_1	190	2.75	0.940	$0.53 * 10^6$
π_2	980	10.78	0.907	$2.34 * 10^6$

Table 5: Sub-IPSs of base example.

π_i	$Q(\pi_i)$	q_i	$\overline{\gamma}_i$	C_i^*
π_1	190	2.75	0.94	$0.53 * 10^6$
π_2	350	3.96	0.92	$0.83 * 10^6$
π_3	220	3.10	0.87	$0.73 * 10^6$
π_4	410	4.70	0.95	$1.23 * 10^6$

In the period of $t = 300$ days, the quantity and quality of the raw materials required are as set out in Table 1.

The transportation distance from the suppliers to the steel group and the initial price are as set out in Table 2.

The QSD of the suppliers' products to meet the needs of subsidiaries are as set out in Table 3.

5.2. The Results. Using the balance factors for the demand satisfaction degree $\alpha = 0.7$ and $\beta = 10e^6$, we get the CPS $\pi = \{\pi_1 = (i_4 \leftrightarrow j_1), \pi_2 = (\{i_1, i_2, i_3\} \leftrightarrow j_2)\}$, and the optimal objective function value is 3.60758. This CPS and its sub-CPSs are seen in Figure 1 and Table 4, respectively.

When we do not adopt the CPS, the optimal independent procurement strategy (IPS) $\pi = \{\pi_1 = (i_4 \leftrightarrow j_1), \pi_2 = (i_1 \leftrightarrow j_2), \pi_3 = (i_3 \leftrightarrow j_3), \pi_4 = (i_2 \leftrightarrow j_5)\}$, and the optimal objective function value $C = 3.47748$. This procurement strategy and its subpolicy are shown in Figure 2 and Table 5, respectively. Compared with the IPS, the optimal objective function value of the CPS is 3.74% higher. On the premise that

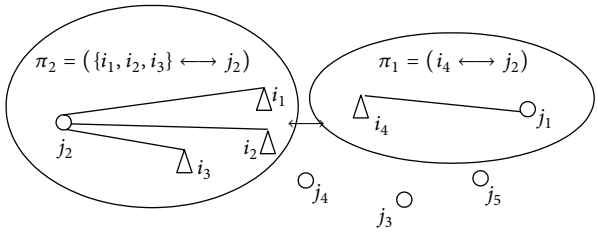


Figure 1: CPS of base example.

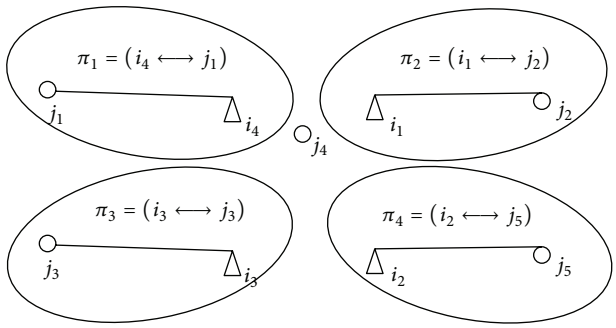


Figure 2: IPS of base example.

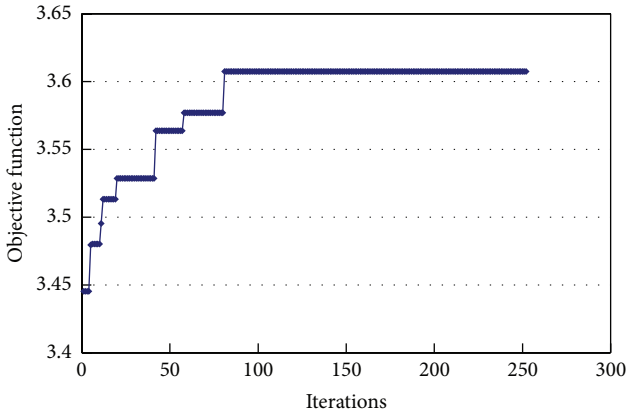


Figure 3: Convergence efficiency.

the procurement quality is met, the procurement cost drops remarkably.

5.3. Analysis of Solving Efficiency. During the solving of the base example, when the number of iterations is increased, the optimal solution varies as shown in Figure 3. As we can see, this algorithm has a good effect and can quickly converge to the optimal solution. The convergence speed and solution quality are both very satisfactory.

6. Results under Other Conditions

Several factors, including the subsidiaries' requirements for the quality of the raw materials, the supply characteristics of the products in the market, the procurement cost structure, and the balance factor α, together have an influence on the CPS. Here, using the base example, we analyse the results when various factors are changed.

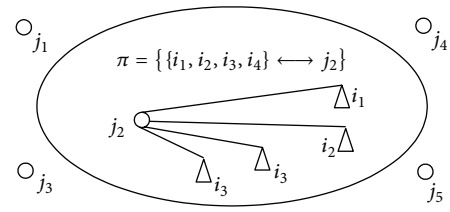


Figure 4: CPS under uniform QSD.

Table 6

Subsidiary	Supplier				
	j_1	j_2	j_3	j_4	j_5
i_1	0.95	0.92	0.87	0.82	0.85
i_2	0.95	0.92	0.87	0.82	0.85
i_3	0.95	0.92	0.87	0.82	0.85
i_4	0.95	0.92	0.87	0.82	0.85

Table 7: Sub-CPS under uniform QSD.

π_i	$Q(\pi_i)$	q_i	$\overline{\gamma}_i$	C_i^*
π	1170	16.57	0.92	$2.79 * 10^6$

Table 8: Sub-CPSs under $C_0 = 0$.

π_i	$Q(\pi_i)$	q_i	$\overline{\gamma}_i$	C_i^*
π_1	1170	16.57	0.90	$2.78 * 10^6$
π_2	980	10.77	0.907	$2.33 * 10^6$

6.1. The Results with a Uniform QSD Condition. In order to find the influence of QSD on the CPS, we take a uniform QSD for each supplier, as set out in Table 6.

We get the optimal CPS $\pi = \{(\{i_1, i_2, i_3, i_4\} \leftrightarrow j_2)\}$, and the optimal objective function value $C = 3.65136$. The results and the specific procurement are shown in Figure 4 and Table 7. Compared with the base example, the optimal objective function value under the uniform QSD condition is increased by 1.21%, and the CPS has an advantage in terms of economies of scale.

6.2. Effects of the Supply Price Discount on the CPS. In this section we analyse the effects of the discount coefficient r on the CPS. The supply price decreases when the order quantity increases, as (6) shows. However, the raw materials for a steel company are both huge in quantity and low in price; the low discount can be provided. When $r = [0, 0.25]$, the CPSs are the same; $\pi = \{\pi_1 = (i_4 \leftrightarrow j_1), \pi_2 = (\{i_1, i_2, i_3\} \leftrightarrow j_2)\}$. The average QSDs of sub-CPSs π_1 and π_2 are 0.94 and 0.907, respectively, and the order quantities are $Q(\pi_1) = 190$, $Q(\pi_2) = 980$, respectively. The objective function values of the CPSs are a little different, with $C = 3.60756$ when $r = 0$ and $C = 3.60761$ when $r = 0.25$. However, the discount coefficient r has an obvious effect on the order quantity of the

Table 9: Optimal CPSs under different balance factors.

α	CPS					
	π	Q_k	q_k	$\overline{\gamma}_k$	C_k (10^6)	F
0.0	$\pi = \{(\{i_1, i_2, i_3, i_4\} \leftrightarrow j_2)\}$	1170	16.5725	0.9	2.78977	3.58452
0.2	$\pi = \{(\{i_1, i_2, i_3, i_4\} \leftrightarrow j_2)\}$	1170	16.5725	0.9	2.78977	3.58762
0.4	$\pi = \{(\{i_1, i_2, i_3, i_4\} \leftrightarrow j_2)\}$	1170	16.5725	0.9	2.78977	3.59071
0.6	$\pi = \{(\{i_1, i_2, i_3, i_4\} \leftrightarrow j_2)\}$	1170	16.5725	0.9	2.78977	3.59381
0.8	$\pi_1 = (i_4 \leftrightarrow j_1)$	190	2.75351	0.94	0.5294	3.62505
	$\pi_2 = (\{i_1, i_2, i_3\} \leftrightarrow j_2)$	980	10.7756	0.907	2.33982	
1.0	$\pi_1 = (i_4 \leftrightarrow j_1)$	190	2.75351	0.94	0.52940	3.7100
	$\pi_2 = (i_3 \leftrightarrow j_3)$	220	3.09546	0.87	0. 733107	
	$\pi_3 = (i_1 \leftrightarrow j_4)$	350	4.578	0.95	1.37965	
	$\pi_4 = (i_2 \leftrightarrow j_5)$	410	4.42068	0.95	0. 976875	

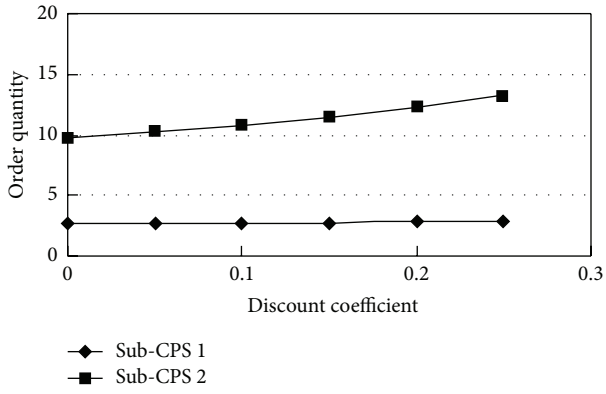


Figure 5: Effect of r on order quantity.

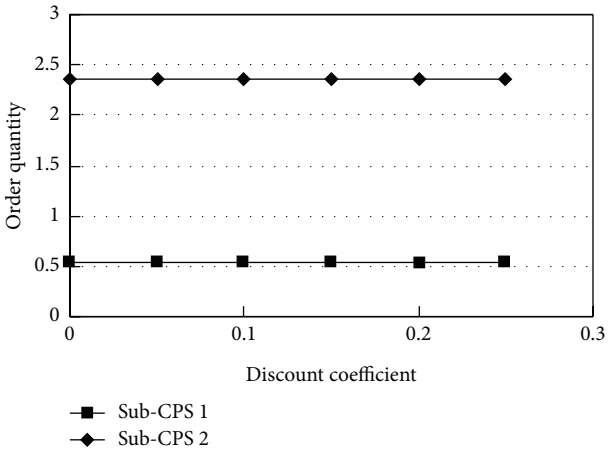


Figure 6: Effect of r on procurement cost.

sub-CPSs π_1, π_2, as Figure 5 shows. In contrast to the order quantity, the difference in the procurement cost of the sub-CPSs is insignificant, as Figure 6 shows. This is because the price discount for the raw materials is limited. At the same time, the economic order quantity has a regulating function to the effect of price discount.

Thus, within a certain range (e.g., $r \in [0, 0.25]$), price discount has little effect on the coordinated procurement strategy.

6.3. Impact of the Coordinated Costs on the CPS. The coordinated costs reflect the operation and the coordinated level of the coordinated procurement department in the group company. Compared with a sub-CPS $\pi_k = \pi(I_k, j)$, the sub-IPSs are composed of each of the subsidiaries $i \in I_k$ and the supplier j; the difference between the logistics costs of the CPS and the IPS is expressed as

$$\Delta C_k^* = C_0\delta_k + 2\sqrt{\frac{t}{2}c^d c^s \sum_{i\in I_k} Q_i - c^d\left(C^y l_{j_k} + r\right)\left[\sum_{i\in I_k} Q_i\right]^2} - \sum_{i\in I_k} 2\sqrt{\frac{t}{2}c^d c^s Q_i - c^d\left(C^y l_{j_k} + r\right)Q_i^2}. \tag{16}$$

From (16), we can see that coordinated costs have a direct effect on the CPS and that $\Delta C_k^* \geq 0$ is a necessary condition for accepting the CPS. $C_0 = 0$ is an ideal situation with the corresponding CPS $\pi = \{\pi_1 = (\{i_1, i_2, i_3, i_4\} \leftrightarrow j_2)\}$ and objective function value $C = 3.599$, and the sub-CPSs are shown in Table 8. Although the CPS when $C_0 = 0$ is still the same as in the base sample, there are some savings in procurement costs.

However, when $C_0 \geq 400000$, the optimal CPS tends to disintegrate. Every subsidiary company has its own supplier and the CPS is similar to the one shown in Figure 2.

6.4. Effects of the Balance Factor on the CPS. The balance factor α of the QSD reflects the weighting relationship between the two objective functions and the procurement requirements of the subsidiaries. Under the extreme condition, when $\alpha = 0$, the company just needs to consider the procurement costs; when $\alpha = 1.0$, the company only needs to

consider the QSD. With different values of α, the CPS varies, as shown in Table 9.

When α is smaller, the purchase logistics of the subsidiaries tend to lead to a higher degree of coordination, and the procurement costs are lower; when $\alpha = 0.8$, the CPS shows evidence of disintegration; and when $\alpha = 1.0$, every subsidiary purchases its own materials from its own suppliers.

7. Conclusions

This paper focused on the optimization of coordinated procurement logistics for a steel group. A simulated annealing algorithm was used to solve this problem. From our analysis of the numerical sample, we can draw the following conclusions.

(1) The CPS can adapt better than the IPS to the internal procurement logistics of the steel company and bring a significant saving in procurement costs.

(2) When the QSD for the quality of the material is not too high or there is no difference between the materials or they are substitutable in the market, the CPS appears to be highly cooperative.

(3) Coordinated costs have a strong effect on the CPS, so a highly advanced coordinated procurement system is the basis for building a significantly efficient coordinated procurement strategy.

Conflict of Interests

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

Acknowledgments

This research is supported by the Science and Technology Research Development Project of the China Railway Corporation (Major Program, 2013X004-A), the Research Fund for the Fok Ying Tong Education Foundation of Hong Kong (Project no. 132017), and the National Natural Science Foundation of China (70901076). This paper is dedicated to the memory of the authors' best friend, Dr. Zhuqiang Qiu.

References

[1] R. N. Roy and K. K. Guin, "Proposed model of JIT purchasing in an integrated steel plant," *International Journal of Production Economics*, vol. 59, no. 1, pp. 179–187, 1999.

[2] Z. Gao and L. Tang, "A multi-objective model for purchasing of bulk raw materials of a large-scale integrated steel plant," *International Journal of Production Economics*, vol. 83, no. 3, pp. 325–334, 2003.

[3] K. Hafeez, M. Griffiths, J. Griffiths, and M. M. Naim, "Systems design of a two-echelon steel industry supply chain," *International Journal of Production Economics*, vol. 45, no. 1–3, pp. 121–130, 1996.

[4] H.-S. Chang, "Coking coal procurement policies of the Japanese steel mills: changes and implications," *Resources Policy*, vol. 23, no. 3, pp. 125–135, 1997.

[5] A. Potter, R. Mason, M. Naim, and C. Lalwani, "The evolution towards an integrated steel supply chain: a case study from the UK," *International Journal of Production Economics*, vol. 89, no. 2, pp. 207–216, 2004.

[6] W. Faes, P. Matthyssens, and K. Vandenbempt, "The pursuit of global purchasing synergy," *Industrial Marketing Management*, vol. 29, no. 6, pp. 539–553, 2000.

[7] H. Akkermans, P. Bogerd, and J. Van Doremalen, "Travail, transparency and trust: a case study of computer-supported collaborative supply chain planning in high-tech electronics," *European Journal of Operational Research*, vol. 153, no. 2, pp. 445–456, 2004.

[8] M. Essig, "Purchasing consortia as symbiotic relationships: developing the concept of 'consortium sourcing'," *European Journal of Purchasing and Supply Management*, vol. 6, no. 1, pp. 13–22, 2000.

[9] J. E. Bishop, "Consortium purchasing," *New Directions For Higher Education*, vol. 120, pp. 81–88, 2002.

[10] E. Tella and V.-M. Virolainen, "Motives behind purchasing consortia," *International Journal of Production Economics*, vol. 93-94, pp. 161–168, 2005.

[11] P. T. Helo, "Dynamic modelling of surge effect and capacity limitation in supply chains," *International Journal of Production Research*, vol. 38, no. 17, pp. 4521–4533, 2000.

[12] M. Türkay, C. Oruç, K. Fujita, and T. Asakura, "Multi-company collaborative supply chain management with economical and environmental considerations," *Computers and Chemical Engineering*, vol. 28, no. 6-7, pp. 985–992, 2004.

[13] P. Kraljic, "Purchasing must become supply management," *Harvard Business Review*, vol. 61, no. 5, pp. 109–117, 1983.

[14] Y. Fu and R. Piplani, "Supply-side collaboration and its value in supply chains," *European Journal of Operational Research*, vol. 152, no. 1, pp. 281–288, 2004.

[15] P. Keskinocak and S. Savaşaneril, "Collaborative procurement among competing buyers," *Naval Research Logistics*, vol. 55, no. 6, pp. 516–540, 2008.

[16] S. K. Goyal and A. T. Satir, "Joint replenishment inventory control: deterministic and stochastic models," *European Journal of Operational Research*, vol. 38, no. 1, pp. 2–13, 1989.

[17] A. Federgruen and Y.-S. Zheng, "The joint replenishment problem with general joint cost structures," *Operations Research*, vol. 40, no. 2, pp. 384–403, 1992.

[18] A. K. Chakravarty and S. K. Goyal, "Multi-item inventory grouping with dependent set-up cost and group overhead cost," *Engineering Costs and Production Economics*, vol. 10, no. 1, pp. 13–23, 1986.

[19] H. Gurnani, "A study of quantity discount pricing models with different ordering structures: order coordination, order consolidation, and multi-tier ordering hierarchy," *International Journal of Production Economics*, vol. 72, no. 3, pp. 203–225, 2001.

[20] K. Lu, C.-H. Yang, and D.-M. Dai, "A Lagrangian-based heuristic algorithm for multi-product capacitated lot sizing with time-varying transportation costs," *System Engineering: Theory and Practice*, vol. 28, no. 10, pp. 47–52, 2008.

[21] J. Q. Xiang, P. Q. Huang, and J. Li, "Optimal order model under periodic order policy of centralized procurement in enterprise group," *Journal of Shang Hai Jiaotong University*, vol. 39, no. 3, pp. 474–478, 2005.

[22] W. Z. Xiao and Z. Q. Qiu, "Optimization of coordinated procurement strategy in steel group," *Journal of Computer Applications*, vol. 7, pp. 1913–1918, 2010.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 143501, 13 pages
http://dx.doi.org/10.1155/2014/143501

Research Article

An Integrative Approach with Sequential Game to Real-Time Gate Assignment under CDM Mechanism

Jun-qiang Liu, Ma-lan Zhang, Peng-chao Chen, Ji-wei Xie, and Hong-fu Zuo

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Correspondence should be addressed to Jun-qiang Liu; liujunqiang@nuaa.edu.cn

Received 26 December 2013; Revised 7 May 2014; Accepted 8 May 2014; Published 1 June 2014

Academic Editor: Hu Shao

This paper focuses on real-time airport gate assignment problem when small-scale or medium- to large-scale flight delays occur. Taking into account the collaborative decision making (CDM) of the airlines and the airport, as well as the interests of multiagent (airlines, airports, and passengers), especially those influenced by flight banks, slot assignment and gate assignment are integrated into mixed set programming (MSP), and a real-time gate assignment model is built and solved through MSP coupled with sequential game. By this approach, the delay costs of multiagent can be minimized simultaneously; the fuel consumption of each airline can be basically equalized; the computation time can be significantly saved by sequential game; most importantly, the collaboration of the airlines and the airport is achieved so that the transferring cost caused by the delay of flight banks can be decreased as much as possible. A case study on small-scale flight delays verifies that the proposed approach is economical, robust, timesaving, and collaborative. A comparison of the traditional staged method and the proposed approach under medium- to large-scale flight delays proves that the integrative method is much more economical and timesaving than the traditional staged method.

1. Introduction

According to the statistical data of China's civilian aviation in 2011 [1], the existing 175 airports include 3 4F-class airports, 30 4E-class airports, 40 4D-class airports, and 85 4C-class airports. The number of civilian airports is expected to increase up to 244 by 2020. Meanwhile, the reconstruction and expansion of most major airports will be performed to meet the increasing demands. In this case, the resource management of these airports will turn out to be increasingly complex. Therefore, a great number of researches have been presented to solve the problem of airport resource management, gate assignment included. Gate assignment problem can be divided into two categories: preassignment and real-time assignment.

In the aspect of preassignment, Bolat developed a gate assignment model that is aimed at minimizing the idle time of the gates both in 1996 and 2000; meanwhile, heuristic algorithm combined with branch-and-bound technique was designed for solution [2]. In 2001, Bolat devised a linearization representation of the same objective and replaced heuristic algorithm with genetic algorithm [3]. In 2003, You et al. [4] were devoted to the study of genetic algorithm, which promoted the multiobjective algorithm for the preassignment of airport gates. In 2005, Lim and Wang constructed a different model that is aimed at reducing the number of conflicts during aircraft operation, and the model was solved by combining tabu search with local search in heuristic algorithm [5]. In 2011, Li et al. proposed an assignment model to obtain the maximum comprehensive efficiency of airlines, airports, and passengers; besides, the solution algorithm was developed based on greedy tabu search [6].

In the aspect of real-time assignment, Cheng designed a gate scheduling knowledge base system [7] by utilizing mathematical programming techniques, in order to support preassignment and real-time assignment. In later studies, they developed two other models to establish a gate assignment expert system: network-based model [8] and rule-based reactive model [9]. In 2010, Li devised an assignment model that was aimed at maximizing the sum of all the flight eigenvalues (aircraft type, flight inbound/outbound time, and the number of passengers) and the gate eigenvalues (idle time

of the gates) [10]. Furthermore, Wei and Liu proposed an approach to perform gate reassignment by minimizing the number of the flights assigned to the apron, the disturbance caused by the assignment, and the walking distance of passengers; meanwhile, a heuristic algorithm combining tabu search was adopted for optimization [11].

The existing research shows that the studies on gate preassignment have evolved and matured over time. However, a comprehensive consideration on the interests of multiagent, namely, airports, airlines, and passengers, has not been given to real-time gate assignment.

Additionally, flight bank [12], a decisive role in determining the satisfaction of passengers and the benefits of airlines, is generally considered in slot assignment rather than gate assignment. Nevertheless, the gate assignment will also be influenced by flight banks, because transfer passengers would have to wait longer if flight banks are delayed, which increases the delay cost of gate assignment. Flight bank is the product of the collaborative decision making (CDM) [13] for airports, airlines, and air traffic control center (ATCC). In CDM mechanism, airports, airlines, and ATCC should work collaboratively to control the entire operation cost. However, almost all kinds of assignment algorithms, such as gate assignment, aircraft sorting, and flight scheduling, are performed separately, which goes against the original intentions of CDM mechanism. In the existing real-time gate assignment [13], airports are given fixed schedule of the delayed flights. Airlines are not informed of the possible gate assignment schemes and not all the information of possible slot rearrangements are delivered to the airport, as a result, the flight banks may be delayed to a great extent. To solve this problem, the CDM should be implemented accurately: ATCC provides updated slots to airlines; airlines choose the slot assignment schemes corresponding to the optimal gate assignment (based on minimum cost principle); finally, the optimal gate assignment is carried out by the airport based on the provided slot assignment schemes. The innovation taking into account the impacts of flight banks while achieving CDM is capable of guaranteeing reasonable slot assignment and gate assignment, thereby saving delay cost for multiagent (airlines, airports, and passengers) significantly, connecting flight banks smoothly, and transferring passengers conveniently.

In order to achieve slot assignment and gate assignment at the same time, integrative recovery strategy [14] is considered as a solution. Generally, traditional integrated recovery strategies, such as MSP, are time-consuming, because the permutation and combination of the optional objects will become enormous with the increasing of the scale of the problem to be solved. However, sequential game has been applied in resource allocation [15, 16] and is proven to be timesaving by deleting the infeasible permutation and combination of the objects from the computation process. Besides, sequential game also guarantees that the benefits of the related game players are maximized. Therefore, sequential game is applied innovatively in real-time gate assignment for timesaving computation and effective cost control.

Consequently, this paper proposes a novel model taking into account the benefits of multiagent, the impact of flight banks, CDM mechanism, fairness principle, and sequential game between the airlines, in order to achieve the following goals. Firstly, the total delay cost, including the walking cost of passengers, the taxiing cost of aircraft, the idle cost of gates, and the waiting cost of transfer passengers should be minimized. Secondly, the taxiing fuel for aircraft of each airline should be balanced. Thirdly, the flight banks containing delayed flights should be connected effectively with minimum adverse impacts on the interests of the airlines. Lastly, unreasonable solutions should be excluded in the first place, thereby saving computation time and resources. To make the model work, MSP combined with sequential game is designed for modeling and solving.

The remainder of this paper is organized as follows. The real-time gate assignment model is established in Section 2. In Section 3, the detailed steps of the solution algorithm are provided. In Section 4, two cases (small-scale flight delays and medium- to large-scale flight delays) are given and the experimental results are analyzed accordingly. Finally, conclusions are drawn in Section 5.

2. Real-Time Gate Assignment Model

2.1. Variable Definition of the Model. The definitions of the variables in the model are given as follows.

(1) N denotes the set of flights and M denotes the set of gates, $i, k \in N$, $j \in M$, $0 < M < N$;

(2) W denotes the set of flight banks, $w \in W$, W_m denotes the set of flight banks for airline m, $W_m \in W$, t_{mw1} is the scheduled arrival time of flight bank w for airline m, and t_{mw2} is the actual arrival time of flight bank w for airline m, $t_{mw2} \geq t_{mw1}$;

(3) Q is the set of aircraft types, A is the set of airlines, $n \in Q$, $m \in A$, Q_i denotes the type of flight i, expressed in numbers, and the bigger Q_i is, the larger flight i is;

(4) S is the set of slots, $s \in S$, S_m is the set of slots for airline m when airline m exchanges slots with other airlines to cut the waiting time of transfer passengers, $S_m \in S$;

(5) X_{ij} is a zero-one variable, and X_{ij} equals 1 when flight i is assigned to gate j, else X_{ij} equals 0; Y_{mw} is a zero-one variable, and Y_{mw} equals 1 when flight bank w belongs to airline m, else Y_{mw} equals 0;

(6) F_i is the fuel consumption of flight i per minute;

(7) $T1_i$ and $T2_i$ denote the taxiing time of flight i before and after the reassignment, respectively;

(8) P_i is the number of passengers in flight i;

(9) $L1_i$ and $L2_i$ denote the walking time for the passengers of flight i before and after the reassignment, respectively;

(10) $I1_j$ and $I2_j$ denote the idle time of gate j before and after the reassignment, respectively;

(11) r_{mw} is the number of transfer passengers in flight bank w of airline m;

(12) $C1$ is the price of jet fuel per kilogram, $C2$ is the walking cost of each passenger per minute, $C3_j$ is the cost of the idle time for gate j per minute, and $C4$ is the waiting cost for each transfer passenger per minute;

(13) R_{ij} is the time when flight i arrives at gate j, L_{ij} is the time when flight i departs from gate j, and R_{kj} is the time when flight k arrives at gate j, $R_{kj} > L_{ij}$;

(14) G_j denotes the type of gate j, expressed in numbers, and the bigger G_j is, the larger gate j is;

(15) ΔT is the buffer time between any two consecutive flights assigned to the same gate;

(16) I_{bj} denotes the beginning of the idle time for gate j and I_{ej} denotes the end of the idle time for gate j;

(17) f_{S_m} is the waiting cost of the transfer passengers for airline m, which is less than the waiting cost of the transfer passengers before the slot exchange of airline m;

(18) X_{S_m} is a zero-one variable, X_{S_m} equals 1 when S_m is the slots combination for airline m after the exchange, else X_{S_m} equals 0;

(19) Ω_m is the set of slot combinations which makes the waiting time of transfer passengers for airline m less after the slot exchange;

(20) ΔK_{nm} is the increased fuel consumption of type n aircraft belonging to airline m; z is the number of airlines, so, $1/z$ is the expected proportion, corresponding to the ideal situation of complete equalization for all airlines.

2.2. Objectives

(1) Minimize the Increased Total Cost. The total cost involves fuel cost, walking cost of passengers, idle cost of gates, and waiting cost of transfer passengers. To achieve the goal of minimizing the increased total cost caused by flight delays, the following four values need to be minimized at the same time: taxiing time of aircraft, walking time of passengers, idle time of gates, and waiting time of transfer passengers. The objective function is formulated as follows:

$$\begin{aligned}\min\ f_1 = &\sum_{i\in N}\sum_{j\in M} X_{ij}\left[F_i\left(T2_i - T1_i\right)C1\right.\\ &\left.+P_i\left(L2_i - L1_i\right)C2\right]\\ &+\sum_{j\in M}\left(I2_j - I1_j\right)C3_j\\ &+\sum_{m\in A}\sum_{w\in W_m} r_{mw}\left(t_{mw2} - t_{mw1}\right)C4.\end{aligned} \tag{1}$$

$F_i(T2_i - T1_i)\,C1$ denotes the increased taxiing cost of flight i; $P_i(L2_i - L1_i)\,C2$ denotes the incremental walking cost for passengers of flight i; $(I2_j - I1_j)\,C3_j$ denotes the change of the idle cost for gate j; $r_{mw}(t_{mw2} - t_{mw1})\,C4$ denotes the increased waiting cost of transfer passengers in flight bank w of airline m.

(2) Minimize the Increased Taxiing Time of Aircraft. Since fuel cost accounts for about 30% of the total operation cost for airlines, fuel cost decrease will make a huge difference in the cost control of airlines. Fuel cost reduction can be achieved by minimizing the incremental taxiing time of all the aircraft according to

$$\min\ f_3 = \sum_{i\in N}\sum_{j\in M} X_{ij}F_i\left(T2_i - T1_i\right). \tag{2}$$

(3) Minimize the Increased Walking Time of Passengers. The reduction of walking distance or time improves the satisfaction of passengers. The purpose of minimizing the walking time of passengers can be realized by

$$\min\ f_2 = \sum_{i\in N}\sum_{j\in M} X_{ij}P_i\left(L2_i - L1_i\right). \tag{3}$$

(4) Minimize the Increased Idle Time of Gates. As airport gates are the core resources of an airport, improving the utilization rate of the idle gates attributes to better airport operation. To realize this goal, the idle time of airport gates can be minimized by

$$\min\ f_4 = \sum_{j\in M}\left(I2_j - I1_j\right). \tag{4}$$

(5) Minimize the Increased Waiting Time of Transfer Passengers. Flight bank has been widely applied in hub airports, so that the transferring efficiency of passengers, the utilization rate of airport resources, and the operational effectiveness of airlines can be improved. Generally, flight bank is capable of connecting flights effectively thus minimizing the waiting time of transfer passengers. However, delayed flight banks will lead to a series of problems, such as increased waiting time of passengers and increased operation cost. The objective function that is aimed at reducing the increased waiting time caused by the delay of flight banks is given by

$$\min\ f_5 = \sum_{m\in A}\sum_{w\in W_m} Y_{mw}r_{mw}\left(t_{mw2} - t_{mw1}\right). \tag{5}$$

Theorem 1. *The increased waiting time of transfer passengers does not necessarily depend on slot assignment.*

Proof of Theorem 1. Suppose V is the set of delay time for all the flights, $V = \{v_1, v_2, \ldots, v_z\}$, $k = 1, 2, \ldots, z$, $v_k = (s, i)$ means slot s is randomly assigned to flight i, and $v_k = |t_s - t_i|$, where t_i denotes the scheduled arrival time of flight i and t_s denotes the time of slot s. Since the number of flights is equal to the number of slots and all the delayed flights are not cancelled, each flight can be assigned to one and only one slot; then, we have

$$\sum_{i\in N}\sum_{s\in S} x_{ij}\left|t_s - t_i\right| = \sum_{k=1}^{z} v_k = \left|\sum_{s\in S} t_s - \sum_{i\in N} t_i\right|. \tag{6}$$

As slot s is randomly assigned to flight i and the number of slots is equal to the number of flights, $|\sum_{s\in S} t_s - \sum_{i\in N} t_i|$ is

a constant. Hence, the increased waiting time of transfer passengers does not necessarily depend on the slot assignment, but the delay of flight banks. □

Theorem 2 (gate assignment hinges on the slot assignment). *Gate assignment depends on the arrival time of flights and the idle time of available gates; therefore, slot assignment plays a decisive role in gate assignment.*

According to Theorems 1 and 2, the slot assignment for delayed flights should be implemented with minimum delay of flight banks, so that the optimal gate assignment with minimum delay cost can be produced. This is why flight bank is involved in the objective of minimizing the waiting cost of transfer passengers in gate assignment.

(6) Minimize the Waiting Cost by Optimizing the Slot Assignment through Non-Zero-Sum Sequential Game. The slots assigned to airlines are exchangeable, so airlines can reduce the waiting time of transfer passengers by exchanging the slots with each other. The process of slot exchange can be achieved through non-zero-sum sequential game between airlines. In sequential game, all the airlines are aware of their previous policy or selection and have to make their current decisions according to their tradeoff of future possibilities. Zero-sum sequential game refers to the situation that the income of one side is equal to the loss of the other side. As our subject is about controlling the loss caused by the flight delays for all the related airlines, non-zero-sum sequential game theory [15] is adopted in this application. The model of non-zero-sum sequential game is given by

$$G = \left\{A, (S_m)_{m\in A}, (\rho_m)_{m\in A}, P(\rho)\right\}, \tag{7}$$

where A denotes the set of airlines; S_m is the set of all optional slot series for airline m, $\forall m \in A$; ρ_m is the realization probability of S_m; $P(\rho)$ denotes the expected revenue matrix.

Theorem 3. *In a sequential game, any realization probability points to a behavior strategy.*

Proof of Theorem 3. $\rho_m(s_m) = \prod_{\alpha\in s_m}\beta_m(\alpha)$, where $\rho_m(s_m)$ is the realization probability for airline m to obtain slot series s_m ($s_m \in S_m$), β_m is the probability distribution of $T(h_m)$, and $T(h_m)$ is the set of optional policies under information set h_m for airline m.

Therefore, any realization probability comes from a corresponding behavior strategy.

The set of behavior sequences for airline m on information set h_m is denoted by $\lambda(h_m)$, γ is an expansion of the behavior sequences denoted by $\lambda(h_m)\gamma$ ($\lambda(h_m)\gamma = \lambda(h_m)\cup\gamma$), and the realization probability of $\lambda(h_m)$ can be denoted by $\rho_m(\lambda(h_m))$.

Consequently, the behavior γ on information set h_m can be confirmed by

$$\beta_m(\gamma) = \frac{\rho_m(\lambda(h_m)\gamma)}{\rho_m(\lambda(h_m))}, \tag{8}$$

where $\rho_m(\lambda(h_m)) > 0$ and $\beta_m(\gamma)$ can be any value, when $\rho_m(\lambda(h_m)) = 0$. □

Suppose that $\beta = (\beta_1, \beta_2, \ldots, \beta_q)$, where β is the behavior strategy set for all airlines, $\rho = (\rho_1, \rho_2, \ldots, \rho_q)$, where ρ is the corresponding realization probability and ρ_0 denotes the realization probability of the virtual player, nature, usually a fixed value, and $S = S_0 \times S_1 \times \cdots \times S_q$, where S denotes sequence space and $s = (s_0, s_1, \ldots, s_q) \in S$, where s is a set of slot series. The expected revenue function can be expressed by

$$P(\rho) = \sum_{s\in S} P(s) \prod_{m=0}^{q} \rho_m(s_m), \tag{9}$$

where $P(s) = P(c)$, $P(c)$ is the revenue of implementing s towards some end note c, and $\prod_{m=0}^{q}\rho_m(s_m)$ is the realization probability of approaching c.

According to the above description, when the revenue $P(\rho)$ for all airlines is maximized, the transferring cost for all the airlines can be achieved, and the objective function is given by

$$\min\ f_6 = \sum_{m\in A}\sum_{S_m\in\Omega_m} f_{S_m} X_{S_m}, \tag{10}$$

where f_{S_m} is the value of the objective function f_5 for airline m, when S_m is the slots combination after the exchange.

(7) Balance the Increased Fuel Consumption for Each Airline. Fairness principle requires that the fuel consumption caused by flight delays should be averaged for airlines to bear. However, airlines are of different scales and aircraft are of different types, so the average fuel consumption should be made from the aspects of both aircraft and airlines. This goal is achieved by averaging the proportion of the fuel consumption change for a certain aircraft type belonging to a certain airline to the fuel consumption change for a certain aircraft type of all airlines. The objective is represented by

$$\min\ f_7 = \sum_{n\in S}\sum_{m\in A}\left|\frac{1}{z} - \frac{\Delta K_{nm}}{\sum_{m\in A}\Delta K_{nm}}\right|, \tag{11}$$

where $\Delta K_{nm}/\sum_{m\in A}\Delta K_{nm}$ is the proportion of the fuel consumption change of type n aircraft belonging to airline m to the fuel consumption change of all airlines' type n aircraft.

2.3. Integrative Assignment Model. According to the above analysis, the real-time gate assignment model based on the principle of minimum delay cost for multiagent can be expressed as follows:

$$\min \quad f = \min\{f_1, f_2, f_3, f_4, f_5, f_6, f_7\} \tag{12}$$

$$\text{S.T} \quad \sum_{i\in N}\sum_{j\in M} X_{ij} = 1 \tag{13}$$

$$X_{ij} \in \{0, 1\} \tag{14}$$

$$\sum_{m\in A}\sum_{S_m\in\Omega_m} X_{S_m} = 1 \tag{15}$$

$$X_{S_m} \in \{0,1\} \tag{16}$$

$$\sum_{j\in M} X_{ij}\left(G_j - Q_i\right) > 0 \tag{17}$$

$$L_{ij} + \Delta T - R_{kj} \le 0 \tag{18}$$

$$R_{ij} - I_{bj} > 0, \qquad L_{ij} - I_{ej} < 0 \tag{19}$$

$$\Delta T, R_{ij}, L_{ij}, I_{bj}, I_{ej}, E_w, t_{mw1}, t_{mw2} > 0, \tag{20}$$

where (12) is the objective function. Equation (13) means every flight is assigned to one and only one gate. Equation (14) is the corresponding relationship between flight and gate. Equation (15) means each slot combination is adopted by one and only one airline. Equation (16) is the corresponding relationship of slot combinations and airlines. Equation (17) enforces that the type of the gate where the aircraft is assigned should match the type of the aircraft. Equation (18) stipulates that the idle time of the gate should be longer than buffer time for the sake of safety. Equation (19) indicates that the beginning of the idle time for any gate should be earlier than the arrival time of the flight which will be assigned to the gate, and the end of the idle time should be later than the departure time of the flight. Equation (20) refers to validity constraint.

Compared with the traditional staged model (the slot assignment is produced before the gate assignment), the advantages of the proposed model are presented as follows: (1) the sequential game helps to obtain better slot combinations for all the airlines and (2) the CDM mechanism contributes to generating gate reassignment with less delay cost of multiagent due to the collaboration of airlines and airports.

To solve the multiobjective optimization problem (MOOP) [17], the objectives are sorted in order of priority, because all the objectives cannot be optimized simultaneously. As the service concept is becoming more and more important, the waiting time and waiting time of passengers are given the highest priority. The second highest priority is the taxiing time of aircraft, because the fuel cost is the direct operation cost of airlines. Following the taxiing time of aircraft is the implicit idle cost of the airport. Fuel equalization is the lowest priority, because the slot exchange between airlines also contributes to the fairness principle when the fuel consumption is not equalized at the very beginning.

3. Solution Algorithm

Mixed set programming (MSP) [18–20] is a logic reasoning algorithm based on first-order logic and set reasoning. In MSP, set operations, quantifiers, Boolean logic, logic functions, date/time reasoning, and numerical constraints are integrated in one system; the reasoning on numeric types, such as reals and integers, is expanded to global reasoning over mixed domains of set types, such as Booleans and references. Most importantly, MSP makes the modeling and solving for constraint satisfaction problems (CSP) realizable. The so-called set programming here is to systematically integrating set reasoning and operational research algorithm, establishing a rigorous and complete set theoretical formulation based on set variables and solving the model by set reasoning algorithm, instead of simply combining set notations with set variables and set constraints. The MSP adopted in this paper involves three major parts, detailed as follows.

Part 1 (carry out optional slot assignment schemes). Consider the following.

Step 1. Sort the flight banks that have not been finished at time t by ascending order of scheduled arrival time, expressed as $B(t) = \{b_1, b_2, \ldots, b_k\}$, where B denotes flight bank set and k is the serial number of flight bank.

Step 2. Let u_k equal the number of flights in flight bank k/the actual arrival time of flight bank k. Then, define the close time of flight bank k corresponding to the maximum u_k as $t1$, $t2$ as the close time of the next flight bank when it has not been finished at time t and can be finished at time $t2$, and $\tau = \min(t1, t2)$.

Step 3. Assign the flights of the flight bank corresponding to the maximum u_k to the time slots before τ by the order of scheduled arrival time. If the number of delayed flights is n, then $n!$ slot assignment plans will be generated.

Step 4. Repeat Steps 1, 2, and 3 for the rest of the flight banks until all the flights are reassigned with slots. It should be noted that, when Part 1 is implemented, a number of slot assignment schemes are produced.

Part 2 (optimize the slot assignment schemes through non-zero-sum sequential game). Consider the following.

Step 1. Input the information needed to implement non-zero-sum sequential game, namely, airlines, delayed flights, and provided slots.

Step 2. Implement non-zero-sum sequential game for all the airlines and calculate the delay cost of the airlines according to the optional slot assignment schemes. The non-zero-sum sequential game between airlines is implemented according to the following.

(1) In order to cut the delay cost of airline A, exchange the ist ($i = 1, 2, \ldots, n$) combination of flight and slot of airline A with the jst ($j = 1, 2, \ldots, m$) combination of flight and slot of airline B, and the exchange meet the requirements of heuristic rules.

(2) Repeat (1) till all the combinations of flights and slots of airline A are exchanged with that of airline B.

(3) In order to cut the delay cost of airline B, exchange the ist ($i = 1, 2, \ldots, n$) combination of flight and slot of airline B with the jst ($j = 1, 2, \ldots, m$) combination

of flight and slot of airline A, and the exchange meet the requirements of heuristic rules.

(4) Repeat (3) till all the combinations of flights and slots of airline B are exchanged with that of airline A.

(5) Combine the above results, and the optional slot exchange schemes are obtained.

(6) Divide the optional slot exchange schemes into two categories: equilibrium schemes and nonequilibrium schemes. In equilibrium schemes, three kinds of situations are included: a win-win situation for airlines A and B, airline A wins and airline B loses, and airline A loses and airline B wins. Nonequilibrium schemes refers to the schemes that make both airlines lose, thereby should be deleted.

(7) Calculate the value of f_6 according to the equilibrium schemes.

Step 3. Repeat Step 2 till the slot assignment scheme corresponding to the minimum delay cost for all the airlines is generated.

Part 3 (carry out the optimal gate assignment scheme). Consider the following.

Step 1. Read the preassignment results of all the flights and obtain the time periods of the available airport gates.

Step 2. For the delayed flights without subsequent flight banks keep their gate assignment as far as possible.

Step 3. For the delayed flights with subsequent flight banks, go to Step 4.

Step 4. Select out the flights which are not delayed but assigned to the gates affected by the delayed flights, and the selection is made on flights which arrive within the span of 50 minutes around the scheduled arrival time of the delayed flights according to [8]. Combine these selected flights with the flights in Step 3, a new flight set is produced. Then, gate assignment for the flights of this flight set can be carried out based on the principle of minimum delay cost according to the types of the flights and the available gates.

Step 5. Combine the assignment results of Steps 2–4, then the real-time gate assignment set is obtained. The real-time gate assignment set includes three parts: gates assigned to the delayed flights with subsequent flight banks, gates assigned to the delayed flights without subsequent flight banks, and gates assigned to the flights which are not delayed but affected by the reassignment of the delayed flights.

To design a solving strategy with preferable performance, heuristic rules are organically integrated in the algorithms. By this method, on one hand, the constraints in the optimization model can be strictly satisfied to ensure the feasibility of the solution; on the other hand, the search process can be flexibly controlled. The heuristic rules are given as follows:

(1) when the scheduled serial number of the flight bank for some delayed flights is k, then the actual serial number of the flight bank should be no less than k;

(2) when airlines exchange their time slots with each other, the serial numbers of the corresponding flight banks should be as similar as possible.

The above three parts as well as the heuristic rules are implanted into depth first search algorithm (DFS) [14], so that the integrative research for slots assignment and gate assignment can be performed. As a result, the slot assignment is optimized to be consistent with the optimal gate assignment which satisfies the multiobjective set previously.

In traditional staged algorithm, the cooperation of airlines is not taken into account, so the slots for delayed flights of airline A can only be adjusted within airline A instead of airline B, and the slots for delayed flights of airline B can only be adjusted within airline B instead of airline A. As a result, the gate assignment may cause losses for both airlines.

Compared with the traditional staged algorithm, the proposed integrative algorithm generates the following advantages: (1) the slots are exchangeable between the airlines, so the transferring cost of airlines can be decreased as much as possible; (2) the slot assignment and gate assignment are integrated into the MSP which supports integrative modeling and solving, so CDM mechanism for the airlines and the airport can be well achieved; (3) based on the software POEM, an integrative MSP method which supports nonzero-sum sequential game is designed, so the gate assignment can be generated much more effectively.

4. Experimental Results

For integrative modeling and solving, the software POEM [14] is taken into application. In order to support sequential game, a game class is added into the program. Four parts are included in the game class: players (airlines), actions (slot exchanges), costs of the players (delay cost of the airlines after the slot exchanges), and total cost of the sequential game. Additionally, rule class, equilibrium class, nonequilibrium class, and result class are designed to run the program. The constraints on the behavior of all the players (rules for slot exchange) are defined in rule class; the equilibrium characteristics for sequential game is included in the equilibrium class; the nonequilibrium characteristics for sequential game is included in the nonequilibrium class; the schemes and the corresponding delay cost for each airline are generated and stored in the result class. By applying those classes in POEM, the sequential game for the airlines in Part 2 can be performed. Parts 1 and 3 are achieved by the original function of the software POEM.

The environment where the experiment is carried out is represented as follows: (1) CPU: Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz; (2) RAM: 8.00 GB; (3) system type: x86-based PC; (4) system manufacturer: Dell Inc.; (5) OS name: Microsoft Windows 7; (6) OS version: 6.1.7601 Service Pack 1 Build 7601.

Table 1: Flight information.

Flight number	Arrival time	Departure time	Aircraft type	Passenger	Flight bank	Airline
1	9:20	10:20	E	300	1	C
2	9:30	10:30	E	300	1	S
3	9:35	10:25	C	100	1	S
4	9:40	10:35	D	200	2	E
5	9:40	10:35	D	200	2	S
6	9:40	10:40	E	300	2	E
7	9:40	10:30	C	100	2	C
8	9:40	10:35	D	200	2	C
9	9:40	10:40	E	300	2	E
10	9:45	10:40	D	200	1	S
11	9:45	10:35	C	100	1	E
12	9:45	10:45	E	300	1	C
13	9:45	10:35	D	200	1	S
14	9:50	10:40	C	100	1	E
15	9:55	10:55	E	300	2	C
16	9:55	10:50	D	200	2	E
17	10:00	11:00	E	300	3	E
18	10:00	10:55	D	200	3	E
19	10:00	11:00	E	300	3	C
20	10:00	10:55	D	200	3	C
21	10:05	10:55	C	100	2	S
22	10:05	11:05	E	300	2	C
23	10:10	11:00	C	100	2	C
24	10:10	11:10	D	200	2	C
25	10:15	11:15	E	300	2	C
26	10:15	11:05	D	300	2	C
27	10:25	11:15	D	200	2	S
28	10:25	11:15	C	100	2	C
29	10:35	11:35	E	300	2	S
30	10:40	11:30	C	100	3	E
31	10:40	11:40	E	300	3	S
32	10:45	11:35	C	100	4	E
33	10:45	11:45	D	200	4	S
34	10:45	11:35	C	100	4	S
35	10:45	11:40	D	200	4	E
36	10:50	11:40	C	100	3	S
37	10:50	11:40	C	100	3	C
38	10:50	11:50	E	300	3	E
39	10:50	11:40	C	100	3	E
40	10:55	11:50	D	200	4	C
41	11:00	12:00	E	300	4	S
42	11:00	12:00	E	300	4	E

4.1. A Case Study on Small-Scale Flight Delays. The data listed in Table 1 is from the 42 operational flights arriving from 9:20 to 11:00 at some major airport, involving three airlines and three types of aircraft. The airlines are Air China (CA), China Eastern (MU), and China Southern (CZ), symbolically denoted by C, E, and S, respectively. The types of the aircraft are small, medium, and large, symbolically denoted by C, D, and E, respectively.

In Table 1, number 17 and number 37 are special flights, meaning the gates should remain the same when flight delays occur and real-time assignment is needed. The gate information is listed in Table 2, 35 gates involved.

The provided arrival times for flights number 13, number 17, and number 37 are 10:05, 10:30, and 11:10. As number 17 and number 37 are special flights, the adjustment should be made on flights arriving within the interval [9:50, 10:50] according to [8]. In other words, a part of the flights in flight bank 1 and flight bank 2 will be influenced by the delay. The original gate assignment is listed in Table 3.

By utilizing the software POEM for the integrative modeling and solving, flights number 13, number 17, and number 37 are delayed to arrive at 10:05, 10:30, and 11:10, respectively, and the real-time gate assignment is produced with results listed in Table 4.

4.1.1. Economic Efficiency. According to the practical operation of most airlines, the fuel consumptions for large aircraft, medium aircraft, and small aircraft are 46 kilograms per minute, 28 kilograms per minute, and 12 kg kilograms per minute, respectively. The idle costs of large gates, medium gates, and small gates are 6 CNY per minute, 4 CNY per minute, and 2 CNY per minute, respectively. In addition, the fuel price is 7 CNY per kilogram, the walking cost of passengers is 3 CNY per minute, and the waiting cost of transfer passengers is 1 CNY per minute.

The total cost is 301,986 CNY in the preassignment, while 305,560 CNY in the reassignment, so it is increased by 3,574 CNY, a small growth of 1.18%. After the reassignment, the increases of all kinds of costs are given in Figure 1.

Fuel cost is increased from 68,306 CNY to 69,860 CNY with a growth of 2.28%, and the fuel consumption increased by the flight delays is equalized for airlines to bear, illustrated in Figure 2. Walking cost is increased from 153,000 CNY to 154,200 CNY, with a growth of 0.78%. Idle cost is decreased from 17,980 CNY to 17,940 CNY, with a drop of 0.22%. Since the gates are of three types, the increased costs of each type are minimized at the same time, results represented in Figure 3. Waiting cost is increased from 62,640 CNY to 63,600 CNY, with a growth of 1.53%. The reason why the waiting cost is increased with just a minor growth of 1.53% is that most of the flight banks are not delayed. The increases of all costs turn out to be quite small after the reassignment; therefore, the real-time assignment is acceptable. Besides, the increased waiting cost accounts for 26% of the total increased cost, which testifies that taking into account the waiting cost of transfer passengers in the cost control is very necessary.

Figure 2 shows that the increased fuel consumption of each type of aircraft is basically equalized for each airline, so the fairness principle is well abided by.

It is demonstrated in Figure 3 that the idle cost of small gates remains the same. For medium gates, the idle cost is increased by 80 CNY; and for large gates, the idle cost is

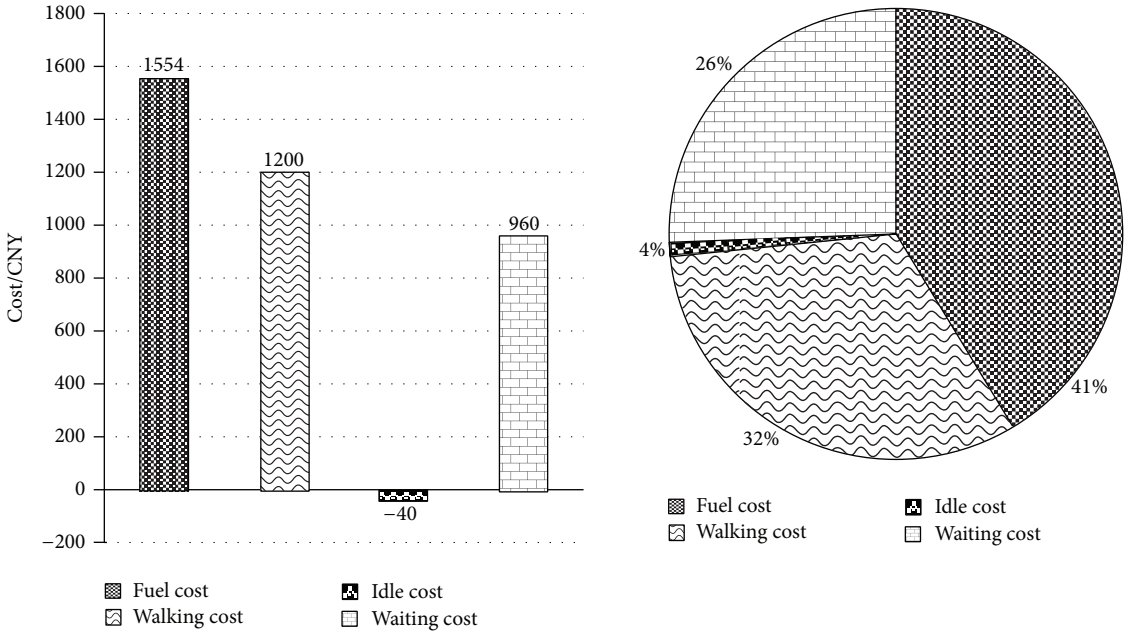


Figure 1: Increases of all kinds of costs.

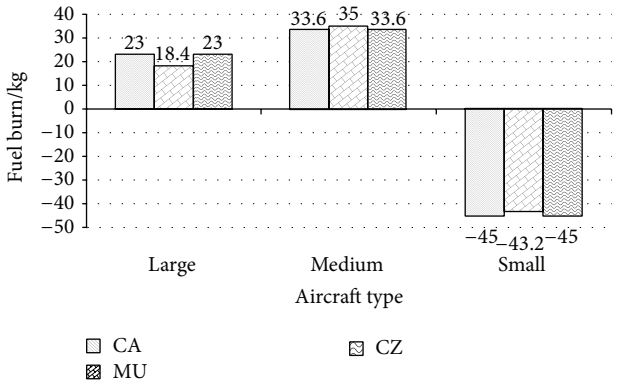


Figure 2: Balanced fuel consumption.

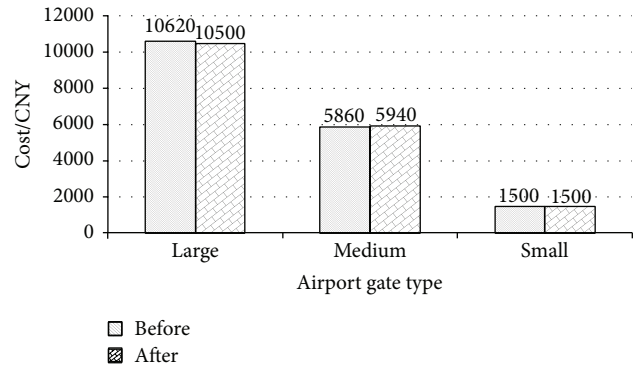


Figure 3: Idle costs for different types of gates.

decreased by 120 CNY. Although the idle cost of medium gates grows, the total cost of all gates turns out to be decreased, because the unit idle cost of large gates is more than that of medium gates.

Consequently, in the circumstance of small-scale flight delays, the real-time gate assignment model proposed in this paper is capable of achieving economic efficiency by adjusting a small number of gates.

4.1.2. Robustness. In fleet assignment [21, 22] and fleet planning [23, 24], robustness has been widely applied, but not in the research of gate assignment. As a complex system, gate assignment should also be robust: on one hand, the gates influenced by the flight delays can be restored in short term; on the other hand, the disturbance brought by the adjustment of the gates can be restricted within a certain scale. Theoretically, the evaluation criteria of the robustness for gate assignment include the utilization of gates, the recoverability of the affected gates and the service quality for passengers. Gate assignment with good performance is supposed to be generated with high utilization rate of gates, small-scale disturbance and convenient service for passengers.

Two major factors are considered to evaluate the robustness of the real-time assignment.

(1) Maximum utilization rate of the gates involves use rate and occupancy rate. Use rate is equal to the number of engaged gates divided by the total number of the gates; occupancy rate is equal to the holding time of the gates divided by the available time of all the available gates.

Table 2: Gate information.

Gate number	Gate type	Walking time/min	Taxiing time/min	Idle period
1	C	8	12	9:00~12:00
2	E	9	11	9:00~12:00
3	E	9	10	9:00~12:00
4	E	8	9	9:00~12:00
5	E	7	8	9:00~12:00
6	E	6	7	9:00~12:00
7	E	8	6	9:00~12:00
8	E	6	5	9:00~12:00
9	E	7	4	9:00~12:00
10	E	8	3	9:00~12:00
11	D	9	2	9:00~12:00
12	D	8	3	9:00~12:00
13	E	7	4	9:00~12:00
14	E	5	5	9:00~12:00
15	E	3	4	9:00~12:00
16	E	2	5	9:00~12:00
17	E	1	6	9:00~12:00
18	E	2	8	9:00~12:00
19	E	4	10	9:00~12:00
20	E	4	9	9:00~12:00
21	D	7	8	9:00~12:00
22	D	8	11	9:00~12:00
23	D	9	10	9:00~12:00
24	D	9	11	9:00~12:00
25	D	7	12	9:00~12:00
26	C	6	13	9:00~12:00
27	D	6	14	9:00~12:00
28	E	5	15	9:00~12:00
29	D	6	20	9:00~12:00
30	D	7	17	9:00~12:00
31	D	8	18	9:00~12:00
32	D	9	18	9:00~12:00
33	C	9	20	9:00~12:00
34	C	9	21	9:00~12:00
35	C	8	22	9:00~12:00

(2) Maximum service level for passengers means minimizing the growth of the walking time and waiting time for passengers.

Table 5 shows that the use rate is increased from 88.57% to 91.43% with a growth of 2.86%; the occupancy rate is increased from 36.75% to 37.86% with a growth of 1.11%; walking time is increased from 51,000 minutes to 51,400 minutes with a growth of 0.78%; waiting time is increased from 62,640 minutes to 63,280 minutes with a growth of 1.02%. The flight delays indeed lower the service quality for passengers by a small decrease; however, the utilization rate of the gates has risen with a big growth. To make a conclusion, the robustness of the real-time assignment scheme is well testified.

Table 3: Original gate assignment information.

Flight	Gate	Flight	Gate	Flight	Gate
1	15	2	17	3	7
4	9	5	27	6	10
7	19	8	31	9	18
10	21	11	29	12	13
13	20	14	26	15	6
16	25	17	3	18	8
19	16	20	22	21	28
22	5	23	23	24	12
25	2	26	4	27	14
28	34	29	15	30	27
31	17	32	20	33	24
34	35	35	10	36	29
37	21	38	18	39	32
40	13	41	8	42	9

Table 4: Real-time gate assignment results.

Flight	Gate	Flight	Gate	Flight	Gate
1	15	2	17	3	7
4	9	5	27	6	10
7	19	8	31	9	18
10	21	11	29	12	13
13	30	14	4	15	16
16	23	17	3	18	6
19	20	20	12	21	24
22	14	23	25	24	5
25	28	26	32	27	22
28	15	29	17	30	26
31	7	32	11	33	10
34	4	35	18	36	35
37	21	38	19	39	1
40	13	41	8	42	9

4.1.3. Timeliness and Collaboration. The cost caused by flight delays can be reduced as much as possible through CDM mechanism. The following part of the case study is taken as an example to make a clear illustration on CDM.

ATCC provides three slots (slot 1 [10:05, 10:55] of flight bank 2, slot 2 [10:30, 11:30] of flight bank 3, and slot 3 [11:10, 12:00] of flight bank 4) for the three delayed flights (A: number 13, B: number 37, and C: number 17) to the airlines. It should be noted that number 13, number 17, and

Table 5: Robustness evaluation.

	Utilization of gates		Service for passengers	
	Use rate (%)	Occupancy rate (%)	Walking time (%)	Waiting time (%)
Growth	2.86	1.11	0.78	1.02

number 37 are flights from three different airlines. As the slots can be exchanged between any two airlines, 6 different slot assignment schemes will be produced theoretically; accordingly, the corresponding gate assignment schemes will also be different. However, if all the schemes are calculated in turn to locate the optimal one, the computation process will be very time-consuming and resource-wasting. So, non-zero-sum sequential game theory is utilized to exclude the infeasible schemes and find out the cost-optimal slot assignment scheme; the process is illustrated in Figure 4 (applying (10)).

Figure 4 shows that only two feasible schemes are carried out from six optional choices, meaning 2/3 of the action sequences are excluded from the set of feasible solutions thereby saving 2/3 of the computation time. Suppose that the realization probability of each slot exchange plan is equal; then, the loss caused by flight delays for each slot assignment plan can be calculated by (11). In plan 1, slots 1, 2, and 3 are directly assigned to flights A, B, and C,respectively, causing a total loss of 1600 CNY. In plan 2, slots 1, 2, and 3 are assigned to flights A, C, and B, respectively, causing a total loss of 960 CNY.

Traditionally, the airlines will adopt plan 1 directly, and the slot assignment will be delivered to the airport without considering the related costs of the airport. However, the gate assignment scheme under this slot assignment is not the optimal choice. According to the calculation, the gate assignment corresponding to plan 2 is better than the gate assignment under plan 1. Under plan 2, the operation cost of both the airport and the airlines can be controlled more effectively; meanwhile, the satisfaction of the passengers can be improved to a greater extent. It can be concluded that real-time gate assignment is produced based on the information of the delayed flights; hence, varying degrees of flight delays (slot reassignment) will lead to different real-time gate assignment schemes, but only one is optimal when comparing the total costs of all the schemes. In turn, the optimal gate assignment scheme can provide a reference for airlines to reassign the delayed flights to updated slots provided by ATCC. In the whole process, CDM between the airlines (and the airport) is effectively achieved, thereby protecting the benefits of airlines, airports, and passengers.

4.2. A Case Study on Medium- to Large-Scale Flight Delays. In the traditional staged method, the slots are reassigned to the flights without consideration of minimizing the delay cost before the gate assignment; as a result, the real-time gate assignment without CDM mechanism may not be optimal. Nevertheless, in the integrative method presented in this paper, the slots can be interchanged between the airlines in the process of real-time gate assignment; therefore, the delay cost of the real-time gate assignment can be minimized as much as possible. To present a significant comparison of the integrative method and the traditional staged method, large-scale flight delays are introduced into the case study. The information on delayed flights before slot exchange is listed in Table 6 and the information on delayed flights after slot exchange is listed in Table 7. The computation results are given in Table 8.

As the equalization of fuel cost and the robustness of the gates almost remain the same, the comparison is mainly

Table 6: Delay information before the slot exchange.

Flight number	Arrival time	Departure time	Flight bank
4	9:50	10:45	2
5	9:50	10:50	2
8	9:55	10:50	2
9	10:15	11:15	3
12	9:50	10:50	2
13	10:00	10:50	1
16	10:10	11:05	2
17	11:00	12:00	4
23	10:40	11:30	2
24	10:30	11:30	2
27	10:40	11:30	2
37	10:55	11:45	3

Table 7: Delay information after the slot exchange.

Flight number	Arrival time	Departure time	Flight bank
4	9:50	10:45	2
5	9:50	10:50	2
8	9:55	10:50	2
9	10:40	11:40	2
12	10:00	11:00	1
13	9:50	10:40	2
16	10:10	11:05	2
17	10:55	11:55	3
23	10:15	11:05	3
24	10:30	11:30	2
27	10:40	11:30	2
37	11:00	11:50	4

Table 8: Comparison of the results generated by traditional staged method and integrative method.

	Fuel cost/CNY	Walking cost/CNY	Idle cost/CNY	Waiting cost/CNY	Total cost/CNY	Total increase/%	Time/min
Original cost	68,306	153,000	17,980	62,640	301,986	/	/
Staged method	70,448	158,400	17,940	65,880	312,668	3.54	10.17
Integrative method	69,860	157,800	17,830	64,080	309,570	2.51	2.66

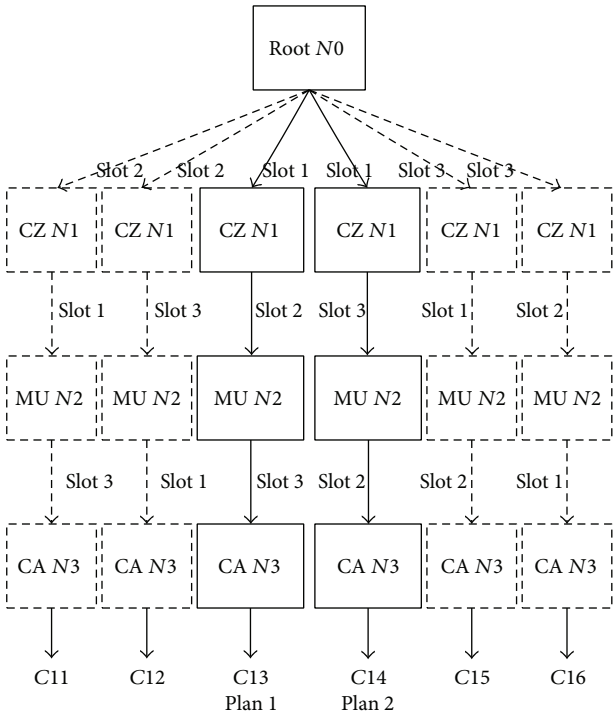


Figure 4: Non-zero-sum sequential game for airlines.

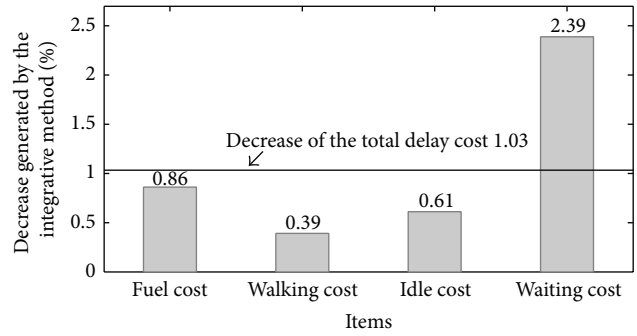


Figure 5: Decreases generated by the integrative method compared with the staged method.

made on the costs and the computation time. Since the number of the optional slot assignment schemes is 132 and the number of the feasible schemes is 32, the computation time of the integrative method should be around 1/4 of the computation time needed in the traditional staged method. Table 8 shows that the staged method is able to reassign the gates appropriately but produces an increase of 3.54% on the total cost, and the computation process consumes 10.17 minutes. However, the integrative method just brings an increase of 2.51% on the total cost, and the computation process only takes 2.66 minutes. Therefore, the integrative method is superior to the traditional staged method in the aspects of cost control and computation time control (2.66/10.17 is approximately equal to 1/4), not only in the circumstance of the aforementioned small-scale flight delays, but also under the situation of medium- to large-scale flight delays.

Figure 5 gives a visualized comparison of each cost item for the staged method and the integrative method. The bar stands for the difference of the increase produced by those two methods, and the greater the value is, the more cost the integrative method can cut. Among these items, the most significant change caused by the integrative method is made in the waiting cost for transfer passengers, which is cut by 2.39%, and this is because (1) flight bank is taken into account in the optimization; (2) sequential game is applied in the slot exchange between airlines; and (3) gate assignment and slot assignment are implemented under the CDM mechanism.

According to the comparison under the condition of medium to large scale flight delays, the integrative method proposed in this paper is much superior to the traditional staged method.

4.3. Conclusions of the case Studies. Based on the experimental results of Sections 4.1 and 4.2, the conclusion are made as follows.

(1) As the approach proposed in this paper is practical, the gate assignment problem is well solved; meanwhile, all the constraints posed in the multiobjective function are satisfied.

(2) The approach is well applied in minimizing delay cost under the situations of both small-scale flight delays and medium- to large-scale flight delays.

(3) The interests of both airlines and airports are taken into account, which contributes to the application of CDM mechanism.

(4) The non-zero-sum sequential game excludes the infeasible slot combinations, so the computation time of the approach is saved to a great extent. In the first case (small-scale flight delays), as the feasible slot assignment schemes account for 1/3 of the total slot combinations, the computation time is saved by

around 2/3. In the second case (medium- to large-flight delays), because the feasible slot assignment schemes account for around 1/4 of the total slot combinations, the computation time is saved by about 3/4. If the proportion of the feasible schemes is smaller, the computation time can be saved much more significantly.

5. Conclusions

This research focuses on the integrative approach with sequential game to the problem of real-time gate assignment. The assignment model is formulated based on CDM mechanism and minimal delay cost principle for multiagent when flight delays occur; meanwhile, MSP combined with sequential game method is designed for calculation. The case studies for both small-scale and medium- to large-scale flight delays verify the validity of the integrative method. Firstly, all kinds of costs are minimized better than the traditional staged method, especially, the waiting cost of transfer passengers. Secondly, the increased fuel burn is basically balanced for each airline. Thirdly, the CDM of the airlines and the airport is well achieved. Lastly, more than half of the computation time needed in the traditional method is saved in the integrative method. In summary, due to the economic efficiency, robustness, collaboration, and timeliness, the integrative approach proposed in this paper is reasonable and feasible in restoring normal airport operation and guaranteeing regular air transportation.

Conflict of Interests

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

Acknowledgments

This research is funded by the Key Program of National Natural Science Foundation of China (no. 61232002 and no. 60939003), China Postdoctoral Science Foundation (nos. 2012M521081 and 2013T60537), the Fundamental Research Funds for the Central Universities (no. NS2014066), and Postdoctoral Science Foundation of Jiangsu Province (no. 1301107C).

References

[1] Planning & Development Department of Civil Aviation Administration of China, *2011 Statistical Data on Civil Aviation of China*, China Civil Aviation Publishing House, Beijing, China, 2011.

[2] A. Bolat, "Procedures for providing robust gate assignments for arriving aircrafts," *European Journal of Operational Research*, vol. 120, no. 1, pp. 63–80, 2000.

[3] A. Bolat, "Models and a genetic algorithm for static aircraft-gate assignment problem," *Journal of the Operational Research Society*, vol. 52, no. 10, pp. 1107–1120, 2001.

[4] J.-J. You, C.-M. Ji, and X. Fu, "New method for solving multi-objective problem based on genetic algorithm," *Journal of Hydraulic Engineering*, no. 7, pp. 64–69, 2003.

[5] A. Lim and F. Wang, "Robust airport gate assignment," in *Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '05)*, pp. 74–81, November 2005.

[6] J.-H. Li, J.-F. Zhu, and Q. Gao, "Airport gate assignment based on Greedy Tabu Search algorithm," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 4, pp. 173–179, 2011.

[7] Y. Cheng, "Network-based simulation of aircraft at gates in airport terminals," *Journal of Transportation Engineering*, vol. 124, no. 2, pp. 188–196, 1998.

[8] C. Yu, "A knowledge-based airport gate assignment system integrated with mathematical programming," *Computers and Industrial Engineering*, vol. 32, no. 4, pp. 837–852, 1997.

[9] Y. Cheng, "A rule-based reactive model for the simulation of aircraft on airport gates," *Knowledge-Based Systems*, vol. 10, no. 4, pp. 225–236, 1998.

[10] W. Li, "Optimized assignment of civil airport gate," in *Proceedings of the International Conference on Intelligent System Design and Engineering Application (ISDEA '10)*, vol. 2, pp. 33–38, October 2010.

[11] D. X. Wei and C. Y. Liu, "Airport gate reassignment problem," *Journal of Nanjing University of Aeronautics and Astronautics*, vol. 41, no. 2, pp. 257–261, 2009.

[12] W. Li, *A Method to Construct Flight Bank for Hub Airports*, Nanjing University of Aeronautics and Astronautics, College of Civil Aviation, Nanjing, China, 2010.

[13] Q. Gao, J. Yan, and J.-F. Zhu, "Airlines' optimization decision of slot allocation in CDM," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 5, pp. 94–98, 2011.

[14] B. Zhu, J. F. Zhu, and Q. Gao, "Constraint programming model of integrated recovery for aircraft and crew," *Journal of Traffic and Transportation Engineering*, vol. 13, no. 1, pp. 77–83, 2013.

[15] D. Gerardi and R. B. Myerson, "Sequential equilibria in Bayesian games with communication," *Games and Economic Behavior*, vol. 60, no. 1, pp. 104–134, 2007.

[16] Z.-J. Li, C.-T. Cheng, F.-X. Huang, and X. Li, "Sequential game-based resource allocation strategy in grid environment," *Journal of Software*, vol. 17, no. 11, pp. 2373–2383, 2006.

[17] H. C. Gomes, F. de Assis das Neves, and M. J. F. Souza, "Multi-objective metaheuristic algorithms for the resource-constrained project scheduling problem with precedence relations," *Computers & Operations Research*, vol. 44, pp. 92–104, 2014.

[18] J. Yan, T. S. Wu, Q. Gao, and J. F. Zhu, "Slot switching model of airlines under cooperative game," *Journal of Traffic and Transportation Engineering*, vol. 12, no. 5, pp. 85–90, 2012.

[19] J. Y. Zhou, "A note on mixed set programming," in *Proceedings of the IEEE The 7th International Symposium on Operations Research and Its Applications*, pp. 131–140, 2008.

[20] J. Zhou, "Introduction to the constraint language NCL," *Journal of Logic Programming*, vol. 45, no. 1–3, pp. 71–103, 2000.

[21] X. H. Zhu, J. F. Zhu, and Q. Gao, "The research on robust fleet assignment problem based on flight purity," *Forecasting*, vol. 30, no. 1, pp. 71–74, 2011.

[22] D. Y. Mou and Z. X. Zhang, "Robust fleet scheduling problem based on probability of flight delay," *Journal of Civil Aviation University of China*, vol. 28, no. 6, pp. 35–39, 2010.

[23] Y. Wang and H. Sun, "Heuristic algorithm to incorporating robustness into airline fleet planning," *Systems Engineering—Theory & Practice*, vol. 33, no. 4, pp. 963–970, 2013.

[24] H. Sun, P. Zhang, and Y. Wang, "Fleet planning approach based on optimized fleet capacity allocation in airline networks," *Journal of Southwest Jiaotong University*, vol. 45, no. 1, pp. 111–115, 2010.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 539596, 12 pages
http://dx.doi.org/10.1155/2014/539596

Research Article

Energy and Environmental Efficiency of China's Transportation Sector: A Multidirectional Analysis Approach

Gongbing Bi, Pingchun Wang, Feng Yang, and Liang Liang

School of Management, University of Science and Technology of China, No. 96 Jinzhai Road, Hefei, Anhui 230026, China

Correspondence should be addressed to Feng Yang; fengyang@ustc.edu.cn

Received 19 March 2014; Accepted 14 April 2014; Published 12 May 2014

Academic Editor: X. Zhang

With the rapid economic development, the transportation sector becomes one of the high-energy-consumption and high-CO_2-emissions sectors in China. In order to ensure efficient use of energy and to reduce CO_2 pollution, it is important to gain the best performance standards in China's transportation sector. Data envelopment analysis (DEA) has been accepted as a popular tool of efficiency measurement. However, previous studies based on DEA are mainly restricted to the radial expansions of outputs or radial contractions of inputs. In this paper, we present a nonradial DEA model with multidirectional efficiency analysis (MEA) involving undesirable outputs for the measurement of regional energy and environmental efficiency of China's transportation sector during the period 2006–2010. We not only evaluate the energy and environmental efficiency level and trend of China's transportation sector but also investigate the efficiency patterns of 30 regions and three major areas of China. Additionally, we identify the energy saving potential and CO_2 emissions reduction potential for each province and area in China in this study.

1. Introduction

After the rapid growth of China's economy in the past three decades, the increasing energy consumption, the carbon dioxide (CO_2) emissions, and environmental pollution are currently hindering the sustainability of China's economic growth. Since 2007, China has already surpassed the USA and become the world's largest energy consumer and contributor of CO_2 emissions [1]. To address this issue, China's 12th five-year plan seeks to establish a "green, low-carbon development concept." [2]. In 2015, China will increase the proportion of nonfossil fuels in energy generation to 11.4%, reduce energy consumption per unit of gross domestic product (GDP) by 16%, and reduce CO_2 emissions per unit of GDP by 17% from the levels in 2010 [2].

Recognizing the great significance of reducing energy consumption and CO_2 emissions as well as sustainable development, more and more researchers have focused on evaluating and improving energy utilization and CO_2 emissions efficiency, which is considered as a crucial way to save energy, reduce greenhouse gas emissions, protect environment, and mitigate global climate change [3–8]. These studies mainly focused on the energy or environmental efficiency of industries such as electricity, iron, and steel. However, few papers have considered the energy and environmental efficiency evaluation of China's transportation sectors.

According to the International Energy Agency (IEA) [9], we found that transport sector consumed 61.2% of the world's oil and approximately 28% of the total final energy in 2007 and the transport sector had become the world's largest oil consumption sector. According to the International Energy Agency (IEA) [10], transport sector became the world's second largest greenhouse gas emitting sector which accounted for 22% of the world's CO_2 emissions. Thus, we take for granted that it is worthwhile analyzing the energy and environmental efficiency of transportation sector in the largest polluting country of the world since it provides a lot of information on energy and environmental policy analysis and decision making.

The indexes of energy or environmental performance are constructed in the form of mathematics programming methods such as data envelopment analysis (DEA). As a nonparametric approach to evaluate the relative efficiency of a set of comparable decision making units (DMUs), DEA, which was developed by Charnes et al. [11], has been widely investigated and popularly applied to many fields,

such as schools, hospital, and banks [12, 13]. Recently, at the macroeconomy level, DEA has been universally used in studying the energy and environmental performance in examining the relative efficiency [14]. For instance, Hu and Wang [15] adopted the traditional Charnes-Cooper-Rhodes (CCR) model to evaluate the total-factor energy efficiency of 29 regions in China during 1995–2002. Zhang et al. [16] analyzed the industrial sectors' ecoefficiency of 30 provinces in China by a DEA model and found that most Chinese regions with higher levels of GDP per capital would have higher ecoefficiency. Yeh et al. [7] incorporated undesirable outputs into calculating the technical efficiency of energy utilization in Chinese mainland and Taiwan during the period of 2002–2007 through employing the traditional Banker-Charnes-Cooper (BCC) model. Shi et al. [17] used three extended DEA models to investigate the energy and environmental overall technical efficiency, pure technical efficiency, and scale efficiency of industry sectors in 28 regions of China during 2000–2006, with the undesirable output of industrial waste gas being treated as inputs in energy and environmental efficiency analysis. Bian and Yang [6] presented several DEA models for calculating resource and environmental efficiency and applied their proposed approach in real data set of 30 provinces in China. Wang et al. [18] adopted a traditional DEA to analyze the industrial sectors' energy efficiency in China's 30 provinces during the period of 2005–2009. Their study showed that the west area had the greatest amount of energy redundancies in the three major areas. Wang et al. [19] established several efficiency models which were capable of integrating undesirable outputs into efficiency measure framework to evaluate the economic efficiency, CO_2 emissions efficiency, and economic-environmental efficiency of 28 provinces in China from 2001 to 2007. Similar studies also can be found in [5, 20–23].

However, most of the previous studies on energy and environmental efficiency mentioned above took advantage of the traditional DEA approach, in which the DMUs under measurement were restricted to the radial constraints on input and output variables. In this study, we take multidirectional efficiency analysis (MEA) instead of the traditional radial DEA approach. MEA selects benchmarks such as the input reductions and the output expansions are proportional to the potential improvements related to each input and output dimension separately. It enables us to have a specific insight of the patterns of efficiencies for each DMU. The traditional MEA approach does not always consider the undesirable outputs. Although Asmild and Matthews [24] and Wang et al. [1] used the approach and took the undesirable outputs as input, it was inappropriate to reflect the actual production process since undesirable outputs were produced as the by-products of production process rather than the input. Thus we adopt the environmental DEA technology proposed by Fare et al. [25] as the alternative approach to model undesirable outputs. Meanwhile, traditional MEA model presents the fact that the input reductions and output expansions have the same proportion to the potential improvements identified considering the improvement potential related to each input and output variable separately. But it is still treated as a radial measure of efficiency. Since the radial model adjusts all variables to efficient targets by the same proportion, it cannot provide information regarding the efficiency of the specific inputs or outputs involved in the production process [26]. At the same time, radial efficiency measures may neglect the slack variables, leading to efficiency overestimation [27]. As a result, recent studies have tried to use the nonradial DEA model [26, 28–30] to overcome these problems. So we follow the same pattern and adjust the MEA model to make input reductions and the output expansions nonproportional to the potential improvements related to each input and output dimension separately.

As for the transportation sector, Ramanathan [31] used DEA to measure the energy efficiency of alternative transport modes in India. Tongzon [32] also assessed the efficiency of four Australian and twelve international container ports based on DEA. However, all these studies in the transportation sector did not involve undesirable outputs in estimating efficiency. In order to investigate the levels and the patterns of efficiency in China's transportation sector and provide additional insights into the energy and emissions efficiency of each China's region, hence our study will use the modified MEA model and consider undesirable outputs.

The rest of this paper is organized as follows. Section 2 introduces the environmental production technologies and outlines the methodology of modified MEA model. In Section 3, we show an empirical study of China's 30 regions in transportation sector during 2006–2010 to illustrate our model. Section 4 gives discussions and conclusions.

2. Methodology

2.1. Environmental Production Technology. Let us consider a production process that consumes a vector of inputs $\mathbf{x}$ to obtain a vector of desirable outputs $\mathbf{y}$ and a set of undesirable outputs denoted by the vector $\mathbf{c}$. Then a production technology is given by

$$T = \{(\mathbf{x}, \mathbf{y}, \mathbf{c}) : \mathbf{x} \text{ can produce } (\mathbf{y}, \mathbf{c})\}. \quad (1)$$

Following Färe et al.'s [33] approach, we treat by-products as outputs and accept the following three assumptions which are imposed on the production technology.

(1) Strong or free disposability of desirable outputs: it implies that if $(\mathbf{x}, \mathbf{y}, \mathbf{c}) \in T$ and $\mathbf{y}^* \leq \mathbf{y}$, then $(\mathbf{x}, \mathbf{y}^*, \mathbf{c}) \in T$. This allows for the assumption that if an observed desirable and undesirable output vector are possible, then each output vector with a smaller desirable output is also feasible. It implies that we can always freely dispose of some desirable outputs without undertaking any cost.

(2) Weak disposability of undesirable outputs: if $(\mathbf{x}, \mathbf{y}, \mathbf{c}) \in T$ and $0 \leq \theta \leq 1$, then $(\mathbf{x}, \theta\mathbf{y}, \theta\mathbf{c}) \in T$. This means that the proportional reduction in desirable and undesirable outputs is possible, whereas it may be costly to reduce undesirable outputs and these abatement activities will usually divert resources away from desirable outputs. The assumption is to introduce the idea that it is not feasible to reduce undesirable outputs solely.

(3) Desirable and undesirable outputs being null-joint: if $(\mathbf{x}, \mathbf{y}, \mathbf{c}) \in T$ and $\mathbf{c} = \mathbf{0}$, then $\mathbf{y} = \mathbf{0}$. This assumption says that it is not technically feasible to produce only desirable outputs in the absence of undesirable outputs. The only way to eliminate all the undesirable outputs is to end the production process.

Once the three assumptions are imposed, T is termed as environmental production technology. Suppose there are $j = 1, 2, \ldots n$ decision making units (DMUs) which consume m inputs to produce s_1 desirable outputs and s_2 undesirable outputs. Now we denote $\mathbf{x}_j = (x_{1j}, \ldots x_{mj})^t$, $\mathbf{y}_j = (y_{1j}, \ldots y_{s_1 j})^t$, and $\mathbf{c}_j = (c_{1j}, \ldots c_{s_2 j})^t$. as the vector of inputs, desirable outputs, and undesirable outputs of DMU_j, respectively. In the DEA framework, the environmental production technology can be modeled as

$$T = \left\{ (\mathbf{x}, \mathbf{y}, \mathbf{c}) : \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{ij},\ i = 1, 2, \ldots, m \right.$$
$$\sum_{j=1}^{n} \lambda_j y_{rj} \le y_{rj},\ r = 1, 2, \ldots, s_1 \qquad (2)$$
$$\left. \sum_{j=1}^{n} \lambda_j c_{kj} = c_{kj},\ k = 1, 2, \ldots, s_2 \right\}.$$

In model (2), $(\lambda_1, \lambda_2, \ldots, \lambda_j)$ denotes a vector of intensity variables that form linear combinations of observed inputs and outputs with constant returns to scale not imposed by the constraint that $\sum_{j=1}^{n} \lambda_j = 1$. The inequality on desirable outputs and equality on undesirable outputs help us to impose the strong or free disposability of desirable outputs and weak disposability of undesirable outputs.

2.2. Multidirectional Efficiency Analysis (MEA). Our aim is to gain a deeper insight into the regional energy and environmental efficiency of China's transportation by investigating the situation and analyzing the efficiency patterns in each region and area. More specifically, we measure the comprehensive efficiency incorporating CO_2 emissions and energy efficiency of China's regional transportation sector. In this study, we will use multidirectional efficiency analysis (MEA) instead of the traditional radial DEA approach, which enables us to have a specific view of the patterns of efficiencies.

Multidirectional efficiency analysis was firstly proposed by Bogetoft and Hougaard [34], further developed by Bogetoft and Hougaard [35] and Asmild and Pastor [36]. Previous studies [24, 37, 38] did not take the undesirable outputs into account by using MEA. In this paper we will make full use of this approach and consider the undesirable outputs simultaneously. It is able to select benchmarks such as the inputs, undesirable outputs reduction, and desirable outputs augmentation, which are not proportional to the actual production, but proportional to the potential improvements related to each input and output variable separately. In addition, since MEA considers the improvement potential in each variable separately, it is very suitable to investigate situations, where the purpose is to reduce the consumptions of some inputs and the emission of some undesirable outputs and to expand the production of some desirable outputs. Furthermore, since efficiency improvements in the present case were derived from a combination of the inputs consumed reduction and the desirable outputs production increment, as well as the undesirable outputs emission abatement, we see it a better way to adopt the MEA approach to achieve the efficiency improvements.

In the general case of considering reductions of all inputs and undesirable outputs simultaneously with augmentations of all of desirable outputs in MEA, we define the unit specific directional distance function on the basis of the unit specific ideal reference point. In order to find the ideal reference point, we firstly solve the following linear programs, one for each of input dimensions, desirable output dimensions, and undesirable output dimensions, respectively:

$$\begin{aligned}
&\min \quad d_{io} \\
&\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j x_{ij} \le d_{io}, \\
&\qquad \sum_{j=1}^{n} \lambda_j x_{-ij} \le x_{-io}, \quad -i = 1, \ldots, i-1, i+1, \ldots m \\
&\qquad \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro}, \quad r = 1, \ldots, s1 \\
&\qquad \sum_{j=1}^{n} \lambda_j c_{kj} = c_{ko}, \quad k = 1, \ldots, s2. \\
&\qquad\quad \lambda_j \ge 0, \quad j = 1, \ldots, n,
\end{aligned} \qquad (3a)$$

$$\begin{aligned}
&\max \quad \delta_{ro} \\
&\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j y_{rj} \ge \delta_{ro}, \\
&\qquad \sum_{j=1}^{n} \lambda_j y_{-rj} \ge y_{-ro}, \quad -r = 1, \ldots, r-1, r+1, \ldots, s1 \\
&\qquad \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io}, \quad i = 1, \ldots, m \\
&\qquad \sum_{j=1}^{n} \lambda_j c_{kj} = c_{ko}, \quad k = 1, \ldots, s2 \\
&\qquad\quad \lambda_j \ge 0, \quad j = 1, \ldots, n,
\end{aligned} \qquad (3b)$$

$$\begin{aligned}
&\min \quad \varphi_{ko} \\
&\text{s.t.} \quad \sum_{j=1}^{n} \lambda_j c_{kj} = \varphi_{ko} \\
&\qquad \sum_{j=1}^{n} \lambda_j c_{-kj} = c_{-ko}, \quad -k = 1, \ldots, k-1, k+1, \ldots, s2
\end{aligned}$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io}, \quad i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro}, \quad r = 1, \ldots, s1$$

$$\lambda_j \ge 0, \quad j = 1, \ldots, n. \tag{3c}$$

By solving the model (3a)–(3c), we are able to determine the ideal reference point $(d_{io}^*, \delta_{ro}^*, \varphi_{ko}^*)$ for (x_{io}, y_{ro}, c_{ko}), where $*$ denotes the optimal solutions of model (3a)–(3c). With the unit specific ideal point determined, we next consider the following linear programming model (4):

$$\begin{aligned}
\max \quad & \beta_o \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \beta_o \left(x_{io} - d_{io}^*\right), \quad i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} + \beta_o \left(\delta_{ro}^* - y_{ro}\right), \quad r = 1, \ldots, s1 \\
& \sum_{j=1}^{n} \lambda_j c_{kj} = c_{ko} - \beta_o \left(c_{ko} - \varphi_{ko}^*\right), \quad k = 1, \ldots, s2 \\
& \lambda_j \ge 0, \quad j = 1, \ldots, n,
\end{aligned} \tag{4}$$

where β_o represents the productive technical inefficiency of DMU_0 and also measures the proportion by which the desirable outputs are added while the undesirable outputs and inputs are contracted in the same proportion. The value of β_o belongs to the interval $[0, 1]$. If $\beta_o = 0$, it means that the evaluated DMU reaches to the frontier of the best practice and is therefore efficient.

From model (4), we find that the inputs and undesirable outputs reduction and desirable outputs expansion have the same proportion to the potential improvements identified considering the improvement potential related to each input and output variable separately. So it may still be treated as a radial measure of efficiency. However, radial efficiency measures overestimate the efficiency when there exist nonzero slacks. Based on this, we will modify model (4) and make the inputs and outputs be adjusted nonproportionally. Next, we will form model (5):

$$\begin{aligned}
\max \quad & \beta_{io} + \beta_{ro} + \beta_{ko} \\
\text{s.t.} \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \le x_{io} - \beta_{io} \left(x_{io} - d_{io}^*\right), \quad i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro} + \beta_{ro} \left(\delta_{ro}^* - y_{ro}\right), \quad r = 1, \ldots, s1 \\
& \sum_{j=1}^{n} \lambda_j c_{kj} = c_{ko} - \beta_{ko} \left(c_{ko} - \varphi_{ko}^*\right), \quad k = 1, \ldots, s2 \\
& \lambda_j \ge 0, \quad j = 1, \ldots, n.
\end{aligned} \tag{5}$$

We are able to get an optimal solution by solving model (5). Then we define the relative variable specific MEA efficiency for every variable separately.

For the input variable x_{io}, we are able to define specific input MEA efficiency as follows:

$$\frac{x_{io} - \beta_{io} \left(x_{io} - d_{io}^*\right)}{x_{io}}. \tag{6}$$

For the desirable output y_{ro}, we are able to define specific desirable output MEA efficiency as follows:

$$\frac{y_{ro}}{y_{ro} + \beta_{ro} \left(\delta_{ro}^* - y_{ro}\right)}. \tag{7}$$

For the undesirable output c_{ko}, we are able to define specific undesirable output MEA efficiency as follows:

$$\frac{c_{ko} - \beta_{ko} \left(c_{ko} - \varphi_{ko}^*\right)}{c_{ko}}. \tag{8}$$

By defining each of MEA efficiencies for every variable, we focus on the pattern of the individual specific variable efficiencies. For instance, we pay attention to the efficiency of energy and environment in China's transportation. Based on the individual variable specific efficiencies, we require a comprehensive measure of efficiency for the measured DMUs. Färe and Knox lovell [39] firstly proposed Russell graph measure which is a comprehensive efficiency measure. This remains as a theoretical contribution and still inspires most of the following comprehensive efficiency measures. For example, many DEA efficiency assessments have been proposed based on the additive model [40], which detect all the technical inefficiency in all dimensions such that the benchmarks lie on the strong efficiency and do not have any slacks [41–44]. Now we will use the same comprehensive measure as the SBM model of Tone [44] to combine the inefficiencies in the different dimensions into one overall value. And the overall score of MEA efficiency for the observation of DMU_0 can be defined as follows:

$$\begin{aligned}
\theta_o = & \left(1 - \frac{1}{m} \sum_{i=1}^{m} \frac{\beta_{io} \left(x_{io} - d_{io}^*\right)}{x_{io}}\right) \\
& \times \left(1 + \frac{1}{s1 + s2} \right. \\
& \quad \times \left(\sum_{r=1}^{s1} \frac{\beta_{ro} \left(\delta_{ro}^* - y_{ro}\right)}{y_{ro}} \right. \\
& \qquad \left. \left. + \sum_{k=1}^{s2} \frac{\beta_{ko} \left(c_{ko} - \varphi_{ko}^*\right)}{c_{ko}} \right) \right)^{-1}.
\end{aligned} \tag{9}$$

Unlike the original MEA efficiency, the comprehensive measure (9) has the desirable characteristic that the resulting value incorporates all variables.

3. Empirical Studies

3.1. Data and Variables. This study will examine the energy and environmental efficiency of China's transportation sector

Table 1: Areas and regions of China (excluding Tibet).

Areas	Regions
East area	Beijing, Tianjin, Shanghai, Liaoning, Hebei, Shandong, Jiangsu, Zhejiang, Fujian, Guangdong, and Hainan
Central area	Heilongjiang, Jilin, Inner Mongolia, Henan, Shanxi, Anhui, Hubei, Hunan, Jiangxi, and Guangxi
West area	Gansu, Guizhou, Ningxia, Qinghai, Shaanxi, Yunnan, Xinjiang, Sichuan, and Chongqing

Table 2: CO_2 emissions factor by major carbonaceous fuel types in China (source: NDRC, 2007 [50]).

Fuels	Coal	Petrol	Kerosene	Diesel	Fuel oil	Nature gas
CCF	27.28	18.9	19.6	20.17	21.09	15.32
HE	192.14	448	447.5	433.3	401.9	0.384
COF (%)	92.3	98.0	98.6	98.2	98.5	99.0

Notes: CCF and HE are expressed in units of tons, carbon/trillion Joules, and trillion Joules/10^4 tons (m^3), respectively.

of 30 provinces in mainland during the 11th five-year plan period (2006–2010). From the perspective of China's development and policy factors, these provinces and autonomous regions are divided into three areas: east, central, and west China [15]. Detailed information is shown in Table 1.

Tibet is excluded since the absence of relevant data in this research. According to China's Statistical Yearbooks, the transportation sector in China is defined as consisting of transport, storage, and post. So we will follow this definition to collect the relevant data. In our study, three inputs, one desirable output, and one undesirable output are considered to measure the energy and environmental efficiency in China's transportation sector. Therefore the labor and capital stocks are used as two nonenergy inputs, the volume of energy consumed in the transportation sector as energy input, a value-added amount in the transportation sector as the only desirable output, and the volume of CO_2 emissions related to fuel used in the sector as undesirable output. Since there is no capital stock statistics data in China, we will use the amount of fixed capital investment to represent capital stock as some authors did [6, 17, 45, 46]. The total energy consumption includes all types of energy, such as coal, oil, natural gas, and electricity. We will convert these energies to the tonne coal equivalent (tce).

The data of nonenergy inputs (labor and the fixed capital investment), energy input, and desirable output (a value-added) are available from China Statistical Yearbook 2007–2011 [47] and China Energy Statistical Yearbook 2007–2011 [48]. Unfortunately, there are still no official statistical data on CO_2 emissions in China. Thereby we will estimate the provincial CO_2 emissions in the transportation sector during 2006–2010 based on a fuel-based carbon footprint model, as described by the Intergovernmental Panel on Climate Change Guidelines [49] for National Greenhouse Gas Inventories for calculating CO_2 data. According to these guidelines for estimating CO_2 emissions, the equation for calculating CO_2 emissions from fossil fuels appeared as follows:

$$CO_2 = \sum_{i=1}^{n} A \times CCF_i \times HE_i \times COF_i \times \left(\frac{44}{12}\right). \quad (10)$$

From (10), we know that CO_2 emissions are related to the amount of all carbonaceous fuel combusted (A), the carbon content factor (CCF), the heat equivalent (HE), and the carbon oxidation factor (COF) of carbonaceous fuel. The number (44/12) is the ratio of the molecular weight of CO_2 (44) to the molecular weight of carbon (12), where $CCF_i \times HE_i \times COF_i \times (44/12)$ is the CO_2 emissions factor of a fuel. As long as we know the amount and the CO_2 emissions factor of each fuel, the total CO_2 emissions of each province can be calculated. The amount of consumption of each fuel of each province in the transportation sector is collected from China Energy Statistical Yearbook 2007–2011. The Energy Research Institute (ERI) of National Development and Reform Commission (NDRC) [50] of China reported the CO_2 emissions factor by major type of carbonaceous fuels in China, as is shown in Table 2. Then we will figure out the total CO_2 emissions based on the formula of IPCC guidelines.

After collecting the related data on the three inputs, one desirable output, and one undesirable output, a data set encompassing 30 provinces during 2006–2010 is prepared for analysis in Table 3. A correlation matrix of all inputs and outputs is calculated for verifying the relationship between the inputs and outputs variables. Table 4 shows the result that all the correlation coefficients in the table are significantly positive, which indicates that a quite high correlation exists among every variable. Thus, energy and environmental efficiency analysis in this case is feasible.

3.2. Efficiency Levels of China's Region in Transportation Sector. First we use models (5) and (9) to evaluate the comprehensive MEA efficiency score of different provinces in China's transportation sector from 2006 to 2010, and the related result is shown in Table 5. We find that the average comprehensive MEA efficiencies of about 70% regions are below 0.5, with Yunnan ranking lowest and Hebei ranking highest during 2006–2010. There are only 3 regions exhibiting MEA efficient at least 1 year during 2006–2010. Only Hebei performed best in keeping efficient during the whole study period. Shandong also performed well and was assessed as efficient region for 4 years except 2010. The result indicates that large parts of the regions in China are not performing efficiently in the transportation sector. Hence, there is a great possibility to reduce the energy consumption and CO_2

Table 3: Descriptive statistics of inputs and outputs for 30 regions.

Inputs and outputs	Year	2006	2007	2008	2009	2010
Labor (10^4 peoples)	Mean	20.4	20.7	20.9	21.1	21.0
	Standard development	11.0	11.6	11.7	12.2	12.4
	Maximum	48.3	48.8	49.3	53.5	56.1
	Minimum	2.9	2.8	2.8	3.0	2.9
Capital (10^8 yuan)	Mean	363.7	409.7	468.5	671.1	821.1
	Standard development	210.9	221.8	246.8	348.0	431.9
	Maximum	865.0	862.5	1108.8	1596.2	1820.0
	Minimum	51.3	48.9	90.1	90.1	120.9
Energy (10^4 TCEs)	Mean	625.4	698.3	764.4	811.9	885.2
	Standard development	456.4	505.1	528.5	556.5	596.3
	Maximum	1851.5	2042.5	2201.1	2311.5	2564.5
	Minimum	39.9	71.1	88.2	98.2	110.5
Value-added (10^8 yuan)	Mean	435.9	498.1	571.2	715.7	826.6
	Standard development	307.3	354.8	429.6	516.7	606.9
	Maximum	1212.3	1399.9	1873.6	1971.0	2328.4
	Minimum	35.3	40.9	40.7	61.3	67.5
CO_2 emissions (10^6 tons)	Mean	13.0	14.5	15.8	16.8	18.2
	Standard development	9.6	10.6	11.1	11.7	12.4
	Maximum	38.6	42.5	45.8	48.0	53.0
	Minimum	0.8	1.5	1.8	201	2.3

Table 4: Correction matrix of inputs and outputs variables.

	Labor	Capital	Energy	Value-added	CO_2
Labor	1				
Capital	0.599^{**}	1			
Energy	0.783^{**}	0.731^{**}	1		
Value-added	0.678^{**}	0.786^{**}	0.804^{**}	1	
CO_2	0.774^{**}	0.722^{**}	0.999^{**}	0.793^{**}	1

**shows significant correlation at 0.01 significance level (2-tailed).

emission and to increase the value-added amounts in each region.

Then the annual average comprehensive MEA efficiency of China and its three major areas during 2006–2010 are calculated and illustrated in Figure 1. From the angle of area, the regions in east China have the highest average comprehensive MEA efficiency, followed by the regions in central China and then the regions in west China. The average comprehensive MEA efficiency of China and its three major areas basically keep first decreasing and then increasing. Also the average comprehensive MEA efficiencies of east China and central China are generally higher than those of China. And west China's average comprehensive MEA efficiency is evidently lower than the average comprehensive MEA efficiency of China. It reflects the materially unbalanced development in China's transportation sector and the gap between east and west. This result is similar to some studies about China's regional energy efficiency and environmental efficiency.

3.3. Efficiency Patterns and Differences of China's Region in Transportation Sector. The above evaluation results manifest that east China has a higher comprehensive MEA efficiency than central and west China in the transportation sector. However, the comprehensive MEA efficiency just displays the efficiency levels and the change trends of different regions and areas. We are not able to take a thorough understanding of the sources of inefficiency and to detect the patterns of efficiency differences among China's regions and three major areas in transportation sector, which are more attractive to policy makers. Therefore, it is necessary to investigate the efficiencies on individual variables of each DMU. Through the definitions (6)–(8), we calculate the relative variable specific MEA efficiencies, which enable us to survey the patterns of efficiencies in different China's regions and areas. Since this study centers on the energy and environmental efficiency, here we only figure out the energy and CO_2 emissions efficiency using the definitions (6) and (8) separately.

Table 6 illustrates the relative variable specific MEA efficiency (i.e., energy and CO_2 emissions efficiency) in each

Table 5: Regional comprehensive MEA efficiency of China (2006–2010).

Region	2006	2007	2008	2009	2010	Mean
Beijing	0.386	0.313	0.239	0.266	0.370	0.315
Tianjin	0.508	0.438	0.391	0.514	0.830	0.536
Hebei	1.000	1.000	1.000	1.000	1.000	1.000
Shanxi	0.653	0.641	0.382	0.306	0.381	0.473
Inner Mongolia	0.767	0.755	0.757	0.747	0.747	0.754
Liaoning	0.328	0.327	0.283	0.343	0.367	0.330
Jilin	0.451	0.388	0.336	0.321	0.327	0.365
Heilongjiang	0.774	0.364	0.326	0.292	0.772	0.506
Shanghai	0.363	0.297	0.281	0.703	0.617	0.452
Jiangsu	0.651	0.619	0.595	0.814	1.000	0.736
Zhejiang	0.474	0.484	0.454	0.423	0.510	0.469
Anhui	0.922	0.874	0.869	0.468	0.525	0.732
Fujian	0.949	0.917	0.848	0.516	0.790	0.804
Jiangxi	0.464	0.482	0.507	0.448	0.450	0.470
Shandong	1.000	1.000	1.000	1.000	0.826	0.965
Henan	0.686	0.714	0.823	0.531	0.499	0.651
Hubei	0.307	0.287	0.726	0.290	0.737	0.469
Hunan	0.480	0.431	0.387	0.383	0.437	0.423
Guangdong	0.462	0.450	0.382	0.380	0.420	0.419
Guangxi	0.344	0.318	0.287	0.261	0.295	0.301
Hainan	0.338	0.330	0.252	0.197	0.703	0.364
Chongqing	0.426	0.323	0.287	0.296	0.312	0.329
Sichuan	0.441	0.393	0.321	0.216	0.232	0.320
Guizhou	0.301	0.301	0.238	0.495	0.526	0.372
Yunnan	0.193	0.189	0.190	0.129	0.113	0.163
Shaanxi	0.331	0.302	0.261	0.271	0.287	0.290
Gansu	0.410	0.505	0.447	0.413	0.402	0.436
Qinghai	0.328	0.225	0.171	0.185	0.218	0.226
Ningxia	0.316	0.327	0.287	0.492	0.608	0.406
Xinjiang	0.283	0.287	0.231	0.233	0.233	0.253

region of China, which is similar to the comprehensive MEA efficiency. Most of the regional energy and CO_2 emissions MEA efficiency are under 0.5. The results make it clear that larger parts of the regions in China are not performing MEA efficiently on energy and CO_2 emissions in the transportation sector. Therefore, there lies a great potential to reduce the energy consumption and CO_2 emission in each region. Also we notice a strange phenomenon that some regions have correspondingly high energy MEA efficiency but a low CO_2 emissions MEA efficiency, for example, Inner Mongolia. It indicates that the resource of inefficiency is from CO_2 emissions inefficiency. From the relative variable specific MEA efficiency illustrated in Table 6, we can detect the reasons of inefficiency in each region, and the decision marker can draw up the relevant policies and measures to improve the relative energy or CO_2 emissions efficiency separately.

After we have the relative variable specific MEA efficiency (i.e., energy and CO_2 emissions relative efficiencies), we are able to investigate the efficiency of three major areas during 2006–2010. Figure 2 demonstrates the annual average energy efficiency and CO_2 emissions efficiency for China and its three major areas in the transportation sector. It seems that both the variable specific efficiencies of China and its three major areas overall experienced a process of first decreasing and then increasing over the study period. In Figure 2, we are able to see that the east area and central area have higher MEA efficiencies than west area on both the two variable specific efficiencies. However, the MEA efficiency difference of energy and CO_2 emissions is mixed between the east China and the central China. We find that the east China performs better than central China on MEA efficiency of CO_2 emissions during 2006–2010. But for MEA efficiency of energy, the east China has lower MEA efficiency score than central China during 2006–2008, and it is just opposite in 2009 and 2010. Therefore it can be concluded that the higher comprehensive MEA efficiency dominance of the east China over the central China and west China results from both the higher energy and CO_2 emissions variable specific efficiency.

3.4. Efficiency Related Reduction Potentials on Energy and CO_2 Emissions of China's Region in Transportation Sector.

Table 6: Regional energy and CO_2 emissions MEA efficiency of China (2006–2010).

Region	Energy					CO_2 emissions				
	2006	2007	2008	2009	2010	2006	2007	2008	2009	2010
Beijing	0.519	0.427	0.340	0.309	0.370	0.510	0.424	0.338	0.306	0.362
Tianjin	0.529	0.550	0.488	0.570	1.000	0.506	0.529	0.468	0.543	0.615
Hebei	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Shanxi	0.725	0.727	0.346	0.324	0.392	0.763	0.768	0.339	0.315	0.391
Inner Mongolia	1.000	1.000	1.000	1.000	1.000	0.392	0.363	0.357	0.322	0.322
Liaoning	0.323	0.292	0.298	0.269	0.306	0.308	0.278	0.282	0.252	0.286
Jilin	0.598	0.458	0.383	0.350	0.371	0.563	0.427	0.364	0.330	0.346
Heilongjiang	1.000	0.437	0.482	0.408	1.000	0.415	0.409	0.461	0.376	0.411
Shanghai	0.300	0.254	0.239	1.000	1.000	0.283	0.239	0.224	0.156	0.247
Jiangsu	0.697	0.634	0.593	1.000	1.000	0.667	0.605	0.568	0.542	1.000
Zhejiang	0.522	0.492	0.463	0.415	0.465	0.500	0.471	0.441	0.392	0.441
Anhui	1.000	1.000	1.000	0.517	0.517	0.831	0.711	0.697	0.491	0.488
Fujian	1.000	1.000	1.000	0.548	1.000	0.893	0.820	0.642	0.515	0.535
Jiangxi	0.636	0.605	0.624	0.532	0.494	0.603	0.573	0.592	0.504	0.469
Shandong	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.578
Henan	0.882	0.784	0.899	0.523	0.493	0.894	0.779	0.895	0.512	0.480
Hubei	0.301	0.277	1.000	0.272	1.000	0.286	0.261	0.252	0.252	0.285
Hunan	0.465	0.421	0.496	0.452	0.480	0.456	0.411	0.483	0.430	0.460
Guangdong	0.416	0.378	0.353	0.333	0.344	0.398	0.361	0.336	0.314	0.325
Guangxi	0.344	0.332	0.339	0.272	0.316	0.324	0.312	0.316	0.252	0.295
Hainan	0.336	0.321	0.225	0.157	1.000	0.318	0.302	0.210	0.145	0.155
Chongqing	0.523	0.387	0.361	0.367	0.339	0.497	0.365	0.340	0.346	0.317
Sichuan	0.471	0.395	0.352	0.237	0.276	0.461	0.386	0.338	0.225	0.267
Guizhou	0.346	0.308	0.244	0.454	0.472	0.350	0.309	0.238	0.445	0.459
Yunnan	0.206	0.189	0.189	0.128	0.111	0.196	0.179	0.180	0.120	0.103
Shaanxi	0.444	0.378	0.311	0.257	0.263	0.446	0.381	0.303	0.252	0.258
Gansu	0.451	0.489	0.419	0.342	0.330	0.459	0.513	0.433	0.348	0.341
Qinghai	0.612	0.354	0.259	0.242	0.268	0.609	0.333	0.251	0.222	0.252
Ningxia	0.280	0.259	0.270	0.396	0.516	0.268	0.252	0.262	0.387	0.514
Xinjiang	0.259	0.237	0.229	0.223	0.222	0.248	0.228	0.217	0.210	0.208

By applying the MEA approach, we are able to measure the energy and CO_2 emissions variable specific efficiency for each region of China in the transportation sector, and, based on the DEA theory, the inefficient regions can become efficient on each of their input and desirable and undesirable output variable so as to reach the benchmark by adjusting improvement potential associated with each variable. Therefore, in this section, we will further use the MEA approach to survey the energy conservation and CO_2 emissions reduction potential for China's different regions during our study period.

From model (5), we know that the potential saving of each input variable is able to be calculated as $\beta_{io}(x_{io} - d^*_{io})$, the target value of each input variables after improvement potential adjustment is $x_{io} - \beta_{io}(x_{io} - d^*_{io})$, the potential redundancy of CO_2 emissions could be calculated as $\beta_{ki}(c_{ko} - \varphi^*_{ko})$, and the target of CO_2 emissions after improvement potential adjustment is $c_{ko} - \beta_{ko}(c_{ko} - \varphi^*_{ko})$. Table 7 documents the energy conservation potential and CO_2 emissions reduction potential for China's each region in transportation sector during 2006–2010. In Table 7, we can see that the energy conservation potential and CO_2 emissions reduction potential of each region are great in the transportation sector. For the energy conservation potential, there are ten out of 30 regions: Hebei, Inner Mongolia, Heilongjiang, Shanghai, Jiangsu, Anhui, Fujian, Shandong, Hubei, and Hainan, which do not have the improvement potential of energy for at least 1 year during our study period. But for CO_2 emissions reduction potential, only two regions, namely, Inner Mongolia and Shandong do not have the improvement potential of CO_2 emissions for at least 1 year during our study period. Only Hebei and Shandong have no improvement potential of energy and CO_2 emissions simultaneously except 2010's improvement potential of CO_2 emissions in Shandong. From an area perspective, we can further investigate the reduction potentials of energy saving and CO_2 emissions reduction of three major areas from 2006 to 2010, which are illustrated in Figure 3. In short, as shown in Figure 3, the total energy conservation potential of China's transportation sector on the whole keeps first increasing and then decreasing

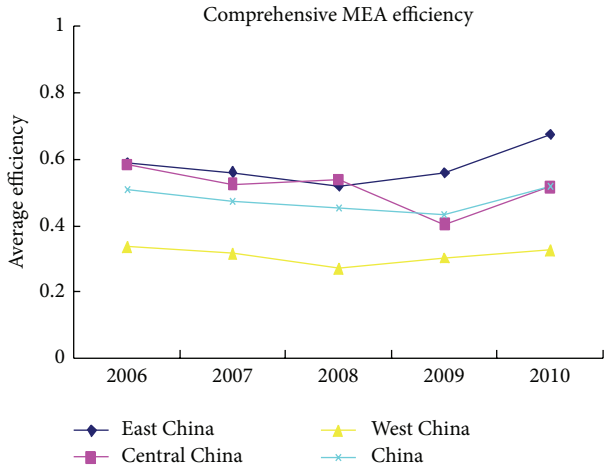


Figure 1: Average comprehensive MEA efficiency of China and its three major areas.

beginning from 2009 during the study period. But the total CO_2 emissions reduction potential of China's transportation sector always keeps an increasing trend. In general, east China has the largest energy conservation and CO_2 emissions reduction potentials, and those of west China are the smallest.

For the sake of a detailed description, we will take the relative data in 2010 as an example to analyze the regional energy conservation potential and CO_2 emissions reduction potential in China's transportation sector. As is shown in Table 7, there are eight regions whose theoretical maximum energy conservation potentials are more than five million TCEs in 2010. In these eight regions, Guangdong has the largest energy conservation potential, followed by Liaoning, Yunnan, Sichuan, Shaanxi, Zhejiang, Beijing, and Guangxi. Furthermore, another ten regions have zero energy conservation potential because they are efficient regions such as Tianjin, Hebei, Inner Mongolia, Heilongjiang, Shanghai, Jiangsu, Fujian, Shandong, Hubei, and Hunan. We see that most of these 10 regions are located in east China and they are all energy specific efficient regions in China. Among the inefficient regions, Guangdong has the largest energy conservation potential, but its energy MEA efficiency was not the lowest. On the contrary, Yunnan, Xinjiang, Shaanxi, Qinghai, Sichuan, Liaoning, Guangxi, Gansu, and Chongqing have lower energy MEA efficiency in 2010. Also we find that these 8 regions are located in west China except Liaoning and Guangxi. The above results point out that these inefficient regions which have a great energy conservation potential and a low energy MEA efficiency should pay more attention to their energy consumption control policy and energy efficiency promotion, so as to increase their energy saving and promote their energy utilization performances and catch up with the high efficiency benchmark regions.

Table 7 also illustrates similar assessment results on the potentials of CO_2 emissions reduction for China's regions in transportation sector in 2010. From Table 7, we see that Guangdong, Shanghai, Liaoning, and Shandong whose potentials of CO_2 emissions reduction are larger than 20 million TCEs rank highly in China in 2010 and Guangdong has the largest potential of CO_2 emissions reduction. But its CO_2 emissions MEA efficiency was not the lowest. On the contrary, Yunnan, Hainan, Xinjiang, Shanghai, Qinghai, and Shaanxi have lower energy MEA efficiency in 2010. Besides, we find that these regions are located in west China except Hainan and Shanghai. The regions with unit CO_2 emissions MEA efficiency and zero CO_2 emissions reduction potentials such as Hebei and Jiangsu are the same as the benchmark regions under energy specific efficiency assessment. The above results show that the regions of greater CO_2 emissions reduction potentials should be given higher priority to CO_2 emissions controls to improve their CO_2 emissions MEA efficiency, reduce CO_2 emissions redundancies, and catch up with the highly performed benchmark regions.

4. Conclusion

In recent years, an increasing number of studies have applied DEA approach to evaluate energy or environmental efficiency; however, there are rare studies which pay close attention to China's transportation sector. In this study, we propose a combination model of nonradial DEA and MEA, focusing on the measurement of regional energy and environmental efficiency of China's transportation sector during the period 2006–2010. Not only the energy and environmental efficiency levels and trend of China's transportation sector are investigated but also the efficiency patterns of China's 30 regions and three major areas are investigated. In addition, the energy saving potential and CO_2 emissions reduction potential for each province and area in China are identified in this study.

With all the models, we can draw the conclusions as follows. (i) Most of regions whose average comprehensive MEA efficiency is below 0.5 during 2006–2010 tell us that many regions in China are not efficient in the transportation sector. (ii) From an area perspective, the regions in east China had the highest average comprehensive MEA efficiency, followed by the regions in central China and then the regions in west China. The comprehensive MEA efficiency of the east area surpassing the other two areas is due to the higher emission and energy specific efficiency except 2006–2008. (iii) During the period from 2006 to 2010, the trend of the comprehensive MEA slightly decreases from 2006 to 2009 and then begins to increase, and the average energy and emission efficiency have the same trend as the comprehensive MEA. (iv) Up to 2010, China's transportation sector still has great energy conservation potential and CO_2 emissions reduction potential; the central government should keep on implementing strict policies for energy consumption and CO_2 emissions to develop more ecofriendly transportation industry.

Our empirical study also has some significant policy implications. First of all, the regional imbalances are narrowed in China's transportation sector, so the government should pay greater attention to the central and the west area whose transportation facilities are relatively undeveloped.

Table 7: Regional energy and CO_2 emissions reduction potential of China (2006–2010).

Region	Energy conservation potential					CO_2 emissions reduction potential				
	2006	2007	2008	2009	2010	2006	2007	2008	2009	2010
Beijing	294.0	415.2	550.3	600.6	586.8	6.1	8.4	11.0	11.9	11.9
Tianjin	155.4	148.0	188.8	171.9	0.0	3.4	3.2	4.0	3.7	3.5
Hebei	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Shanxi	114.4	116.1	505.6	526.9	490.0	1.9	1.9	10.3	10.8	9.6
Inner Mongolia	0.0	0.0	0.0	0.0	0.0	9.1	11.0	12.8	15.4	17.4
Liaoning	821.4	957.2	943.6	1035.3	1019.1	17.5	20.4	20.2	22.1	21.8
Jilin	110.2	200.6	288.5	306.1	307.1	2.5	4.5	6.2	6.6	6.7
Heilongjiang	0.0	289.5	233.3	303.5	0.0	6.3	6.4	5.0	6.8	6.4
Shanghai	1080.3	1304.9	1377.3	0.0	0.0	23.4	28.2	29.6	32.5	30.7
Jiangsu	279.4	369.7	469.1	0.0	0.0	6.4	8.3	10.3	11.3	0.0
Zhejiang	398.7	469.0	537.4	603.7	599.6	8.7	10.2	11.6	13.0	12.9
Anhui	0.0	0.0	0.0	211.0	238.3	1.2	2.4	2.6	4.6	5.2
Fujian	0.0	0.0	0.0	299.4	0.0	0.9	1.8	4.5	6.7	7.2
Jiangxi	121.2	135.6	129.7	168.0	221.0	2.8	3.1	2.9	3.7	4.8
Shandong	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21.4
Henan	68.4	146.6	72.4	362.9	434.8	1.2	3.0	1.5	7.4	8.9
Hubei	683.5	770.6	0.0	830.4	0.0	14.6	16.6	19.2	18.0	17.9
Hunan	339.4	404.2	298.6	412.7	436.0	7.0	8.4	6.2	8.8	9.2
Guangdong	1081.6	1270.5	1423.2	1541.2	1681.4	23.2	27.2	30.4	32.9	35.7
Guangxi	344.0	384.8	392.9	490.0	503.4	7.5	8.4	8.6	10.6	10.9
Hainan	100.8	112.5	186.3	229.8	0.0	2.2	2.4	4.0	5.0	5.2
Chongqing	163.4	259.0	307.3	289.8	368.3	3.6	5.7	6.7	6.2	7.9
Sichuan	349.9	481.8	585.8	809.1	730.0	7.3	9.9	12.3	17.0	14.9
Guizhou	175.6	227.0	314.2	232.4	259.7	3.4	4.5	6.4	4.7	5.4
Yunnan	470.1	518.0	532.7	588.6	751.1	10.0	11.0	11.2	12.4	15.9
Shaanxi	244.0	315.5	437.6	591.6	642.0	4.8	6.2	9.0	11.9	12.9
Gansu	143.0	135.8	164.0	198.8	222.7	2.8	2.5	3.1	3.8	4.2
Qinghai	15.5	45.9	65.4	74.4	80.8	0.3	1.0	1.3	1.6	1.7
Ningxia	89.8	98.1	96.5	84.7	72.9	1.9	2.0	2.0	1.7	1.5
Xinjiang	327.4	351.1	361.6	351.9	378.1	6.9	7.3	7.7	7.4	8.0

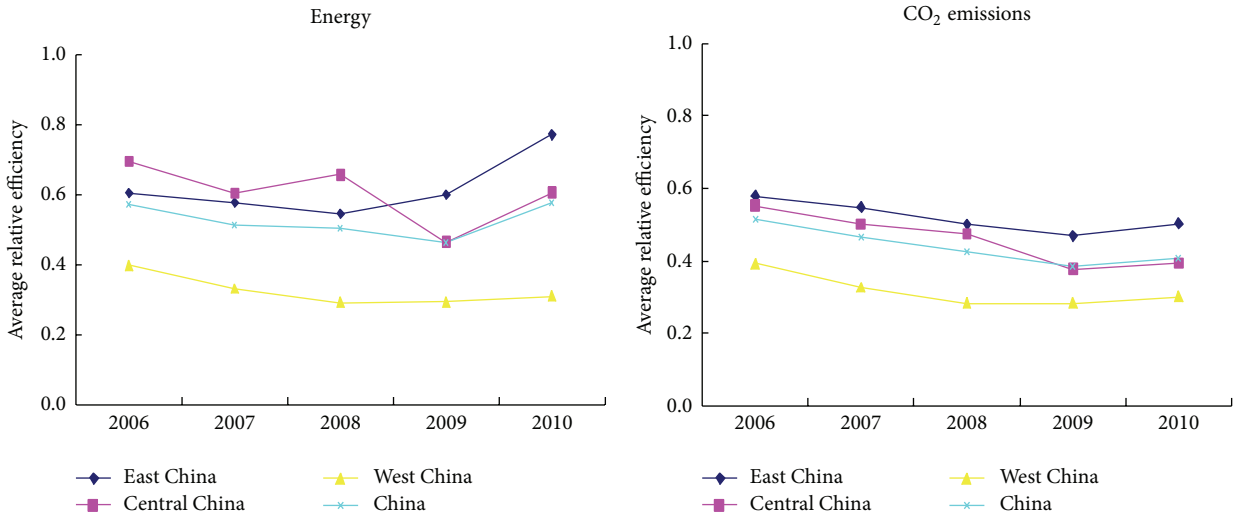


Figure 2: Average relative variable specific efficiencies of China and its three areas (2006–2010).

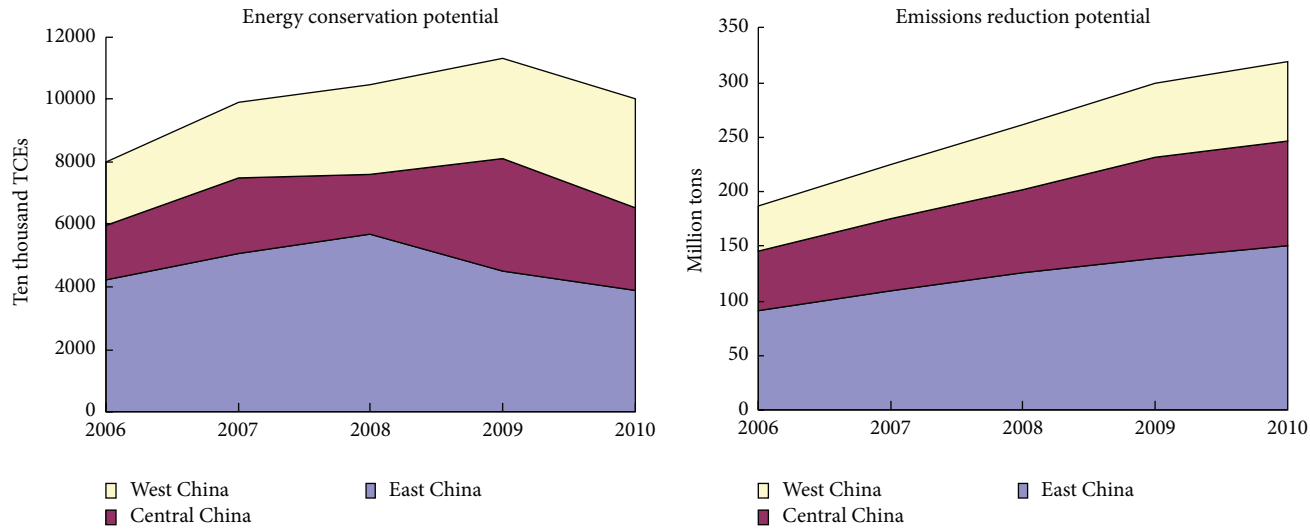


Figure 3: MEA relative energy conservation and CO_2 emissions reduction potentials of China's three areas (2006–2010).

Secondly, the related policies to encourage public transportation and control individual transportation guide huge energy conservation potential and CO_2 emissions reduction potential in some big cities in east China. Furthermore, we should give impetus to transportation technological innovation and promote vehicle emission standards and explore alternative transport energy which has a great advantage to reduce energy consumption and CO_2 emissions reduction.

This study only compares energy and environmental efficiency of the transportation sector among the Chinese provinces. If the data includes other advanced countries, for instance, OECD countries, it may provide more information on the level of China's transportation sector. Also this research can be combined with Malmquist Productivity Index to investigate the technical efficiency change of China's transportation sectors. All these remain avenues for future research.

Conflict of Interests

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

Acknowledgments

The authors thank the anonymous reviewers for their valuable comments. In addition, the financial support from the National Natural Science Foundation of China (71171181, 71322101, 71271195, and 71121061), the Ministry of Education of China (NCET-10-0993), and USTC Fund for Innovation Teams (WK2040160008) is acknowledged.

References

[1] K. Wang, Y. M. Wei, and X. Zhang, "Energy and emissions efficiency of Chinese regions: a multidirectional efficiency analysis," *Applied Energy*, vol. 104, pp. 105–116, 2013.

[2] State Council of the People's Republic of China (SCPRC), The 12th Five Year Plan of China, 2011, (Chinese).

[3] W. Lu, "Potential energy savings and environmental impacts of energy efficiency standards for vapor compression central air conditioning units in China," *Energy Policy*, vol. 35, no. 3, pp. 1709–1717, 2007.

[4] Y. Hu, "Implementation of voluntary agreements for energy efficiency in China," *Energy Policy*, vol. 35, no. 11, pp. 5541–5548, 2007.

[5] M. Watanabe and K. Tanaka, "Efficiency analysis of Chinese industry: a directional distance function approach," *Energy Policy*, vol. 35, no. 12, pp. 6323–6331, 2007.

[6] Y. Bian and F. Yang, "Resource and environment efficiency analysis of provinces in China: a DEA approach based on Shannon's entropy," *Energy Policy*, vol. 38, no. 4, pp. 1909–1917, 2010.

[7] T.-L. Yeh, T.-Y. Chen, and P.-Y. Lai, "A comparative study of energy utilization efficiency between Taiwan and China," *Energy Policy*, vol. 38, no. 5, pp. 2386–2394, 2010.

[8] K. Wang, S. Yu, and W. Zhang, "China's regional energy and environmental efficiency: a DEA window analysis based dynamic evaluation," *Mathematical and Computer Modelling*, vol. 58, no. 5-6, pp. 1117–1127, 2011.

[9] International Energy Agency, *CO_2 Emissions from Fuel Combustion*, IEA, Paris, France, 2009.

[10] International Energy Agency, *CO_2 Emissions from Fuel Combustion Highlights*, IEA, Cancun, Mexico, 2011.

[11] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[12] W. W. Cooper, L. M. Seiford, and J. Zhu, Eds., *Handbook on Data Envelopment Analysis*, Kluwer Academic, London, UK, 2004.

[13] W. D. Cook and L. M. Seiford, "Data envelopment analysis (DEA)—thirty years on," *European Journal of Operational Research*, vol. 192, no. 1, pp. 1–17, 2009.

[14] P. Zhou, B. W. Ang, and K. L. Poh, "A survey of data envelopment analysis in energy and environmental studies," *European Journal of Operational Research*, vol. 189, no. 1, pp. 1–18, 2008.

[15] J.-L. Hu and S.-C. Wang, "Total-factor energy efficiency of regions in China," *Energy Policy*, vol. 34, no. 17, pp. 3206–3217, 2006.

[16] B. Zhang, J. Bi, Z. Fan, Z. Yuan, and J. Ge, "Eco-efficiency analysis of industrial system in China: a data envelopment analysis approach," *Ecological Economics*, vol. 68, no. 1-2, pp. 306–316, 2008.

[17] G.-M. Shi, J. Bi, and J.-N. Wang, "Chinese regional industrial energy efficiency evaluation based on a DEA model of fixing non-energy inputs," *Energy Policy*, vol. 38, no. 10, pp. 6172–6179, 2010.

[18] K. Wang, Y.-M. Wei, and X. Zhang, "A comparative analysis of China's regional energy and emission performance: Which is the better way to deal with undesirable outputs?" *Energy Policy*, vol. 46, pp. 574–584, 2012.

[19] Q. Wang, P. Zhou, and D. Zhou, "Efficiency measurement with carbon dioxide emissions: the case of China," *Applied Energy*, vol. 90, no. 1, pp. 161–166, 2012.

[20] L. Wang, L. Xu, and H. Song, "Environmental performance evaluation of Beijing's energy use planning," *Energy Policy*, vol. 39, no. 6, pp. 3483–3498, 2011.

[21] Z.-H. Wang, H.-L. Zeng, Y.-M. Wei, and Y.-X. Zhang, "Regional total factor energy efficiency: an empirical analysis of industrial sector in China," *Applied Energy*, vol. 97, pp. 115–123, 2012.

[22] F. Wu, L. W. Fan, P. Zhou, and D. Q. Zhou, "Industrial energy efficiency with CO_2 emissions in China: a nonparametric analysis," *Energy Policy*, vol. 49, pp. 164–172, 2012.

[23] Y. Choi, N. Zhang, and P. Zhou, "Efficiency and abatement costs of energy-related CO_2 emissions in China: a slacks-based efficiency measure," *Applied Energy*, vol. 98, pp. 198–208, 2012.

[24] M. Asmild and K. Matthews, "Multi-directional efficiency analysis of efficiency patterns in Chinese banks 1997–2008," *European Journal of Operational Research*, vol. 219, no. 2, pp. 434–441, 2012.

[25] R. Fare, S. Grosskopf, C. A. K. Lovell, and C. Pasurka, "Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach," *The Review of Economics and Statistics*, vol. 71, pp. 90–98, 1989.

[26] P. Zhou, B. W. Ang, and H. Wang, "Energy and CO_2 emission performance in electricity generation: a non-radial directional distance function approach," *European Journal of Operational Research*, vol. 221, pp. 625–635, 2012.

[27] H. Fukuyama and W. L. Weber, "A directional slacks-based measure of technical inefficiency," *Socio-Economic Planning Sciences*, vol. 43, no. 4, pp. 274–287, 2009.

[28] P. Zhou, K. L. Poh, and B. W. Ang, "A non-radial DEA approach to measuring environmental performance," *European Journal of Operational Research*, vol. 178, no. 1, pp. 1–9, 2007.

[29] H. Fukuyama, Y. Yoshida, and S. Managi, "Modal choice between air and rail: a social efficiency benchmarking analysis that considers CO_2 emissions," *Environmental Economics and Policy Studies*, vol. 13, no. 2, pp. 89–102, 2011.

[30] C. P. Barros, S. Managi, and R. Matousek, "The technical efficiency of the Japanese banks: non-radial directional performance measurement with undesirable output," *Omega*, vol. 40, no. 1, pp. 1–8, 2012.

[31] R. Ramanathan, "A holistic approach to compare energy efficiencies of different transport modes," *Energy Policy*, vol. 28, no. 11, pp. 743–747, 2000.

[32] J. Tongzon, "Efficiency measurement of selected Australian and other international ports using data envelopment analysis," *Transportation Research A: Policy and Practice*, vol. 35, no. 2, pp. 107–122, 2001.

[33] R. Färe, S. Grosskopf, D.-W. Noh, and W. Weber, "Characteristics of a polluting technology: theory and practice," *Journal of Econometrics*, vol. 126, no. 2, pp. 469–492, 2005.

[34] P. Bogetoft and J. L. Hougaard, "Efficiency evaluations based on potential (non-proportional) improvements," *Journal of Productivity Analysis*, vol. 12, no. 3, pp. 233–247, 1999.

[35] P. Bogetoft and J. L. Hougaard, "Super efficiency evaluations based on potential slack," *European Journal of Operational Research*, vol. 152, no. 1, pp. 14–21, 2004.

[36] M. Asmild and J. T. Pastor, "Slack free MEA and RDM with comprehensive efficiency measures," *Omega*, vol. 38, no. 6, pp. 475–483, 2010.

[37] M. Asmild, T. Holvad, J. L. Hougaard, and D. Kronborg, "Railway reforms: do they influence operating efficiency?" *Transportation*, vol. 36, no. 5, pp. 617–638, 2009.

[38] T. Holvad, J. L. Hougaard, D. Kronborg, and H. K. Kvist, "Measuring inefficiency in the Norwegian bus industry using multi-directional efficiency analysis," *Transportation*, vol. 31, no. 3, pp. 349–369, 2004.

[39] R. Färe and C. A. Knox Lovell, "Measuring the technical efficiency of production," *Journal of Economic Theory*, vol. 19, no. 1, pp. 150–162, 1978.

[40] A. Charnes, W. W. Cooper, B. Golany, L. Seiford, and J. Stutz, "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *Journal of Econometrics*, vol. 30, no. 1-2, pp. 91–107, 1985.

[41] C. A. Knox Lovell, J. T. Pastor, and J. A. Turner, "Measuring macroeconomic performance in the OECD: a comparison of European and non-European countries," *European Journal of Operational Research*, vol. 87, no. 3, pp. 507–518, 1995.

[42] W. W. Cooper, K. S. Park, and J. T. Pastor, "RAM: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA," *Journal of Productivity Analysis*, vol. 11, no. 1, pp. 5–42, 1999.

[43] J. T. Pastor, J. L. Ruiz, and I. Sirvent, "An enhanced DEA Russell graph efficiency measure," *European Journal of Operational Research*, vol. 115, no. 3, pp. 596–607, 1999.

[44] K. Tone, "A slacks-based measure of super-efficiency in data envelopment analysis," *European Journal of Operational Research*, vol. 143, no. 1, pp. 32–41, 2002.

[45] J. Zhu, "Data envelopment analysis vs. principal component analysis: an illustrative study of economic performance of Chinese cities," *European Journal of Operational Research*, vol. 111, no. 1, pp. 50–61, 1998.

[46] C.-C. Lee, "Energy consumption and GDP in developing countries: a cointegrated panel analysis," *Energy Economics*, vol. 27, no. 3, pp. 415–427, 2005.

[47] National Bureau of Statistics of China (NBSC), *China Energy Statistical Year Book*, China Statistics Press, Beijing, China, 2007b–2011b.

[48] National Bureau of Statistics of China (NBSC), *China Statistical Year Book*, China Statistics Press, Beijing, China, 2007a–2011a.

[49] IPCC, "IPCC Guidelines for National Greenhouse Gas Inventories," 2006, http://www.ipcc-nggip.iges.or.jp/public/2006gl/vol2.html.

[50] National Development and Reform Commission (NDRC), *National Greenhouse Gas Inventory of the People's Republic of China*, Chinese Environmental Science Press, Beijing, China, 2007, (Chinese).

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 246791, 12 pages
http://dx.doi.org/10.1155/2014/246791

Research Article

Synthetic Optimization Model and Algorithm for Railway Freight Center Station Location and Wagon Flow Organization Problem

Xing-cai Liu,[1,2] Shi-wei He,[1] Rui Song,[1,2] Hao-dong Li,[1] Long Wang,[1] and Yang Sun[3]

[1] *School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China*
[2] *MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China*
[3] *Integrated Transport Research Center, China Academy of Transportation Sciences, Beijing 100029, China*

Correspondence should be addressed to Xing-cai Liu; xcliu88@163.com and Shi-wei He; shwhe@bjtu.edu.cn

Received 27 November 2013; Revised 8 April 2014; Accepted 10 April 2014; Published 11 May 2014

Academic Editor: X. Zhang

The railway freight center stations location and wagon flow organization in railway transport are interconnected, and each of them is complicated in a large-scale rail network. In this paper, a two-stage method is proposed to optimize railway freight center stations location and wagon flow organization together. The location model is present with the objective to minimize the operation cost and fixed construction cost. Then, the second model of wagon flow organization is proposed to decide the optimal train service between different freight center stations. The location of the stations is the output of the first model. A heuristic algorithm that combined tabu search (TS) with adaptive clonal selection algorithm (ACSA) is proposed to solve those two models. The numerical results show the proposed solution method is effective.

1. Introduction

In the past decade, China railway has invested many railway freight center stations, which are equipped comprehensive transportation facilities and logistics facilities for the purpose of centralized transportation. At the same time, most of the stations with small transport demand were closed. Those center stations can help to centralize transport demand and gain economic of scale, while the transport demand of closed stations may be lost, if the transport cost of the railway is higher than the other transport modes such as road and water transport. In order to attract the transport demand, railway must improve the wagon flow organization to provide service with high level.

The center station location and wagon flow organization are interacted and interconnected. In order to organize products with high level of service, the amount of goods must meet the train size limitation. Otherwise, long waiting time may be caused. At the same time, the efficiency of wagon flow organization can decide the operation efficiency of stations. For example, the limited storage capacity of station will be occupied, if there is no suitable train to serve the demand.

In order to solve the complicated station location and wagon flow organization problem, a two-stage method is proposed to optimize station location and wagon flow organization together. This paper is organized as follows. Section 2 briefly reviews the relevant literature. Section 3 introduces the optimal models in a two-stage method. In Section 4, a heuristic algorithm is proposed to solve the models. Finally, a numerical example is provided to illustrate the application of the models and algorithm.

2. Literature Review

The location of railway freight center stations is similar to hub location problem, which is among the most critical management decisions. Many models have been proposed

such as set covering problem (SCP), p-center problem, and p-median problem [1–5].

The SCP model was presented by Caprara et al. [5] and solved by a heuristic method based on Lagrangian relaxation. Campbell [6] came up with formulations for the multiple and single allocation p-hub median problem with two heuristic methods. Besides, Campbell et al. [7] concluded that there are two types of hub location problem: single allocation and multiple allocation. Skorin-Kapov et al. [8] studied the uncapacitated multiple and single allocation p-hub median problems. O'Kelly [9] analyzed the models for both single-hub and two-hub problems. And O'Kelly [10] presented a quadratic integer program for the hub location. The model is linearized and solved by a heuristic algorithm. Furthermore, O'Kelly and Bryan [11] proposed a reliable model by considering the scale economies effect of traffic. Racunica and Wynter [12] studied the location problem in a hub-and-spoke network, aiming to increase the share of rail in intermodal transport.

Recently, fuzzy theory and dynamic environment were introduced into the location problem. Batanović et al.'s study [13] concerned the uncertain parameters in maximum covering location problem, which were modeled by fuzzy set. Sáez-Aguado and Trandafir [14] improved the p-median model by considering a coverage constraint. A dynamic uncapacitated hub location problem was present in Contreras et al.'s paper [15]. Correia et al. [16] proposed four mixed-integer linear programming formulations in order to study the extension of classical capacitated single-allocation hub location problem.

For the wagon flow organization problem, Bodin et al. [17] developed a nonlinear, mixed-integer programming model for the railroad blocking problem. Martinelli and Teng [18] applied neural network to solve the railway operation problem. Some research described the railway blocking problem with network theory. Newton et al. [19] took the railroad blocking problem as a network design problem and developed a column generation and branch-and-bound algorithm to solve it. Similarly, Barnhart et al. [20] proposed a Lagrangian relaxation approach for optimizing railway blocking problem. Fukasawa et al. [21] considered both the loaded and empty cars in the network and proposed an integer multicommodity flow model for the problem. Woxenius [22] provided six principles of rail operation and applied them into intermodal rail freight transport network system. Jeong et al. [23] formulated a linear integer programming model in a hub-and-spoke network. Jha et al. [24] compared arc-based with path-based formulations of the block-to-train assignment problem and proved that the latter is smaller in scale. Keaton [25] presented a mixed-integer programming for railroad blocking problem.

In China, Wu [26] studied the organization scheme for through trains. Lin et al. [27] formulated a nonlinear 0-1 through train formation model considering the different cost parameters and block capacity constraint. Nonlinear 0-1 programming models were established in Cao et al.'s study [28, 29] to determine the freight train scheduling plan based on analyzing the logistics system costs.

Most of the research in literature focused either on the location of stations or the wagon flow organization, while the influence of the location on wagon flow organization should be considered. In this paper, a two-stage method is used to solve this problem.

3. The Two-Stage Programming

3.1. Model of Optimal Railway Freight Center Station Location Problem. As the location of railway freight center stations is similar to the hub location problem, customer specifies the ordinary railway freight station, while service point specifies the candidate railway freight center station.

3.1.1. Decision Variables. The objective of this model is to find the optimal location of service points and the assignment between customer and service point. The location decision and assignment are treated as decision variables. Those are

x_{ij} equals 1 if customer i is assigned to service point j. Otherwise, it equals 0.

y_j equals 1 if service point is located at candidate service point j. Otherwise, it equals 0.

3.1.2. Objective Function. Customers hope that the service points are located close to themselves, so as to send the goods to the station quickly and cheaply, while the planners want to maximize the railway coverage with limited center stations, so as to reduce the total investment. Both the transport cost from costumers to service points and construction costs of stations are considered by the following objective function:

$$\min Z = \mu_1 \sum_{i \in I} \sum_{j \in J} \lambda d_{ij} q_i' x_{ij} + \mu_2 \sum_{j \in J} C_j y_j, \tag{1}$$

where

λ is unit cost to transport the goods from customer to service point;

μ_1 and μ_2 are weight of transport cost and construction cost in objective function. The values are defined in advance. And $\mu_1 + \mu_2 = 1$, $\mu_1 \geq 0$, $\mu_2 \geq 0$;

d_{ij} is distance between customer i and candidate service point j;

q_i' is transport demand of customer i;

C_j is fixed cost to construct a service point at candidate service point j;

I is set of customers, $i \in I$;

J is set of candidate service points, $j \in J$.

3.1.3. Constraints. The coverage constraint of a service point is considered. The whole distance which is greater than the preestablished coverage distance at a service point should not exceed a previously chosen value. This constraint is related to the maximum covering location problem (MCLP), whereas the investment of service point may change the transport demand. The model also takes this situation into account.

(i) Each customer must be assigned to one service point

$$\sum_{j\in J} x_{ij} = 1 \quad \forall i \in I. \tag{2}$$

(ii) Candidate service point j cannot serve any customer, if j is not chosen as a service point

$$x_{ij} \le y_j \quad \forall i \in I, \quad j \in J. \tag{3}$$

(iii) The total number of chosen service points should be constrained

$$\sum_{j\in J} y_j \le p. \tag{4}$$

p is the maximum number of the stations.

(iv) Sum of the distance which is greater than coverage distance DC at a service point should not exceed δ. Both DC and δ are prespecified

$$\sum_{i\in I} l_{ij} x_{ij} \le \delta \quad \forall j \in J. \tag{5}$$

l_{ij} is defined as follows:

$$l_{ij} = \begin{cases} d_{ij} & d_{ij} > \text{DC} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

(v) The goods serviced by point j cannot exceed its capacity Cap_j.

$$\sum_{i\in I} q_i' x_{ij} \le \text{Cap}_j \quad \forall j \in J. \tag{7}$$

q_i' is defined as follows:

$$q_i' = \begin{cases} \varepsilon q_i & \text{point } i \text{ is not a service point} \\ q_i & \text{otherwise} \end{cases} \quad \forall i \in I, \tag{8}$$

where ε is a discount coefficient of demand at point i if i is not a service point, which describes the change of transport demand. q_i is the expected transport demand at point i.

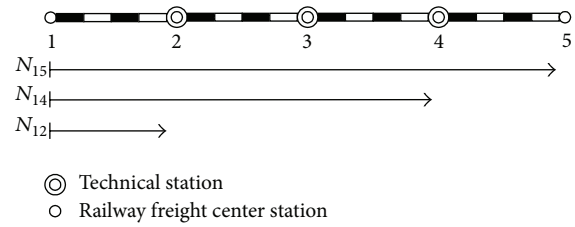


Figure 1: A simple rail network.

3.1.4. Mathematical Model. The location model of railway freight center stations can be stated as

(M-I)

$$\begin{aligned} \min \quad & Z = \mu_1 \sum_{i\in I}\sum_{j\in J} \lambda d_{ij} q_i' x_{ij} + \mu_2 \sum_{j\in J} C_j y_j \\ \text{s.t.} \quad & \sum_{j\in J} x_{ij} = 1 \quad \forall i \in I \\ & x_{ij} \le y_j \quad \forall i \in I, \ j \in J \\ & \sum_{j\in J} y_j \le p \\ & \sum_{i\in I} l_{ij} x_{ij} \le \delta \quad \forall j \in J \\ & \sum_{i\in I} q_i' x_{ij} \le \text{Cap}_j \quad \forall j \in J \\ & x_{ij}, y_j \in \{0,1\} \quad \forall i \in I, \quad j \in J. \end{aligned} \tag{9}$$

3.2. Model of Wagon Flow Organization Problem. The outputs of M-I include the location of railway freight center stations and the transport demand of each station. Based on the average loaded weight of a wagon, the transport demand can be turned into wagon flows. Those are the input data of wagon flow organization.

3.2.1. An Example of Wagon Flow Organization Problem. The wagon flow organization problem can be illustrated by a simple network (see Figure 1), which has three technical stations, two freight center stations, and three shipments (N_{12}, N_{14}, and N_{15}).

There are 12 combinations for routing all the shipments on potential train services (see Figure 2). Arcs in the network specify the available train connections. The OD pair will be reclassified at the technical stations in the itinerary if it is not served by through train. If transport demand between two adjacent stations is positive, the train service must be provided. So shipment N_{12} has only one service strategy that is served by the train service $1 \to 2$.

Strategy 1 (see Figure 2(a)). The OD pairs N_{15}, N_{14}, and N_{12} are consolidated in train service $1 \to 2$ at the origin station 1.

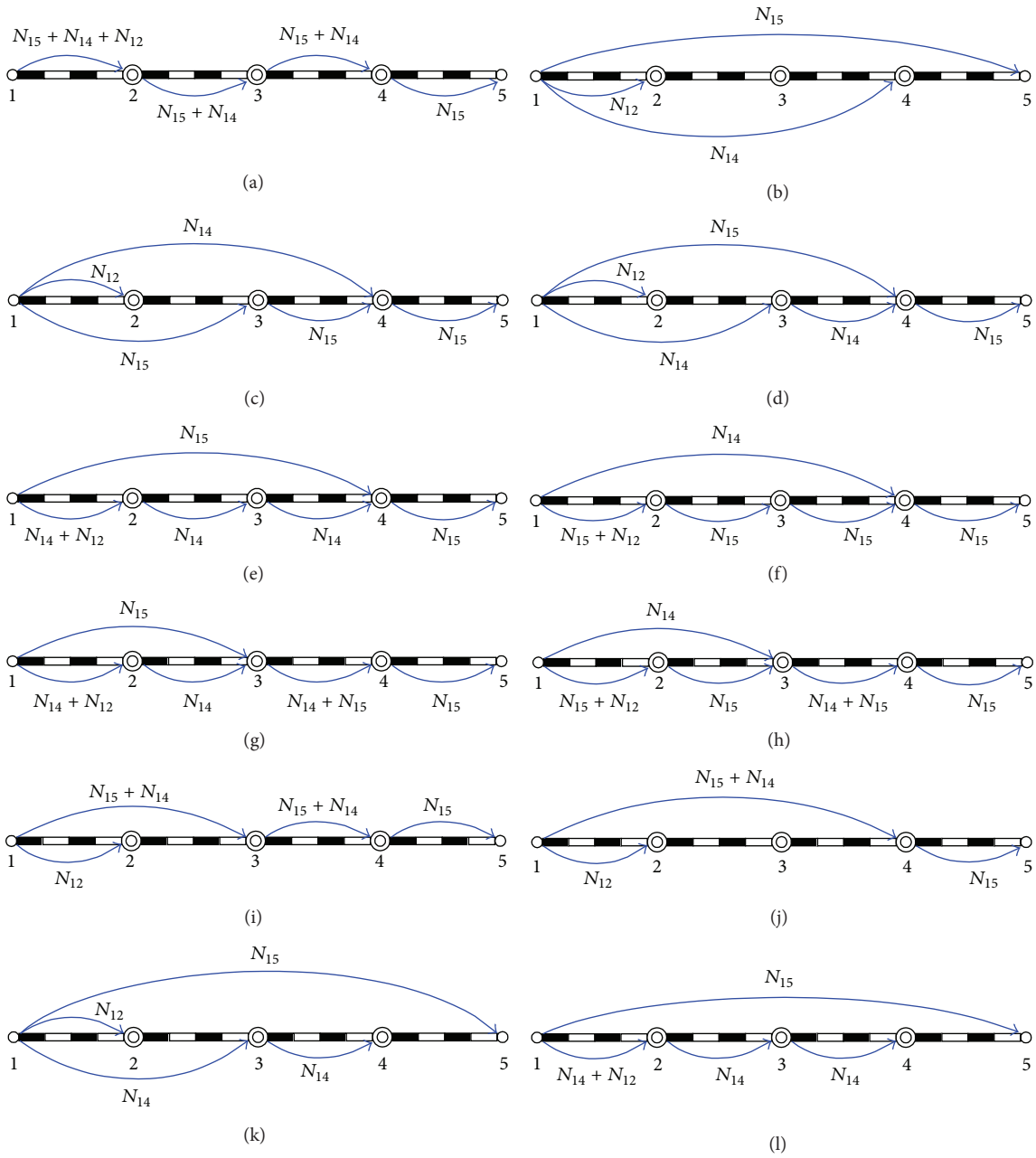


Figure 2: Twelve combinations for routing all the shipments on potential train services.

In this case, N_{15} is reclassified at technical stations 2, 3, and 4. N_{14} is reclassified at technical stations 2 and 3.

Strategy 2 (see Figure 2(b)). There are two through train services $1 \rightarrow 4$ and $1 \rightarrow 5$. In this case, both N_{15} and N_{14} are directly shipped to the destination.

Strategy 3 (see Figure 2(c)). There are two through train services $1 \rightarrow 3$ and $1 \rightarrow 4$. In this case, N_{14} is directly shipped to the destination. While N_{15} is directly shipped to technical station 3, and N_{15} still needs to be reclassified at technical stations 3 and 4.

Strategy 4 (see Figure 2(d)). Strategy 3 and Strategy 4 share the common train service set, but the train services are different. N_{15} is directly shipped to technical station 4 and reclassified there. N_{14} is directly shipped to technical station 3 and reclassified there.

Strategy 5 (see Figure 2(e)). There is one through train service $1 \rightarrow 4$. In this case, N_{15} is directly shipped to the technical station 4 and reclassified there. N_{14} and N_{12} are consolidated in train service $1 \rightarrow 2$. N_{14} is reclassified at technical stations 2 and 3.

Strategy 6 (see Figure 2(f)). Strategy 5 and Strategy 6 share the common train service set, but the train services are different. N_{14} is directly shipped to technical station 4 by train service $1 \to 4$. N_{15} and N_{12} are consolidated in train service $1 \to 2$. N_{15} is reclassified at technical stations 2, 3, and 4.

Strategy 7 (see Figure 2(g)). There is one through train service $1 \to 3$. In this case, N_{15} is directly shipped to the technical station 3 and reclassified at technical stations 3 and 4. N_{14} and N_{12} are consolidated in train service $1 \to 2$. N_{14} is reclassified at technical stations 2 and 3.

Strategy 8 (see Figure 2(h)). Strategy 7 and Strategy 8 share the common train service set, but the train services are different. N_{14} is directly shipped to technical station 3 by train service $1 \to 3$ and reclassified there. N_{15} and N_{12} are consolidated in train service $1 \to 2$. N_{15} is reclassified at technical stations 2, 3, and 4.

Strategy 9 (see Figure 2(i)). There is one through train service $1 \to 3$. In this case, N_{15} and N_{14} are directly consolidated together and shipped to the technical station 3 by train service $1 \to 3$. N_{14} is reclassified at technical station 3. N_{15} is reclassified at technical station 3 and 4.

Strategy 10 (see Figure 2(j)). There is one through train service $1 \to 4$. In this case, N_{15} and N_{14} are consolidated together and directly shipped to the technical station 4 by train service $1 \to 4$. N_{15} is reclassified there.

Strategy 11 (see Figure 2(k)). There are two through train services $1 \to 3$ and $1 \to 5$. In this case, N_{15} is directly shipped to destination by train service $1 \to 5$. N_{14} is directly shipped to the technical station 3 by train service $1 \to 3$ and reclassified there.

Strategy 12 (see Figure 2(l)). There is one through train service $1 \to 5$. In this case, N_{15} is directly shipped to destination by train service $1 \to 5$. N_{14} and N_{12} are consolidated in train service $1 \to 2$. N_{14} is reclassified at technical station 2 and 3.

3.2.2. Analysis of Wagon Flow Organization Problem. Among the twelve strategies above, there are two extreme ones, that is, Strategy 1 and Strategy 2.

(1) In Strategy 1, train connection services are only provided between adjacent stations. In this way, the number of through train services reaches minimum; that is, the total waiting time at the dispatching station reaches minimum. However, each shipment will be reclassified at each technical station on its itinerary. Thus the reclassification fee and time consumption will increase significantly. The increasing number of reclassified wagons may cause congestion and time delay at some technical stations.

(2) In Strategy 2, transport demands are all delivered to their destinations without reclassification, which will reduce the reclassification fee and time consumption. However, this strategy may be unworkable. Such strategy requires enough classification tracks to store outbound trains, and single OD pair must be large enough to dispatch a through train service. Waiting time at the dispatching station will increase if the OD pair is small.

The solution of wagon flow organization problem is a tradeoff between Strategy 1 and Strategy 2, aiming at minimizing the total shipping and handling cost for all shipments. The time consumption is used to weigh the cost.

3.2.3. Decision Variables. From the example above, it can be concluded that there are three train services for a wagon flow: the first one is shipped to the destination by nonstop shipping scheme at the origin station, the second one is served by nonthrough shipment, and the third one is shipped to a technical station in the itinerary by nonstop shipping scheme consolidated with other OD pairs. These variables are

$$x_{st} = \begin{cases} 1 & N_{st} \text{ is shipped by nonstop scheme} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{st}^{k} = \begin{cases} 1 & N_{st} \text{ is shipped to first technical station } k \text{ in the itinerary and reclassified there} \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in V(s) \tag{10}$$

$$y_{st}^{k} = \begin{cases} 1 & N_{st} \text{ is shipped to tcehnical station } k \text{ and reclassified there} \\ 0 & \text{otherwise} \end{cases} \quad \forall k \in P(s),$$

where

N_{st} is number of wagon flow from origin s to destination t;

s is origin station;

t is destination station;

$V(s)$ is set of technical stations next to the origin station s, $k \in V(s)$;

$P(s)$ is set of technical stations in the itinerary of wagon flow N_{st}; exclude the technical station next to the origin station, $P(s) \cap V(s) = \emptyset$, $k \in P(s)$.

3.2.4. Objective Function. The optimal goal is to obtain the wagon flow organization plan with minimum time consumption, which means the level of service. The travel time is not considered for it is identical. In order to simplify the problem, the following assumptions are proposed: (1) the supply of empty wagons is not considered, (2) the storage capacity of the origin and destination station are not considered, (3)

the shipping routes for all the service are given, and (4) the train formation plans (TFP) of the technical stations in the itinerary are known. This information can be obtained previously.

Time consumption of three train services for N_{st} is as follows.

(i) For the nonstop service, the service time equals the loading and unloading times As there is no intermediate service such as reclassification

$$F_1 = N_{st} w_{st} x_{st}, \tag{11}$$

where

w_{st} is loading and unloading time at the origin and destination for a wagon served by nonstop shipping scheme.

(ii) Nonthrough shipment cost, which includes the reclassified time in the itinerary and loading and unloading time

$$F_2 = \sum_{k \in V(s)} N_{st} \overline{\overline{w}}_{st} x_{st}^k + \sum_{k \in V(s)} \sum_{g \in G(k,t)} N_{st} t_g x_{st}^k, \tag{12}$$

where

$\overline{\overline{w}}_{st}$ is loading and unloading time at the origin and destination for a wagon shipped by non-through shipment;

t_g is reclassified time at technical station g;

$G(k, t)$ is set of technical stations in the itinerary after N_{st} has been reclassified at technical station k, and $g \in G(k, t)$.

(iii) The wagon flow is shipped to a technical station in the itinerary by nonstop shipping scheme. This item includes the reclassified time consumption in the itinerary and time consumption for loading and unloading

$$F_3 = \sum_{k \in P(s)} N_{st} \overline{w}_{st} y_{st}^k + \sum_{k \in P(s)} \sum_{g \in G(k,t)} N_{st} t_g y_{st}^k, \tag{13}$$

where

$\overline{w}_{st}$ is loading and unloading time at the origin and destination for a wagon shipped to a technical station in the itinerary by nonstop shipping scheme.

Then, the total time consumption for N_{st} can be formulated as

$$F_{st} = F_1 + F_2 + F_3. \tag{14}$$

The objective function of wagon flow organization can be stated as

$$\min \sum_{s \in S} \sum_{t \in Q(s)} F_{st}, \tag{15}$$

where S is set of origin stations and $s \in S$. $Q(s)$ is set of destinations of the wagon flows that originate from s. And $Q(s) = \{t : N_{st} \neq 0\}$. $t \in Q(s)$.

3.2.5. Constraints. The wagon flow organization problem has two constrains.

(1) Loading and Unloading Capacity Constraint for Nonstop Train. The loading and unloading capacity must reach the minimum size of loading or unloading a whole nonstop train. If there are several wagon flows consolidated together to be shipped to a technical station in the itinerary by nonstop shipping scheme, the sum of their loading or unloading capacity must reach the average marshaling number of wagons in a nonstop shipping scheme

$$\overline{m}_{st} I\left(\sum_{t \in Q(s)} y_{st}^k\right) - \sum_{t \in Q(s)} m_{st} y_{st}^k \leq 0 \quad \forall s \in S, \ k \in P(s), \tag{16}$$

where

m_{st} is the smaller one between maximum loading capacity of station s and maximum unloading capacity of station t at certain period;

$\overline{m}_{st}$ is the average marshaling number of wagons in nonstop shipping scheme;

$I(x)$ is a step function, $I(x) = \begin{cases} 1 & \text{if } x>0 \\ 0 & \text{otherwise} \end{cases}$.

(2) Each Wagon Flow Is Served by One and Only One Train Service.

$$I_{st} x_{st} + \sum_{k \in V(s)} x_{st}^k + \sum_{k \in P(s)} y_{st}^k = 1 \quad \forall s \in S, \quad t \in Q(s), \tag{17}$$

where $I_{st} = \begin{cases} 1 & \text{if } m_{st} \geq \overline{m}_{st} \\ 0 & \text{otherwise} \end{cases}$ $\forall s \in S$, $t \in Q(s)$, which ensures that only the wagon flow that reaches the average marshaling number of wagon in nonstop shipping scheme can be shipped by it.

3.2.6. Constraint Linearization. Constraint (16) can be linearized by introducing a variable y_{sk}

$$y_{sk} = I\left(\sum_{t \in Q(s)} y_{st}^k\right) = \begin{cases} 1 & \sum_{t \in Q(s)} y_{st}^k > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall s \in S, \quad k \in P(s). \tag{18}$$

Then constraint (16) can be formulated as follows:

$$\overline{m}_{st} y_{sk} - \sum_{t \in Q(s)} m_{st} y_{st}^k \leq 0 \quad \forall s \in S, \ k \in P(s). \tag{19}$$

Constraint (20) is also introduced to ensure that y_{sk} equals 1 if and only if $y_{st}^k = 1$

$$M y_{sk} \geq \sum_{t \in Q(s)} y_{st}^k \quad \forall s \in S, \ k \in P(s), \tag{20}$$

where M is a large positive constant.

3.2.7. Mathematical Model. The wagon flow organization model can be stated as

$$\text{(M-II)}$$

$$\min \quad \sum_{s\in S}\sum_{t\in Q(s)} F_{st}$$

$$\text{s.t.} \quad I_{st}x_{st} + \sum_{k\in V(s)} x_{st}^{k} + \sum_{k\in P(s)} y_{st}^{k} = 1 \quad \forall s \in S,\ t \in Q(s)$$

$$\overline{m}_{st}y_{sk} - \sum_{t\in Q(s)} m_{st}y_{st}^{k} \le 0 \quad \forall s \in S,\ k \in P(s)$$

$$My_{sk} \ge \sum_{t\in Q(s)} y_{st}^{k} \quad \forall s \in S,\ k \in P(s)$$

$$x_{st}, x_{st}^{k} \in \{0,1\} \quad \forall s \in S,\quad t \in Q(s),\quad k \in V(s)$$

$$y_{sk}, y_{st}^{k} \in \{0,1\} \quad \forall s \in S,\quad t \in Q(s),\quad k \in P(s). \tag{21}$$

4. Solution Algorithm

Tabu search (TS) [30, 31] uses the tabu list to record the information of local optimal solutions, which can help to enlarge the search space and avoid the locally optimal solutions. The adaptive clonal selection algorithm (ACSA) [32–34] is shown to be an evolutionary strategy capable of high convergence rate and diversified. In this section, TS and ACSA are combined into a new heuristics method, called T-ACSA, to solve the proposed models.

4.1. The Detail Techniques of T-ACSA

4.1.1. Affinity Measure. Affinity measure of the algorithm is the objective of the model; the smaller the better. In order to extend the space of solution, the algorithm accepts solutions which fail to satisfy the constraints. However, penalty coefficient will be added to the affinity measure.

4.1.2. The Design of Antibody. The antibody of model M-I is designed as follows.

The length of antibody is equal to the amount of customers in I. The codes in antibody are in J, and the amount of codes is p. To better understand the antibody, a simple example consisting of seven customers and four service points is proposed. p equals three (see Figure 3). Service point 2 is not included in the antibody, which means that it is not chosen as a service point.

The antibody of model M-II is designed as follows.

According to the optimal result of model M-I, the antibody of each service point can be designed. A piece of antibody is proposed (see Figure 4).

The number of y_{st}^{k} in each piece is decided by the amount of technical stations in $P(s)$ in the itinerary $s \to t$, which can be obtained when the route is set. The sum value of each piece is 1. The pieces are separated between each other. And their amount equals the number of wagon flows.

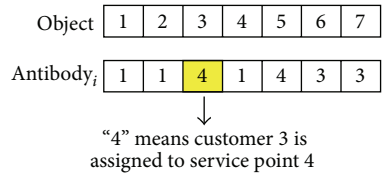


Figure 3: The design of antibody for M-I.

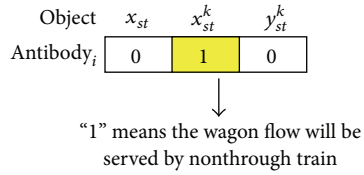


Figure 4: One piece of antibody for M-II.

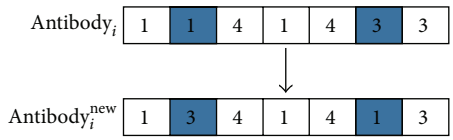


Figure 5: Neighborhood search operation of model M-I.

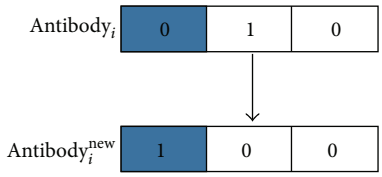


Figure 6: Neighborhood search operation of model M-II.

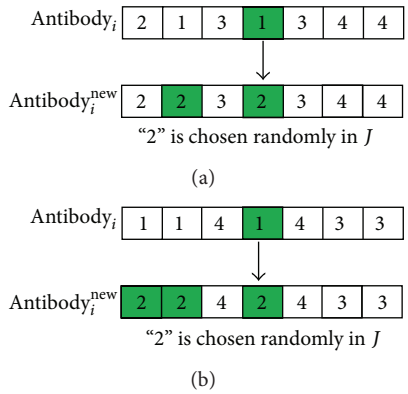


Figure 7: The mutation operation of model M-I.

4.1.3. Neighborhood Search Operation. The neighborhood search operation of model M-I is as follows: choose two positions in the antibody randomly and exchange the values. The neighborhood search operation of the antibody in Figure 3 is shown in Figure 5.

Table 1: The investment for each candidate station.

Candidate point	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C_j	150	300	170	160	210	170	200	230	180	210	150	60	110	250	180	50

Unit: ten thousand CNY.

Table 2: The parameters of model M-I and T-ACSA.

Parameters of model M-I									
	μ_1	μ_2	DC	p	δ	Cap_j	λ	ε	
Value	0.3	0.7	182 (km)	8	200 (km)	1752000 (t)	0.67 (CNY/t)	0.8	
Parameters of T-ACSA									
	$w_{\max}$	$w_{\min}$	L	D	α	N	H	R	U
Value	8	2	100	18	0.6~0.8	20~30	4~6	10~15	8

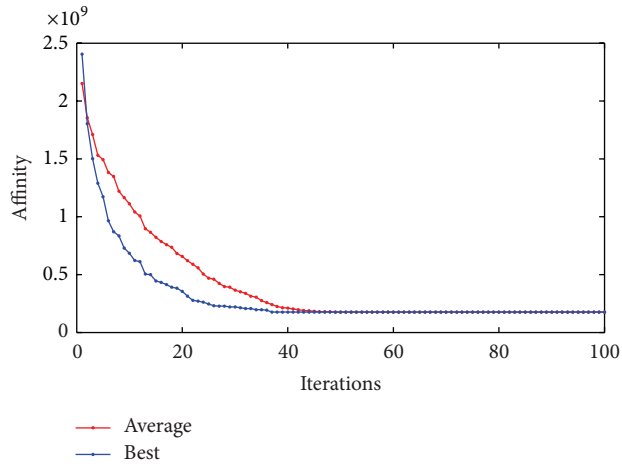


Figure 8: The iterative process of T-ACSA for M-I.

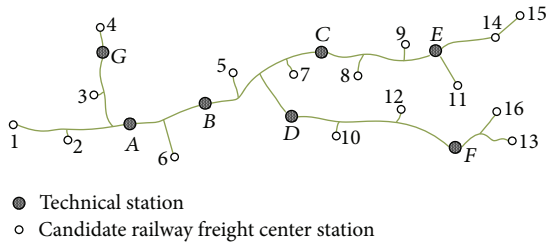


Figure 9: The railway network and the location of candidate freight center stations.

The neighborhood search operation of model M-II is as follows: choose a piece in antibody randomly. Find a code whose value is 0 and assign it to 1. Change the other codes into 0. The other pieces of antibody remain the same. The neighborhood search operation of the antibody in Figure 4 is shown in Figure 6.

4.1.4. Mutation Operation. Mutation operation of model M-I is shown in Figure 7: the maximum number of codes is four. If the amount of the chosen candidate points reaches the maximum, choose a code e in the antibody randomly. Change both e and the codes whose value is the same as e (see Figure 7(a)). Else choose a code e and change its value randomly (see Figure 7(b)).

The mutation operation of model M-II is the same as the neighborhood search operation.

Table 3: The reclassified time consumption at a technical station.

Technical station	t_A	t_B	t_C	t_D	t_E	t_F	t_G
t_g	4	6	4.5	5	4	5	4.5

Unit: car hour.

Table 4: The wagon flows between railway freight center stations.

Origin	Destination					Sum of wagon flows
	3	5	6	13	14	
3	—	24	8	42	21	95
5	14	—	20	41	22	97
6	23	7	—	14	40	84
13	19	28	15	—	25	87
14	28	7	6	34	—	75

4.1.5. TS Operation. The tabu strategy is first in first out. The length of tabu list is H; the search times are R. The number of the candidate solutions is U. TS operation procedure is as follows.

Step 1. Choose an antibody and set it as the current solution. Initialize the tabu list and the best solution.

Step 2. Use the neighborhood search operation to update the current solution and generate U candidate solutions. Then sort them by the affinity measure values. Choose the best solution.

Step 3. Judge whether the best solution is tabued. If it is tabued, go to Step 2. Else judge whether the length of tabu list reaches H. If not, add the best solution into the tabu list and tabu it. Else remove the tabu information of the first solution

Table 5: The parameters of model M-II and T-ACSA.

	Parameters of model M-II								
	w_{st}	$\overline{w}_{st}$	$\overline{\overline{w}}_{st}$	$\overline{m}_{st}$	m_{st}				
Value	20 (car hour)	15 (car hour)	12 (car hour)	40	40				
	Parameters of T-ACSA								
	$w'_{\max}$	$w'_{\min}$	L'	D'	α'	N'	H'	R'	U'
Value	6	2	50	10	0.6~0.8	16~20	4	5~15	6

Table 6: The optimal result of wagon flow organization.

Origin	Destination				
	3	5	6	13	14
3	—	$y^B_{3,5} = 1$	$x^A_{3,6} = 1$	$x_{3,13} = 1$	$y^B_{3,14} = 1$
5	$x^B_{5,3} = 1$	—	$x^B_{5,6} = 1$	$x_{5,13} = 1$	$x^C_{5,14} = 1$
6	$x^A_{6,3} = 1$	$x^B_{6,5} = 1$	—	$x^B_{6,13} = 1$	$x_{6,14} = 1$
13	$y^B_{13,3} = 1$	$y^D_{13,5} = 1$	$y^B_{13,6} = 1$	—	$y^D_{13,14} = 1$
14	$y^B_{14,3} = 1$	$x^E_{14,5} = 1$	$y^B_{14,6} = 1$	$x^E_{14,13} = 1$	—

in the tabu list and add the best solution into the tabu list. Go to Step 4.

Step 4. Judge whether the current best solution is the best one in the history. If not, go to Step 5. Else set it as the best one in the history. Go to Step 5.

Step 5. Judge whether the search times reach R. If not, go to Step 2. Else update the antibody with the best solution in the history.

4.2. The Process of T-ACSA. Based on the aforementioned detailed analysis, T-ACSA approach is designed as follows.

Step 1. Initialize the population of antibody. Generate N antibodies and constitute the species group P.

Step 2. Count the affinities and sort the antibodies according to their affinities in an ascending order.

Step 3. Clone each antibody in P then get a new species group C. The number of clone is $n_i = [w_{\max}(1-((i-1)/N))]$ and $n_i \geq w_{\min}$, where i is the order of the antibody after sorting. $w_{\max}$ is the maximum clone number and $w_{\min}$ is the minimum clone number.

Step 4. Apply the TS operation to update each antibody in C and get the new species group C'.

Step 5. Use mutation operation to update each antibody in C'. And get the new species group C''. The probability of mutation is inversely proportional to the evolution generation $\varphi_l = [\alpha(1 - (l/L))]$. Where α is the coefficient, l is the current generation and L is the maximum generation.

Step 6. Choose the first d_l antibodies in C'' and replace the worst d_l antibodies in P by them; $d_l = [(\overline{f} - f_{\min})(D/\overline{f})]$. Where D is the coefficient, $\overline{f}$ is the average value of affinities in C'' and $f_{\min}$ is the minimum value of affinities in C''.

Step 7. If current status does not meet the terminal condition (the maximum searching times), go to Step 2. Otherwise, go to Step 8.

Step 8. Output the best solution, that is, the optimal location of service points or the wagon flow organization.

5. Numerical Experiments

The models and algorithm are tested by a railway network with 48 railway freight stations. Station 1 to station 16 is candidate freight center stations. The parameters of this physical network are provided in Tables 7 and 8, which include the distance information and the transport demand of each customer. Table 1 shows the investment to locate a center station at the candidate sites. Parameters of M-I model and T-ACSA are shown in Table 2.

We test the algorithm ten times by different combination of parameters. The best affinity value of objective function is 176776352.8 CNY. And the final location plan is 3, 5, 6, 13, and 14. The assignments for the forty-eight freight stations are 3, 3, 3, 3, 5, 6, 5, 5, 14, 6, 13, 13, 13, 14, 14, 13, 3, 3, 3, 3, 3, 5, 5, 5, 6, 5, 6, 5, 6, 6, 6, 5, 5, 5, 6, 13, 6, 14, 14, 6, 13, 14, 13, 13, 13, 14, 14, and 13. The iterative process of the algorithm is shown in Figure 8.

For the second step, we can plan wagon flow organization with outputs of model M-I, such as the location of freight center stations, the transport volume between each center station. Assuming the average loaded weight of a rail wagon is 50 t, and the stations operate for 365 days per year. Then the total transport demand can converted into daily transport volume. The average volumes of railway freight center station 3, 5, 6, 13, and 14 are 95, 97, 84, 87, and 75, respectively. The railway network and the location of railway freight center stations are shown in Figure 9. The reclassified time

Table 7: The distance between customer and service point.

Customer	The candidate service point															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	119	133	196	278	292	348	403	536	415	561	521	602	704	762	741
2	119	0	60	156	159	182	229	284	417	329	443	422	489	585	643	630
3	133	60	0	96	170	218	237	290	421	378	457	461	513	588	646	657
4	196	156	96	0	223	292	277	324	445	461	494	531	564	605	662	708
5	278	159	170	223	0	89	70	125	258	260	288	311	345	426	484	489
6	292	182	218	292	89	0	113	153	273	174	278	243	310	435	492	449
7	348	229	237	277	70	113	0	55	188	257	221	277	287	356	414	431
8	403	284	290	324	125	153	55	0	133	267	170	261	245	301	359	388
9	536	417	421	445	258	273	188	133	0	331	77	269	179	168	226	300
10	415	329	378	461	260	174	257	267	331	0	293	128	263	457	506	370
11	561	443	457	494	288	278	221	170	77	293	0	209	102	165	219	228
12	521	422	461	531	311	243	277	261	269	128	209	0	147	358	401	242
13	602	489	513	564	345	310	287	245	179	263	102	147	0	216	255	144
14	704	585	588	605	426	435	356	301	168	457	165	358	216	0	58	254
15	762	643	646	662	484	492	414	359	226	506	219	401	255	58	0	262
16	741	630	657	708	489	449	431	388	300	370	228	242	144	254	262	0
17	54	173	183	233	332	344	402	457	590	458	614	569	654	758	816	792
18	71	48	77	163	207	225	277	332	465	361	491	461	534	633	691	675
19	252	222	162	66	273	349	319	360	472	521	528	584	605	626	680	748
20	219	184	124	28	243	316	294	338	456	486	507	553	581	613	669	724
21	172	124	64	32	202	266	260	310	435	433	480	507	546	598	655	690
22	158	104	44	52	190	250	252	302	430	415	472	492	535	594	651	679
23	172	53	80	165	106	138	176	231	364	299	391	381	440	532	590	582
24	240	121	135	197	38	97	108	163	296	270	325	334	378	464	522	522
25	279	161	180	240	24	65	74	127	259	237	282	291	334	427	485	477
26	306	187	196	244	28	93	42	97	230	257	261	296	321	398	456	465
27	327	209	225	277	56	78	36	82	212	228	235	261	290	379	437	434
28	352	234	229	252	86	156	50	74	195	307	242	324	320	359	417	462
29	284	170	199	267	59	30	92	138	265	203	278	264	319	430	488	461
30	311	211	255	335	139	50	156	187	293	126	285	211	300	448	504	434
31	337	223	252	317	97	48	87	114	226	172	230	214	266	387	444	407
32	405	287	285	307	131	180	68	40	139	306	191	300	275	304	361	415
33	413	298	286	292	154	218	105	90	161	354	224	350	316	314	370	452
34	451	332	337	367	173	195	103	48	85	285	128	256	214	253	311	353
35	354	249	287	360	147	69	141	159	252	117	239	173	252	403	458	387
36	506	387	392	417	228	245	158	103	30	313	89	261	188	198	256	317
37	388	296	341	421	216	131	212	225	298	45	268	136	253	433	484	372
38	584	465	469	490	306	319	236	181	48	363	79	287	174	120	178	278
39	556	437	444	470	279	287	209	154	27	331	56	259	157	149	207	274
40	452	360	405	483	273	192	257	258	304	46	259	82	220	419	467	324
41	487	391	432	507	290	215	264	256	284	88	231	40	182	387	432	282
42	649	530	533	552	371	382	301	246	113	412	119	322	189	55	113	258
43	572	457	478	524	309	282	247	203	135	260	60	159	45	200	246	184
44	546	437	468	527	305	255	256	226	202	190	135	78	73	280	323	196
45	650	536	558	605	389	358	328	283	196	311	122	191	50	191	221	106
46	707	588	598	624	430	427	361	307	179	425	148	316	170	62	85	192
47	711	594	607	639	438	426	370	318	197	408	151	294	147	102	117	152
48	698	584	604	647	434	407	371	324	223	359	156	237	100	176	194	82

Unit: km.

Table 8: The transport demand of each customer.

Customer	1	2	3	4	5	6	7	8	9	10
Demand	219000	109500	197100	262800	219000	153300	131400	262800	438000	175200
Customer	11	12	13	14	15	16	17	18	19	20
Demand	306600	219000	109500	131400	175200	262800	438000	175200	219000	175200
Customer	21	22	23	24	25	26	27	28	29	30
Demand	328500	219000	219000	175200	131400	153300	394200	109500	219000	109500
Customer	31	32	33	34	35	36	37	38	39	40
Demand	175200	262800	262800	153300	109500	240900	284700	438000	197100	131400
Customer	41	42	43	44	45	46	47	48	—	—
Demand	175200	350400	219000	131400	175200	175200	21900	131400	—	—

Unit: ton per a year.

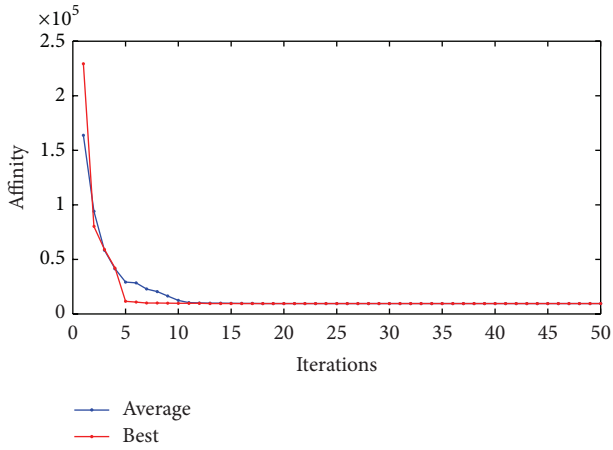


Figure 10: The iterative process of T-ACSA for M-II.

at technical stations is listed in Table 3. The wagon flows between railway freight center stations are shown in Table 4. And the parameters of model M-II and T-ACSA are shown in Table 5.

We test the algorithm five times by different combination of parameters. The final result of wagon flow organization is shown in Table 6. The iterative process of the algorithm is shown in Figure 10.

To verify the algorithm, we solved the models M-I and M-II by ILOG Cplex at the same time. The final results of the present algorithms and Cplex are the same. The run time of models M-I and M-II is less than 2 s and 1 s.

6. Conclusion

A two-stage programming is proposed to describe and solve the location of railway freight center stations and wagon flow organization problem. The first stage determines the optimal location with the objective to minimize the total cost of service and investment. The second stage optimizes the wagon flow organization among different stations. Different from the research in literature, the first model considered the coverage distance constraint and the change of transport demand. The second model analyzed the cost of different schemes. A heuristic algorithm that combines TS with ACSA is designed. The numerical example of a network with 48 stations demonstrates that the method is workable.

While the microoperations in wagon flow organization, like wagon flow route decision and the supply of empty wagons, have not been considered in the scheduling process, these aspects can be considered in the future research.

Conflict of Interests

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

Acknowledgments

This work was supported by National Basic Research Program of China (no. 2012CB725403), National Natural Science Foundation of China (no. 61374202), Research Project of China MOR (nos. 2012X012-E and 2012X006-C), and Research Project of China Railway Company (nos. 2013X005-A and 2013F021).

References

[1] C. Toregas, R. Swain, C. ReVelle, and L. Bergman, "The location of emergency service facilities," *Operations Research*, vol. 19, no. 6, pp. 1363–1373, 1971.

[2] O. Kariv and S. L. Hakimi, "An algorithmic approach to network location problems. I. The p-centers," *SIAM Journal on Applied Mathematics*, vol. 37, no. 3, pp. 513–538, 1979.

[3] O. Kariv and S. L. Hakimi, "An algorithmic approach to network location problems. II. The p-medians," *SIAM Journal on Applied Mathematics*, vol. 37, no. 3, pp. 539–560, 1979.

[4] S. L. Hakimi, "Optimum locations of switching centers and the absolute centers and medians of a graph," *Operations Research*, vol. 12, no. 3, pp. 450–459, 1964.

[5] A. Caprara, M. Fischetti, and P. Toth, "A heuristic method for the set covering problem," *Operations Research*, vol. 47, no. 5, pp. 730–743, 1999.

[6] J. F. Campbell, "Hub location and the p-hub median problem," *Operations Research*, vol. 44, no. 6, pp. 923–935, 1996.

[7] J. F. Campbell, A. T. Ernst, and M. Krishnamoorthy, *Facility Location: Applications and Theory*, Springer, New York, NY, USA, 2002.

[8] D. Skorin-Kapov, J. Skorin-Kapov, and M. O'Kelly, "Tight linear programming relaxations of uncapacitated p-hub median problems," *European Journal of Operational Research*, vol. 94, no. 3, pp. 582–593, 1996.

[9] M. E. O'Kelly, "The location of interacting hub facilities," *Transportation Science*, vol. 20, no. 2, pp. 92–106, 1986.

[10] M. E. O'Kelly, "A quadratic integer program for the location of interacting hub facilities," *European Journal of Operational Research*, vol. 32, no. 3, pp. 393–404, 1987.

[11] M. E. O'Kelly and D. L. Bryan, "Hub location with flow economies of scale," *Transportation Research B: Methodological*, vol. 32, no. 8, pp. 605–616, 1998.

[12] I. Racunica and L. Wynter, "Optimal location of intermodal freight hubs," *Transportation Research B: Methodological*, vol. 39, no. 5, pp. 453–477, 2005.

[13] V. Batanović, D. Petrović, and R. Petrović, "Fuzzy logic based algorithms for maximum covering location problems," *Information Sciences*, vol. 179, no. 1-2, pp. 120–129, 2009.

[14] J. Sáez-Aguado and P. C. Trandafir, "Some heuristic methods for solving p-median problems with a coverage constraint," *European Journal of Operational Research*, vol. 220, no. 2, pp. 320–327, 2012.

[15] I. Contreras, J. F. Cordeau, and G. Laporte, "The dynamic uncapacitated hub location problem," *Transportation Science*, vol. 45, no. 1, pp. 18–32, 2011.

[16] I. Correia, S. Nickel, and F. Saldanha-da-Gama, "Single-assignment hub location problems with multiple capacity levels," *Transportation Research B: Methodological*, vol. 44, no. 8-9, pp. 1047–1066, 2010.

[17] L. D. Bodin, B. L. Golden, A. D. Schuster, and W. Romig, "A model for the blocking of trains," *Transportation Research B. Methodological*, vol. 14, no. 1-2, pp. 115–120, 1980.

[18] D. R. Martinelli and H. Teng, "Optimization of railway operations using neural networks," *Transportation Research C: Emerging Technologies*, vol. 4, no. 1, pp. 33–49, 1996.

[19] H. N. Newton, C. Barnhart, and P. H. Vance, "Constructing railroad blocking plans to minimize handling costs," *Transportation Science*, vol. 32, no. 4, pp. 330–345, 1998.

[20] C. Barnhart, H. Jin, and P. H. Vance, "Railroad blocking: a network design application," *Operations Research*, vol. 48, no. 4, pp. 603–614, 2000.

[21] R. Fukasawa, M. V. P. de Aragão, O. Porto, and E. Uchoa, "Solving the freight car flow problem to optimality," *Electronic Notes in Theoretical Computer Science*, vol. 66, no. 6, pp. 42–52, 2002.

[22] J. Woxenius, "Alternative transport network designs and their implications for intermodal transshipment technologies," *European Transport*, vol. 35, pp. 27–45, 2007.

[23] S. J. Jeong, C. G. Lee, and J. H. Bookbinder, "The European freight railway system as a hub-and-spoke network," *Transportation Research A: Policy and Practice*, vol. 41, no. 6, pp. 523–536, 2007.

[24] K. C. Jha, R. K. Ahuja, and G. Şahin, "New approaches for solving the block-to-train assignment problem," *Networks*, vol. 51, no. 1, pp. 48–62, 2008.

[25] M. H. Keaton, "Designing optimal railroad operating plans: lagrangian relaxation and heuristic approaches," *Transportation Research B*, vol. 23, no. 6, pp. 415–431, 1989.

[26] H. L. Wu, "Study on the organization scheme for through trains from car loading localities," *Journal of the China Railway Society*, vol. 9, no. 4, pp. 17–26, 1987.

[27] B. L. Lin, S. N. Zhu, D. Y. Shi, and S. W. He, "The optimal model of the direct train formation plan for loading area," *China Railway Science*, vol. 16, no. 2, pp. 108–114, 1995.

[28] X. M. Cao, B. L. Lin, and H. X. Yan, "Optimization of direct freight train service in the loading place," *Journal of the China Railway Society*, vol. 28, no. 4, pp. 6–11, 2006.

[29] X. M. Cao, B. L. Lin, H. X. Liu, and H. X. Yan, "Optimization of car flow organization in nonstop train loading base," *Journal of the China Railway Society*, vol. 29, no. 1, pp. 16–20, 2007.

[30] F. Glover, E. Taillard, and E. Taillard, "A user's guide to tabu search," *Annals of Operations Research*, vol. 41, no. 1, pp. 1–28, 1993.

[31] F. Glover and S. Hanafi, "Tabu search and finite convergence," *Discrete Applied Mathematics*, vol. 119, no. 1-2, pp. 3–36, 2002.

[32] L. N. de Castro and F. J. von Zuben, "The clonal selection algorithm with engineering application," in *Proceedings of GFCCO, Workshop on Artificial Immune Systems and Their Application*, pp. 36–37, Las Vegas, Nev, USA, 2000.

[33] L. N. de Castro and F. J. von Zuben, "The clonal selection algorithm with engineering applications," in *Proceedings of the GECC*, pp. 36–37, 2000.

[34] A. Adnan, "Clonal selection algorithm with operator multiplicity," *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 2, pp. 1909–1915, 2004.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 608315, 12 pages
http://dx.doi.org/10.1155/2014/608315

Research Article

SIRS Model of Passengers' Panic Propagation under Self-Organization Circumstance in the Subway Emergency

Haifeng Zhao, Jie Jiang, Rongyu Xu, and Yang Ye

School of Economic and Management, Tongji University, Shanghai 200092, China

Correspondence should be addressed to Haifeng Zhao; hfzhao@tongji.edu.cn

Received 31 January 2014; Revised 13 April 2014; Accepted 14 April 2014; Published 8 May 2014

Academic Editor: X. Zhang

Subway emergency may lead to passengers' panic, especially under self-organizing circumstance, which will spread rapidly and have an adverse impact on the society. This paper builds an improved SIRS model of passengers' panic spread in subway emergency with consideration of passengers' density, the characteristic of subway car with the confined space, and passengers' psychological factors. The spread of passengers' panic is simulated by use of Matlab, which draws the rules of how group panic spreads dynamically. The trend of stable point of the infection ratio is analyzed by changing different parameters, which help to draw a conclusion that immunization rate, spontaneous immune loss rate, and passenger number have a great influence on the final infected ratio. Finally, we propose an integrated control strategy and find the peak of passengers' panic and the final infected ratio is greatly improved through the numerical simulation. The research plays a vital role in helping the government and subway administration to master the panic spread mechanism and reduce the panic spread by improving measures and also provides certain reference significance for rail system construction, emergency contingency plans, and the construction and implementation of emergency response system.

1. Introduction

Since the first subway, named LMS (London metropolitan subway), was built in 1863, nowadays, there are more than 6000 kms lines in more than 100 cities of 35 countries and regions in the world that have been constructed. With the growing population and the increasing speed of urbanization, the subway is becoming the first choice for the public transportation of passengers to travel due to the advantage of the large volume, fast speed, convenience, and so forth. However, because of the large population density and the small enclosed space of the metro, the panic caused by emergency would be spread rapidly and result in the confusion, which would further magnify the impact of the entire event. In 2003, up to 25 million passengers were trapped in the subway because of a major blackout in London and caused panic, leading to great dissatisfaction. What is more, the spreading panic of passengers may cause a number of secondary accidents. For example, in the evening of September 2, 2013, Guangzhou Metro Line 2 subway suddenly braked and another one in the back caught up, the passengers in the back of carriage mistook it as a rear-end accident and fled toward the front of the cars with shouting, and some passengers were injured in the stampede. At noon of March 4, 2014, two youngsters played antiwolf spray in the back of Guangzhou Metro Line 5 and suddenly pungent odor was emitted, panicking passengers rushed to the front of the cars continuously and caused stampede, wounds, and luggage scattering, seriously disrupting social order and public transport. Even more, those incidents will further lead to people's psychological panic and it will not only affect the individual physical and mental health but also cause serious damage to the politics, economy, and social life. Therefore, the study of passengers' psychology and emotion in the emergency has got extensively concerned.

The related research on subway emergencies began in the 1990s, and the majority concentrated in the emergency was evacuation capability assessment [1, 2], emergency evacuation strategies [3], emergency location [4], and establishing emergency system [5, 6]. In addition, some scholars have established emergency evacuation models [7–10] to simulate subway emergencies scenarios. But most of these studies only

considered the safety of passenger; few scholars considered the psychological impact of the passengers caused by the subway emergencies. When subway emergencies occur, the normal social environment will be disrupted and inner tension will be expanded, and when the psychological tension reaches a certain level, it will cause group psychological panic, the inherent performance, which is panic emotion. Earlier studies suggested that people would lose essential humanity and fall into fear beast in the face of terrible disaster. Quarantelli [11] believed that panic is a collection of selfish behavior; when psychological panic occurs, people are more concerned with their own destiny rather than collective one. Le Bon [12] thought that people are impulsive and irrational and lack accountability, due to the factors such as anonymity, infection, and hint. The individual will lose rationality and responsibility, once he entered the masses, and then shows impulsive, brutal antisocial behavior. There are some researches on factors for panic; Mawson [13] pointed out that the panic comes from awareness and has relationship with social organization, culture, environment, situational factors, and social control. Aguirre [14] pointed out that the generation of panic is influenced by architecture structure, group members, group density, the relationship between the groups, the resource situation, and the amount of information. Panic can be described as the psychological panic as well as the panic behavior. Panic in behavioral performance is panic behavior. On the study of panic behavior, Kelley et al. [15] provided the simulation study about cluster behavior under the panic environment. Ebihara et al. [16] explored the behavior of individual panic. Saloma et al. [17] considered the existence of the self-organization queue behavior and freedom scale behavior. Low [18] assumed that groups are made up of different individuals with ideas and the ability, establishing a quantitative model to study the characteristics of irrational group behavior. He thought those group behaviors are generated because the widespread impact occurred under the situation of relatively spontaneous behavior and disorganized situation and it is dependent on stimulation of each participant. Helbing et al. [19] studied simulation dynamic characteristic of the panic to escape. As a special group behavior, the fugitive groups behavior in emergencies shows imitation, no purpose, spontaneity, vulnerability, and other nine characteristics. Because panic is reflected by the panic behavior, therefore, we can conclude that panic can be infectious.

Two of the most common and far-reaching models, SIS model [20] and SIR model [21], which were originally used for propagation mechanism of the virus, could be used and thoroughly researched for studying the infection and communication processes of crisis information [20–33]. Pastor-Satorras et al. [34–37] classified the complex nodes in the network according to their value and established the SIS model. Moreno et al. [38, 39] confirmed that there exist a certain number of infected nodes in the end, even if the initial infection is very low by applying the SIR model. Li et al. [40] believed that, in real life, there are some viruses that cannot be immune for all life and built a complex heterogeneous network SIRS epidemic model to a more realistic portrayal of the spread of infection. Based on this work, Zhao et al. [24, 26, 39] applied the epidemic model to the study of spread of rumors issues. Yuanyuan et al. [41] applied SIR epidemic model to the stock market crisis communication research. Scholars have also studied the problem for the specific context and the classic epidemic model has been improved. In the study of the propagation of the disease, taking into account the nonuniform interaction between nodes, Dybiec [23] extended the classical SIR model. Sekiguchi et al. [40, 42, 43] studied the distributed delay characteristics of infectious diseases in the model. Tchuenche et al. [44–47] believed that the total population is changing in real life due to the birth and death rates. Li et al. [48–51] thought that the infection rate and cure rate in the spread of disease are nonlinear. Zhao et al. [24] used infectious disease model to study the rumor spread issue. Considering the characteristics of rumors spread and social networks, they added forgetting factor in the model to describe the node spontaneous autoimmune conditions and then concluded that forgetting rate coefficient and immunization rates have a significant impact on the spread of rumors in the social network. Between the listed companies and the main stock holders, Yuanyuan et al. [41] established a susceptible-infected-removed model of crisis spreading (SIR) in the stock markets by taking the mutual influences into account which resulted from reduced cash flows or the fracture of capital chain. Then, a numerical simulation is used to analyze the crisis spreading in the correlated networks when the networks meet the random failure or the intentional attacks. These models did not consider the crowd density, dynamic infection rates, and some other conditions, whereas these conditions are very important in the study of the panic spreading in subway emergencies. This paper simulates the panic spread of the passengers under the subway emergency based on the epidemic models.

In summary, with the occurrence of unexpected events, people's ability to think will draw down. It will be more likely for them to accept the implied information, and they will be much more thirsty for information than usual. At this time, people's psychological emotions are in extreme tension and become panicked, and the most outstanding performance is the herd mentality of the individual. When some of the passengers appear confused in verbal expression, actions, or abnormal panic expression, panic will infect neighboring passengers. Then, the neighboring passengers will probably become panicked and then spread panic mood in the whole metro. When the emergency calms down, part of passengers may calm down, and then individuals become immune with a certain probability. However, immunity is not permanent; if their surroundings are still in a state of panic due to their own poor mentality, the group will become susceptible with certain probability, which makes it very similar to the propagation mechanism of SIRS model to a certain extent. Therefore, we chose SIRS model as an analysis model to study the spread of panic in the subway emergency. It is worth noting that, comparing with the classical SIRS model which describe the spread of viruses or rumors, the SIRS model describing the passengers also exist three corresponding state that are healthy state (S), infection status (I) and immune status (R), but this one is established based on the actual existence of exchanges

between the panic passengers in the subway. First of all, we need to verify whether it is peak period. Secondly, panic may spread from one of the cars to another in the narrow closed space, and therefore the probability to be infected or lose immunity between the passengers is not fixed. But the infection rate and the immunity loss rate are nonlinear functions rather than constants in the situation that panic spreads, and the immunization rate is related to the state of the emergency situation. Thirdly, the infected passengers are likely to be spontaneously immune because of their own mentality, and immune passengers are likely to spontaneously lose immunity and become susceptible groups due to the psychological factors of passengers. Besides, it should be emphasized that we only studied the self-organizing behavior without considering the participation of the government and the media.

It is important to grasp the mechanism of panic spread and improving measures to reduce the spread of panic, which will provide a reference for providing the rail system emergency plans and emergency information system construction and implementation. In this research, the study method on virus spreading is introduced to the subway emergencies to analyze the psychological and behavioral research, and then the SIRS model of emotional panic spread of subway passengers in emergencies is built, and the spreading process is analyzed by quantitative analysis and numerical simulation to reveal its spread rule and predict its spreading trends. The main contents are as follows. In Section 2, the subway passengers in the specific context in subway emergencies are classified; model assumptions are provided; then the model with improved parameters is established. Section 3 analyzes the model stability and Section 4 gives the simulation of the model. Finally, in Section 5, the findings of this study are summarized and the direction of future research is pointed out.

2. The Propagation Model of Panic in the Subway Emergency

2.1. The Definition of the Passenger State Node. In the event of the emergency, passengers in the cars may be in three kinds of states, S state, I state, and R state.

Susceptible state (S state) is as follows: the susceptible ones are comprised of individuals who are not in panic and are susceptible to become panicked, who are also called susceptible persons. The ratio of susceptible person is $S(t)$, indicating the ratio of passengers that have not been infected and remain calm at the time t.

Infection state (I state) is as follows: the infected ones are those who are in panic and spread the panic to others; panic can be caused by the emergency itself or the fact that they are infected by the surroundings. And the ratio of infected ones is $I(t)$, indicating the ratio of passengers that have become panicked with the ability to spread panic at the time t.

Immune state (R state) is as follows: the passengers who were affected by panic but later become patient and not afraid in the eased situation are known as the immune ones. And the ratio of immune ones is $R(t)$, indicating the ratio of passengers who are in immune state from the infection state at the time t. However, these passengers are also likely to become susceptible once again.

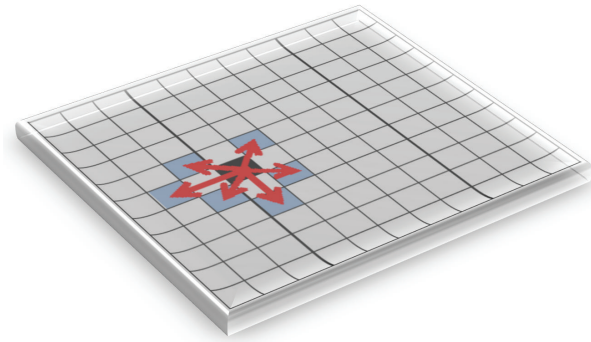


Figure 1: The individual contact model in dimensional lattice diagram.

2.2. The Model Assumptions

(1) The total passenger number in the subway is always maintained at a constant N.

(2) The passengers in the subway cars are uniformly distributed and the average degree of mutual contact between individuals is $\langle k \rangle$ during the normal driving which means at the off-peak or off-peak time.

(3) There are no birth or death issues in the process of panic spreading.

(4) Suppose that the probability of a susceptible passenger being infected by an infected one is constant, so the probability of an immune person losing immunity after contact with an infected one.

(5) At the initial period of the emergency, there are only the passengers who are in susceptible state and infection state but no immune state.

2.3. Improved Model Parameters

2.3.1. Population Density. The variable of passenger density is ρ, which may vary from time to time; the moving crowd can be simulated by using the two-dimensional regular lattices; we assume that the metro is an area of $L * W$ zone, where L is the length of the metro and W is the width of it. In the subway, the total number of this passenger group is N and the group makes random motion in the two-dimensional lattice. In the graph, each square represents an individual. The state to which he belongs is not taken into account and it just shows the mutual contact between the individual in the middle and the surrounding one in the graph. Besides, the number of the contacted individuals is $\langle k \rangle$, which equals 6, and it is the degree to the node. The simulated passenger contact model is shown in Figure 1.

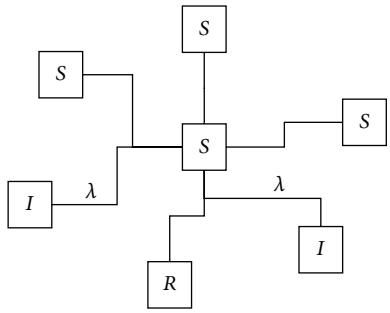


Figure 2: The contact rendering in unit space.

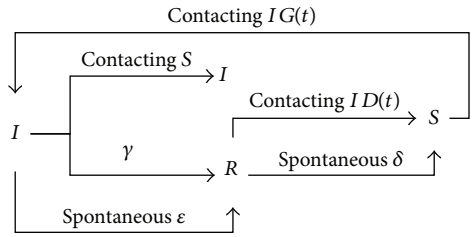


Figure 3: Structure of spreading process.

Therefore, the passenger density can be expressed as follows:

$$\rho = \frac{N}{L * W}. \tag{1}$$

2.3.2. The Infection Rate. As is known about the assumption about the same infection rate among individuals and the constant probability λ to converse to the panic state when the healthy person contacts the infected person and susceptible people contact an infected person, the probability of susceptible individuals infected is related to the number of infected individuals around the susceptible passengers. As the susceptible individual degree is $\langle k\rangle$, $\rho\langle k\rangle I(t)$ means the number of infected state individuals whom the susceptible individuals may contact at time t; $(1-\lambda)$ represents the no-infection probability after the healthy person contacts an infected person, and the number of the infections around the healthy is $\rho\langle k\rangle I(t)$, so the no-infection probability after the healthy person contacts all the infected persons around is $(1-\lambda)^{\rho\langle k\rangle I(t)}$ and the probability of being infected is $1-(1-\lambda)^{\rho\langle k\rangle I(t)}$; the contact rendering in unit space is shown in Figure 2.

As a result, we can get the infection rate of the susceptible individuals, which is as follows:

$$G(t) = \left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right]. \tag{2}$$

2.3.3. The Immunization Rate. Immune rate represents the recovery ability of the infected individual. As we only study self-organization without considering the impact on passengers panic by another organization, the immunization rate in the model is related to the calm-down speed and the dramatic degree of the emergency. As some emergencies are caused by the rumor from some passengers in the subway, so it can become immune state when finally the scared passengers discover the truth of events; we assume that this kind of immunization rate is γ. Besides, the immunization rate is related to the psychological state, educational level, age, and so forth, and the panic passengers can spontaneously become immune with the spontaneous immune probability of ε.

2.3.4. The Immune Loss Rate. Based on the assumption (4) in Section 2.2, the immune losing rate among individuals is the same when the immune contact with the infected person, we assume that this kind of immune loss rate is β. The probability of immune individual becoming susceptible is related to the number of infected individuals around. So the probability function of the immune loss rate of immune individuals is

$$D(t) = \left[1-(1-\beta)^{\rho\langle k\rangle I(t)}\right]. \tag{3}$$

At the same time, even when there is noninfected individual, there are still a handful of individuals becoming susceptible to the factors such as the psychological mentality, marking this state of immunization rate as spontaneous immunization rate δ.

2.3.5. The Model. This kind of spreading process is shown in Figure 3.

Based on the assumption and conditions, we established a subway emergencies propagation model which improves SIRS epidemic model, and it is shown as follows:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right]S(t) \\
&\quad + \left[1-(1-\beta)^{\rho\langle k\rangle I(t)}\right]R(t) + \delta R(t) \\
\frac{dI(t)}{dt} &= \left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right]S(t) - (\gamma+\varepsilon)I(t) \\
\frac{dR(t)}{dt} &= (\gamma+\varepsilon)I(t) - \left[1-(1-\beta)^{\rho\langle k\rangle I(t)}\right]R(t) - \delta R(t).
\end{aligned} \tag{4}$$

3. The Lyapunov Stability for the New Epidemic Model

The model represented above is established on the assumption that the population size N is a constant. The condition $S(t)+I(t)+R(t)=1$ can omit the equation of $R(t)$ by $S(t)$ and $R(t)$, so the two-dimensional system is given by

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right]S(t) \\
&\quad + \left[1-(1-\beta)^{\rho\langle k\rangle I(t)} + \delta\right](1-S(t)-I(t)) \\
\frac{dI(t)}{dt} &= \left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right]S(t) - \gamma I(t).
\end{aligned} \tag{5}$$

Let $dS(t)/dt = 0$ and $dI(t)/dt = 0$; then

$$\begin{aligned} &-\left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right] S(t) \\ &\quad+\left[1-(1-\beta)^{\rho\langle k\rangle I(t)}+\delta\right](1-S(t)-I(t))=0 \\ &\left[1-(1-\lambda)^{\rho\langle k\rangle I(t)}\right] S(t)-(\gamma+\varepsilon) I(t)=0. \end{aligned} \tag{6}$$

Theorem 1. *When $-\langle k\rangle\rho\ln(1-\lambda)/(\gamma+\varepsilon) < 1$, there is one and only disease-free equilibrium for Model (4) in the positive invariant set, and the equilibrium is $S(t)$, $I(t) = (1, 0)$.*

Proof. The Jacobian matrix J is as follows:

$$J=\begin{pmatrix} -\delta, & \langle k\rangle\,\rho\ln(1-\lambda) \\ 0, & -(\gamma+\varepsilon)-\langle k\rangle\,\rho\ln(1-\lambda) \end{pmatrix}. \tag{7}$$

When $|J| > 0$, it means the trace of matrix $\mathrm{tr}(J) = -(\gamma + \varepsilon) - \langle k\rangle\rho\ln(1-\lambda) < 0$, and it also means when $-\langle k\rangle\rho\ln(1-\lambda)/(\gamma+\varepsilon) < 1$, $(s(t), I(t)) = (1, 0)$ is stable. □

Theorem 2. *When $1 < -\langle k\rangle\rho\ln(1-\lambda)/(\gamma+\varepsilon) < 1/S^*$, there exists the local asymptotically stable equilibrium point (S^*, I^*).*

Assuming that $(S^*, I^*) \neq (1, 0)$ is an another positive equilibrium state of system (6), the objective is to prove if there exist one pair or more pairs (S^*, I^*) to ensure the existence of the solution when $-\langle k\rangle\rho\ln(1-\lambda)/(\gamma+\varepsilon) > 1$.

Let $G(I) = [1-(1-\lambda)^{\rho\langle k\rangle I(t)}]$ and $D(I) = [1-(1-\beta)^{\rho\langle k\rangle I(t)}]$ for calculating easily. So (6) can be shown as

$$\begin{aligned} -G(I)S+(D(I)+\delta)(1-S-I)&=0 \\ G(I)S-(\gamma+\varepsilon)I&=0. \end{aligned} \tag{8}$$

Substituting the point (S^*, I^*) in system (8), we can rewrite this system as

$$\begin{aligned} -G(I^*)S^*+(D(I^*)+\delta)(1-S^*-I^*)&=0 \\ G(I^*)S^*-(\gamma+\varepsilon)I^*&=0. \end{aligned} \tag{9}$$

We use the elimination method as

$$-(\gamma+\varepsilon)I^*+(D(I^*)+\delta)\left(1-\frac{(\gamma+\varepsilon)I^*}{G(I^*)}-I^*\right)=0. \tag{10}$$

Let $H(I) = -(\gamma+\varepsilon)I + (D(I)+\delta)(1-((\gamma+\varepsilon)I/G(I)) - I)$, where $f(0) = g(0) = 0$, $f'(0) > 0$, $g'(0) > 0$, $f''(0) < 0$, and $g''(0) < 0$; then

$$H(0)=\delta\left(1-\frac{\gamma+\varepsilon}{G'(0)}\right)>0$$

$$H(1)=-(\gamma+\varepsilon)-\frac{\gamma+\varepsilon}{G(1)}(D(1)+\delta)<0$$

$$\begin{aligned} H'(I) &= -(\gamma+\varepsilon)+D'(I)\left(1-\frac{(\gamma+\varepsilon)I}{G(I)}-I\right) \\ &\quad+(D(I)+\delta)\left(1-I-\frac{(\gamma+\varepsilon)\left[G(I)-IG'(I)\right]}{G^2(I)}\right)<0. \end{aligned} \tag{11}$$

So a point makes the system stable exist, and the point is not equal to zero. The Jacobian matrix of (8) is

$$J=\begin{pmatrix} -G(I^*)-D(I^*)-\delta & -G'(I^*)S^*-D'(I^*)(1-S^*-I^*)-D(I^*)-\delta \\ G(I^*) & G'(I^*)S^*-(\gamma+\varepsilon) \end{pmatrix}. \tag{12}$$

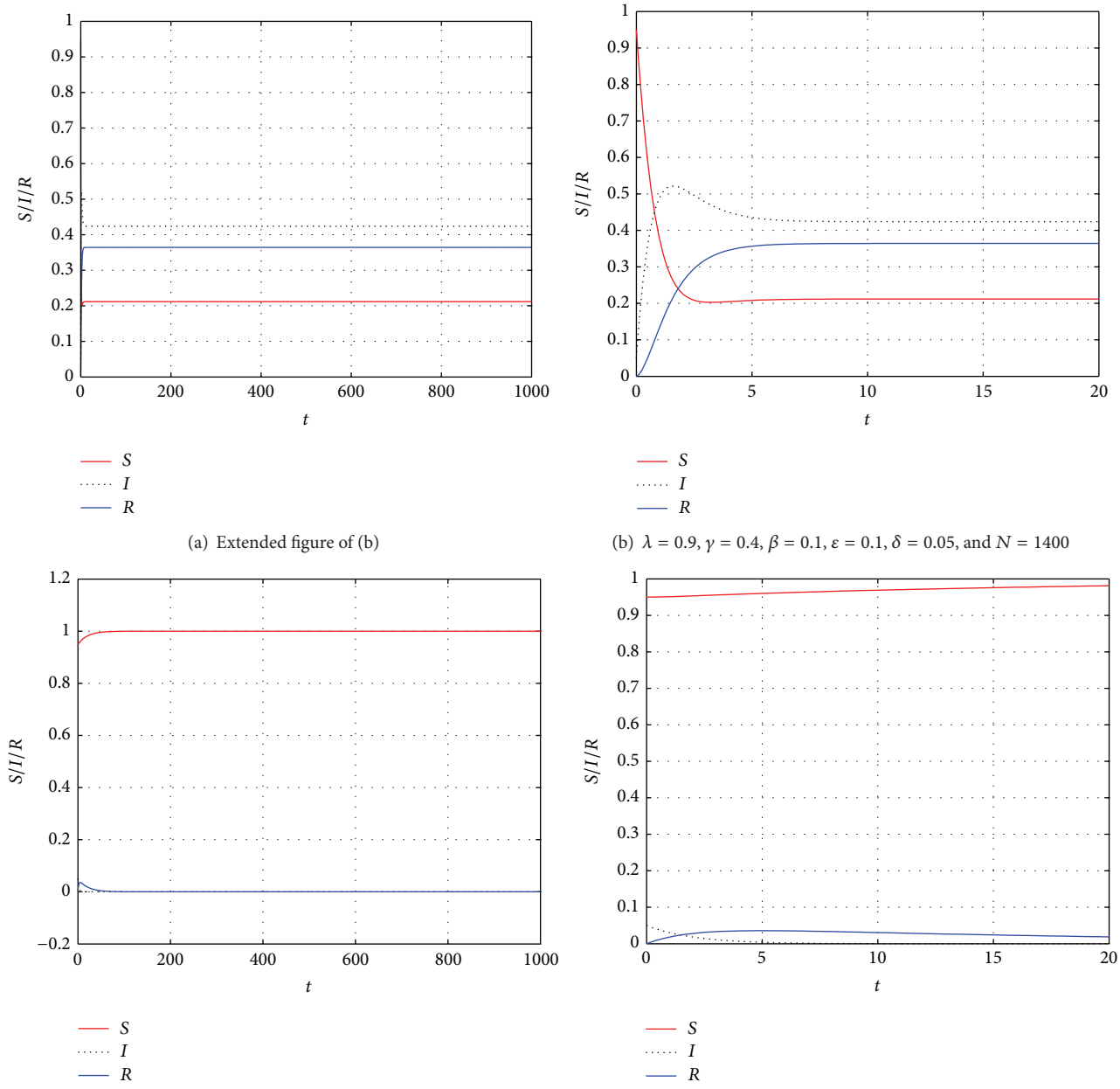
When the trace of matrix $G(I^*) + D(I^*) + \delta - G'(I^*)S^* + \gamma + \varepsilon > 0$, it can be simplified as $G'(I^*)S^* < \gamma + \varepsilon$.

$G'(I)$ is decreasing, and $G'(I) > 0$, so if it satisfied the condition $G'(0)S^* < \gamma + \varepsilon$, that is, $-\langle k\rangle\rho\ln(1-\lambda)/(\gamma+\varepsilon) < 1/S^*$, the system (6) can achieve a globally asymptotically stable, and, at this time, the infected people will not gone, but the proportion of it can asymptotically stable as a constant (S^*, I^*).

4. Numerical Simulation

4.1. Dynamic Simulation of Panic Spreading. After actual survey, we make numerical simulation of the panic spread model built in Section 2 and set model parameters as number of passengers $N = 1400$, length and width of the subway car $L = 164.78$ meters and $W = 3$ meters, the average degree of mutual contact between individuals $\langle k\rangle = 6$, and initial proportion of the susceptible and the infected $S(t) = 0.95$, $I(t) = 0.05$. Given the confined space in the subway car, panic will spread quickly once emergency happens, so we set the infected rate a higher value $\lambda = 0.9$. As the immunization rate is related to the development, γ is set to be 0.4. The immune loss rate is the rate of passengers who become susceptible again, influenced by the infected passengers around, so we set $\beta = 0.1$. Meanwhile, the probability of the infected persons turning to be immune by self-mentality is generally higher than that of the immune persons turning to be susceptible, and we set $\varepsilon = 0.1$ and $\delta = 0.05$. By solving the model equations using the ODE45 arithmetic of Matlab, we change the proportion of the susceptible, the infected, and the recovered persons in the spread process of panic, as shown in Figures 4(a) and 4(b).

According Figure 4(b), we can find out that, with the rapid spread of panic in the subway car, the proportion of the infected persons increases quickly from the initial 5% to the maximum value 52.14% with 730 infected passengers



(a) Extended figure of (b)

(b) $\lambda = 0.9$, $\gamma = 0.4$, $\beta = 0.1$, $\varepsilon = 0.1$, $\delta = 0.05$, and $N = 1400$

(c) Extended figure of (b)

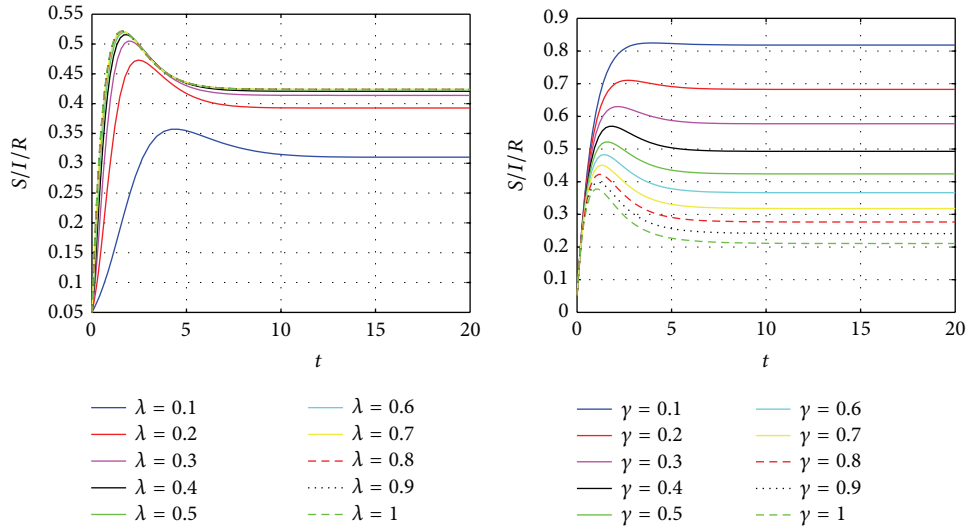
(d) $\lambda = 0.000001$, $\gamma = 0.4$, $\beta = 0.1$, $\varepsilon = 0.1$, $\delta = 0.05$, and $N = 1400$

Figure 4: The dynamic variation of S, I, and R (the figure on the right is the top 20 steps of the left).

when $t = 1.502$ and reaches steady state on 42.38% with 593 infected passengers when $t = 12.39$. With time step going on, the proportion of the susceptible persons $S(t)$ descends quickly while the proportion of the recovered persons$R(t)$ increases. And the trends of both curves of $S(t)$ and $R(t)$ finally reach steady state with $S^* = 0.2119$. All the model parameters meet $1 < -\langle k\rangle \rho \ln(1-\lambda)/(\gamma+\varepsilon) < 1/S^*$ and Theorem 2 is proved.

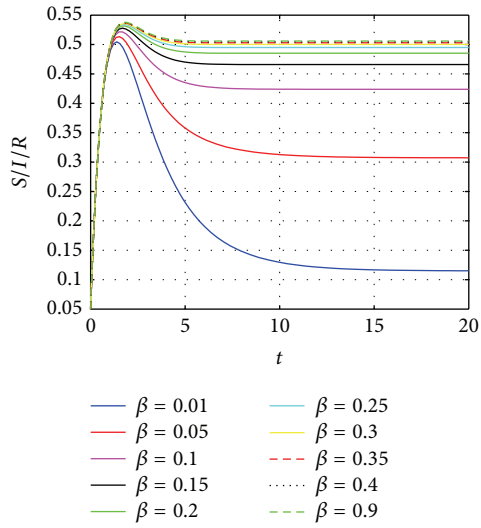
In order to verify Theorem 1, we randomly assign model parameters as $\lambda = 0.000001$, $\gamma = 0.4$, $\beta = 0.1$, $\varepsilon = 0.1$, and $\delta = 0.05$ to meet $-\langle k\rangle \rho \ln(1-\lambda)/(\gamma+\varepsilon) < 1$. Simulation results are showed in Figures 4(c) and 4(d). The model finally reaches its stable point at $(S, I) = (1, 0)$ and Theorem 1 is proved.

4.2. The Impact of Different Parameters on Panic. This section will use control variate method to analyze how model parameters influence the number of the infected people by separately changing infection rate, immunization rate, immune loss rate, and passenger number on Matlab simulation, based on the model parameters setting of Figures 4(a) and 4(b). The simulation results are shown in Figure 5.

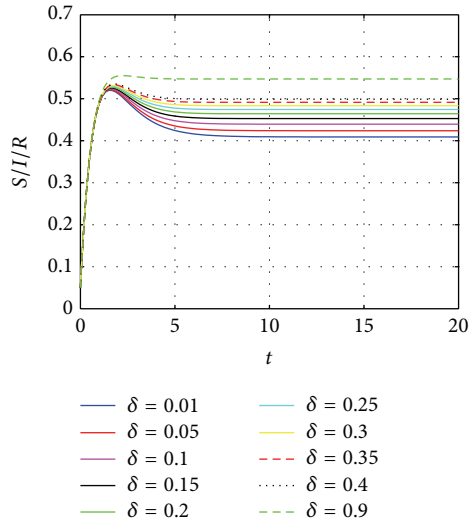


(a) Changing infection rate

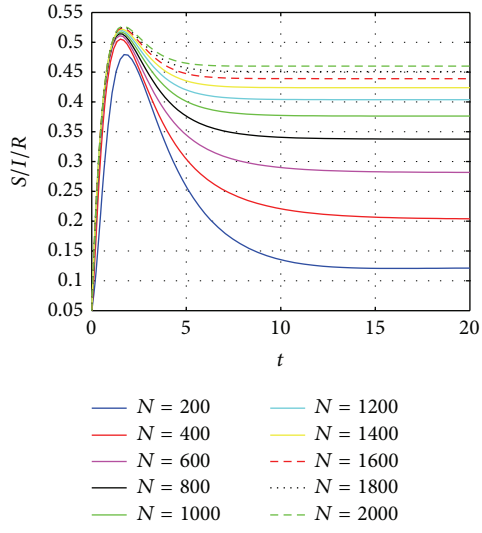
(b) Changing immunization rate and spontaneous immunization rate



(c) Changing immune loss rate

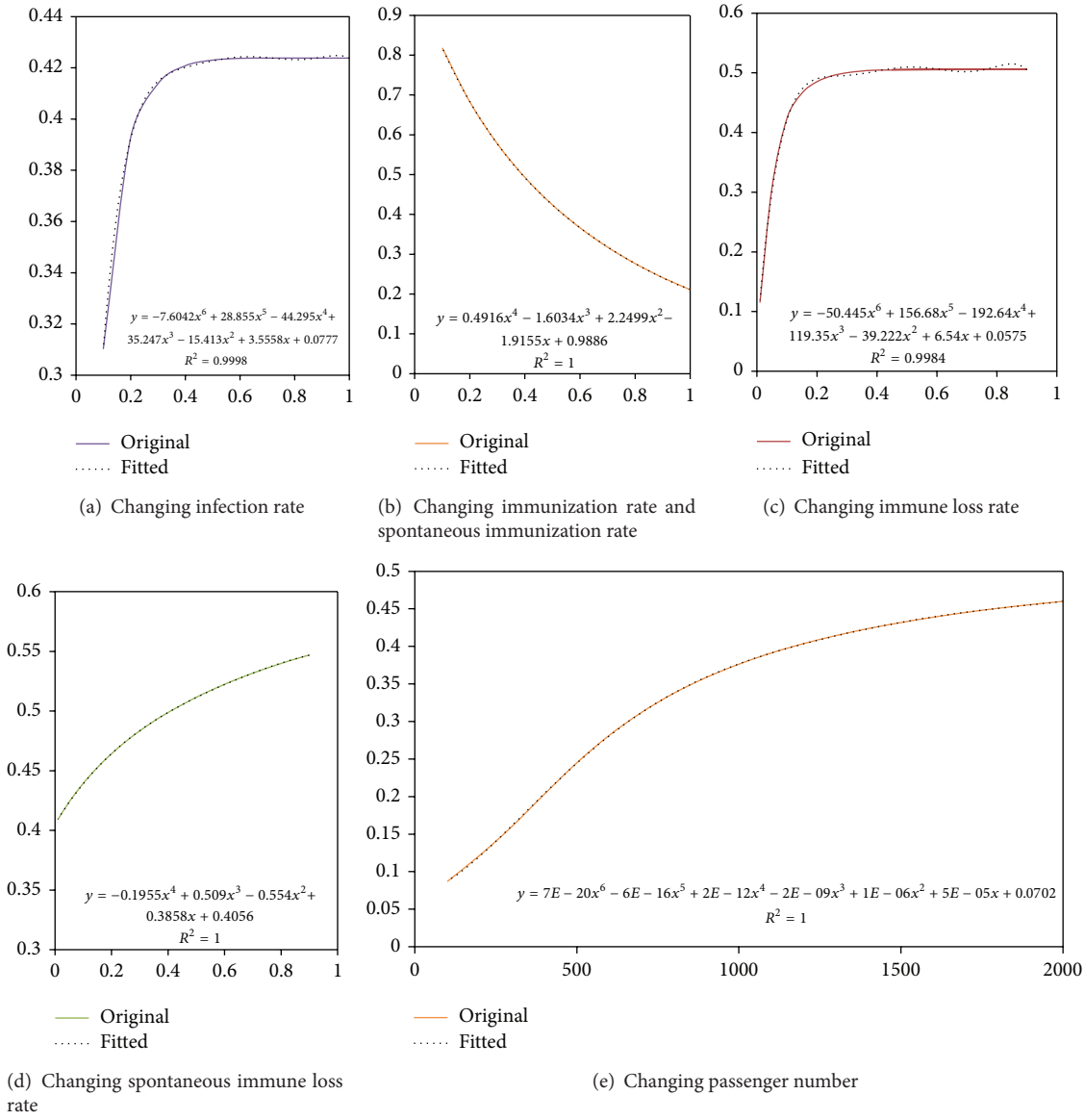


(d) Changing spontaneous immune loss rate



(e) Changing passenger number

Figure 5: Influence of the infected passengers number by different parameters.



(a) Changing infection rate

(b) Changing immunization rate and spontaneous immunization rate

(c) Changing immune loss rate

(d) Changing spontaneous immune loss rate

(e) Changing passenger number

Figure 6: Changing curves of the stable point by different parameters.

According to Figure 5, we can figure out that the changes of some parameters make the number of the infected passengers in steady state change greatly while the others change little. In order to reflect the relationships between parameters range and stable point, we draw the relationships in curves and fit them into equations, as shown in Figure 6.

In conclusion, according to Figures 5 and 6, we can figure that the infected proportion have a great influence on the final panic passengers number when the infected rate rises to 0.3 or the immunization rate goes to 0.2, and after that the curve becomes gentle. Figure 6(d) shows that changing the spontaneous immune loss rate does not have so many effects on the final number of panic passengers; the proportion of infected people will be changed from 0.4 to 0.55. However, if the immunization rates and spontaneous immunization rates are in a state of low value, passengers' panic spreads quickly and cannot be restored, which will make the proportion of infected passengers in a high dangerous state which is shown in Figure 6(b). Changing the total number of passengers, in other words, is changing passenger density; in this study, we set the maximum available vehicle capacity of the subway to 2000 persons according to statistics. As shown in Figure 6(e), it is obvious that the more the number of passengers in the car, the greater the proportion of panic at last.

4.3. Simulation of the Effect of Comprehensive Control Strategy on Panic Spreading. Table 1 displayed the impact of final infection rate and the amount of infected people by, respectively, changing multiple model parameters, because we have a lot of dates, so we cut out only intercept part of them. As you can see, certain priority relation between the parameters exists, such as 1st–10th, 11th–20th, and 21th–26th lines of

Table 1: The impact of final infection rate and the amount of infected people by respectively changing multiple model parameters.

No.	λ	$\delta + \varepsilon$	β	δ	N	Stable point	Infected amount
1	0.9	0.5	0.1	0.05	1400	0.4238	593
2	0.8	0.5	0.1	0.05	1400	0.4238	593
3	0.9	0.6	0.1	0.05	1400	0.3663	513
4	0.8	0.6	0.1	0.05	1400	0.3663	513
5	0.9	0.5	0.05	0.05	1400	0.3073	430
6	0.8	0.5	0.05	0.05	1400	0.3095	433
7	0.9	0.5	0.1	0.03	1400	0.4167	583
8	0.8	0.5	0.1	0.03	1400	0.4167	583
9	0.9	0.5	0.1	0.05	1000	0.3764	376
10	0.8	0.5	0.1	0.05	1000	0.3762	376
11	0.9	0.6	0.05	0.05	1400	0.2456	344
12	0.8	0.6	0.05	0.05	1400	0.2454	344
13	0.9	0.6	0.1	0.03	1400	0.3581	501
14	0.8	0.6	0.1	0.03	1400	0.358	501
15	0.9	0.6	0.1	0.05	1000	0.3154	315
16	0.8	0.6	0.1	0.05	1000	0.3152	315
17	0.9	0.5	0.05	0.05	1000	0.2418	242
18	0.8	0.5	0.05	0.05	1000	0.2415	242
19	0.9	0.5	0.1	0.03	1000	0.3647	365
20	0.8	0.5	0.1	0.03	1000	0.3646	365
21	0.9	0.6	0.05	0.03	1400	0.2211	310
22	0.8	0.6	0.05	0.03	1400	0.2209	309
23	0.9	0.6	0.1	0.03	1000	0.3017	302
24	0.8	0.6	0.1	0.03	1000	0.3016	302
25	0.9	0.5	0.05	0.03	1000	0.2085	209
26	0.8	0.5	0.05	0.03	1000	0.2078	208
27	0.9	0.6	0.05	0.03	1000	0.1512	151
28	0.8	0.6	0.05	0.03	1000	0.1499	150

comparison which shows that the immune loss rate is the most significant impact on the proportion of final infected proportion. In addition, comparing lines 1, 2, 4, 12, 22, and 28 to each other, we conclude that the passenger density also has great influence on the infection proportion. However, the infection rate almost had no effects on the proportion of infected people between 0.3 and 0.9.

The reduction of the degree of passengers' panic depends on the measures that the subway departments take. The safety management of the subway should be strengthened, national educational activities for subway emergencies should be carried out, and the passengers' number in the subway cars should be controlled strictly. Figure 7 is the propagation curve simulation with the parameters $\lambda = 0.8$, $\gamma = 0.6$, $\beta = 0.01$, $\varepsilon = 0.01$, $\delta = 0.05$, and $N = 1000$.

The final numbers of the infected and the immune passengers have an obviously big difference by comparing Figure 7 with Figure 4(b). When the infected passengers' number in Figure 4(b) is 686, it reaches the stable state. The final infected passengers' proportion is 8.56% in Figure 7. They use different group numbers. While the former uses $N = 1400$, the latter uses $N = 1000$, so the number of infected passengers is about 86. Without the control strategies, the panic peak reaches 57% and the number of panic passengers is 798. By the use of the control strategies, the panic peak reaches 46.66% and the number of panic passengers is 467, which suggests that the control strategies have greatly improved the panic peak and the final number of panic passengers.

5. Conclusions

This paper takes the spread characteristic of passengers panic under subway emergency into consideration and improves traditional SIRS model and parameters.

(1) The parameter, passengers' density ρ, which may change significantly in different time, is added in the model.

(2) In the subway car with strait and confined space, when the panic happens, it is probably not integrated panic, but the panic that spreads from the emergency car to another, which means that the distribution of panic passengers is not homogeneous. The infection rate and immune loss rate between passengers are determined by the surroundings. Therefore, it is more

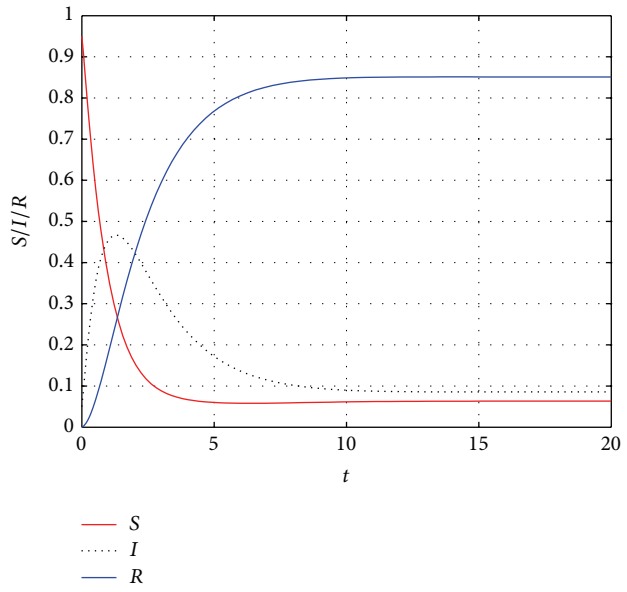


Figure 7: Result of integrated control strategies.

conformed to describe the infection rate and immune loss rate as

$$G(t) = \left[1 - (1-\lambda)^{\rho\langle k\rangle I(t)}\right],$$
$$D(t) = \left[1 - (1-\beta)^{\rho\langle k\rangle I(t)}\right]. \quad (13)$$

(3) The influence of passenger psychological factors is needed to be considered, because the infected passengers possibly become spontaneously immune by their own psychological mentality while the immune passengers possibly become susceptible again by their own spontaneous immune loss characteristic. Therefore, the spontaneous immune parameters ε and δ are added to the model to represent the spontaneous immune probability of the infected passengers and spontaneous immune loss probability of the immune passengers.

According to the three aspects above, the SIRS model of panic spread of passengers under subway emergency is built and is to be used to simulate the panic spread of the passengers, which reveals the rules of how group panic dynamic spread and verified the model stability. The trend of stable point of the infection rate is analyzed by changing different parameters and comes to a conclusion that immunization rate, spontaneous immune loss rate, and passengers' number had a great influence on the final infected passengers' number, rapidly reducing the effect of panic spread. Finally, this paper proposed integrated control strategies to strengthen the safety management of the subway, carry out national educational activities for subway emergencies, and strictly control the passengers' number in the subway car and made simulation to find that the passenger panic peak and the final infected passenger number were greatly improved. Currently, the model and conclusions are established under self-organizing subway emergencies. However, there are more factors that need to be considered, such as government or subway administration and some control strategies made by other organizations, which should be taken into account and it deserves more attention in the future research.

Conflict of Interests

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

Acknowledgments

This research was supported by the NSFC Major Program (Grants nos. 71090404/71090400) and NSFC General Program (Grants nos. 71272045 and 71101107).

References

[1] S. L. Zhang, M. L. Liu, and E. M. Sun, "Comprehensive assessment on emergency evacuation capability of the subway station based on fuzzy network analysis," *Applied Mechanics and Materials*, vol. 357, pp. 2935–2939, 2013.

[2] T. Okumura, K. Suzuki, A. Fukuda et al., "The Tokyo subway sarin attack: disaster management, part 1: community emergency response," *Academic Emergency Medicine*, vol. 5, no. 6, pp. 613–617, 1998.

[3] L. Yan, W. Tong, Z. Wei, and W. Zongzhi, "Computer simulation of human emergency evacuation in subway station," in *Proceedings of the 2nd International Conference on Risk Analysis and Crisis Response*, pp. 513–519, 2009.

[4] Z. Zou and H. Guo, "Research on a fuzzy neural network based medical institutions selection in subway station emergency," *Advanced Materials Research*, vol. 361–363, pp. 1204–1210, 2012.

[5] P. Li, R. Lu, and F. Wang, "Study on supervision, alarm and emergency system for subway in big city," in *Proceedings of the 2nd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS '11)*, pp. 286–289, August 2011.

[6] L. Hu, L. Wu, K. Lu et al., "Optimization of emergency ventilation mode for a train on fire stopping beside platform of a metro station," *Building Simulation*, vol. 7, no. 2, pp. 137–146, 2014.

[7] C. S. Jiang, Y. Ling, C. Hu, Z. Yang, H. Ding, and W. K. Chow, "Numerical simulation of emergency evacuation of a subway station: a case study in Beijing," *Architectural Science Review*, vol. 52, no. 3, pp. 183–193, 2009.

[8] P. Yang, C. Li, and D. Chen, "Fire emergency evacuation simulation based on integrated fire-evacuation model with discrete design method," *Advances in Engineering Software*, vol. 65, pp. 101–111, 2013.

[9] J. Wan, J. Sui, and H. Yu, "Research on evacuation in the subway station in China based on the Combined Social Force Model," *Physica A: Statistical Mechanics and Its Applications*, vol. 394, pp. 33–46, 2014.

[10] W. Lei, A. Li, R. Gao et al., "Simulation of pedestrian crowds' evacuation in a huge transit terminal subway station," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 22, pp. 5355–5365, 2012.

[11] E. Quarantelli, "The behavior of panic participants," *Sociology and Social Research*, vol. 41, no. 3, pp. 187–194, 1957.

[12] G. Le Bon, *The Crowd: A Study of the Popular Mind*, Macmillan, 1897.

[13] A. R. Mawson, "Is the concept of panic useful for study purposes," in *Proceedings of the 2nd International Seminar on Behavior in Fire Emergencies*, p. 29, National Bureau of Standards, Special Publication, 1978.

[14] B. E. Aguirre, "Emergency evacuations, panic, and social psychology," *Psychiatry*, vol. 68, no. 2, pp. 121–129, 2005.

[15] H. H. Kelley, J. C. Condry Jr., A. E. Dahlke, and A. H. Hill, "Collective behavior in a simulated panic situation," *Journal of Experimental Social Psychology*, vol. 1, no. 1, pp. 20–54, 1965.

[16] M. Ebihara, A. Ohtsuki, and H. Iwaki, "A model for simulating human behavior during emergency evacuation based on classificatory reasoning and certainty value handling," *Computer-Aided Civil and Infrastructure Engineering*, vol. 7, no. 1, pp. 63–71, 1992.

[17] C. Saloma, G. J. Perez, G. Tapang, M. Lim, and C. Palmes-Saloma, "Self-organized queuing and scale-free behavior in real escape panic," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 11947–11952, 2003.

[18] D. J. Low, "Following the crowd," *Nature*, vol. 407, no. 6803, pp. 465–466, 2000.

[19] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.

[20] M. Kermack and A. Mckendrick, "Contributions to the mathematical theory of epidemics: part II," *Proceedings of the Royal Society A*, vol. 138, pp. 55–83, 1932.

[21] M. Kermack and A. Mckendrick, "Contributions to the mathematical theory of epidemics: part I," *Proceedings of the Royal Society A*, vol. 115, pp. 700–721, 1927.

[22] C. H. Li, C. C. Tsai, and S. Y. Yang, "Analysis of epidemic spreading of an SIRS model in complex heterogeneous networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 19, no. 4, pp. 1042–1054, 2014.

[23] B. Dybiec, "SIR model of epidemic spread with accumulated exposure," *European Physical Journal B*, vol. 67, no. 3, pp. 377–383, 2009.

[24] L. Zhao, Q. Wang, J. Cheng, Y. Chen, J. Wang, and W. Huang, "Rumor spreading model with consideration of forgetting mechanism: a case of online blogging Live Journal," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 13, pp. 2619–2625, 2011.

[25] L. Zhao, J. Wang, Y. Chen, Q. Wang, J. Cheng, and H. Cui, "SIHR rumor spreading model in social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 7, pp. 2444–2453, 2012.

[26] L. J. Zhao, H. X. Cui, X. Y. Qiu, and X. L. Wang, "SIR rumor spreading model in the new media age," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 4, pp. 995–1003, 2013.

[27] F. R. Durand, M. L. F. Abbade, E. Moschim et al., "SIR optimization in wavelength-hopping time spreading optical code routed networks," *Optik*, vol. 124, no. 18, pp. 3208–3214, 2013.

[28] Y. Maki and H. Hirose, "Infectious disease spread analysis using stochastic differential equations for SIR model," in *Proceedings of the 4th IEEE International Conference on Intelligent Systems Modelling & Simulation (ISMS '13)*, pp. 152–156, 2013.

[29] D. F. Bernardes, M. Latapy, and F. Tarissan, "Relevance of SIR model for real-world spreading phenomena: experiments on a large-scale p2p system," in *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM '12)*, pp. 327–334, IEEE Computer Society, 2012.

[30] S. Bargmann and P. M. Jordan, "A second-sound based, hyperbolic SIR model for high-diffusivity spread," *Physics Letters A: General, Atomic and Solid State Physics*, vol. 375, no. 5, pp. 898–907, 2011.

[31] S.-K. Zhang, S.-R. Gong, and Z.-M. Cui, "Study on spreading of virus infection with SIRS characteristic in wireless sensor networks," in *Proceedings of the WRI International Conference on Communications and Mobile Computing (CMC '09)*, vol. 3, pp. 512–517, January 2009.

[32] H. Fukś, A. T. Lawniczak, and R. Duchesne, "Effects of population mixing on the spread of SIR epidemics," *European Physical Journal B*, vol. 50, no. 1-2, pp. 209–214, 2006.

[33] H. W. Hethcote, "Mathematics of infectious diseases," *SIAM Review*, vol. 42, no. 4, pp. 599–653, 2000.

[34] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.

[35] A. Ramani, A. S. Carstea, R. Willox, and B. Grammaticos, "Oscillating epidemics: a discrete-time model," *Physica A: Statistical Mechanics and Its Applications*, vol. 333, no. 1–4, pp. 278–292, 2004.

[36] C. Piccardi and R. Casagrandi, "Inefficient epidemic spreading in scale-free networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 77, no. 2, Article ID 026113, 2008.

[37] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics and endemic states in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 63, no. 6, Article ID 066117, 2001.

[38] Y. Moreno, J. B. Gómez, and A. F. Pacheco, "Epidemic incidence in correlated complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 3, Article ID 035103, 2003.

[39] Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of rumor spreading in complex networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 69, no. 6, Article ID 066130, 2004.

[40] C.-H. Li, C.-C. Tsai, and S.-Y. Yang, "Analysis of the permanence of an SIR epidemic model with logistic process and distributed time delay," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 9, pp. 3696–3707, 2012.

[41] M. Yuanyuan, Z. Xintian, and L. Linguan, "Susceptible-Infected-Removed (SIR) model of crisis spreading in the correlated network of listed companies and their main stockholders.," *Journal of Management Sciences in China*, vol. 16, no. 7, pp. 80–94, 2013.

[42] M. Sekiguchi and E. Ishiwata, "Global dynamics of a discretized SIRS epidemic model with time delay," *Journal of Mathematical Analysis and Applications*, vol. 371, no. 1, pp. 195–202, 2010.

[43] J. M. Tchuenche, A. Nwagwo, and R. Levins, "Global behaviour of an SIR epidemic model with time delay," *Mathematical Methods in the Applied Sciences*, vol. 30, no. 6, pp. 733–749, 2007.

[44] J. M. Tchuenche and C. Chiyaka, "Global dynamics of a time delayed SIR model with varying population size," *Dynamical Systems*, vol. 27, no. 2, pp. 145–160, 2012.

[45] A. D'Onofrio, "On pulse vaccination strategy in the SIR epidemic model with vertical transmission," *Applied Mathematics Letters*, vol. 18, no. 7, pp. 729–732, 2005.

[46] S. Gao, D. Xie, and L. Chen, "Pulse vaccination strategy in a delayed SIR epidemic model with vertical transmission," *Discrete and Continuous Dynamical Systems B*, vol. 7, no. 1, pp. 77–86, 2007.

[47] Z. Lu, X. Chi, and L. Chen, "The effect of constant and pulse vaccination on SIR epidemic model with horizontal and vertical transmission," *Mathematical and Computer Modelling*, vol. 36, no. 9-10, pp. 1039–1057, 2002.

[48] J. Li and N. Cui, "Dynamic behavior for an SIRS model with nonlinear incidence rate and treatment," *The Scientific World Journal*, vol. 2013, Article ID 209256, 5 pages, 2013.

[49] Y. Lin and D. Jiang, "Dynamics of a multigroup SIR epidemic model with nonlinear incidence and stochastic perturbation," *Abstract and Applied Analysis*, vol. 2013, Article ID 917389, 12 pages, 2013.

[50] Z. Hu, W. Ma, and S. Ruan, "Analysis of SIR epidemic models with nonlinear incidence rate and treatment," *Mathematical Biosciences*, vol. 238, no. 1, pp. 12–20, 2012.

[51] J. Li, N. Cui, and H. Sun, "Dynamic behavior for an SIRS model with nonlinear incidence rate," *Advanced Materials Research*, vol. 479–481, pp. 1495–1498, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 845752, 11 pages
http://dx.doi.org/10.1155/2014/845752

Research Article

Berth Allocation Problem with Quay Crane Assignment for Container Terminals Based on Rolling-Horizon Strategy

Ling Xiao[1] and Zhi-Hua Hu[2]

[1] *Logistics Research Center, Shanghai Maritime University, Shanghai 200135, China*
[2] *School of Economics and Management, Tongji University, Shanghai 200092, China*

Correspondence should be addressed to Zhi-Hua Hu; zhhu@shmtu.edu.cn

Received 13 January 2014; Accepted 27 March 2014; Published 7 May 2014

Academic Editor: Hu Shao

In order to solve the large-scale integral dynamic scheduling of continuous berths and quay cranes problem, a method based on rolling-horizon strategy is proposed. A multiobjective optimization model that is established minimizes the total penalty costs considering vessels' deviations to their preferred berthing positions, delayed times for berthing comparing to their estimated arrival times, and delayed times for departure comparing to their estimated departure times. Then, the scheduling process was divided into a set of continual scheduling interval according to the dynamic arrival sequences. Meanwhile, rolling-horizon strategies for setting rolling and frozen windows and the parameter updating strategy are designed. The input parameters of the model in the next rolling window are updated according to the optimal results of each time window which have been obtained. The model is solved by choosing appropriate rolling and freezing window lengths that represents the numbers of adjacent vessels in the sequence of calling vessels. The holistic optimal solution is obtained by gradually rolling and combining the results of each window. Finally, a case study indicated that the rolling schedule can solve large-scale scheduling problems, and the efficiency of the proposed approach relates to the size of rolling window, freeze ship quantity, and rolling frequency.

1. Introduction

Along with the rapid development of large-scale container terminals and large container ships, the container handling efficiency plays an increasingly important role in competition environment of container terminals, which increases the difficulty and complexity of large-scale integral dynamic scheduling of berths and quay cranes. Berths and quay cranes are critical resources in container terminals. So, the berth allocation problem (BAP) and quay crane assignment problem (QCAP) are important issues which improve the whole efficiency of a container terminal. A reasonable arrangement of berth and quay cranes reduces the stay time of vessels at berths as well as the operational cost and improves the turnover rate of vessels at then container terminal and profit and service ability of the container terminal. For most large-scale container terminals in China, many methods mentioned in the literature cannot be applied in practice because of unacceptable computational performance, even for solving a 24-hour berth schedule. While rolling-horizon strategy is widely used in real operational plan of container terminals, especially suitable for dynamic continuous berth allocation problem (DCBAP) which conforms well with real situations. Rolling-horizon scheduling method is a decomposition strategy for large-scale problem, which means to solve original problem by decomposing it into small subproblems. It is an effective method to solve large-scale complex problems [1–3]. Berthing plan is a kind of multiprocessor task scheduling [4, 5], so the rolling strategy can be used for reference.

Compared with the existing literature, including that of [6], which formulated the continuous CBAP, the main contributions of this paper are elucidated as follows. First, this paper formulated two constraints including the adjustment of the amount of quay cranes allocated to a vessel and the limitation of total number. A mixed-integer multiobjective optimization model was established which minimizes the total penalty cost that consists of deviations of the solution to the preferred berthing position and estimated berthing and

departure times. Then, the scheduling process was divided into a set of continual scheduling intervals according to the vessels' arrival sequences. Third, this paper developed the rolling-horizon strategies for setting rolling and frozen windows and updating methods for the present window based on past windows. Through large-scale numerical case study, it is indicated that the model and rolling-horizon strategy proposed in our paper can solve large-scale scheduling problems and that the relationship between the efficiency and the amount of vessels is almost linear; it can be applied to decision support system for DCBAP large-scale container terminals.

2. Related Studies

BAP and QCAP are two important issues in operations optimization of container terminals. According to whether mooring points are continuous, BAPs can be divided into discrete BAP (DBAP) and continuous BAP (CBAP) [7]. In DBAP, the available berths are regarded as a limited set of mooring points. Vessels cannot berth across more than one berth [8], while CBAP takes berths as a continuous mooring space. Vessels can moor and be handled anywhere in this space, not subjected to the limitations of berthing positions [9]. According to the vessel's arrival time, BAPs can be classified into static BAP and dynamic BAP. Static BAP means all the vessels have already arrived in port before berth allocation [10]. While in Dynamic BAP, some vessels are yet in terminal at the beginning of the berth allocation, but others will arrive at a certain time during the allocation [11]. The dynamic arrival sequence of the vessels provides the foundation for the rolling scheduling in this work.

According to the literature, the CBAP can be divided into two categories. The first is to consider BAP and QCAP as two independent stages. In berth allocation, the duration of a vessel is estimated by its capacity and distance between the actual berthing position and the preference berthing position. Then, the berthing position, berthing time, and departure time can be obtained. However, It may cause problems when BAP and QCAP are studied independently. For instance, if the container terminal is busy, the limited quay cranes cannot meet the demand of operations and it may result in increasing waiting times of vessels at the terminal, while if the container terminal is idle, it may result in wasting resources [5, 10, 12, 13]. Considering that the operation time of a vessel depends on the amount of cranes assigned to it, the second method integrated scheduling raised a lot of attention. In integrated scheduling, the amount of quay cranes is taken into consideration in BAP. Imai et al. [13] established a model for DBAP that optimized the berth allocation and quay cranes assignment with the consideration of path optimization of quay cranes. Legato et al. [14] designed a model in stochastic dynamic environment with the minimization of vessel's operation time and quay cranes. Though many models and algorithms are developed for BAPs, the quay-crane constraints (the amount of quay cranes allocated to a vessel determines the vessel's handling time and that the total number of quay cranes is limited) are simplified for computing performance problem. For instance, Zhen et al. [15] conducted a scenario analysis that aimed at the uncertainty of berth allocation by heuristic method. An experiment which has a scale of 40 vessels was performed, while the effect of the quantity of cranes on the operating time was not considered. Sammarra et al. [16] solved a QCAP with Tabu search algorithm and also ignored this effect. To our knowledge, few literature studied large-scale scheduling problems (e.g., over 80 vessels) that considered the two quay-crane constraints.

The large-scale problems of dynamic CBAP with quay assignment are difficult to be solved in acceptable computation times. Previous researches used heuristic algorithms, for example, Tabu search and simulated annealing algorithm, to solve the problems [14]. Exact algorithms are generally used to solve small-scale problem [16]. Rolling-horizon scheduling has been widely used in manufacture, and the main principle is to solve a series of small optimization problems instead of solving large-scale scheduling problem [17, 18]. Raa et al. [19] solved the integral scheduling of dynamic DBAP and QCAP by rolling-horizon scheduling, and the length of rolling window was optimized by a hybrid heuristic algorithm. Chang et al. [20] considered a CBAP from the perspective of energy consumption and proposed the rolling-horizon optimization strategy but ignored the impact of the assignment of quay cranes on the results.

In this work, DCBAP considered dynamic berthing, continuous berth allocation, and quay cranes assignment at the same time. DCBAP conforms well with the actual situation of dynamic optimization in container terminals. Few literature have considered DCBAP. Kim and Park [21] designed a two-phase solution procedure for DCBAP. The first phase determines the berthing position and time of each vessel as well as the number of cranes assigned to each vessel at each time segment. In the second phase, the subgradient optimization technique is applied to obtain a near-optimal solution of the first phase. In this phase, a detailed schedule for each quay crane is constructed. But in the experiment, the method can only obtain approximate solutions for small- or medium-sized problems with 9 quay cranes, 40 vessels, and 1200 meters berthing line. Zhang et al. [22] considered the coverage ranges of quay cranes and allowed for limited adjustments of quay cranes during loading and discharging based on Park and Kim [6]. Lim [23] transformed BAP to a restricted form of the two-dimensional packing problem. A graph-based representation is used to capture the problem, and an effective heuristic is proposed to solve the problem.

3. DCBAP Problem

The scheduling of continuous berth allocation problem with quay assignment needs to determine the berthing positions, berthing times, and the handling cranes for all vessels at each time. Generally, before the vessel moor alongside, the container terminal will allocate a preferred berth that the distance to the yard is the shortest. To improve the operating efficiency and reduce the operation cost, the real berthing time should be close to the expected berthing time. If the berthing time delays, the vessel should speed down or wait

in the anchorage, which will increase fuel consumption and affect the sail plan. The handling operations of a vessel should be completed in time so it can depart in time; delays will cause penalties and will reduce the service quality of the container terminal.

The berth in most of modern large container terminals is continuous. Park and Kim [6] solved DCBAP by dividing berth and time into several intervals and established a two-dimensional coordinate system with time as X-axis and berth as Y-axis; the X-axis denotes the berthing times and the Y-axis denotes the berthing positions. Both of the positions and times are continuous, and each vessel can only occupy one rectangle in the X-Y space. The horizontal length of rectangle depends on vessel length, and the width of rectangle depends on the container handling time that varies inversely to the number of cranes assigned to this vessel. Each rectangle represents a berth plan for a vessel. Based on the model proposed by Park and Kim [6], the DCBAP model built in this work considers the following reasonable situations. First, each vessel has a preferred berth that is determined in advance. Both of the position and time are continuous, and each vessel can only occupy one rectangle. The width of the rectangle is only related to vessel length and is irrelevant to vessel type. Second, the horizontal length of rectangle depends on the container handling time that varies inversely to the cranes assigned to this vessel. Third, the quay cranes available for each vessel are constrained by a minimum quantity and maximum quantity. The two quantities usually are determined by the contract between the shipping company and the terminal. Fourth, the ship handling tasks must be finished without interruption (when no crane is allocated to the vessel).

4. Formulation

4.1. Notations

(1) Sets are as follows:

(a) $\mathrm{sl} = \{1, 2, \ldots, \mathrm{sls}\}$: the set of vessels;
(b) $\mathrm{sm} = \{1, 2, \ldots, \mathrm{sms}\}$: the set of berthing positions;
(c) $\mathrm{sn} = \{1, 2, \ldots, \mathrm{sns}\}$: the set of time segments.

(2) Parameters are as follows:

(a) e_k: the expected arrival time of vessel k;
(b) a_k: the total operation time of cranes, which means the total amount of time segments required to handle all containers for vessel k;
(c) b_k: the length of vessel k;
(d) d_k: the due time for the departure of vessel k;
(e) s_k: the preferred berthing position of vessel k;
(f) c_k^1: the penalty cost per unit distance of vessel k between the berthing position and preferred berthing position;
(g) c_k^2: the penalty cost of vessel k per unit time of arrival after e_k;
(h) c_k^3: the penalty cost of vessel k per unit time of delay beyond the due time;
(i) l_k: the minimum quantity of cranes that should be assigned to vessel k;
(j) u_k: the maximum quantity of cranes that can be assigned to vessel k;
(k) c: the total number of available cranes ($C > u_k$);
(l) $A_{i,j}$: if the cell (i, j) in the time-space grid is occupied, it equals 1; otherwise, it is equal to 0;
(m) D_j: the number of cranes that is assigned at time j (the available cranes number equals $c - D_j$);
(n) M: a large positive real number.

(3) Decision variables are as follows:

(a) $x_{i,j,k}$: if the grid square (i, j) is covered by the rectangle for vessel $k \in sv$, it is equal to 1; otherwise, it is equal to 0;
(b) $z_{i,j,k}$: 1, if the preference point of vessel $k \in sv$ is located at (i, j) in the lower-left comer point of the rectangle corresponding to the vessel; otherwise, it is equal to 0;
(c) v_{kj}: 1, if vessel k is operated at time segment j; otherwise, it is equal to 0;
(d) u_{ki}: 1, if vessel k is berthed at berthing position i; otherwise, it is equal to 0;
(e) Y_{kj}: the number of cranes allocated to vessel $k \in sv$ at time j;
(f) C_k: the completion time of container handling for vessel k;
(g) BL_k: the left most berthing position of a vessel k;
(h) BR_k: the right most berthing position of a vessel k;
(i) TL_k: the delayed amount of berthing time of vessel k;
(j) DL_k: the delayed amount of departure time of vessel k.

The following variables can be expressed by the above variables and parameters:

(1) $B_k = \sum_{i \in \mathrm{sm}, j \in \mathrm{sn}} (Z_{k,i,j} \cdot i)$: the berthing position of vessel $k \in sv$;
(2) $T_k = \sum_{i \in \mathrm{sm}, j \in \mathrm{sn}} (Z_{k,i,j} \cdot j)$: the berthing time of vessel $k \in sv$;
(3) $\mathrm{POS}_{i,j} = \sum_{k \in \mathrm{sn}} (Z_{k,i,j} \cdot k)$: the vessel that occupies the cell (i, j) in the time-space grid.

4.2. The Model. The goals of the DCBAP include two aspects: first, the optimization of berthing position, which means that the berthing position should be as close as possible to the preferred berthing position so as to reduce the time to move the container from vessel to container yard and increase the operation efficiency; second, the optimization of berthing time, which means the operations should be completed as

early as possible in due time. Aimed at the optimization of berth and time, a multiobjective model proposed as (M1) minimizes the total penalty cost including deviation of berthing, berthing delay, and departure delay. The objective functions are defined in (1)–(4), and the constraints are defined in (5)–(29). Consider the following:

$$\text{(M1) Minimize} \quad f = \left(f^1, f^2, f^3\right), \tag{1}$$

$$f^1 = \sum_k \left(BL_k + BR_k\right) c_k^1, \tag{2}$$

$$f^2 = \sum_k \left(DL_k \cdot c_k^2\right), \tag{3}$$

$$f^3 = \sum_k \left(TL_k \cdot c_k^3\right), \tag{4}$$

$$\text{s.t.} \quad BL_k \geq B_k - S_k, \quad \forall k, \tag{5}$$

$$BR_k \geq S_k - B_k, \quad \forall k, \tag{6}$$

$$TL_k \geq T_k - e_k, \quad \forall k, \tag{7}$$

$$T_k \geq e_k, \quad \forall k, \tag{8}$$

$$DL_k \geq C_k - d_k, \quad \forall k, \tag{9}$$

$$C_k \geq V_{k,j}\left(j+1\right), \quad \forall k, \tag{10}$$

$$\sum_k X_{k,i,j} \leq 1, \quad \forall i, j, \tag{11}$$

$$\sum_k Y_{k,j} \leq c, \quad \forall j, \tag{12}$$

$$\sum_j Y_{k,j} \geq a_k, \quad \forall k, \tag{13}$$

$$V_{k,j} \leq Y_{k,j}, \quad \forall j, k, \tag{14}$$

$$Y_{k,j} \leq M \cdot V_{k,j}, \quad \forall j, k, \tag{15}$$

$$Y_{k,j} + M\left(1 - V_{k,j}\right) \geq l_k, \quad \forall j, k, \tag{16}$$

$$Y_{k,j} \leq u_k, \quad \forall j, k, \tag{17}$$

$$Y_{k,j} \leq \sum_i X_{k,i,j}, \quad \forall j, k, \tag{18}$$

$$M \cdot V_{k,j} \geq \sum_i X_{k,i,j}, \quad \forall j, k, \tag{19}$$

$$U_{k,i} \leq \sum_j X_{k,i,j}, \quad \forall i, k, \tag{20}$$

$$M \cdot U_{k,i} \geq \sum_j X_{k,i,j}, \quad \forall i, k, \tag{21}$$

$$ib - ia + 1 \leq \sum_{ia \leq i \leq ib} U_{k,i} + M\left(2 - U_{k,ia} - U_{k,ib}\right), \quad \forall k,\ 1 < ia < ib < \text{sms}, \tag{22}$$

$$jb - ja + 1 \leq \sum_{ja \leq j \leq jb} V_{k,j} + M\left(2 - V_{k,ja} - V_{k,jb}\right), \quad \forall k,\ 1 < ja < jb < \text{sns}, \tag{23}$$

$$V_{k,j} \leq \sum_{i, ja \leq j} Z_{k,i,ja}, \quad \forall j, k, \tag{24}$$

$$\sum_{i,j} Z_{k,i,j} = 1, \quad \forall k, \tag{25}$$

$$\sum_{i,j,\left(i<ia, \text{ or, } i \geq ia+b_k\right)} X_{k,i,j} \leq M\left(1 - \sum_j Z_{k,ia,j}\right), \quad \forall k,\ 2 \leq ia \leq \left(\text{sms} - b_k\right), \tag{26}$$

$$\sum_{i,j,i>b_k} X_{k,i,j} \leq M\left(1 - \sum_j Z_{k,1,j}\right), \quad \forall k, \tag{27}$$

$$\sum_{i,j,i \geq \text{sms}-b_k} X_{k,i,j} \leq M\left(1 - \sum_j Z_{k,\text{sms}-b_k+1,j}\right), \quad \forall k, \tag{28}$$

$$b_k - \sum_i X_{k,i,j} \leq M\left(1 - V_{k,j}\right), \quad \forall k, j. \tag{29}$$

The objective equation (1) specifies the total penalties including the cost for deviation of the actual berthing position to the preferred berthing position, the cost for vessels' berthing delay, and the cost for vessels' departure delay, as, respectively, defined in (2)-(4). Equations (5)-(6) indicate that the berthing position deviates to the preferred berthing position. Equation (7) constrains the delayed amount of berthing time of vessel. Equation (8) constrains that the arrival time must be earlier than the berthing time. Equation (9) calculates the delayed amount of departure time. Equation (10) constrains that the departure time of vessel must be later than the completion time. Equation (11) indicates that each time-space grid can only be occupied by one vessel. Equation (12) represents the fact that the total number of cranes allocated to vessels at each time segment is restricted by c. Equation (13) constrains that the total operation time of the vessel should excess its total operation time of cranes. Equations (14)-(15) restrict the sustainability of vessel handling operation; that is, if the loading and unloading operations of a vessel start, more than one crane should be

allocated to it during the operation time until all loading and unloading operations are completed. Equations (16)-(17) indicate that the quantity of cranes available for each vessel is restricted by the maximum and minimum quantities. Equations (18)-(19) represent the relations between $V_{k,j}$ and $X_{k,i,j}$. Equations (20)-(21) represent the relations between $U_{k,i}$ and $X_{k,i,j}$. Equations (22)–(24) constrains that the vessel occupies continuous time and space after berthing. Equation (25) ensures that each vessel has only one preference point. Equations (26)–(29) ensure that the time-space grid can only be occupied by a vessel that is berthed and handled in the schedule.

The multiobjective model (M1) is transformed into a single objective model. The objective of the new model is the weighted sum of three objectives, namely, $f = \sum_*(W^* \cdot f^*)$, where a weight vector W^* ($* \in \{1, 2, 3\}$) represents the significance of the three subobjectives and the preferences of decision-makers. Therefore, a new model (M2) is devised as follows:

$$\begin{aligned} &\text{(M2) Minimize} && f = W^* \cdot f^* \\ &\text{s.t.} && \text{Constraints (2)–(10) and (13)–(29).} \end{aligned} \tag{30}$$

5. Rolling-Horizon Strategy

In the process of the operation of the rolling scheduling, all vessels that need to be allocated are classified as three sets including vessels berthed set, set of vessels waiting for berthing, and set of vessels waiting for scheduling. Vessels berthed represent those whose handling operations have already completed or are ongoing. The vessels waiting for berthing are those who are in scheduling but have not yet started the handling operations. The vessels waiting for scheduling are those who have arrived at the port but are not yet in the schedule. During the operations of the container terminal, the vessels arrive to the port according to the expected berthing time, and the vessels stay in the set of vessels waiting for scheduling in order of arrival times. In each optimization of rolling-horizon schedule, the vessels that have finished the handling operation can be removed from the rolling window and the vessels waiting for scheduling will be added into the rolling window. Then, the schedule of the present rolling window can be obtained through optimal algorithm.

In the process of rolling scheduling, the input parameters of vessels waiting for berthing are updated by the scheduling information of vessels berthed. Based on the DCBAP model (M2), a new model (M3) was proposed as (31)–(33). Equation (32) expands (11), which indicates that a time-space grid can be occupied by present scheduling vessels and the vessels that have already berthed but the operation is still ongoing right now. The parameter A is set as the time-space grids that have been occupied in past rolling-horizon schedule (frozen ship). That is, A represents the time-space grids occupied by those vessels that have already been berthed. If the time-space grid is occupied, set the element of A to 1; otherwise, set it to 0. Equation (33) expands (12), and the parameter D is introduced to calculate the quantity of quay cranes occupied by vessels berthed in the past rolling-horizon schedules in each time so as to update the available quantity of quay cranes in each time. Consider

$$\text{(M3) Minimize} \quad f = W^1 \cdot f^1 + W^2 \cdot f^2 + W^3 \cdot f^3, \tag{31}$$

$$\text{s.t.} \quad \sum_k \left(X_{k,i,j} + A_{i,j}\right) \le 1, \quad \forall i, j, \tag{32}$$

$$\sum_k Y_{k,j} \le c - D_j, \quad \forall j,$$

$$\text{Constraints (2)–(10) and (13)–(29).} \tag{33}$$

The rolling-horizon strategy decomposes the scheduling horizon into a series of scheduling windows, and optimal results for each window are obtained as a small-scale optimization of (M3). Then, the scheduling window is shifted to the next and the input parameters of (M3) are updated. The transition interval length of a shift and the size of the rolling window are basic factors of the rolling-horizon strategy. The transition interval is the interval between present rolling window and next rolling window. In general, decreasing the transition interval length can make the optimal result close to the result of global optimization; the number of iterations of rolling-horizon optimization increases the computation time. The rolling window for the tth time of rolling contains a set of top k of vessels waiting for scheduling, defined as $k(t)$. k is the size of the rolling window. When k is normal, the number of the vessels in each rolling window equals to k except the last rolling window which may be less than k. A smaller rolling window indicates less vessels scheduled for each time and less computation time, while the number of iterations of rolling optimization improves the result compared to the result of global optimization. Figure 1 shows the rolling strategy from tth time to $t + 1$th time.

In the optimal results for the tth window of the rolling-horizon optimization, the first λ vessels are frozen from the rolling window $k(t)$ according to the sequence of scheduled berthing times. λ is the size of the frozen window in the rolling-horizon strategy, that is, the transition interval. The scheduled vessels in next time $(t + 1)$ for rolling-horizon optimization are obtained by merging the λ scheduled vessels into $V(t-1)$ which is the vessels scheduled at $(t-1)$th time of rolling-horizon optimization. The other vessels in $k(t)$ which are not frozen still stay in $\overline{V}(t)$ and wait for the next time of rolling-horizon optimization. Then, a new time of rolling-horizon optimization, $k(t + 1)$, is obtained by choosing the first k vessels from $\overline{V}(t)$ according to their berthing times. If the vessels left in $\overline{V}(t)$ are less than k, we choose all vessels left. When all vessels are in $V(t)$, all the vessels are believed to have been scheduled, and the entire rolling-horizon optimization is finished. The holistic optimal solution is obtained by the rolling-horizon optimization process and combining the results returned at each rolling window.

Based on the above analysis, the rolling-horizon strategy for DCBAP involves the following four steps.

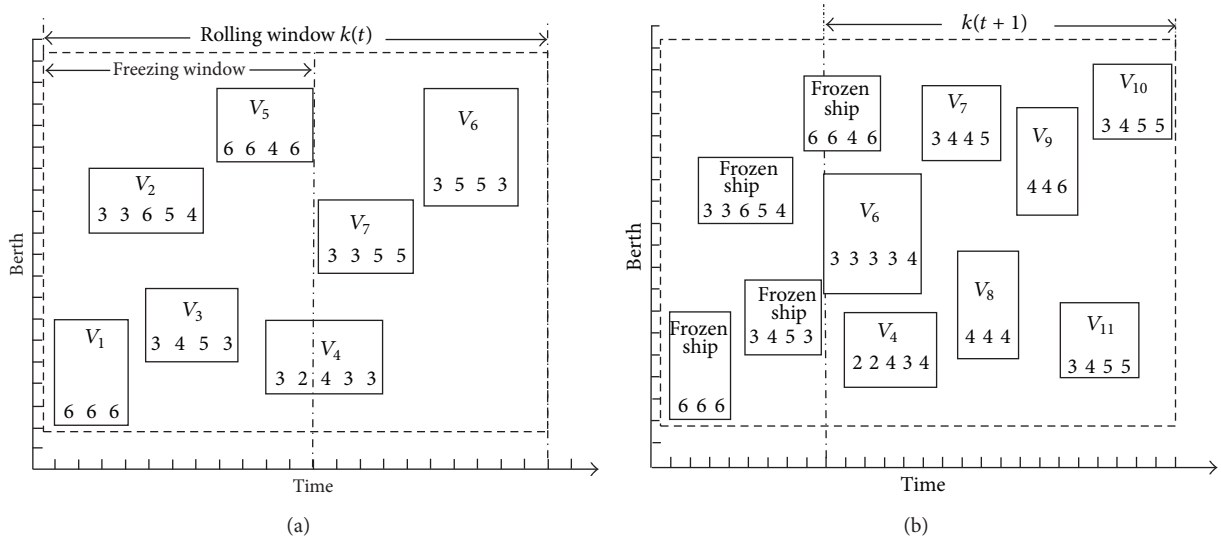


Figure 1: An illustration of rolling strategy.

Step 1. Sort all vessels that are to be scheduled according to the arrival times, initialize the total vessels set as sl, and set $t = 1$.

Step 2. Initialize the set of scheduled vessels by setting $V(t) = \Phi$ and initialize the set of vessels waiting for scheduling by setting $\overline{V}(t) = \text{sl}$, obtain the rolling window $k(t)$ by choosing the first k vessels in set $\overline{V}(t)$, and update $\overline{V}(t)$ by setting $\overline{V}(t) = \overline{V}(t) \setminus k(t)$. Then, the parameter A is updated according to whether the time-space grid is occupied by those vessels that have already berthed; update the parameter D by calculating the quantity of quay cranes occupied in each time segment.

Step 3. After $\text{sl} \leftarrow k(t)$, solve (M2) and obtain the berthing schedule of $k(t)$; freeze the first λ vessels from the rolling window $k(t)$ according to the sequence of scheduled berthing times of vessels in optimal results; set the vessels scheduled as $V(t+1)$ by merging the λ scheduled vessels into $V(t)$; set the vessels waiting for scheduling as $\overline{V}(t+1)$ by merging the vessels left in rolling window $(k - \lambda)$ into $\overline{V}(t)$.

Step 4. If $|V| = \text{sls}$ (all vessels have been scheduled), stop and combine the results of each window as the optimal solution; or else, set $t = t + 1$ and go to Step 2.

6. Experiments

The known data and the initialized schedule are set as follows.

(1) The length of the quay line is 1200 meters (m), which is divided into 24 grids by the unit of 50 m, while the planning horizon is divided into 48 grids (i.e., 48 time segments). During this period, 20 vessels arrive at the terminal for vessel handling operations successively.

(2) For vessel k, e_k is generated from the uniform distribution $U[1, 37]$. Similarly, r_k is generated from $U[1, 18]$; a_k is generated from $U[10, 30]$. d_k is generated based on e_k and a_k. The length of the vessel (b_k) is generated from $U[4, 8]$ based on the setting of berth axis. The maximum and minimum numbers of cranes are set as Table 1.

(3) The penalty cost per unit is set to the same value; $c_*^1 = 100$, $c_*^2 = 200$, and $c_*^3 = 300$ for all vessels.

(4) The maximum available crane number is set to $c = 9$. Set $M = 9999$.

(5) The size of the rolling window is set to $k = 5$. And the length of the transition interval is set to $\lambda = 2$.

Based on Table 1, CPLEX is used as the optimization engine for solving the mixed-integer linear programming model (M2), and the rolling-horizon strategy was implemented with *C#*. A berthing schedule for 20 vessels in 48 hours is obtained, as shown in Figure 2. The berthing time, departure time, berthing position, and quay cranes assigned are indicated in Figure 2. Each rectangle represents a set if time-space grids are occupied by berthing vessels, and the number represents the quay cranes assigned to the vessel at that time segment.

Two important parameters in the rolling strategy are the size of rolling window (k) and transition interval (λ). Table 2 presents a series of results of adjusting k and λ. When $\lambda = 4$, the total penalties decrease first and then increase along with the increasing of k which changes from 5 to 8. The minimum objective function value is $\lambda = 15600$ when $k = 7$. When k changes from 5 to 7, the total penalty decreases with the increase of the size of the rolling window (k). This is mainly due to the fact that with the increase of the size of the rolling window, the quantity of vessels scheduled at a time increases, so that the result of the rolling-horizon optimization gets close to that of the global optimization. The decrease of total penalty decreases the berthing delay penalty and departure delay penalty, because the penalty for

Table 1: Input data of vessels.

Vessels	ETA	Operations	Vessel length	Due time	Ref. position	Min. cranes	Max. cranes
	e_k	a_k	b_k	d_k	r_k	l_k	u_k
1	1	15	8	11	1	3	6
2	2	10	4	14	16	2	4
3	4	27	6	16	3	3	6
4	4	11	4	17	10	2	4
5	5	24	7	17	5	3	6
6	9	30	6	16	14	3	5
7	10	21	5	20	7	3	5
8	10	16	4	20	13	3	5
9	13	23	5	23	10	3	5
10	16	18	6	26	15	3	6
11	17	15	6	27	2	3	6
12	20	20	6	30	10	3	5
13	21	21	6	31	15	2	5
14	26	18	4	36	8	3	6
15	27	20	4	37	6	3	6
16	27	17	6	37	15	3	5
17	27	22	5	37	9	3	5
18	28	15	7	38	1	3	6
19	34	18	6	44	6	3	5
20	36	24	6	46	7	3	6

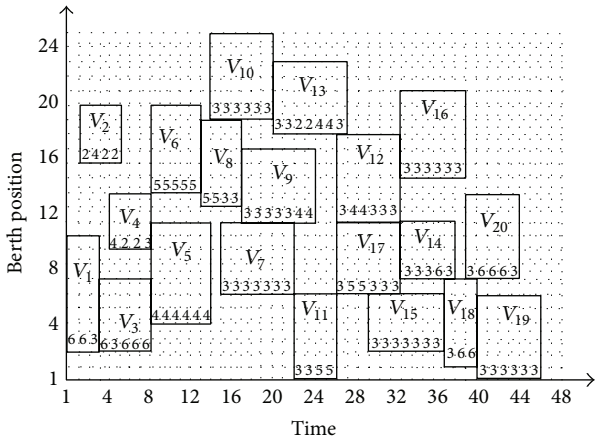


Figure 2: The schedule of a DCBAP.

berthing delay or departure delay is much more than the penalty for deviating from the preferred positions. So when the size of the rolling window gets bigger, the optimal result prefers to be deviated from the preferred position or assigns more quay cranes instead of delay to keep the total penalty lower. Meanwhile, with the increase of k, the utilization rate of the quay cranes also increases, the fluctuation of quay cranes (the variance yields of the quantity of quay cranes assigned to vessels at each time) gets lower, the quantity of quay cranes assigned to the vessel at each time increases the balancing degree, the operation cost of quay cranes decreases, and the total stay time at the terminal increases slightly. A reasonable and efficient rolling-horizon schedule needs the cooperation between the container terminal operators and the shipping company. Moreover, optimal results can be reached under the balance of the stakeholders.

In the rolling-horizon strategy, decreasing the transition interval length will increase the rolling times but the result can be improved. Based on the comparisons of the objectives and the vessel's total stay times when λ changes from 5, 7, and 8 in Table 2, it is found that decreasing λ can improve the result. For instance, when $k = 7$ and λ changes from 5 to 4, f, f^1, f^2, and f^3 can be improved by 13.8%, 25%, 13.7%, 0%, and 5%, respectively. Obvious effects of optimization can also be seen in other experiments.

When the size of the rolling window increases from 7 to 8, the rolling times increase, and the total penalty cost increases. The size of the rolling window and the quantity of the frozen vessels have opposite effects on the rolling times. The former has a negative correlation with rolling times while the latter's correlation with rolling times is positive. Therefore, when the container terminal operator makes a rolling-horizon schedule, the size of the rolling window and quantity of the frozen vessels should be set reasonably, so as to improve the efficiency of the rolling-horizon scheduling process.

Still as presented in Table 2, the quantity of frozen vessels is varied to study the effect of it on the result. With the increase of the number of the frozen vessels, the penalty increases and the performance of rolling-horizon strategy decreases. With the increase of the frozen vessels, the number of the fixed vessels in each time of rolling scheduling increases. The fixed vessels are not involved in the next times of optimization, so that the rolling-horizon optimization

Table 2: The results under different λ.

k	λ	f	f^1	f^2	f^3	TSP (h)	FC	LRW
5	3	17300	2400	13400	**1500**	80	1.597	6
5	4	17600	1800	14000	1800	100	1.436	5
6	4	16800	1700	13600	**1500**	100	1.388	5
7	4	15600	1500	12600	1500	95	1.344	5
7	5	18100	2000	14600	**1500**	100	1.320	**4**
8	4	18100	2000	14600	**1500**	105	**1.313**	**4**
8	5	17600	2100	14000	**1500**	105	1.350	**4**

Note: TST: total time staying at terminal; FC: fluctuations of cranes; LRW: length of the rolling window.

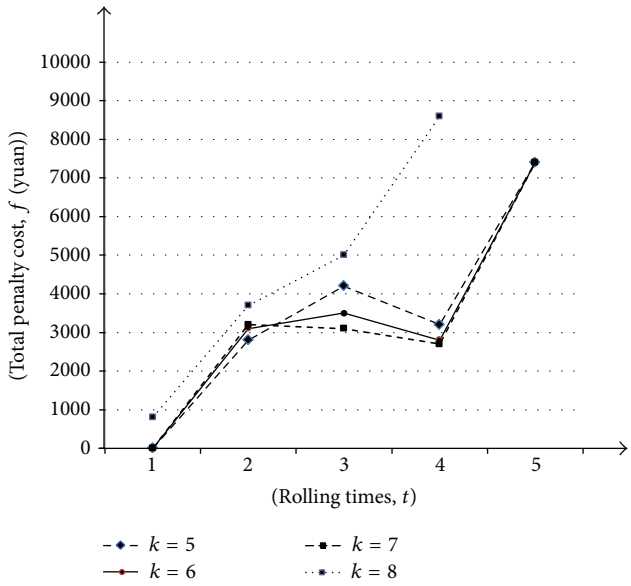


Figure 3: The penalty cost of each rolling time under different size of rolling window.

method cannot obtain the best result. Setting the quantity of the frozen ships to a small value may result in rolling too many times and computing for a long time. Therefore, in an actual schedule, a reasonable configuration of quantity of frozen vessels can decrease the effect of the number of fixed vessels and optimize the computation time as well as optimize the performance of the rolling-horizon strategy.

The coupling degree of vessels' arrival times has a big impact on the efficiency of the rolling strategy. As shown in Figure 3, when the size of the rolling window changes, the similar tendency can be seen that the penalty costs increase sharply in the third and fifth times of rolling under different size of rolling windows. Due to the high coupling degree in vessels' arrival in the last time of rolling, vessels arrive intensely and bring too many handling operations, so resources including quay cranes and berths are occupied for a long time and the available resources for the next rolling schedule are insufficient. Therefore, the successor ships have to delay the times for berth and departure. As a result, the efficiency of the rolling schedule decreases and the penalty increases.

The factors that affect the rolling strategy mainly include the size of rolling window, quantity of the frozen vessels, and the rolling times. So appropriate size of rolling window and quantity of frozen vessels need to be chosen to improve the performance of the rolling-horizon strategy. When choosing the size of the rolling window, both operation penalty cost and computation time should be considered. On the one hand, the size should be bigger and the quantity of vessels in the rolling window should be sufficient enough so that the local optimal solution (returned by solving (M3)) will be closer to the result of global optimization. Moreover, it will decrease the operations cost of the container terminal and improve the service quality. On the other hand, if the size is too big and the computation time of each time of the rolling window gets too long, which may lead to delay of the berthing time of vessels and affect the normal operations of container terminals and increase the cost inversely. So the size of the rolling window should balance both cost and computation times to improve the operation efficiency of the container terminal. And based on this, an appropriate quantity of frozen vessels should be chosen to achieve the best performance of the rolling-horizon strategy.

In this paper, the multiobjective function is changed to a single-objective function by using weighting efficient. In order to analyze the tradeoff between berth deviation and time delay, the objectives are divided into two categories, berth-varied objective f^1, and time-varied objectives f^2 and f^3. The weights of f^2 and f^3 are set as equal. By adjusting the weight of the objects ($W^1 = 0.1, 0.2, \ldots, 0.9$, $W^2 = 1 - W^1$), 10 experimental variables can be obtained based on M2. Then, the objective function value can be calculated based on each experimental variable. The change trend of the two objective functions is shown in Figure 4. With the increase of the weight W^1, the berthing deviation is reduced from 7200 to 0, and when $W^1 = 0.9$ each vessel can be assigned to the preferred berth. But the result is that the waiting time for handling is increasing and the number of vessels that delayed departure and the delay time is also increasing, which lead to the fact that the penalty cost including berthing delay and departure delay increases to 152000, the waiting time is added by 76%, and the departure delay is added by 57%. Therefore, controlling weight W^1 reasonably can seek a balance between the deviation from the preferred berth and the waiting time.

As shown in Table 2, the optimal solution is obtained when $k = 7$, $\lambda = 4$. So a comparative analysis between

Table 3: Comparison between rolling-horizon and normal schedule for different vessel quantity.

Quantity of vessels	Scheduling method	f	f^{c1}	f^{c2}	f^{c3}	Computation time (s)
10	Rolling-horizon	4400	900	3200	300	564.4
	Normal schedule	4000	900	2800	300	1133.3
11	Rolling-horizon	5300	900	4000	0	763.2
	Normal schedule	5300	900	4400	0	2054.8
12	Rolling-horizon	6200	1200	5000	0	778.2
	Normal schedule	6000	1000	5000	0	7871.1
13	Rolling-horizon	7000	1000	6000	0	853.3
	Normal schedule	7000	1000	6000	0	21092.1
14	Rolling-horizon	7500	1300	6200	0	954.9
	Normal schedule	—	—	—	—	—
15	Rolling-horizon	8500	1300	7200	0	1125.7
	Normal schedule	—	—	—	—	—

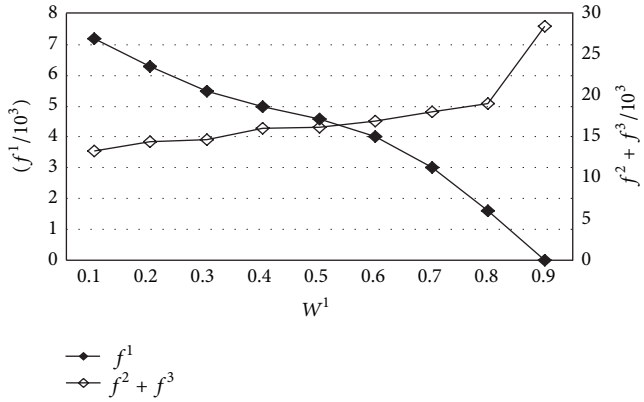


Figure 4: Tradeoff between penalty of berthing deviation and time delay.

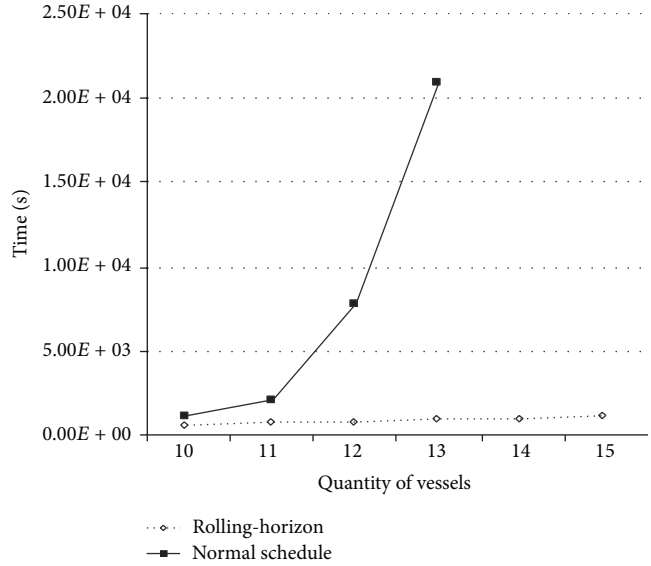


Figure 5: The computation time tendency of rolling-horizon and normal schedule.

normal schedule and rolling schedule is conducted as shown in Table 3, where "—" indicates that the result cannot afford the CPLEX solver running out of memory.

As shown in Table 3, when using rolling-horizon strategy, optimal results are obtained for all the six examples. While using normal schedule, the solution cannot be obtained when the quantity of vessels is increasing to 14. It indicates that the rolling-horizon strategy can decrease the complexity of computation compared to normal schedule and solve large-scale problems.

When the quantity of vessels increases from 10 to 15, the optimal result of the rolling-horizon strategy is similar to that of the normal schedule. The rolling-horizon strategy performs well. As shown in Figure 5, from the tendency of computation time of the rolling-horizon strategy and normal schedule, the time increases sharply with the increase of the total quantity of vessels. When the total quantity of vessels is 10, the computation time of normal schedule is twice that of the rolling schedule. When the total quantity is 13, the time of the normal schedule is 25 times that of the rolling schedule. The computation time of normal schedule increases exponentially with the number of vessels, and when the total quantity is 14, the optimal result cannot be obtained. On the contrary, the computation time of the rolling-horizon strategy increases slightly, which is directly in proportion to the total quantity. The normal schedule searches for a large amount of nodes once, which burdens the CPU and memory much. The rolling schedule has better computational efficiency than normal schedule, and it can solve large-scale problems in the limited time.

Though in the above experiment only 20 vessels are considered in the problems, the quantity of the vessels is directly in proportion to the computation time. So the rolling-horizon strategy proposed in this paper can be used to solve large-scale problems and meet practical requirements. Based on the above experiments, the algorithm was performed to solve large-scale problems whose quantities of vessels range from 40 to 200 and the planning horizons range from 48 to 400 hours. The results show that the computation performance is also almost directly in proportion to the quantity.

7. Conclusions

To solve the large-scale integral scheduling of DCBAP, a method based on the rolling-horizon strategy is developed and a mixed-integer multiobjective linear programming model is established. The model minimizes the total penalty including deviation of berthing, berthing delay, and departure delay. Moreover, the model is revised for the rolling-horizon strategy. The experimental results indicate that the method can be applied to solve large-scale DCBAP problem with the computation time linear to the quantity of involved vessels. In the rolling-horizon strategy, the parameters include the sizes of the rolling window and transition intervals have big influences on the optimality and the overall performance of the rolling-horizon strategy. The increase of the length of the rolling window can improve the results, while the computation time increases exponentially. The decrease of the length of the transition interval increases the rolling times and improves the results. The experiments prove that the settings of $k = 7$, $\lambda = 4$, can obtain better results than other settings. Meanwhile, the high coupling degree of vessels' arrival times will decrease the efficiency of the rolling strategy. The actual scheduling should choose appropriate combination of the size of rolling window and transition interval by a series of experiments on the balance of optimality and computation performance to improve the efficiency of berthing schedule and the service quantity of the container terminal and decrease the operations cost. The rolling-horizon strategy in this paper considers the dynamic characteristics of the vessels' expected berthing time. Compared to the model proposed by Park and Kim [6], the problem scale (vessels involved) is greatly expanded, but the possibility that vessels may arrive and be scheduled in advance is not considered. Besides, though the on-the-shelf solver for mixed-integer linear programming model is convenient for researching the effectiveness of the rolling-horizon strategy, this solver becomes a bottleneck that restricts improvement of the computing performance. To meet the demands of practical applications, the work on the way is to design a heuristic algorithm to solve the model (M2) effectively and merge it into the rolling-horizon strategy for further improving the optimality and computational performance.

Conflict of Interests

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

Acknowledgments

This study is partially supported by the National Nature Science Foundation of China (71101088, 71171129, and 71390521), the Science Foundation of Ministry of Education of China and Shanghai (20113121120002, 14YZ100, 20123121110004, and 13SG48), the Science and Technology Commission of Shanghai (12ZR1412800 and 12510501600), the high-tech research and development program of China (2013A2041106), and Graduate Innovation Fund Project of Shanghai Maritime University (GK2013022).

References

[1] S. Chand, R. Traub, and R. Uzsoy, "Rolling horizon procedures for the single machine deterministic total completion time scheduling problem with release dates," *Annals of Operations Research*, vol. 70, pp. 115–125, 1997.

[2] J. Fang and Y. Xi, "A rolling horizon job shop rescheduling strategy in the dynamic environment," *The International Journal of Advanced Manufacturing Technology*, vol. 13, no. 3, pp. 227–232, 1997.

[3] I. M. Ovacik and R. Uzsoy, "Rolling horizon procedures for dynamic parallel machine scheduling with sequence-dependent setup times," *The International Journal of Production Research*, vol. 33, no. 11, pp. 3173–3192, 1995.

[4] Y. Guan, W.-Q. Xiao, R. K. Cheung, and C.-L. Li, "A multiprocessor task scheduling model for berth allocation: heuristic and worst-case analysis," *Operations Research Letters*, vol. 30, no. 5, pp. 343–350, 2002.

[5] C. L. Li, X. Q. Cai, and C. Y. Lee, "Scheduling with multiple-job-on-one-processor pattern," *IIE Transactions*, vol. 30, no. 5, pp. 433–445, 1998.

[6] Y.-M. Park and K. H. Kim, "A scheduling method for berth and quay cranes," *OR Spectrum*, vol. 25, no. 1, pp. 1–23, 2003.

[7] C. Bierwirth and F. Meisel, "A survey of berth allocation and quay crane scheduling problems in container terminals," *European Journal of Operational Research*, vol. 202, no. 3, pp. 615–627, 2010.

[8] I. F. A. Vis and R. de Koster, "Transshipment of containers at a container terminal: an overview," *European Journal of Operational Research*, vol. 147, no. 1, pp. 1–16, 2003.

[9] K. H. Kim and K. C. Moon, "Berth scheduling by simulated annealing," *Transportation Research B: Methodological*, vol. 37, no. 6, pp. 541–560, 2003.

[10] Y.-M. Park and K. H. Kim, "Berth scheduling for container terminals by using a sub-gradient optimization technique," *Journal of the Operational Research Society*, vol. 53, no. 9, pp. 1054–1062, 2002.

[11] A. Imai, E. Nishimura, and S. Papadimitriou, "The dynamic berth allocation problem for a container port," *Transportation Research B: Methodological*, vol. 35, no. 4, pp. 401–417, 2001.

[12] G. G. Brown, S. Lawphongpanich, and K. P. Thurman, "Optimizing ship berthing," *Naval Research Logistics*, vol. 41, no. 1, pp. 1–15, 1994.

[13] A. Imai, H. C. Chen, E. Nishimura, and S. Papadimitriou, "The simultaneous berth and quay crane allocation problem," *Transportation Research E: Logistics and Transportation Review*, vol. 44, no. 5, pp. 900–920, 2008.

[14] P. Legato, D. Gullí, and R. Trunfio, "The quay crane deployment problem at a maritime container terminal," in *Proceedings of the 22nd European Conference on Modelling and Simulation (ECMS '08)*, pp. 196–209, Nicosia, Cyprus, June 2008.

[15] L. Zhen, L. H. Lee, and E. P. Chew, "A decision model for berth allocation under uncertainty," *European Journal of Operational Research*, vol. 212, no. 1, pp. 54–68, 2011.

[16] M. Sammarra, J.-F. Cordeau, G. Laporte, and M. F. Monaco, "A tabu search heuristic for the quay crane scheduling problem," *Journal of Scheduling*, vol. 10, no. 4-5, pp. 327–336, 2007.

[17] I. M. Ovacik and R. Uzsoy, "Rolling horizon algorithms for a single-machine dynamic scheduling problem with sequence-dependent setup times," *The International Journal of Production Research*, vol. 32, no. 6, pp. 1243–1263, 1994.

[18] M. Singer, "Decomposition methods for large job shops," *Computers & Operations Research*, vol. 28, no. 3, pp. 193–207, 2000.

[19] B. Raa, W. Dullaert, and R. van Schaeren, "An enriched model for the integrated berth allocation and quay crane assignment problem," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14136–14147, 2011.

[20] D. Chang, J. He, and H. Zhang, "A rule-based joint berth allocation and quay crane assignment," in *Proceedings of the 2nd International Conference on Industrial and Information Systems (IIS '10)*, pp. 464–467, IEEE, Dalian, China, July 2010.

[21] K. H. Kim and Y.-M. Park, "A crane scheduling method for port container terminals," *European Journal of Operational Research*, vol. 156, no. 3, pp. 752–768, 2004.

[22] C. Zhang, L. Zheng, Z. Zhang, L. Shi, and A. J. Armstrong, "The allocation of berths and quay cranes by using a sub-gradient optimization technique," *Computers & Industrial Engineering*, vol. 58, no. 1, pp. 40–50, 2010.

[23] A. Lim, "The berth planning problem," *Operations Research Letters*, vol. 22, no. 2-3, pp. 105–110, 1998.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 851697, 7 pages
http://dx.doi.org/10.1155/2014/851697

Research Article

Day-to-Day Scheduling Travel Time Adjustment Behavior and Simulation

Hua-Min Li,[1,2] David Z. W. Wang,[2] Hai-Jun Huang,[3] and Xiao-Jun Yu[4]

[1] *College of Automation, Chongqing University, Chongqing 400044, China*
[2] *School of Civil and Environmental Engineering, Nanyang Technological University, Singapore 639798*
[3] *School of Economics and Management, Beihang University, Beijing 100191, China*
[4] *School of Mathematics & Statistics, Guizhou University of Finance & Economics, Guiyang 550004, China*

Correspondence should be addressed to Hua-Min Li; huaminli@cqu.edu.cn

Received 19 February 2014; Revised 6 April 2014; Accepted 13 April 2014; Published 30 April 2014

Academic Editor: X. Zhang

This paper proposes a modeling framework to study the day-to-day scheduling travel time adjustment behavior on the base of past experiences. Scheduling travel time is defined as the difference between the requested arrival time and the departure time. Mathematical equations are established to formulate every traveler's dynamic adjustment on his/her departure time. In the adjustment process, the last day's scheduling time is an essential component, while the numbers of late arrivals, punctual arrivals, and early arrivals in a previous day are used to reflect the past experiences. Simulation results are presented to illustrate the effectiveness of the proposed modeling framework.

1. Introduction

Travel time has always been a fundamental component of the transportation investigation and has been applied in various forms. A large number of researches and literature reviews are concerned with the formula and estimation of travel time, which constitutes the essential work of studying various traffic assignments based on the principles of user equilibrium, system optimization, and their stochastic and/or dynamic counterparts. Meanwhile, travel time is an important index to evaluate the performance of a transportation system and hence is regularly surveyed in almost all cities. Travel time is also information widely used in the advanced traffic management systems and in-vehicle route guidance systems.

Travel time can be estimated by using different approaches, depending upon the applications with different purposes. The point delay models are commonly used to estimate the travel times of signalized links (Webster [1]; Allsop [2]). In the Highway Capacity Manual [3], the average travel time of a link is calculated as the sum of the link running time and the intersection delay which is given by a deterministic point delay model. Skabardonis and Geroliminis [4] proposed an analytical model to estimate the travel times of arterial streets, using the 15–30-second flow and the occupancy data provided by loop detectors and traffic signal settings. These studies mainly focused on the estimation of average travel time.

Travel time is affected by some unknown factors and is thus stochastic, especially in the complex urban environment. This makes the investigation of travel time uncertainty become a hot topic in recent years. Guo et al. [5] proposed a multistate model which is employed to fit a mixture of Gaussian distributions into travel time observations of an expressway corridor. Each normal distribution is associated with an underlying traffic state providing quantitative uncertainty evaluation. The multistate mixture model leads to a better fitting, revealing that travel time distribution usually has more than one mode which is entirely dependent on the time horizon of the study, the demand, the topology, and others. Using multistate models, a recent work by Park et al. [6] tried to quantify the impact of traffic incidents. Feng et al. [7]

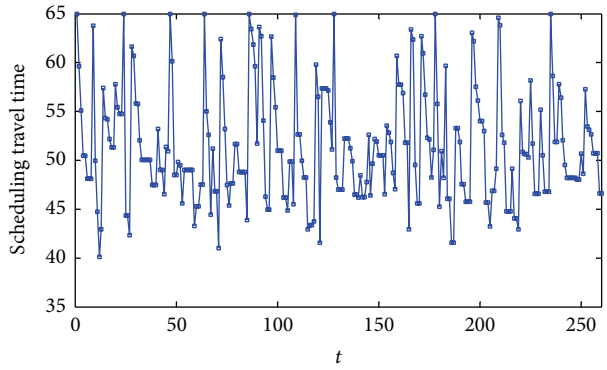


Figure 1: Day-to-day scheduling travel time when $Toltime = 5$ and $Rv = (b - a)/6$.

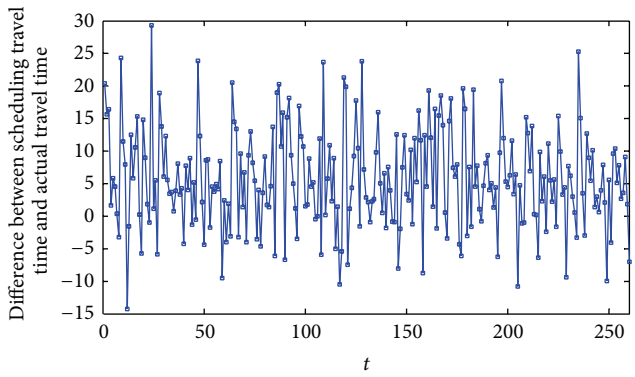


Figure 2: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 5$ and $Rv = (b - a)/6$.

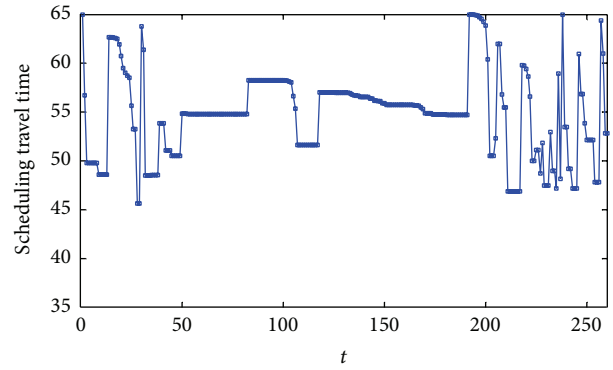


Figure 3: Day-to-day scheduling travel time when $Toltime = 10$ and $Rv = (b - a)/6$ (scenario 1).

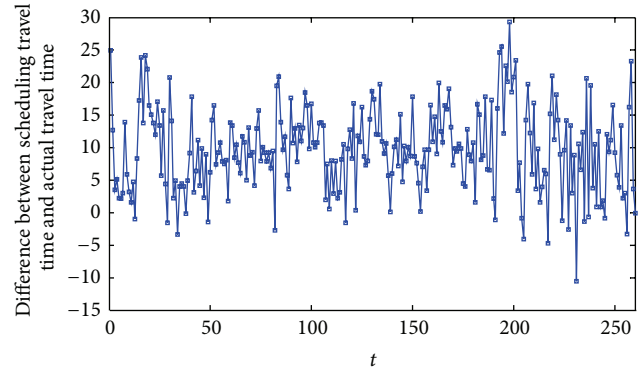


Figure 4: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 10$ and $Rv = (b - a)/6$ (scenario 1).

conducted a similar investigation, utilizing mixtures of normal distributions to estimate mean travel times for arterial routes.

As mentioned above, the actual travel time in an urban traffic environment is highly stochastic and time-dependent due to random fluctuations in interruptions caused by traffic control devices, weather conditions, and day-to-day events, such as vehicle stall, minor accident, traffic signal, bus stop, merge and diverge bottleneck, truck and bus platoon, and train-crossing. Considering these factors is essential for improving the estimation accuracy and expanding the relevant studies. It has been increasingly recognized that the actual travel time should be assumed to be stochastic.

On the other hand, travelers, especially commuters, learn the travel times based on the past experiences and then determine their departure times. They realize that traffic participants, vehicles, and such facilities as roads and signals are the main components of traffic system. Various factors which affect the operation of the components would influence the final travel time. For instance, different drivers and road conditions could result in large variation of travel times. Even in the same time interval and on the same link, different vehicles can have quite different travel times (Li and McDonald [8]). Travelers recognize that the actual travel time is stochastic and should adjust their departure time according to their past experiences on late arrival, punctual arrival, and early arrival. More importantly, due to the randomness of human choices, these adjustments should not be in general deterministic but random.

In Section 2 of this paper, we make a synthetic consideration about the scheduling travel time and the actual travel time. In Section 3, a model for formulating the day-to-day scheduling travel time adjustment behavior is proposed. In Section 4, simulation results are presented to illustrate the effectiveness of the proposed model. Section 5 concludes the paper.

2. Specifications of Scheduling Travel Time and Actual Travel Time

We define the scheduling travel time, TP_t, as the time consumption on day t required for traveling from an origin to a destination (e.g., from a commuter's residence to his/her working place) and exactly arriving at the requested time. Thus, the scheduling travel time TP_t is equal to the requested arrival time minus the departure time. Obviously the scheduling travel time should vary within a certain range of time defined by an upper bound b and a lower bound a that is corresponding to the free-flow travel time.

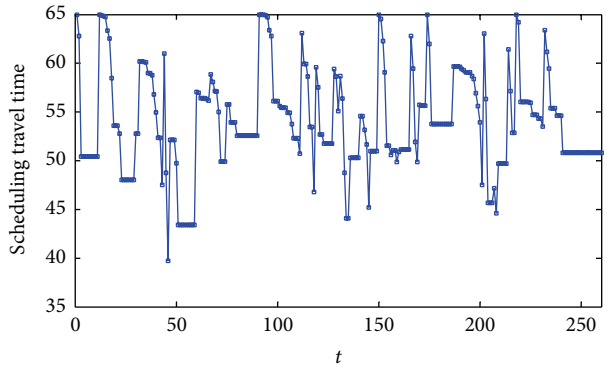


FIGURE 5: Day-to-day scheduling travel time when $Toltime = 10$ and $Rv = (b - a)/6$ (scenario 2).

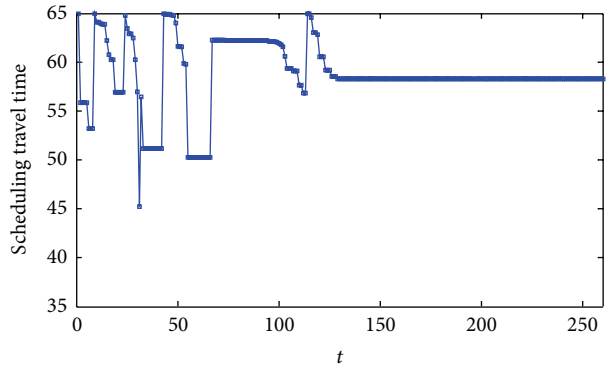


FIGURE 7: Day-to-day scheduling travel time when $Toltime = 15$ and $Rv = (b - a)/6$.

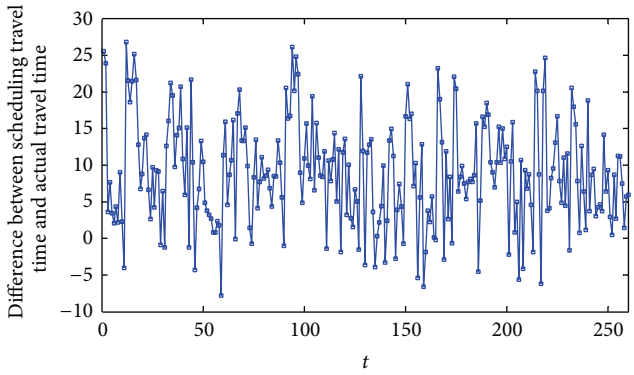


FIGURE 6: Day-to-day difference between scheduling and actual travel times when $Toltime = 10$ and $Rv = (b - a)/6$ (scenario 2).

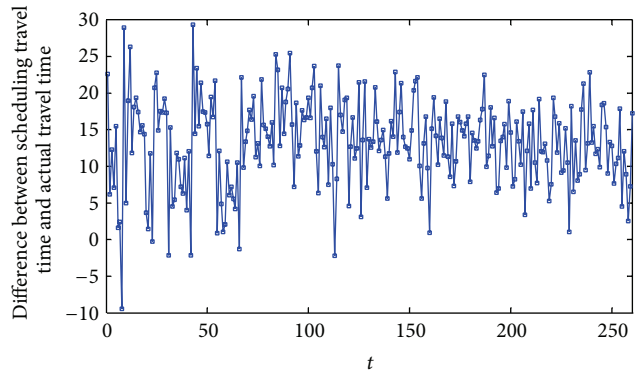


FIGURE 8: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 15$ and $Rv = (b - a)/6$.

In reality, it is easily observed that a commuter will modify his departure time on day t if he arrives lately or early on day $t - 1$. This modification should simultaneously take into account his past experiences, such as the numbers of late arrivals, punctual arrivals, and early arrivals in previous days. A commuter is willing to retain his departure time on day t if he arrives on time on day $t-1$. In this regard, a commuter will adjust his departure time whenever he arrived late or early. The adjustment process and its magnitude should depend on each commuter's preferences and recallable experience. However, the mechanism by which each individual handles his experiences and carries out the adjustment is not known (Chang and Mahmassani [9]). Therefore, it is reasonable to assume this adjustment to be a random variant related to the past experiences. Consequently, according to the preceding analysis, the scheduling travel time on day t can be defined as the scheduling travel time on day $t-1$ plus a corresponding random variable that grounds on the last situation.

The actual travel time, denoted as TR_t, is the time a commuter actually travels from his origin to the destination on day t. As mentioned at the foregoing paragraph, numerous factors affect the actual travel time. The actual travel time is highly stochastic. Thus, we can presume that the actual travel time is a random variable which has a lower bound a given by the free-flow travel time but has no upper bound.

3. Day-to-Day Scheduling Travel Time Adjustment Behavior

In this study, we propose a model that incorporates dynamic and randomness into consumer choice of departure time. In our day-to-day decision-making framework for scheduling travel time (corresponding to departure time), a commuter will adjust his scheduling travel time whenever he experienced the difference between last scheduling travel time and last actual travel time. The model allows for an increase and a decrease of the scheduling travel time in a particular fashion that is interrelated with the numbers of previous late arrivals, punctual arrivals, and early arrivals. If the commuter arrives late, he will be induced to increase the scheduling travel time on the subsequent day. If the commuter arrives early, he will be more likely to decrease the scheduling travel time. Otherwise, the commuter will retain the scheduling travel time on the subsequent day if he arrives on time at present day.

An assumption is made that not only does the researcher not observe all factors in the consumer's adjustment, but also, even in the same situation, the same commuter can make quite different adjustment. Therefore, a random term is supposed to incorporate into the adjustment to represent this unobserved component. On the other hand, to reflect

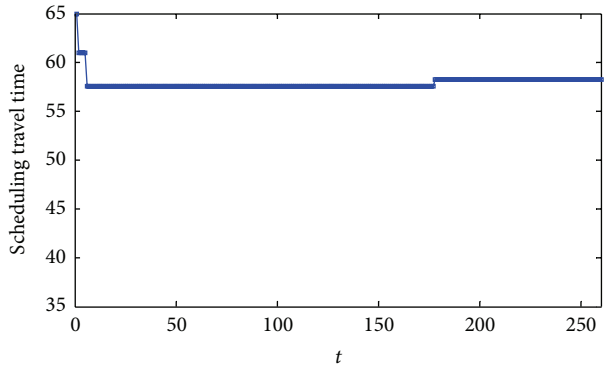


Figure 9: Day-to-day scheduling travel time when $Toltime = 20$ and $Rv = (b - a)/6$.

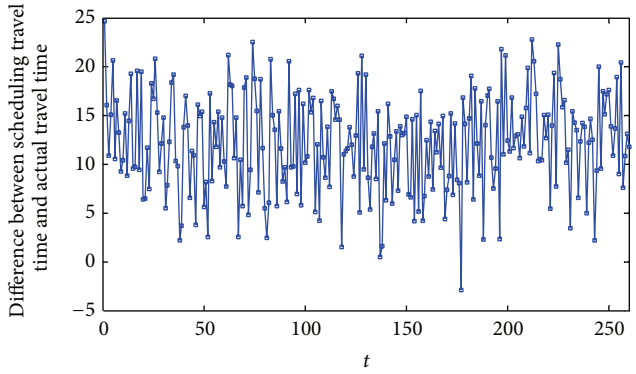


Figure 10: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 20$ and $Rv = (b - a)/6$.

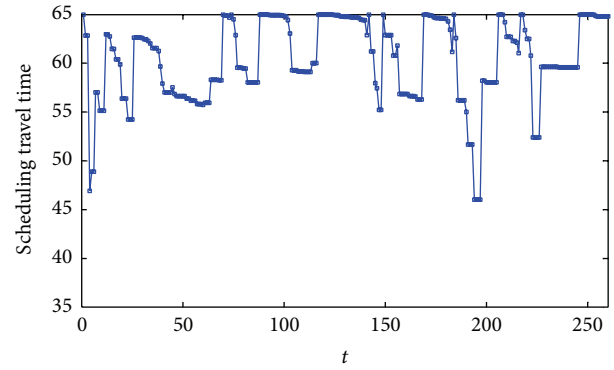


Figure 11: Day-to-day scheduling travel time, when $Toltime = 15$ and $Rv = (b - a)/3$.

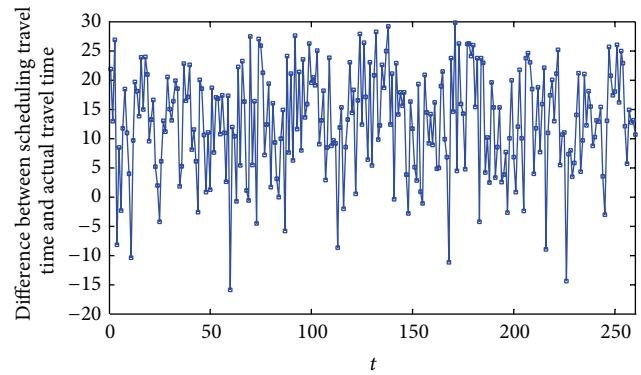


Figure 12: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 15$ and $Rv = (b - a)/3$.

the commuter's past experiences more appropriately, the approach considers the numbers of late arrivals, punctual arrivals, and early arrivals on a previous day.

Summing up the above description and analysis to mathematical formulation, the scheduling travel time of day t is given by

$$\begin{aligned} TP_t = {} & I_{L,t-1} \\ & \cdot \min\left(TP_{t-1} + \left(1 + \frac{e^{N_{Lt}}}{e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}}}\right)\right. \\ & \left. \cdot \left[(b-a) \cdot \left(1 - \sqrt{2(1-\mu_1)}\right)\right], b\right) \\ & + I_{E,t-1} \cdot \max\left(TP_{t-1} + \frac{e^{N_{Et}}}{e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}}}\right. \\ & \left. \cdot \left[(TP_{t-1} - \mathrm{TR}_{t-1}) \cdot \left(\sqrt{2\mu_2} - 1\right)\right], a\right) \\ & + I_{P,t-1} \cdot TP_{t-1}, \end{aligned} \tag{1}$$

where TP_t is the scheduling travel time selected by the commuter on day t, TR_t is the actual travel time experienced by the commuter on day t, $I_{L,t}$ is an indicator parameter whose value is one if the commuter is late for his commute on day t (i.e., $TP_t < TR_t$), otherwise zero, and $I_{P,t}$ is another indicator parameter whose value is one if the commuter arrives at destination punctually on day t, otherwise zero. Note that the punctuality here means that the commuter arrives before the requested arrival time within a tolerable time range. This says that $0 \le TP_t - TR_t \le Toltime$, where $Toltime$ is the longest time that is acceptable for the commuter to arrive early, that is, the maximum time the commuter can tolerate if he arrives earlier than the requested time. $I_{E,t}$ is an indicator parameter whose value is one if the commuter arrives early on day t; that is, $TP_t - TR_t \ge Toltime$. N_{Lt}, N_{Pt}, and N_{Et} are the numbers of late arrivals, punctual arrivals, and early arrivals before the day t, respectively. μ_1 is a random parameter following a uniform distribution between 0.5 and 1; μ_2 is also a random parameter distributed uniformly between 0 and 0.5. a and b are, respectively, the lower and upper bounds of the scheduling travel time. Note again that the actual travel time TR_t is assumed to be a random variable that ranges from a to infinity governed by a normal density distribution with mean Rm and standard deviation Rv.

Model (1) is explained as follows. If the commuter arrives punctually on day $t - 1$, he spontaneously remains with

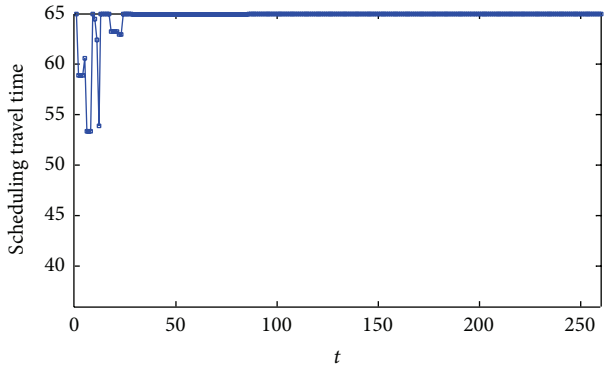


Figure 13: Day-to-day scheduling travel time when $Toltime = 20$ and $Rv = (b - a)/3$.

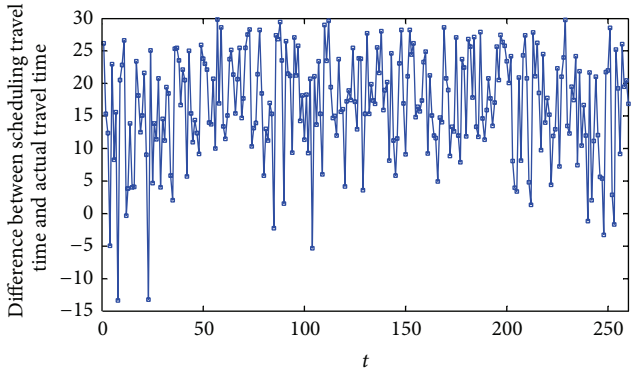


Figure 14: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 20$ and $Rv = (b - a)/3$.

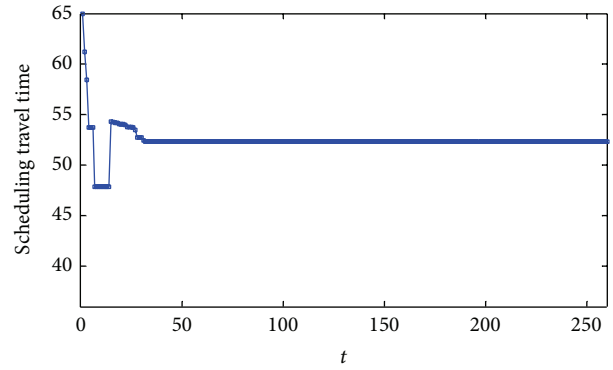


Figure 15: Day-to-day scheduling travel time when $Toltime = 10$ and $Rv = (b - a)/12$.

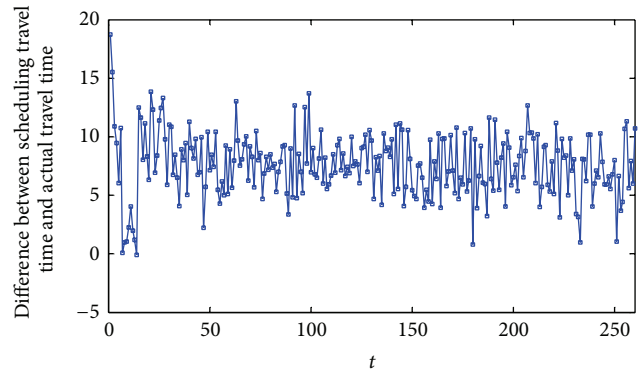


Figure 16: Day-to-day difference between scheduling travel time and actual travel time when $Toltime = 10$ and $Rv = (b - a)/12$.

the same scheduling travel time on day t. If the commuter arrives late on day $t-1$, he naturally intends to depart earlier on day t than on day $t-1$. So, the scheduling travel time should increase on day t, which is reflected by

$$TP_{t-1} + \left(1 + \frac{e^{N_{Lt}}}{e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}}}\right) \cdot \left[(b-a) \cdot \left(1 - \sqrt{2\left(1-\mu_1\right)}\right)\right]. \tag{2}$$

In the above, the second item represents the extension of the scheduling travel time for considering the commuter's past experiences and randomness. $(1 + (e^{N_{Lt}}/(e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}})))$ states that the numbers of late arrivals, punctual arrivals, and early arrivals are involved into the adjustment. $(b-a)$ denotes the potential interval of the adjustment and $(1 - \sqrt{2(1-\mu_1)})$ displays the randomness characteristics of human choices. On the other hand, it is impossible for a commuter to leave home too early. For instance, if a commuter is required to arrive at a destination at 8:00 am, he or she certainly will not leave home before 1:00 am. Then, TP_t should logically have an upper bound. Therefore, when the commuter arrives late on day $t-1$, $\min(TP_{t-1} + (1 + (e^{N_{Lt}}/(e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}}))) \cdot [(b-a) \cdot (1 - \sqrt{2(1-\mu_1)})], b)$ can correctly depict the adjustment of scheduling travel time on day t. If the commuter arrives early, the second item in (1) gives the adjustment on scheduling travel time. The major difference between the first and the second items of the equation is that the adjustment of scheduling travel time in the case of arriving late mainly considers $(b-a)$ for avoiding penalty but $TP_{t-1} - TR_{t-1}$ in the case of arriving early.

Equation (1) can be rewritten as

$$\begin{aligned} TP_t = {} & I\left(TP_{t-1} < TR_{t-1}\right) \\ & \cdot \min\left(TP_{t-1} + \left(1 + \frac{e^{N_{Lt}}}{e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}}}\right) \cdot \left[(b-a) \cdot \left(1 - \sqrt{2\left(1-\mu_1\right)}\right)\right], b\right) \\ & + I\left(TP_{t-1} - TR_{t-1} \geq Toltime\right) \\ & \cdot \max\left(TP_{t-1} + \frac{e^{N_{Et}}}{e^{N_{Lt}} + e^{N_{Et}} + e^{N_{Pt}}} \cdot \left[\left(TP_{t-1} - TR_{t-1}\right) \cdot \left(\sqrt{2\mu_2} - 1\right)\right], a\right) \\ & + I\left(0 \leq TP_{t-1} - TR_{t-1} \leq Toltime\right) \cdot TP_{t-1}, \end{aligned} \tag{3}$$

where $I(\cdot)$ is a step function whose value is one if the event in the parenthesis occurs, zero otherwise.

4. Simulation

This section aims at illustrating the model's capability of depicting the commuters' scheduling travel time adjustment behavior and revealing some properties of the adjustment behavior by changing the parameter values. The evolution of the day-to-day scheduling travel times and the trend of the day-to-day departure times will be presented. The maximum value of the time index t is set to be 260 (basically the number of working days in one year).

Suppose the lower bound of the scheduling travel time a is 35 minutes and the upper bound b is 65 minutes. The actual travel time is assumed to be a random variable that ranges from 35 to infinity with density the same as the normal density, with mean Rm and standard deviation Rv within this range. Because there exist differences between actual travel time distribution and perceived travel time distribution (Chen et al. [10]), the mean of the actual travel time distribution is generally less than the mean of the perceived travel time distribution. Consequently, if the scheduling travel time and the actual travel time are both considered to be random variables, the mean of the actual travel time should be less than the mean of the scheduling travel time distribution. For simplicity, Rm is assumed to be less than $0.5(a + b) - 5$ in this study. Furthermore, some events (e.g., wide geographical bumper-to-bumper traffic jams) are with small probability and can be handled by the three-sigma rule or an empirical rule which states that nearly all values of a random variable are in a range covering 3 standard deviations from the mean. So, Rv can be logically presumed to take $(b - a)/6$ (this is from $0.5(a + b) - a = 3Rv$).

First, let *Toltime* be equal to 5. Figure 1 displays the scheduling travel time and Figure 2 the difference between scheduling travel time and actual travel times. It can be seen that there are wide variations concerning the changes of scheduling travel time against day index. This implies that the departure time is selected nearly at random. Simulation results show that the numbers of late arrivals, punctual arrivals, and early arrivals in these 260 days are 56, 73, and 131, respectively. So, wide variations indeed occur. The reason is that the value of *Toltime* is too small, which makes commuters obligated to arrive punctually. Therefore, small tolerance leads to big variations in departure time.

Second, let *Toltime* be 10. Figures 3 to 6, respectively, exhibit two scenarios that are mostly likely to occur under the assumptions; $Toltime = 10$ and $Rv = (b - a)/6$. Because the tolerable time, *Toltime*, is increased, there are relatively small variations in the scheduling travel time and a relatively narrow range of the difference between scheduling travel time and actual travel time, compared to the results shown in Figures 1 and 2. As shown in Figures 3 and 4, the scheduling times are established in a relatively consistent daily schedule but still disrupted by small events due to the randomness of actual travel time. Accordingly, the numbers of late arrivals, punctual arrivals, and early arrivals are around 19, 121, and 120, respectively. Nevertheless, the similar result shown in Figures 1 and 2 likely occurs in Figures 5 and 6, where the numbers of late arrivals, punctual arrivals, and early arrivals are around 30, 122, and 108, respectively. Both scenarios may occur, which depends on the actual travel time sourced from reality.

Third, further enlarge *Toltime* to 15. As shown in Figures 7 and 8, the scheduling travel time is likely to be fixed and be disrupted occasionally, and the difference between scheduling travel time and actual travel time is not as large as that when *Toltime* is small. In addition, the number of late arrivals is only around 6; the number of punctual arrivals stands at more than 150.

Fourth, let *Toltime* be 20. Figures 9 and 10 show the simulation results. It can be seen that the scheduling travel time is almost fixed and hardly disrupted. Due to the use of a large *Toltime*, the number of late arrivals is very small, but the number of punctual arrivals is almost up to 244.

In the previous analyses, we set Rv as $(b - a)/6$. When Rv is changed to $(b - a)/3$, there are two representative scenarios. Let *Toltime* be equal to 15 and 20, respectively. As the standard deviation of actual travel time distribution increases, the results shown in Figures 11 and 12 are analogues to those in Figures 5 and 6 (where $Toltime = 10$ and $Rv = (b-a)/6$), respectively. The numbers of late arrivals, punctual arrivals, and early arrivals in these two cases are likely to be equal, respectively. The appearance exhibited in Figure 13 is like that in Figure 9, but the scheduling travel time in the former approaches 65 while 58 in the later, and there exist significant differences between the numbers of late arrivals, punctual arrivals, and early arrivals in these two cases (in the case of Figures 13 and 14, the number of punctual arrivals is about 144).

Finally, let Rv be $(b - a)/12$, which means that the actual travel time varies a little. The scheduling travel time approaches 53 even if *Toltime* is 10, as shown in Figure 15. The difference between scheduling travel time and actual travel time varies within a narrow range, as exhibited in Figure 16. The reason is that the actual travel time remains at a stable state, so that commuters easily adjust their scheduling travel times to match the actual travel time. The number of punctual arrivals is about 209.

5. Conclusions

In this paper, the day-to-day scheduling travel time adjustment behavior (also, the day-to-day departure time adjustment behavior) is studied. A modeling framework is constructed to formulate the adjustment. The experiences of the last day's arrival (late arrival, punctual arrival, or early arrival) are used to influence today's behavior. The randomness of the commuters' behavior is also considered. Simulation results support the effectiveness of the modeling framework. Two parameters, that is, the maximum time that a commuter can tolerate if he or she arrives earlier than the requested time and the standard deviation of the actual travel time distribution, are numerically investigated. Results show that they have big impacts on scheduling travel time adjustment behavior.

Conflict of Interests

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

Acknowledgments

This study has been supported by the National Natural Science Foundation of China (Grant no. 71201178) and the Fundamental Research Funds for the Central Universities (Grant no. CDJZR12170009). And it was partly funded by the National Natural Science Foundation of China (Grant no. 71161005) and the Nomarch Foundation of Guizhou Province (Grant no. 2011067).

References

[1] F. Webster, "Traffic signal settings," Road Research Technical Paper no. 39, Road Research Laboratory, Berkshire, UK, 1958.

[2] R. E. Allsop, "Delay at fixed time traffic signals—I: theoretical analysis," *Transportation Science*, vol. 6, no. 3, pp. 260–285, 1972.

[3] National Research Council, *Highway Capacity Manual*, Transportation Research Board, Washington, DC, USA, 2000.

[4] A. Skabardonis and N. Geroliminis, "Real-time estimation of travel times on signalized arterials," in *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, pp. 387–406, University of Maryland, College Park, Md, USA, 2005.

[5] F. Guo, H. Rakha, and S. Park, "Multistate model for travel time reliability," *Transportation Research Record*, vol. 2188, pp. 46–54, 2010.

[6] S. Park, H. Rakha, and F. Guo, "Multi-state travel time reliability model: impact of incidents on travel time reliability," in *Proceedings of the 14th IEEE International Intelligent Transportation Systems Conference (ITSC '11)*, pp. 2106–2111, Washington, DC, USA, October 2011.

[7] Y. Feng, G. Davis, and J. Hourdos, "Arterial travel time characterization and real-time traffic condition identification using GPS-equipped probe vehicles," in *Proceedings of the 90th Transportation Research Board Annual Meeting*, Washington, DC, USA, 2011.

[8] Y. Li and M. McDonald, "Link travel time estimation using single GPS equipped probe vehicle," in *Proceedings of the 5th IEEE International Conference on Intelligent Transportation Systems*, pp. 932–937, Singapore, 2002.

[9] G.-L. Chang and H. S. Mahmassani, "Travel time prediction and departure time adjustment behavior dynamics in a congested traffic system," *Transportation Research B: Methodological*, vol. 22, no. 3, pp. 217–232, 1988.

[10] A. Chen, Z. Zhou, and W. H. K. Lam, "Modeling stochastic perception error in the mean-excess traffic equilibrium model," *Transportation Research B: Methodological*, vol. 45, no. 10, pp. 1619–1640, 2011.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 508039, 11 pages
http://dx.doi.org/10.1155/2014/508039

Research Article

Traffic Incident Clearance Time and Arrival Time Prediction Based on Hazard Models

Yang beibei Ji,[1] Rui Jiang,[2] Ming Qu,[2] and Edward Chung[2]

[1] *School of Management, Shanghai University, Shangda Road 99, Shanghai, China*
[2] *Smart Transport Research Centre, Queensland University of Technology, Level 8, P Block, Brisbane, QLD 4001, Australia*

Correspondence should be addressed to Yang beibei Ji; jybb@shu.edu.cn

Received 25 February 2014; Accepted 31 March 2014; Published 17 April 2014

Academic Editor: X. Zhang

Accurate prediction of incident duration is not only important information of Traffic Incident Management System, but also an effective input for travel time prediction. In this paper, the hazard based prediction models are developed for both incident clearance time and arrival time. The data are obtained from the Queensland Department of Transport and Main Roads' STREAMS Incident Management System (SIMS) for one year ending in November 2010. The best fitting distributions are drawn for both clearance and arrival time for 3 types of incident: crash, stationary vehicle, and hazard. The results show that Gamma, Log-logistic, and Weibull are the best fit for crash, stationary vehicle, and hazard incident, respectively. The obvious impact factors are given for crash clearance time and arrival time. The quantitative influences for crash and hazard incident are presented for both clearance and arrival. The model accuracy is analyzed at the end.

1. Introduction

Traffic incident is considered as one of the major factors which cause traffic congestion and delay. At the same time traffic incident has wide negative impact on both traffic system and related social activities. Many studies have been done to predict, estimate, or try to decrease traffic duration time just because the accurate prediction of incident time can (1) reduce incident duration time, (2) associate Traffic Incident Management (TIM) to quickly respond to incident to mitigate the impact of incidents (Chou [1]), and (3) improve travel time reliability by predicting travel time while occurrence of incident (Tsubota et al. [2]).

Traffic incident is nonrecurrent events which cause a capacity reduction or an abnormal increase in traffic demand, such as crash accident, stalled vehicles, debris, fire, construction, and sporting events. A general incident timeline as shown in Figure 1 reveals that incident duration can be divided into verification, response, clearance, and recovery period by recording timestamps at various stage of an incident.

However, most of the prediction models did not include all four parts or did not give the exact definition of incident time. Each part of incident time has statistic distribution and has different influence factors. An example is given in Figure 2.

Figure 2 is the occupancy contour plot from 3:00 a.m. to 11:00 a.m., in which the y-axis denotes the distance from start point of the freeway and x-axis indicates the time. The green indicates low occupancy (free flow conditions), the yellow indicates increasing congestion, and the red represents much congestion on the links. The incident is marked as the red area where the incident occurred time (7:00 a.m.), cleared time (7:53 a.m.), and traffic recovered time (8:20 a.m.) are clearly marked. According to the incident data base (SIMS), a multiple vehicle crash incident on 22 July in 2011 was verified on 7:39 a.m. and was cleared on 7:53 a.m., and the tow assistance arrived on 7:47 a.m.. The total duration time was 80 minutes calculated based on real traffic flow data, while the duration time was only 14 minutes according to incident data base. The real duration time was almost 6 times of incident data base, resulting in errors of prediction model.

However, the accuracy of the incident prediction model cannot be improved, which is partly caused by the definition of the incident duration. It is difficult to get real incident duration time as shown in Figure 1, because the incident

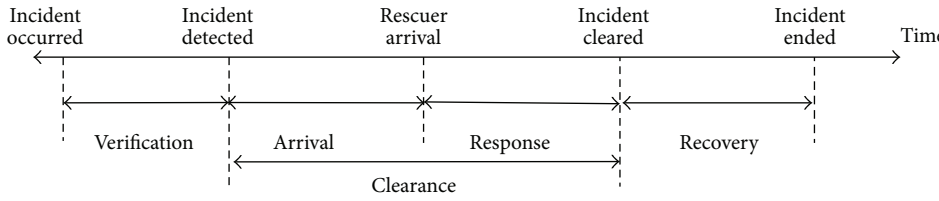


Figure 1: Traffic incidents timeline.

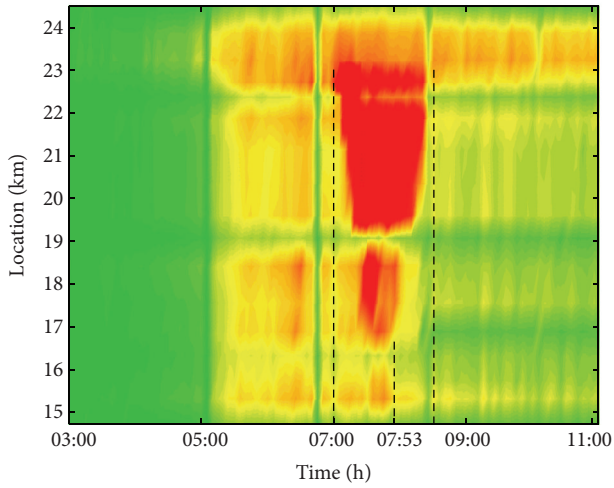


Figure 2: Contour plot on the incident scene.

occurred time and the incident ended time cannot be collected exactly. In this paper, the clearance time and arrival time are investigated individually. The major impact factors for the clearance time and arrival time are studied, and two prediction models are then developed. The comparison of different impact factors is put forward.

The rest of this paper is structured as follows. The introduction of Incident Management System (SIMS) in the Pacific Motorway, Brisbane, Australia, and the incident parameter properties are represented in Section 3. Section 4 describes model development: firstly the hazard based model is introduced; secondly, the best statistic fitting distribution for 3 types of incident is given for both clearance time and arrival time; thirdly, the obvious impact factors of the hazard based model are given separately after data filter; finally, the quantitative impact evaluation and precision are obtained. The conclusions are summarized in Section 5.

2. Literature Review

Over the past several decades, various models have been developed to predict traffic incidents. Most of these studies are based on statistics theory. These representative prediction models can be classified into the following categories: probabilistic distribution, linear regression analyses, time sequential methods, decision tree models, neural network models, Bayesian classifier, hazard analysis model, and so forth.

Golob et al. [3] provided the first probabilistic distribution of incident duration. They found that each part of duration time was related to the previous duration time and demonstrated that the duration fits a lognormal distribution. Giuliano [4] extended Golob's work by applying a lognormal distribution among similar type of incidents and reported that whether involving truck was the key impact factors of incident duration. Garib et al. [5] and Sullivan [6] did similar research. Jones et al. [7] developed double logarithmic distribution of incident duration time. Nam and Mannering [8] found that the Weibull distribution can also be used to describe incident data.

Linear regression models have been widely used due to their simplicity and effectiveness. The relationship between impact factors and incident duration can be obviously reflected in regression models [5, 9]. Ozbay and Kachroo [10] developed incident clearance time based on regression method using 121 traffic incident data collected from Chicago, USA. Nine statistically significant variables were put forward, but the accuracy and effectiveness of the model were not given. Kau [11] estimated freeway incident duration using multiple linear regression method, with confidence interval of 95%. He defined the incident clearance time as the duration from the time that a police vehicle or freeway service patrol vehicle arrives at the scene until the vehicles are cleared from the scene.

The advantage of time sequential model is that it can do regression prediction using a few incident property variables at the early stage of the incident and can update the prediction result when more information is collected. Khattak et al. [12] found that incident type and severity were the most significant factors. He applied time sequential model based on small incident sample, but the practicability had not been demonstrated because of poor accuracy and little incident data.

The decision tree model is independent of distribution of dependent variables. Sethi et al. [13] indicated that the average duration time was 21 minutes based on 801 incident data, including data set of traffic interruption, disable vehicle, and severity. The results showed that the incident type was the most obvious factor on duration. Ji et al. [14] set up the decision tree model based on Bayesian using 1853 incident data in Utrecht, the Netherlands. The precision was improved to 73.39% compared to the decision model. Stephen et al. [15] developed an incident duration model based on a naïve Bayesian classifier. He emphasized that incident duration was

a highly variable quantity and although the model performed better than a linear regression, its classification was still correct only in half of the time.

Artificial neural networks (ANN) have been widely used for prediction and pattern classification problems. Lopes et al. [16] presented an adaptive model to forecast the clearance time of real time traffic incidents. The solutions included four models which were calibrated and tested by incident records from Portuguese highways. The performance showed that it was able to estimate 72% of incident with less than 10 minutes error and about 92% with less than 20 minutes error. Some other examples can be found in [17, 18].

The hazard analysis model has been used in traffic engineering, which is a common topic in many fields such as life sciences, biomedical, and reliability engineering. The model is more effective to analyze time-related problem, which is generally used to describe the analysis of data in the form of time from a well-defined time origin until the occurrence of some particular event of an end point [19]. Examples include the time between incident occurrence and its clearance [8, 20], the time between planning and execution of an activity [21], and the analysis of urban travel time [22].

Furthermore, hazard-based model has been used to model incident duration time. Chung [23] presented an accident duration model using 2-year-accident dataset from 2006 to 2007 in Korean freeway systems, and the Log-logistic distribution was selected for accelerated failure time metric model. Although the model had large prediction error, statistical test results indicated that this model was stable over time. Tavassoli et al. [24] developed parametric accelerated failure time survival models of incident duration. They found that the duration of each type of incident is uniquely different and responds to different factors.

One distinctive feature of hazard based model is that the model precision will be improved if the best fitting distribution of time variable is chosen. In this study, hazard based models, in particular the accelerated failure time (AFT) metric, are utilized to model both incident clearance time and arrival time.

3. Description of Incident Base

Incident data was collected by Queensland Department of Transport and Main Roads' STREAMS Incident Management System (SIMS) for South East Queensland urban networks from November 2009 to November 2010. SIMS is an incident management system which is used throughout Queensland, Australia, to capture traffic incidents which cause an impact on traffic flow on the road network. There are total 35103 incident data for one year, which can be classified into 9 types: alert, congestion, crash, fault, flood, hazard, planned incident, road works, and stationary vehicles. There are many detailed properties in SIMS incident data base, but not all of them are closely related to incident time prediction, such as location, SIMS ID, and status. Hence, the major properties of each incident data are shown in Table 1. However, not all these properties are recorded for each incident occurrence. For example, the parameters are only applicable for crash data which are "number of vehicles involved," "number of people injured," and "number of fatalities".

Only 3 types of incidents: crash, stationary vehicle, and hazard are used to model development though 9 types of incident recorded in SIMS data base. Other incident type data are rare recorded. Consequently, the clearance time and arrival time prediction model are only developed for 3 types of incident.

4. Model Development

Hazard based time models were originally used for problems in biomedical, engineering, and social sciences, which are a class of statistical methods for studying the occurrence and timing of events. Recently, they were used to model time related issues in transportation. A review of the application of the hazard based duration models in transportation up to the early 1990s [25].

The incident time in hazard based model is a realization of a continuous random variable T, with a cumulative distribution function $F(t)$, which is called the failure function. A probability density function $f(t)$, survival function $S(t)$, and hazard function $h(t)$ are given as (2)–(4). The relationships between these four functions are formulated in (1)–(4), and $P(\cdot)$ means probability. The function of a random variable T is given by

$$F(t) = \int_0^t f(u)\,du = P(T < t), \tag{1}$$

$$f(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t)}{\Delta t}, \tag{2}$$

$$S(t) = P(T \ge t) = 1 - F(t), \tag{3}$$

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \lim_{\Delta t \to 0} \frac{P(t + \Delta t \ge T \ge t)}{\Delta t}. \tag{4}$$

In (4), with fully parametric models, three distributional alternatives were considered, namely: Gamma, Log-logistic, and Weibull, for the hazard function and are tested to find the best fit to the incident clearance time and arrival time. The functional forms of the hazard function for each model can be derived by using each distribution model and general function.

4.1. Gamma Distribution. The Gamma distribution is briefly described as a two-parameter family of continuous probability distributions. The scale parameter is λ and the shape parameter is $\widetilde{k}$, where $\widetilde{k} > 0$ and $\lambda > 0$. The Gamma function is mathematically defined as [26]

$$\Gamma\left(\widetilde{k}\right) = \int_0^\infty t^{\widetilde{k}-1} e^{-t} dt. \tag{5}$$

After algebra transform, the p.d.f. (probability density function) of the Gamma distribution, generally written as $f[t; T \sim \Gamma(\lambda, \widetilde{k})]$, is given by

$$f\left[t; T \sim \Gamma\left(\lambda, \widetilde{k}\right)\right] = \frac{\lambda(\lambda t)^{\widetilde{k}-1} e^{-\lambda t}}{\Gamma\left(\widetilde{k}\right)}, \quad t > 0. \tag{6}$$

Table 1: Incident property information.

Parameter	Description information
Priority	
Classification	The classification to further define the type of incidents
Blockage type	The type of blockage caused by the traffic incident. Possible values are unknown, no blockage, partially blocked, blocked, both direction blocked
Number of lanes blocked	The number of lanes blocked by the traffic incident
Fire hazard	If hazard resulted in smoke/poor/visibility and/or fire
Intersection	Whether nearby the intersection
Major incident	Whether the incident is major
Lane blocked	How many lanes are blocked
Heavy traffic	Whether heavy traffic involved
Towing	Whether towing is needed
Number of vehicles involved	The number of vehicles involved in the crash. (only applicable for crash incident data)
Number of people injured	The number of people injured in the crash. (only applicable for crash incident data)
Number of fatalities	The number fatalities from the crash
Medical emergency	Yes/no
Weather	Possible values are cloudy, electrical storm, fine, fog, hail, heavy rain, showers, snow, unknown, wind. This column will always display "Unknown" for weather hazard incident.
Opp lanes blocked	The number of lanes blocked by the traffic incident, on the opposite link.
Diversion required	Indicates whether a diversion was required due to the traffic incident
Lateral position	Possible values are in bay, in lanes, in median, left shoulder, right shoulder, and unknown.
Chemical spill	This is only applicable to crash incidents and debris/obstruction/spill.

When $\widetilde{k} = 1$, the Gamma density function reduces to the exponential density function, and the exponential distribution is also a special case of the Gamma distribution.

When $\lambda = 1$, (5) reduces to the one-parameter Gamma distribution, also referred to as the standard Gamma distribution of T, written as

$$f\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right] = \frac{t^{\widetilde{k}-1}e^{-t}}{\Gamma\left(\widetilde{k}\right)}, \quad t > 0 \tag{7}$$

with its c.d.f. (cumulative distribution function) defined as

$$F\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right] = \int_0^t \frac{t^{\widetilde{k}-1}e^{-t}}{\Gamma\left(\widetilde{k}\right)}, \quad t > 0. \tag{8}$$

Specification of the survival and hazard functions for the Gamma distribution are based on (8), which is called the incomplete Gamma function. The survival function $S(t)$ is given by the following equation:

$$S\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right] = 1 - F\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right] = 1 - \int_0^t \frac{t^{\widetilde{k}-1}e^{-t}}{\Gamma\left(\widetilde{k}\right)}. \tag{9}$$

The Gamma distribution hazard function can be expressed as

$$\begin{aligned} h\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right] &= \frac{f\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right]}{S\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right]} \\ &= \frac{\lambda(\lambda t)^{\widetilde{k}-1}e^{-\lambda t}}{\Gamma\left(\widetilde{k}\right)\left\{1 - F\left[t; T \sim \Gamma\left(\widetilde{k}\right)\right]\right\}}. \end{aligned} \tag{10}$$

4.2. Weibull Distribution. The Weibull distribution model is almost the most widely applied parametric function in survival analysis because of its flexibility and simplicity among all the families of parametric time distributions [26].

The Weibull probability of event time T, a continuous function, is featured by the use of two parameters: a scale parameter λ and a shape parameter $\widetilde{p}$. The survival function with the Weibull distribution is given by

$$S\left(t; \lambda, \widetilde{p}\right) = \exp\left[-(\lambda t)^{\widetilde{p}}\right], \quad t > 0. \tag{11}$$

Given the intimate associations among various lifetime measures, the hazard function in the Weibull distribution can be readily derived from the above equation.

Consider

$$h(t;\lambda,\widetilde{p}) = -\frac{(d/dt)\,e^{-(\lambda t)^{\widetilde{p}}}}{e^{-(\lambda t)^{\widetilde{p}}}} = \frac{\lambda\widetilde{p}t^{\widetilde{p}-1}\exp\left[-(\lambda t)^{\widetilde{p}}\right]}{\exp\left[-(\lambda t)^{\widetilde{p}}\right]} = \lambda\widetilde{p}(\lambda t)^{\widetilde{p}-1}. \tag{12}$$

The cumulative hazard function $H(t)$ can be expressed in terms of $S(t)$, given by

$$H(t) = -\log S(t) = -\log\left\{\exp\left[-\int_0^t h(u)\,du\right]\right\}. \tag{13}$$

Therefore, the cumulative hazard function $H(t;\lambda,\widetilde{p})$ can be written as

$$H(t;\lambda,\widetilde{p}) = -\log S(t;\lambda,\widetilde{p}) = -\log\left\{\exp\left[-(\lambda t)^{\widetilde{p}}\right]\right\} = (\lambda t)^{\widetilde{p}}. \tag{14}$$

Taking natural Log values on both sides of (14), (14) can be written as

$$\log\left[-\log S(t;\lambda,\widetilde{p})\right] = \log\lambda + \widetilde{p}\log t. \tag{15}$$

Specifications of $S(t)$ and $h(t)$ lead to the following equation for the Weibull p.d.f. function:

$$f(t;\lambda,\widetilde{p}) = h(t)\,S(t) = \lambda\widetilde{p}(\lambda t)^{\widetilde{p}-1}\exp\left[-(\lambda t)^{\widetilde{p}}\right]. \tag{16}$$

Likewise, the c.d.f. at time t is derived by

$$F(t;\lambda,\widetilde{p}) = 1 - S(t;\lambda,\widetilde{p}) = 1 - \exp\left[-(\lambda t)^{\widetilde{p}}\right]. \tag{17}$$

Given $\lambda^* = 1/\lambda$, the Weibull hazard function can be reexpressed as

$$h(t;\lambda^*,\widetilde{p}) = \frac{\widetilde{p}}{\lambda^*}\left(\frac{t}{\lambda^*}\right)^{\widetilde{p}-1}. \tag{18}$$

4.3. Log-Logistic Distribution. The lognormal distribution is widely used to describe events whose rate increases initially and decreases consistently afterwards. The Log-logistic distribution of T is the antilogarithm of the familiar logistic distribution. Let $Y = \log T$. The density function of Y is defined as the familiar logistic distribution [26]:

$$f(y) = \frac{\widehat{b}^{-1}\exp\left[(y-\widetilde{\mu})/\widehat{b}\right]}{\left\{1+\exp\left[(y-\widetilde{\mu})/\widehat{b}\right]\right\}^2}, \quad y\in(-\infty,\infty), \tag{19}$$

where $\widetilde{\mu}$ and $\widehat{b}$ are parameters for the logistic function of Y, described as $Y \sim \text{Logist}(\widetilde{\mu},\widehat{b})$. Let $\lambda = \exp(\widetilde{\mu})$ and $\widehat{p} = \widehat{b}^{-1}$. The antilogarithm of (19) is the density function of T:

$$f(t) = \frac{(\widehat{p}/\lambda)(t/\lambda)^{\widehat{p}-1}}{\left[1+(t/\lambda)^{\widehat{p}}\right]^2}, \quad t>0, \tag{20}$$

where λ and $\widehat{p}$ are parameters of the Log-logistic distribution, written as $T \sim \text{LLogist}(\lambda,\widehat{p})$. The c.d.f. of T is then given as

$$F\left[t;T\sim\text{LLogist}(\lambda,\widehat{p})\right] = \frac{1}{1+(t/\lambda)^{-\widehat{p}}}. \tag{21}$$

Therefore, the survival and hazard rate functions of T can then be readily derived as follows:

$$S\left[t;T\sim\text{LLogist}(\lambda,\widehat{p})\right] = 1 - F\left[t;T\sim\text{LLogist}(\lambda,\widehat{p})\right] = \frac{1}{1+(t/\lambda)^{\widehat{p}}}$$
$$h\left[t;T\sim\text{LLogist}(\lambda,\widehat{p})\right] = \frac{f\left[t;T\sim\text{LLogist}(\lambda,\widehat{p})\right]}{S\left[t;T\sim\text{LLogis}t(\lambda,\widehat{p})\right]} = \frac{(\widehat{p}/\lambda)(t/\lambda)^{\widehat{p}-1}}{1+(t/\lambda)^{\widehat{p}}}. \tag{22}$$

5. Model Result

5.1. The Fitness of Distribution. Understanding of incident characteristics and patterns is essential to establish an appropriate prediction model; therefore, the statistical analysis is carried out firstly. There are 4966 crash records, 15791 stationary vehicle data, and 3847 hazard records for clearance time which are used to do distribution fitting analysis. Four probability density functions, which are Gamma, Log-logistic, Weibull, and lognormal, are fitted to the clearance time for crash, stationary vehicle, and hazard incidents, respectively, (see Figures 3(a), 3(b), and 3(c)). Thick full lines indicate the best fitness distribution. The figures indicate that each incident classification has its respective best fitness distribution function. Four parameters estimates of clearance time probability density distribution: Log likelihood, domain mean, and variance are listed in Table 2. Log likelihood and variance statistics indicate the goodness of fit distribution.

The less the variance is, the better the distribution fitting will be. For example the Gamma distribution variance for crash clearance time is 967.94, which is the least one comparing other distributions. It is clearly shown that Gamma distribution is best fit for crash clearance time, Log-logistic for stationary vehicle, and Weibull for the hazard.

There are totally 4569 crashes, 14665 stationary vehicle data, and 3382 hazard records for arrival time which are used in this distribution fitting. The number of arrival time record is less than the counterpart for clearance time, because there exists an abundant of invalid arrival time data records in SIMS. All the invalid and defective data are filtered. Figures 4(a), 4(b), and 4(c) represent the probability density distributions of arrival time for crash, stationary vehicle, and hazard separately. Table 3 lists the parameters estimates of arrival time probability density distribution for each incident type. Both the estimate parameters and the figure indicate that the Gamma distribution is best fit for crash arrival time, Log-logistic for stationary vehicle, and Weibull for the hazard.

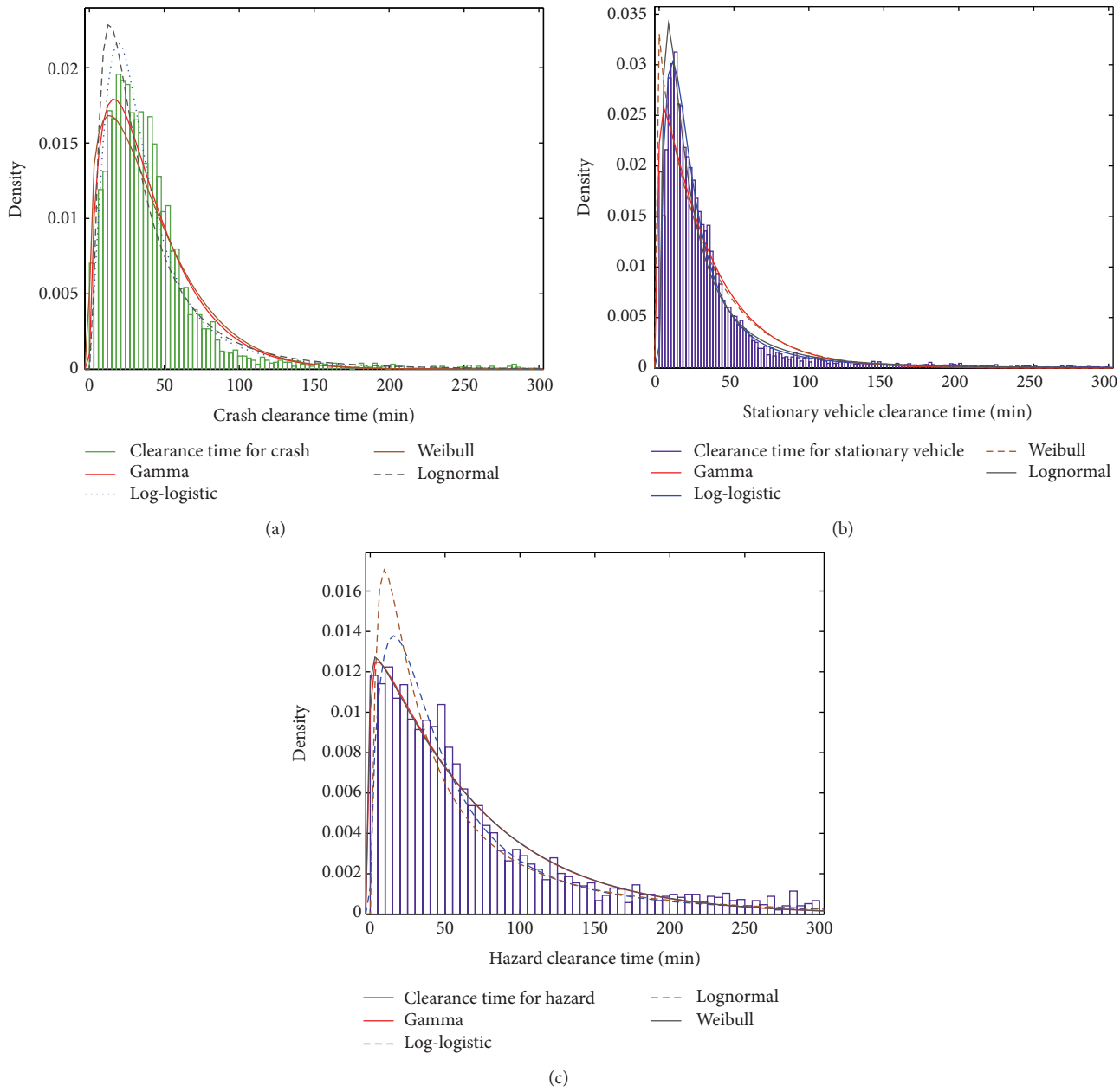


Figure 3: Probability density distribution of clearance time for crash (a), stationary vehicle (b), and hazard (c).

In summary, the clearance time and arrival time for the same incident classification follow the same probability distribution, but with different estimated parameters. Different incident classification has different probability distribution characteristics for both clearance time and arrival time. The best fitting model was selected for each incident type based on the results above to develop hazard based prediction model.

5.2. Hazard Based Model for Crash Clearance Time and Arrival Time. The crash clearance time and arrival time are developed based on the Gamma distribution survival model. Data filter is carried out before model development in order to improve model precision. For example, the clearance times longer than 3 hours, accounting for less than 5% of the total crash dataset, are filtered as shown in the green bars in Figure 5 in order to improve the model accuracy.

Tables 4 and 5 list the parameter estimate results for model estimated for crash incident clearance time and arrival time. A positive sign of an estimate coefficient suggests an increase in the incident clearance time and a decrease in hazard function associated with an increase in that property variable.

All variables are statistically significant at a 95% confidence level. Therefore, all 14 significant property variables for crash clearance time are listed in Table 4. However, there are

Table 2: Parameters estimates of clearance time probability density distribution fitness for each incident type.

Incident type	Parameters	Gamma	Log-logistic	Weibull	Lognormal
Clearance time for crash	Log likelihood	**−23032.7**	−23007.3	−23128.9	−23202.9
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$
	Mean	40.3562	45.586	40.5906	42.8821
	Variance	**967.94**	11380.3	1020.94	2143.83
Clearance time for stationary vehicle	Log likelihood	−71707.57	**−70991.8**	−71724.7	−71306.9
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$
	Mean	33.2276	39.749	33.189	35.3951
	Variance	1032.58	**975.23**	1142.8	2897.44
Clearance time for hazard	Log likelihood	−20244.6	−20437.6	**−20245.9**	−20527.5
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$
	Mean	68.5289	103.329	68.5441	83.078
	Variance	4382.43	4478.4	**4357.8**	22634.2

Table 3: Parameters estimates of arrival time probability density distribution fitness for each incident type.

Incident type	Parameters	Gamma	Log-logistic	Weibull	Lognormal
Arrival time for crash	Log likelihood	**−18182.7**	−18211	−18215.4	−18207.7
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$
	Mean	19.9232	25.3389	19.979	21.4866
	Variance	**323.564**	Inf	343.475	885.809
Arrival time for stationary vehicle	Log likelihood	−58947.1	**−58415.1**	−58769.5	−58097.4
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$1 < y < \text{Inf}$
	Mean	21.8141	32.3138	21.5822	23.2076
	Variance	559.408	**Inf**	628.02	1895.03
Arrival time for hazard	Log likelihood	−17132.8	−17269.8	**−17127.8**	−17250
	Domain	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$	$0 < y < \text{Inf}$
	Mean	58.4637	102.204	58.4066	70.5117
	Variance	3728.08	Inf	**3914.51**	20842.2

Table 4: Estimation results of survival AFT model for clearance time for crash.

Variable	Estimate coefficient	Wald Chi-square	Pr > ChiSq
Intercept	4.7907	8.74	0.0031
Priority	−0.1495	27.7239	<0.0001
Intersection	0.1737	5.3656	0.0205
Blockage type	0.1878	13.7469	0.0081
Weather	0.1103	15.1529	0.0341
Total opp lanes block	0.0753	11.5657	0.0412
Traffic disrupted	0.0628	4.6412	0.0312
Heavy traffic	0.1356	10.4797	0.0012
Major incident	0.2158	17.7047	<0.0001
Diversion required	0.2775	33.0936	<0.0001
Towing required	−0.1518	49.2159	<0.0001
Number of vehicles involved	0.2862	70.2871	<0.0001
Number of people injured	0.7411	26.1993	0.0101
Number of fatalities	1.0412	30.9592	<0.0001
Chemical spill	0.3045	17.4566	<0.0001
Log likelihood	−5509.6196		

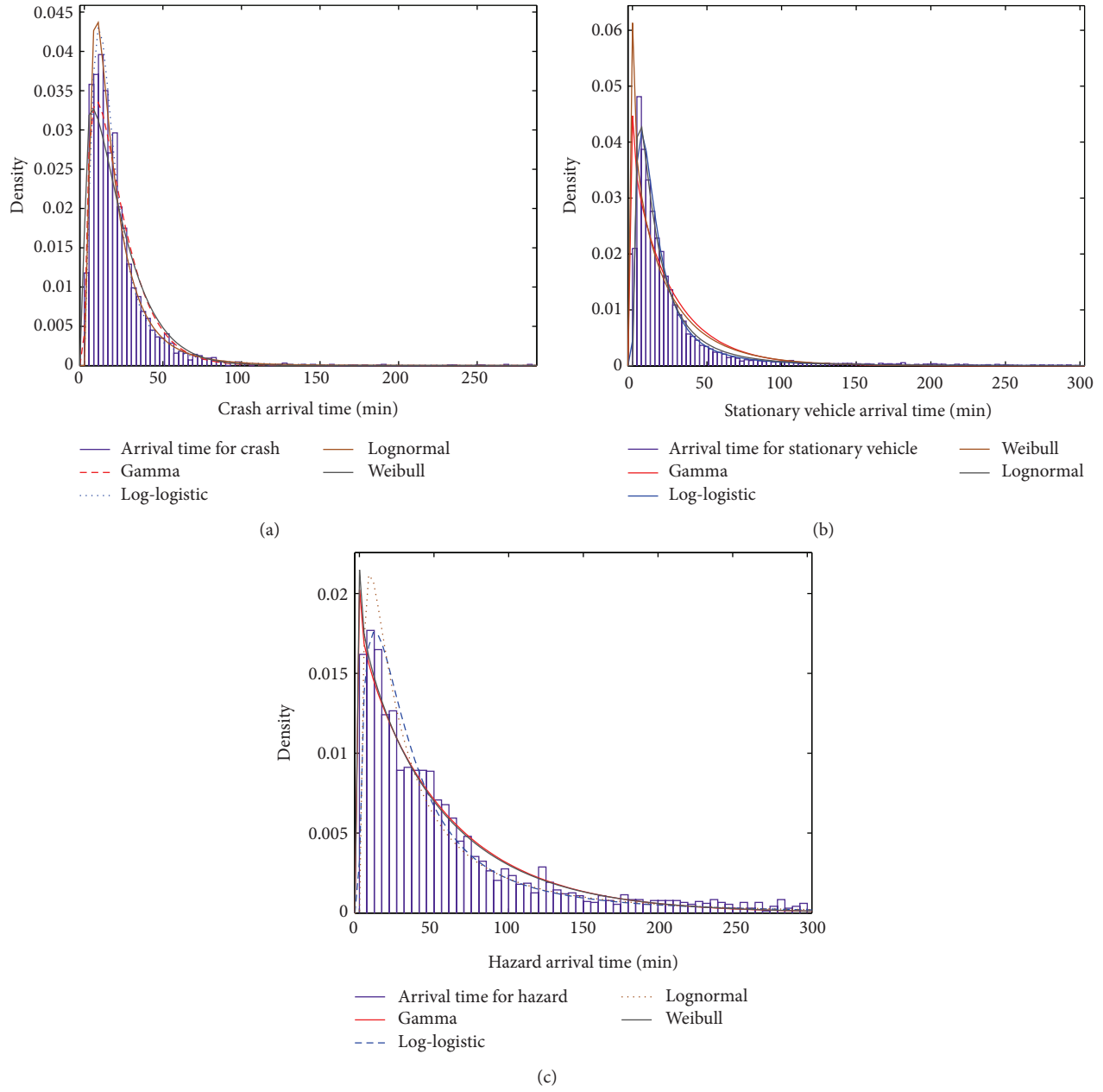


Figure 4: Probability density distribution of arrival time for crash (a), stationary vehicle (b), and hazard (c).

only 8 significant variables for arrival time, which is obviously less than that for clearance time. For example, priority, blockage type, weather, and so forth have a significant effect on the clearance time but not on the arrival time, because of different incident time characteristics. Another reason is that the information recorded in the SIMS for arrival time is less than clearance time.

5.3. Influence of Property Parameters on the Prediction Time. Table 6 represents the percentage change in clearance time and arrival time for crash and hazard incident. A negative percentage indicates a decrease in the clearance time with an increase in that property variable. Line "—" means that the variable has no significant effect on the incident time. For instance, when priority is increased by one, the crash clearance time is 16.13% shorter, and the hazard clearance time is 7.03% shorter, but no significant influence on arrival time. As the number of injured people is increased by one, the crash clearance time is 109.83% longer, while crash arrival time is 57.02% longer, but no significant influence on hazard clearance time and arrival time.

The evaluation of the prediction accuracy for crash incident clearance time is given in Table 7 as an example. There are 4966 crash clearance time data which are used to set up hazard based model. 30% of them are used to evaluate the prediction accuracy. Results in Table 7 indicate the absolute

Table 5: Estimation results of survival AFT model for arrival time for crash.

Variable	Estimate coefficient	Wald Chi-square	Pr > ChiSq
Intercept	−0.5248	0.06	0.81
Intersection	0.246	3.6168	0.0472
Lateral position	0.2479	14.2799	0.0039
Total lanes blocked	0.3002	10.6649	0.0493
Traffic disrupted	0.0732	2.8418	0.0218
Towing required	0.2474	48.519	<0.0001
Number of vehicles involved	0.617	57.2253	<0.0001
Number of people injured	0.4512	18.3643	0.0351
Number of fatalities	0.5321	13.7241	0.0082
Log likelihood	−6416.712		

Table 6: Percentage change in clearance time and arrival time for crash and hazard.

Variable	Clearance time for crash	Clearance time for hazard	Arrival time for crash	Arrival time for hazard
Priority	−16.13%	−7.03%	—	—
Intersection	18.97%	73.13%	27.89%	71.77%
Lateral position	—	25.54%	28.13%	—
Blockage type	20.65%	47.38%	—	106.49%
Weather	11.66%	11.17%	—	6.56%
Total lanes blocked	—	—	35.01%	—
Total opp lanes blocked	7.82%	—	—	33.99%
Traffic disrupted	6.48%	—	7.59%	—
Heavy traffic	14.52%	—	—	—
Major incident	24.08%	81.56%	—	—
Diversion required	31.98%	—	—	130.23%
Towing required	−16.40%	—	28.07%	—
Number of vehicles involved	33.14%	—	85.34%	377.41%
Number of people injured	109.83%	—	57.02%	—
Number of fatalities	183.26%	—	70.25%	—
Chemical spill	35.59%	—	—	—

Table 7: Summary of evaluation of the prediction accuracy.

Performance measure	Value	Percentage
<10 min	543 incidents	39.7%
<15 min	768 incidents	56.1%
<20 min	963 incidents	70.4%
<30 min	1158 incidents	84.6%
<50 min	1280 incidents	93.5%

value of the difference between prediction clearance time and measured clearance time. For example there are 543 incidents whose difference between prediction and measured time is less than 10 minutes, which account for 39.68% of the total evaluation incident data. The accuracy of the model is similar with that of Chung [23].

6. Conclusions

Hazard based prediction model for both incident clearance time and arrival time are developed. Three types of incidents are investigated based on data collected from SIMS. Before model development, the best fitting model was selected for each incident type based on the results of the likelihood ratio and variance. The results show the following.

(1) Clearance time and arrival time follow the same distribution with the different estimated parameters for each incident type.

(2) Gamma, Log-logistic, and Weibull distribution are best fit for crash, stationary vehicle, and hazard incident, respectively. After data filter, the hazard based prediction model is developed for crash incident as example.

There are 14 significant incident property variables for clearance time, while there are only 8 significant variables for arrival time. It clearly indicates that clearance time and arrival time have different impact factors.

The percentage changes in clearance time and arrival time for crash and hazard incident are given. The impact of each property variable on clearance time and arrival time is quantitatively provided. The model accuracy is given at the end of paper.

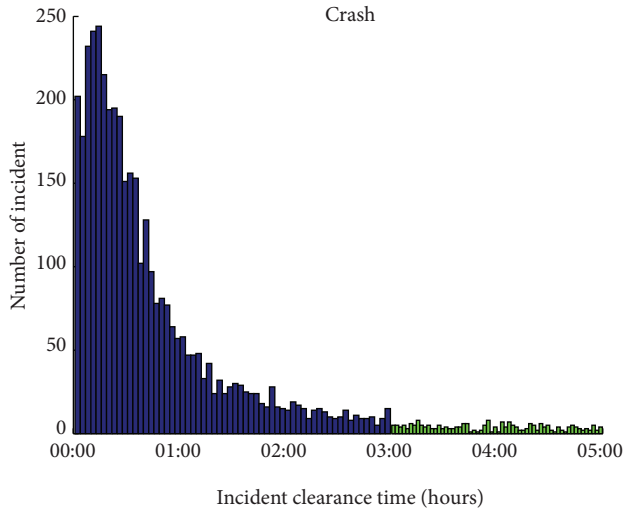


Figure 5: Distribution of crash clearance time.

Conflict of Interests

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

Acknowledgments

This study has been substantially supported by the National Natural Science Foundation Council of China (no. 71201103). The first author finished the work during the postdoctoral research in Tongji University, and the support of a Project of Shanghai Shuguang Program (13SG23) is acknowledged. The first author thanks the support of the Smart Transport Research Center at the Queensland University of Technology by providing real traffic incident data.

References

[1] C.-S. Chou, *Understanding the impact of incidents and incident management programs on freeway mobility and safety [Dissertation]*, University of Maryland, College Park, Md, USA, 2010.

[2] T. Tsubota, H. Kikuchi, K. Uchiumi, H. Warita, and F. Kurauchi, "Benefit of accident reduction considering the improvement of travel time reliability," *International Journal of Intelligent Transportation Systems Research*, vol. 9, no. 2, pp. 64–70, 2011.

[3] T. F. Golob, W. W. Recker, and J. D. Leonard, "An analysis of the severity and incident duration of truck-involved freeway accidents," *Accident Analysis & Prevention*, vol. 19, no. 5, pp. 375–395, 1987.

[4] G. Giuliano, "Incident characteristics, frequency, and duration on a high volume urban freeway," *Transportation Research A: General*, vol. 23, no. 5, pp. 387–396, 1989.

[5] A. Garib, A. E. Radwan, and H. A. Deek, "Estimating magnitude and duration of incident delays," *Journal of Transportation Engineering*, vol. 123, no. 6, pp. 459–466, 1997.

[6] E. C. Sullivan, "New model for predicting freeway incidents and incident delays," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 267–275, 1997.

[7] B. Jones, L. Janssen, and F. Mannering, "Analysis of the frequency and duration of freeway accidents in Seattle," *Accident Analysis & Prevention*, vol. 23, no. 4, pp. 239–255, 1991.

[8] D. Nam and F. Mannering, "An exploratory hazard-based analysis of highway incident duration," *Transportation Research A: Policy and Practice*, vol. 34, no. 2, pp. 85–102, 2000.

[9] S. Cohen and C. Nouveliere, "Modelling incident duration on an Urban expressway," in *Proceedings of the 8th International Federation of Automatic Control Symposium on Transportation Systems (IFAC '97)*, Chania, Greece, 1997.

[10] K. Ozbay and P. Kachroo, *Incident Management in Intelligent Transportation Systems*, Artech House, Boston, Mass, USA, 1999.

[11] V. H. L. Kau, *Estimating Freeway Incident Clearance Duration Using Multiple Linear Regression*, The University of Texas at Arlington, Arlington, Tex, USA, 2007.

[12] A. J. Khattak, J. L. Schofer, and M. H. Wang, "A simple time sequential procedure for predicting freeway incident duration," *Journal of Intelligent Transportation Systems*, vol. 2, no. 2, pp. 113–138, 1995.

[13] V. Sethi, F. S. Koppelman, C. P. Flannery, N. Bhandari, and J. L. Schofer, "Duration and travel time impacts of incidents-advance project," Tech. Rep. TRF-ID-202, Northwestern University, Evanston, Ill, USA, 1994.

[14] Y. B. B. Ji, X. N. Zhang, and L. J. Sun, "Traffic incident duration prediction grounded on Bayesian decision method-based tree algorithm," *Journal of Tongji University: Natural Science*, vol. 36, no. 3, pp. 319–324, 2008.

[15] B. Stephen, D. Fajardo, and S. T. Waller, "A naïve bayesian classifier for incident duration prediction," in *Proceedings of the 86th TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2007.

[16] J. Lopes, J. Bento, F. C. Pereira, and M. B. Akiva, "Dynamic forecast of incident clearance time using adaptive artificial neural network models," in *Proceedings of the 92nd TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2013.

[17] C.-H. Wei and Y. Lee, "Sequential forecast of incident duration using Artificial Neural Network models," *Accident Analysis & Prevention*, vol. 39, no. 5, pp. 944–954, 2007.

[18] V. Gaetano, M. Lelli, and D. Cucina, "A comparative study of models for the incident duration prediction," *European Transport Research Review*, vol. 2, no. 2, pp. 103–111, 2010.

[19] D. Collett, *Modelling Survival Data in Medical Research*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2003.

[20] A. Stathopoulos and M. G. Karlaftis, "Modeling duration of urban traffic congestion," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 587–590, 2002.

[21] C. R. Bhat and A. R. Pinjar, "Duration modeling," in *Handbook of Transport Modeling*, D. A. Hensher and K. J. Button, Eds., Elsevier, Amsterdam, The Netherlands, 2nd edition, 2008.

[22] P. C. Anastasopoulos, J. E. Haddock, M. G. Karlaftis, and F. L. Mannering, "An analysis of urban travel times: a random parameters hazard-based approach," in *Proceedings of the 91st TRB Annual Meeting Compendium of Papers*, Washington, DC, USA, 2012.

[23] Y. Chung, "Development of an accident duration prediction model on the Korean Freeway Systems," *Accident Analysis & Prevention*, vol. 42, no. 1, pp. 282–289, 2010.

[24] H. A. Tavassoli, L. Ferreira, S. Washington, and P. Charles, "Hazard based models for freeway traffic incident duration," *Accident Analysis & Prevention*, vol. 52, pp. 171–181, 2013.

[25] D. A. Hensher and F. L. Mannering, "Hazard-based duration models and their application to transport analysis," *Transport Reviews*, vol. 14, no. 1, pp. 63–82, 1994.

[26] L. Xian, *Survival Analysis: Models and Applications*, Higher Education Press, Beijing, China, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 806916, 10 pages
http://dx.doi.org/10.1155/2014/806916

Research Article

Transportation Network Design considering Morning and Evening Peak-Hour Demands

Hua Wang,[1] Gui-Yuan Xiao,[2] Li-Ye Zhang,[3] and Yangbeibei Ji[4]

[1] *School of Economics and Management, Tongji University, Shanghai 200096, China*
[2] *College of Civil and Architectural Engineering, Guilin University of Technology, Guilin 541004, China*
[3] *School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410076, China*
[4] *School of Management, Shanghai University, Shanghai 200444, China*

Correspondence should be addressed to Li-Ye Zhang; liyezhang.uf@gmail.com

Received 18 December 2013; Accepted 17 January 2014; Published 10 March 2014

Academic Editor: X. Zhang

Previous studies of transportation network design problem (NDP) always consider one peak-hour origin-destination (O-D) demand distribution. However, the NDP based on one peak-hour O-D demand matrix might be unable to model the real traffic patterns due to diverse traffic characteristics in the morning and evening peaks and impacts of network structure and link sensitivity. This paper proposes an NDP model simultaneously considering both morning and evening peak-hour demands. The NDP problem is formulated as a bilevel programming model, where the upper level is to minimize the weighted sum of total travel time for network users travelling in both morning and evening commute peaks, and the lower level is to characterize user equilibrium choice behaviors of the travelers in two peaks. The proposed NDP model is transformed into an equivalent mixed integer linear programming (MILP), which can be efficiently solved by optimization solvers. Numerical examples are finally performed to demonstrate the effectiveness of the developed model. It is shown that the proposed NDP model has more promising design effect of improving network efficiency than the traditional NDP model considering one peak-hour demand and avoids the misleading selection of improved links.

1. Introduction

The transportation network design problem (NDP) is characterized as one of the most important and challenging optimization problems in the transportation system [1]. It aims to improve the network system efficiency by expanding link capacities on existing roads or building new roads/lanes for the network. In general, the NDP problem can be well formulated by a bilevel programming model, in which the Stackelberg behavior between network planner and network users is characterized. The upper level problem is to optimize a set of design objectives (e.g., minimizing total travel time) for the urban transportation system by setting link capacity expansion scheme with some necessary constraints (e.g., financial budget constraint). The lower level problem is to describe the user's behavior in terms of path/mode choice, departure time choice, and origin/destination choice.

After a pioneering work of Abdulaal and Leblanc [2], more and more researchers have paid attention to the NDP problem with focus on advanced model formulation and effective algorithm design. The achievements made before last century on the NDP studies can be found in two comprehensive reviews: Magnanti and Wong [3] and Yang and Bell [4]. In recent decade, we have also witnessed a large number of emerging advances on the NDP studies. These new advances are pertinent to uncertain parameters (e.g., stochastic demand and capacity variation), sustainability constraints, and distinct time dimension. Uncertainty is one underlying and important feature of travel activity and plays a critical role in network design and planning policy. To avoid unnecessary risks and possibly misleading outcomes, demand uncertainty and/or capacity variation have been taken into account in the NDP problem (e.g., [5, 6]), by introducing expected value model and robust optimization model. Two typical

approaches are used to model the stochastic traffic flows under uncertain traffic conditions. One is to develop a two-stage formulation by means of scenario-construction approach, where a number of finite scenarios of uncertain parameters with known probability distribution will be generated (e.g., [7, 8]). Another way, called probability-analytical method, is to derive a reliability-based traffic assignment, where the users follow probabilistic user equilibrium under uncertain traffic conditions (e.g., [9–11]).

Time dimension, as another important factor, has also attracted the researchers' attention in recent NDP studies. Three scales of time dimension can be considered in the NDP problem: short-term/real-time (seconds), medium-term (days), and long-term (years). The NDP considering time dimension is defined as time-dependent NDP, which can be specifically classified into multiperiod NDP and dynamic NDP according to different scales of time dimension. The former is generally to analyze the NDP problem taking into account a long-term spanning construction and maintenance of transport infrastructure [12–15]. The latter is to precisely characterize the real-time traffic dynamics and unsteady-state conditions in the NDP by introducing dynamic traffic assignment. Hopefully it can be used to estimate microinteraction among adjacent links and thereby identify and examine the possible bottleneck [6, 16, 17]. A comprehensive review of recent developments on the NDP studies can be referred to Meng et al. [18].

By retrieving and reviewing the NDP literature, especially the valuable reviews across several decades [3, 4, 18, 19], it can be found that almost all NDP studies assume that an NDP model based on one peak-hour O-D demand distribution is capable and effective to explain the real traffic patterns. For urban traffic, there always are two peaks of a daily traffic flow corresponding to morning commute and even commute, respectively, on weekday [20]. For commuting travel trips, the repeatability of hourly fluctuation of traffic flow gives rise to a comparative stable peak-hour demand. But such repeatability of hourly variation of traffic flow does not mean the same O-D demand distributions in the morning and evening peaks.

Although the daily commuters have high repetitive travel activities, the peak-hour O-D demand distribution on the network will be different in the morning and evening peaks. This fact has been recognized and emphasized in the transportation network modeling problem, such as commuting pattern analysis [21] and road tolling and parking fee optimization [22, 23]. It is always assumed that the O-D demand matrix is asymmetric diagonal for the morning and even peaks, namely, $q_{od} = q_{do}$, if only daily regular commuters are considered. The O-D demand matrix used in the network planning and design comes up with the peak-hour traffic survey data (e.g., link traffic volumes) in either morning or evening peak. In practice, the traffic volumes in the morning peak may largely differ from that of evening peak commute, which has been demonstrated by some empirical studies (e.g. [20, 24]). The O-D demand matrices thus may also be different for two peaks because they are estimated based on the corresponding peak-hour traffic volumes. In other words, the travelers in two commuting peaks would make different travel decision in terms of both path choice and departure time choice [25]. Therefore, the NDP considering one peak-hour O-D demand might be misleading and even incorrect. On the one hand, the O-D demand distribution on a real network is asymmetric due to different traffic patterns in the morning and evening peaks, though the morning and evening round commuting trips can be assumed to be equal. We should overlook demand uncertainty and diverse spatial-temporal characteristic of travel activities in two peaks. For example, the noncommuting trips in morning peak (e.g., shopping, going to hospital, and tourism) may not return back to the origin places in evening peak. Similarly, part of activities for the travelers who travel in evening peak do not commence in the morning and evening peaks but other traveling periods. Therefore, which peak-hour commute pattern should be used to estimate the O-D demand matrix is a controversial and intractable issue.

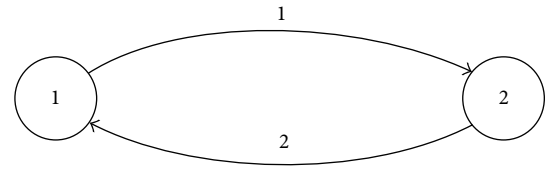


Figure 1: An illustrative example.

On the other hand, network structure and link sensitivity could have important impact on the network performance evaluation for two peaks. Even though the O-D pairs and their demands can be assumed to be symmetric in two peaks, namely, $q_{od} = q_{do}$, the path choices in two commute peaks may be still different, which largely depends on the network topological structure. If different paths/links are chosen in two peaks, network situation in terms of traffic congestion would show large difference between two peaks. In this respect, the NDP considering one peak-hour O-D demand may generate misleading and even wrong outcomes. More importantly, link sensitivity of the NDP effect should not be underestimated and overlooked under different peaks. Take a simple network comprising of two links shown in Figure 1 as an example. Assume that q_{od} in the morning peak is equal to q_{do} in the evening peak. Consider two O-D pairs, $q_{12} = 30$, $q_{21} = 10$ in the morning peak and $q_{12} = 10$, $q_{21} = 30$ in the evening peak. Links 1 and 2 have equal capacities of 20 veh/h and free flow travel time of 1 minute. The available budget invested on link capacity expansion is set as 20 equivalently. BPR function is used to measure the link travel time performance.

It is not difficult to see that link 1 will be improved by adding 20 units in the NDP considering morning peak-hour demand, and link 2 will be improved by expanding 20 units in the NDP considering evening peak-hour demand. The total travel times on the network for these two NDPs are equal, namely, 104.305 minutes. In the NDP simultaneously considering morning and evening peak-hour demands, both links 1 and 2 will be improved by adding 10 units, respectively, and thereby the total travel time on the network will be reduced to 89.037 minutes. Evidently, the NDP considering one peak-hour demand matrix is not the promising scheme

to improve transportation system performance. Therefore, it is necessary to develop a new network design model in order to avoid misleading decision-making.

To address the above problems, we develop a network design model simultaneously taking into account morning and evening peak-hour demands. The traditional NDP model considering one peak-hour O-D demand has been regarded as a special case of the proposed model. The proposed model is formulated as a bilevel programming model with objective to minimize the weighted sum of total travel times of the traffic patterns in the morning and evening peaks. In the lower level, network users are assumed to follow the user equilibrium principle of their route choice decision. The proposed model is finally transformed into an equivalent mixed integer linear programming (MILP) so that a global optimum can be obtained by means of MIP solvers (e.g., CPLEX).

The rest of the paper is organized as follows. In the next section, the new NDP model is formulated by simultaneously considering the morning and evening peal-hour O-D demand distributions. Then, an equivalent MILP model is introduced. Numerical examples are discussed in Section 4. Finally, conclusions and further studies are given in Section 5.

2. The Model

This section builds a network design model simultaneously considering morning and evening peak-hour demand matrices. Consider a directed transportation network, $G(N, A)$, comprising of a set N of nodes and a set A of directed links. Since traffic flow patterns in two peaks will be involved, the following notations are defined for two peaks, respectively, which are summarized in the Appendix (Notations).

The link travel time in the morning or evening peak is assumed to be continuous, convex, and strictly increasing function of its own link flow, such as the widely-applied BRP function. Meanwhile, our focus is put on the deterministic NDP problem with fixed demand.

2.1. Model Formulation. Similar to the traditional NDP model taking into account one peak-hour demand matrix, a bilevel program is formulated to model the proposed NDP problem considering two peak-hour demand matrices in that the Stackelberg behavior between network planner and network users can be well characterized. In the upper level, the network design objective is to minimize the weighted sum of total travel times (TTC) of the users traveling in the morning and evening peaks. The lower level is the user equilibrium traffic assignment problem, which is used to characterize the user's route choice behavior. The NDP formulation is given below.

Upper Level

$$\min_{\mathbf{f}^m,\mathbf{f}^e,\mathbf{x}} \mathrm{TTC} = \alpha \sum_{w\in W^m} \mu_w^m\left(\mathbf{f}^m,\mathbf{x}\right) q_w^m + \beta \sum_{w\in W^e} \mu_w^e\left(\mathbf{f}^e,\mathbf{x}\right) q_w^e, \quad (1)$$

subject to

$$x_a \geq 0, \quad a \in \overline{A}, \quad (2)$$

$$\sum_{a\in\overline{A}} b_a(\mathbf{x}) \leq B. \quad (3)$$

Constraint (2) means a nonnegative link capacity expansion and (3) is the total invested budget constraint. The construction cost for each improved link, $b_a(x_a)$, can be approximately estimated by a linear function:

$$b_a(x_a) = \gamma_a x_a, \quad a \in \overline{A}. \quad (4)$$

The traffic flow patterns, $f^m(\mathbf{x})$, $f^e(\mathbf{x})$, $\mu_w^m(\mathbf{x})$, and $\mu_w^e(\mathbf{x})$ can be obtained by solving the lower level traffic assignment problems in the morning and evening peaks.

Lower Level

(1) User equilibrium model in the morning peak is as below:

$$\min_{\mathbf{x}} \sum_{a\in A} \int_0^{v_a^m} t_a^m(\mathbf{x},\omega)\, d\omega, \quad (5)$$

subject to

$$v_a^m = \sum_{w\in W^m} \sum_{r\in R_w^m} f_{r,w}^m \delta_{ar,w}^m, \quad a \in A,$$

$$\sum_{r\in R_w^m} f_{r,w}^m = q_w^m, \quad w \in W^m, \quad (6)$$

$$f_{r,w}^m \geq 0, \quad r \in R_w^m,\ w \in W^m.$$

(2) User equilibrium model in the evening peak is as below:

$$\min_{\mathbf{x}} \sum_{a\in A} \int_0^{v_a^e} t_a^e(\mathbf{x},\omega)\, d\omega, \quad (7)$$

subject to

$$v_a^e = \sum_{w\in W^e} \sum_{r\in R_w^e} f_{r,w}^e \delta_{ar,w}^e, \quad a \in A,$$

$$\sum_{r\in R_w^e} f_{r,w}^e = q_w^e, \quad w \in W^e, \quad (8)$$

$$f_{r,w}^e \geq 0, \quad r \in R_w^e,\ w \in W^e.$$

It should be noted that all travelers in the morning and evening peaks will complete their travel journeys on the same transportation network; that is, the link capacity expansion scheme would be designed and served for the travelers in both morning and evening peaks. Hopefully, it is intuitive that the travelers in both morning and evening peaks will all benefit from the NDP scheme.

We can easily see that, if $\alpha = 1$, $\beta = 0$, the proposed model is equivalent to the traditional NDP model considering morning peak-hour demand distribution in that the evening peak-hour traffic patterns are not considered in the objective function. Similarly, if $\alpha = 0$, $\beta = 1$, the proposed model is equivalent to the traditional NDP model considering evening peak-hour demand distribution, since the morning commute pattern has no impact on the objective function.

2.2. Extensions of considering Elastic Demand and Uncertain Demand. The proposed model can be also extended to account for the elastic demand and/or uncertain demand. We here give a brief discussion of these extensions. In reality, whether a potential traveler decides to finish her/his travel journey in great extent depends on the traffic congestion on the network, no matter what commute peak is considered. The O-D demand thus can be assumed to be a function of the O-D travel cost. Take the O-D travel demand in the morning peak as an example:

$$q_w^m = d_w\left(\mu_w^m\right), \quad w \in W^m, \tag{9}$$

where $d_w(\cdot)$ is the demand function between O-D pair $w \in W^m$ and its inverse function is represented by $d_w^{-1}(\cdot)$. In general, $d_w(\cdot)$ is assumed to be a positive, continuously differentiable, and strictly decreasing function with respect to shortest path cost μ_w^m, such as negative exponential function.

It is not difficult to deal with the proposed NDP model with elastic demand since the traffic assignments for two peaks with elastic demand can be easily solved by a supernetwork method or an improved Frank-Wolfe algorithm in Sheffi [26].

Another interesting extension is to take into account of the uncertain travel demands in two peaks. The importance of considering uncertainties in the NDP is to avoid the unnecessary risks and misleading policies. The O-D demand in either morning commute peak or evening commute peak will fluctuate in future. To capture these impacts, we can develop a general stochastic NDP according to the below framework:

$$\min_{\mathbf{V},\mathbf{x}} F(\mathbf{V}, \mathbf{x}, \Xi) \tag{10}$$

subject to

$$\begin{aligned} &\mathbf{V} \in \Omega(\mathbf{x}, \Xi) \\ &(\mathbf{V}, \mathbf{x}) \in \Theta. \end{aligned} \tag{11}$$

The bold notations, $\mathbf{V}$, Ξ, are random variables, which denote stochastic link flows and uncertain demands in the morning and evening peaks. Notation Θ denotes the set of additional constraints of variables $\mathbf{V}$ and $\mathbf{x}$, and $\Omega(x, \Xi)$ is the set of feasible link flows on stochastic network.

Two modeling methods can be used to characterize the morning and evening peak-hour demand uncertainties in the NDP problem. One is to develop the probabilistic or reliability-based user equilibrium model of deriving the stochastic traffic flow patterns in the morning and evening peaks and then embed them into the upper level optimization of the NDP problem. For such probabilistic user equilibrium model, it can be referred to Lo et al. [9], Shao et al. [10], and Wang et al. [11]. Another way is to formulate a two-stage NDP model, where the uncertainties of two peak-hour demands can be captured by the random samples generated by some scenario-construction methods, such as sample average approximation [7, 8]. The main difference between two modeling approaches and the detailed formulation framework can be found in Meng et al. [18].

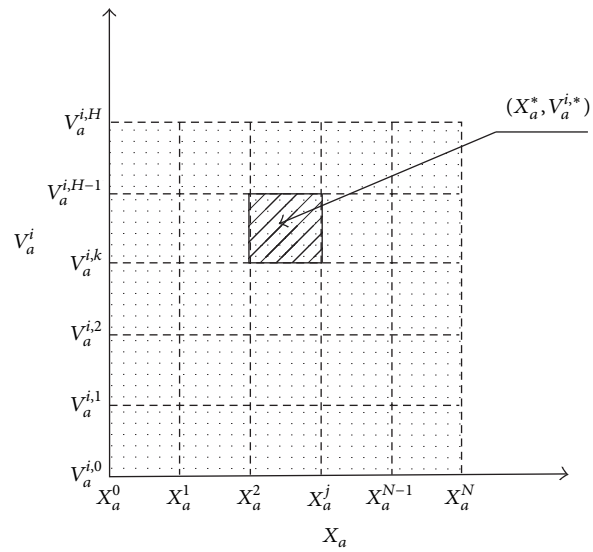


Figure 2: Discretize binary space into feasible regions.

3. The Equivalent MILP Model

This study focuses on the NDP problem with two peak-hour demand impacts and on examining the design effects between the proposed model and the conventional one with one peak-hour demand matrix. In order to precisely compare the network design effects, it would be better to solve the NDP model by a global solution algorithm. We here use a global solution algorithm proposed by Wang and Lo [27] by transforming the bilevel NDP model into single-level MILP. The transformation of MILP includes linearization of design objective function, link travel time function, and other side constraints.

For the sake of simplicity, the fixed demand is considered and the BRP function is used to measure the congestion effect of the link travel time, namely,

$$t_a^i\left(v_a^i, x_a\right) = t_a^0\left(1 + 0.15\left(\frac{v_a^i}{c_a^0 + x_a}\right)^4\right), \quad i = \{m, e\}, \tag{12}$$

where t_a^i and v_a^i are used to represent the link travel time and link flow for the morning peak if $i = m$ or for the evening peak if $i = e$. Hereafter, we briefly revisit the transformation of MILP.

3.1. Linearization of Link Travel Time Function. It can be seen in (12) that the link travel time function for each commute peak is a function of link flow in the peak and design variable, namely, link capacity expansion. Similar to Wang and Lo [27], Luathep et al. [28], and Zhang and Van [29], the binary space in terms of link flow and link capacity expansion can be divided into $H \times N$ feasible regions as shown in Figure 2. Let $v_a^{i,l}$, $v_a^{i,u}$ be the lower and upper bounds of v_a^i, and let x_a^l, x_a^u be the lower and upper bounds of x_a. For all discretized intervals, we have $v_a^{i,l} < K_{a,h}^i < K_{a,h+1}^i < v_a^{i,u}$ and $x_a^l < L_{a,n} < L_{a,n+1} < x_a^u$, $h \in \{1, 2, \ldots, H\}$, $n \in \{1, 2, \ldots, N\}$, and $i = \{m, e\}$.

For any feasible region $[h, n]$, the link travel time function (12) can be approximated as a linear function by Taylor expansion:

$$t_a^i\left(v_a^i, x_a\right) = a_a^{i,h,n} v_a^i + b_a^{i,h,n} x_a + c_a^{i,h,n}, \quad \text{if } K_{a,h}^i \le v_a^i \le K_{a,h+1}^i,\ L_{a,n} \le x_a \le L_{a,n+1}, \tag{13}$$

where coefficients $a_a^{i,h,n}$, $b_a^{i,h,n}$, $c_a^{i,h,n}$ can be obtained by determining the partial derivatives of the BPR link performance function:

$$a_a^{i,h,n} = \left.\frac{\partial t_a^i}{\partial v_a^i}\right|_{(K_{a,h}^i, L_{a,n})}$$

$$b_a^{i,h,n} = \left.\frac{\partial t_a^i}{\partial x_a}\right|_{(K_{a,h}^i, L_{a,n})} \tag{14}$$

$$c_a^{i,h,n} = t_a^i\left(K_{a,h}^i, L_{a,n}\right) - K_{a,h}^i \cdot a_a^{i,h,n} - L_{a,n} \cdot b_a^{i,h,n}.$$

It is clear that link travel time can be precisely estimated by (13), if very large H, N are set to guarantee sufficient number of binary space splits. The link travel time function thus can be replaced by an equivalent set of mixed integer linear constraints. For link a, $a \in A$ ($h = 1, \ldots, H$; $n = 1, \ldots, N$; $i = m, e$), we have

$$\begin{aligned}
L \cdot \xi_{a,h}^i &\le v_a^i - K_{a,h}^i \le U \cdot \left(1 - \xi_{a,h}^i\right) - \varepsilon \\
\kappa_{a,h}^i &= \xi_{a,h+1}^i - \xi_{a,h}^i \\
L \cdot \tau_{a,n} &\le x_a - L_{a,n} \le U \cdot \left(1 - \tau_{a,n}\right) - \varepsilon \\
\lambda_{a,n} &= \tau_{a,n+1} - \tau_{a,n} \\
\psi_a^{i,h,n} &= \kappa_{a,h}^i + \lambda_{a,n} \\
L \cdot \left(2 - \psi_a^{i,h,n}\right) &\le t_a - \left(a_a^{i,h,n} \cdot v_a^i + b_a^{i,h,n} \cdot x_a + c_a^{i,h,n}\right) \\
&\le U \cdot \left(2 - \psi_a^{i,h,n}\right) \\
\text{integer } \xi_{a,h}^i, \tau_{a,n} &\in \{0, 1\},
\end{aligned} \tag{15}$$

where L, U are, respectively, a very large negative constant and a very large positive constant; ε is a very small positive constant. It is not difficult to prove that (15) is equivalent to the linear approximation function (13). We here do not attempt to present the proof again because the detailed proof can be found in [27].

3.2. Linearization of the Constraints. The proposed NDP model includes three kinds of constraints: deterministic user equilibrium constraint, definitional constaints (e.g., demand conservation) and other side constraints (invested budget constaint and boundary constraints of design variables).

(1) Deterministic User Equilibrium Constraint. Recall that the deterministic user equilibrium principle can also be expressed by complementary constraint, which is derived from the first-order condition of the lower level traffic assignment problem:

$$f_{r,w}^i \cdot \left(c_{r,w}^i - \mu_w^i\right) = 0, \qquad c_{r,w}^i - \mu_w^i = 0, \quad \forall r, w, i. \tag{16}$$

The "if-then" complementary constraint can be transformed into an equivalent set of constraints by introducing a set of binary variables, shown as:

$$\begin{aligned}
L \cdot \sigma_{r,w}^i + \varepsilon &\le f_{r,w}^i \le U \cdot \left(1 - \sigma_{r,w}^i\right) \\
L \cdot \sigma_{r,w}^i &\le c_{r,w}^i - \mu_w^i \le U \sigma_{r,w}^i \\
c_{r,w}^i - \mu_w^i &\ge 0 \\
\text{integer } \sigma_{r,w}^i &\in \{0, 1\}, \quad i = \{m, e\}.
\end{aligned} \tag{17}$$

Evidently, in (17), if $\sigma_{r,w}^i = 0$, we have $f_{r,w}^i > 0$ and $c_{r,w}^i - \mu_w^i = 0$; otherwise, $\sigma_{r,w}^i = 1$, we have $f_{r,w}^i = 0$ and $c_{r,w}^i - \mu_w^i \ge 0$. That is, the user equilibrium condition holds.

(2) Definitional Constaints. We have

$$\begin{aligned}
\sum_{r \in R_w^i} f_{r,w}^i &= q_w^i \\
v_a^i &= \sum_{w \in W^i} \sum_{r \in R_w^i} f_{r,w}^i \delta_{ar,w}^i \\
c_{r,w}^i &= \sum_{a \in A} t_a^i\left(v_a^i, x_a\right) \delta_{ar,w}^i
\end{aligned} \tag{18}$$

$$v_a^i \ge 0, \qquad f_{r,w}^i \ge 0,$$

$$t_a^i \ge t_a^0, \quad i \in \{m, e\}. \tag{19}$$

The definitional constraints in (18) are all linear constraints due to their additive properties.

(3) Other Side Constraints. We have

$$0 \le x_a \le x_a^u, \qquad 0 \le \sum_{a \in \bar{A}} \gamma_a x_a \le B. \tag{20}$$

So far, we have completely transformed the lower level user equilibrium traffic assignment problem into an equivalent set of mixed integer linear constraints. Since the design objective function (1) is also linear for the NDP with fixed demand, the bilevel NDP model simultaneously considering morning and evening peak-hour demands can be perfectly transformed into the equivalent MILP.

3.3. Solution Algorithm. Comparing to a nonlinear and nonconvex bilevel NDP, it is simple and effective to solve the transformed MILP problem. A more attractive merit of solving the NDP problem by transforming into MILP is that a global solution can be guaranteed. The global optimal solution is helpful and convincing for exploring

Table 1: O-D demands in two peaks.

O-D pair	$1 \to 3$	$3 \to 1$	$2 \to 3$	$3 \to 2$
Commute demand in the morning peak (q_w^m)	25	10	20	10
Commute demand in the evening peak (q_w^e)	10	25	10	20

Table 2: Parameters used for numerical examples.

Link number	1	2	3	4
t_a^0	5.0	10.0	15.0	10.0
c_a^0	15	20	20	25
γ_a	5.0	10.0	15.0	10.0

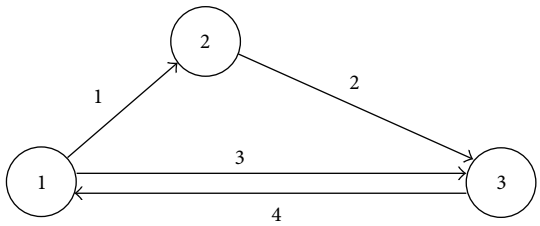


Figure 3: The transportation network used in numerical examples.

the design effect of the proposed NDP model. The MILP model can easily be solved by off-the-shelf MIP solvers, including IBM ILOG CPLEX, LINGO, and GUROBI. It has been demonstrated that the bilevel optimization problem can be fully transformed into equivalent MILP and solved by the MIP solvers efficiently and precisely (e.g., [27]) [29]. In this paper, the proposed bilevel NDP model will be solved as the equivalent MILP by the CPLEX solver.

4. Numerical Examples

4.1. Preliminary. The numerical examples are used to illustrate the difference between the proposed NDP model and the traditional one considering traffic patterns in one peak. In the numerical study, a small network shown in Figure 3 is used to demonstrate the property of the proposed model. This transportation network comprises of 3 nodes, 4 links, and 4 O-D pairs in each commute peak. All 4 links are considered in the candidate set of the capacity improvement scheme, namely, $\overline{A} = A$. The travel demands for each O-D pair in morning and evening peaks are given in Table 1. Table 2 provides the link performance parameters, t_a^0 and c_a^0, and the link capacity expansion cost coefficient γ_a. The weighted parameters in design objective function are set as $\alpha = 0.5$, $\beta = 0.5$. The total budget invested on the link capacity improvement scheme is 300. The commercial optimization package CPLEX-12.5 is used to solve the MILP model with a gap tolerance of 0.1%. All experiments run on Windows 7 system with the following attributes: Intel Core i5-2520 2.5 GHz × 2 and 4 GB RAM.

4.2. Comparison of the NDP Schemes. We investigate three NDP schemes and make a comparison of them in terms of network design effect. These NDP schemes are the proposed NDP simultaneously considering morning and evening peak-hour commuting demands, the traditional NDP only considering morning peak-hour demand matrix, and the traditional NDP only considering evening peak-hour demand matrix. The outcomes of three NDP schemes are provided in Tables 3, 4, and 5, respectively.

In the NDP considering morning peak-hour demand, the total travel time in the morning peak is 860.551. These four links will be improved by adding capacities of 9.749, 18.202, 2.998, and 2.427, respectively. Once this NDP scheme is implemented, the total travel time for the network users commuting in the evening peak is 1296.560. It can be found that travelers of O-D pairs 1-3 and 2-3 will largely benefit from Scheme A in the morning and evening peaks. But the travelers of another two O-D pairs obtain little benefit from the NDP Scheme A.

In the NDP considering evening peak-hour demand, the total travel time in the evening peak is 851.822. Links 1 and 4 will be expanded by adding capacities of 7.637 and 26.181, respectively, and links 2 and 3 maintain their initial capacities. Once the NDP Scheme B is performed, the total travel time for the network users commuting in the morning peak is 958.753. We can see that travelers of O-D pairs 1-3, 3-1, and 3-2 will benefit from Scheme A in two commute peaks in that their path travel times will be reduced by the NDP scheme. But Scheme B does not bring any benefit for the travelers of O-D pair 2-3.

In the NDP simultaneously considering two peak-hour demands, the total travel time in the morning peak is 902.092 and the total travel time in the evening peak is 877.038. Links 1, 2, and 4 will be expanded by adding capacities of 6.152, 6.201, and 20.725, respectively, and link 3 will not be considered to be improved. It is easy to see that travelers of all O-D pairs will benefit from the NDP Scheme C in two commute peaks.

By comparing the NDPs considering one peak-hour demand, namely, Schemes A and B, it is shown that Scheme B is better than Scheme A in terms of reducing total travel time on the network. Specifically, the sum of total travel time of two peaks in Scheme B is far less than that in Scheme A in the sense that the travelers commuting in the evening peak largely reduce their travel costs in Scheme B. This clearly indicates that the effects of the NDPs considering one peak-hour demand will be affected by the network structure and link sensitivity, which bring some troubles in choosing appropriate peak-hour demand matrix. The fact that Scheme A is far inferior to Scheme B on design effect also gives us a reminder that it should be careful to select the commute peak in performing traffic data collection.

We also compare the performance of the NDP simultaneously considering two peak-hour demands (Scheme C) with the NDP considering one peak-hour demand (Scheme B). As expected, the sum of total travel time of two peaks in Scheme C is less than that in Scheme B, although total travel time in the evening peak of Scheme C is slightly more than that of Scheme B. The reason is that Scheme B overlooks

Table 3: Network design considering morning peak-hour demand (Scheme A).

Commute peak	TTC	Link			Path			
		Number	Expanded capacity (x_a)	Flow (v_a^i)	Number	Link component	Flow ($f_{r,w}^i$)	Equilibrium cost (μ_w^i)
Morning peak	860.551	1	9.749	18.108	1	1-2	8.108	15.655
		2	18.202	28.108	2	3	16.892	15.655
		3	2.998	16.892	3	4	10.000	10.424
		4	2.427	20.000	4	2	20.000	10.440
		—	—	—	5	1–4	10.000	15.639
Evening peak	1296.560	1	9.749	20.000	1	1-2	0.000	15.327
		2	18.202	10.000	2	3	10.000	15.080
		3	2.998	10.000	3	4	25.000	20.873
		4	2.427	45.000	4	2	10.000	10.007
		—	—	—	5	1–4	20.000	26.193
Sum	2157.111	—	—	—	—	—	—	—

Table 4: Network design considering evening peak-hour demand (Scheme B).

Commute peak	TTC	Link			Path			
		Number	Expanded capacity (x_a)	Flow (v_a^i)	Number	Link component	Flow ($f_{r,w}^i$)	Equilibrium cost (μ_w^i)
Morning peak	958.753	1	7.637	13.551	1	1-2	3.551	17.980
		2	0.000	23.551	2	3	21.449	17.979
		3	0.000	21.449	3	4	10.000	10.035
		4	26.181	20.000	4	2	20.000	12.884
		—	—	—	5	1–4	10.000	15.131
Evening peak	851.822	1	7.637	20.000	1	1-2	0.000	15.551
		2	0.000	10.000	2	3	10.000	15.141
		3	0.000	10.000	3	4	25.000	10.896
		4	26.181	45.000	4	2	10.000	10.094
		—	—	—	5	1–4	20.000	16.353
Sum	1810.575	—	—	—	—	—	—	—

the traffic congestion in the morning peak; that is, it does not take into account the benefits of the travelers in morning peak while designing NDP scheme. In reality, Schemes A and B can be regarded as two special cases of the proposed model if one of the weighted parameters is zero. In summary, the proposed NDP model simultaneously considering two peak-hour demands can well characterize the practical traffic situation and also bring about promising design effect in terms of improving the transportation system performance. Note that the NDP model can be extended to consider more than two peaks. But the NDP simultaneously considering two peak-hour demands is believed to be good enough when the data collection costs in each peak and model flexibility are taken into account.

4.3. Impact Analysis of Weighted Parameters. We here conduct the impact analysis of the weighted parameter setting for the design objective function. Without loss of generality, it is assumed that $\alpha + \beta = 1.0$ by normalization. The weighted parameter β is set to be increased from 0.0 to 1.0 with each increment of 0.1. The variation of total travel time for each peak with different weighted parameter setting is depicted in Figure 4. As shown in Figure 5, the total travel time of two peaks changes with the weighted parameter setting.

It is shown in Figure 4 that, as expected, the total travel time of evening peak decreases monotonically with the increasing weight of β, since an increasing priority will be put on improving the traffic congestion in the evening peak. In turn, the total travel time of morning peak continuously increases with the weighted parameter β; that is, less emphasis will be paid on reducing the traffic congestion in the morning peak. It should be stressed that, the improvement effect of the NDP scheme in great measure depends on the network structure and demand distribution. In this regard, we repeat that the NDP scheme A greatly overlooks the social welfare of the travelers in the evening peak. Therefore, the network planner should avoid implementing the NDP scheme A.

In Figure 5, we can clearly see how important it is to account for the traffic congestions in both commuting peaks. Although the weight of considering the traffic pattern in the evening peak is small (e.g., $\beta = 0.1$), the network performance for whole daily commuting will be greatly improved. That is, the NDP only considering morning peak-hour

Table 5: Network design simultaneously considering two peak-hour demands (Scheme C).

Commute peak	TTC	Link			Path			
		Number	Expanded capacity (x_a)	Flow (v_a^i)	Number	Link component	Flow ($f_{r,w}^i$)	Equilibrium cost (μ_w^i)
Morning peak	902.092	1	6.152	16.198	1	1-2	6.198	16.757
		2	6.201	26.198	2	3	18.804	16.758
		3	0.000	18.804	3	4	10.000	10.055
		4	20.725	20.000	4	2	20.000	11.499
		—	—	—	5	1–4	10.000	15.313
Evening peak	877.038	1	6.152	20.000	1	1-2	0.000	15.631
		2	6.201	10.000	2	3	10.000	15.141
		3	0.000	10.000	3	4	25.000	11.407
		4	20.725	45.000	4	2	10.000	10.032
		—	—	—	5	1–4	20.000	17.007
Sum	1779.130	—	—	—	—	—	—	—

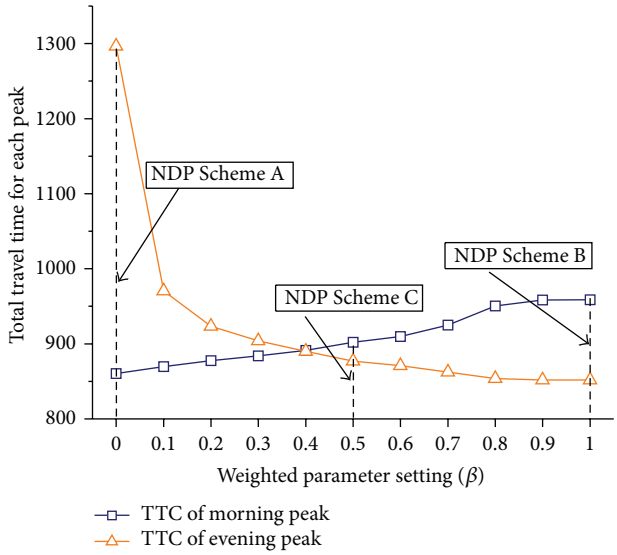


Figure 4: Total travel time for each peak with different weighted parameter settings.

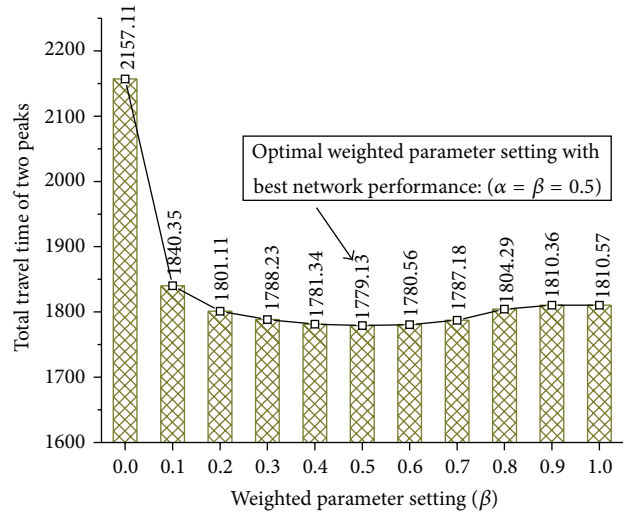


Figure 5: Total travel time of two peaks with different weighted parameter settings.

demand is far inferior to the NDPs taking into account the traffic congestion in the evening peak. Meanwhile, it can be found that how to determine the weighted parameter setting is important for developing a reasonable NDP scheme. Recall that the network decision-maker only concerns the traffic congestions in two commuting peaks, although some travelers might be more sensitive to the traffic congestion in the morning. To achieve the design objective of improving the traffic situation of two peaks as much as possible, the weighted parameter setting is preferred to set as $\alpha = \beta$, $\alpha, \beta > 0$, no matter whether the weighted parameters are normalized or not. In other words, if $\alpha = \beta$, $\alpha, \beta > 0$, the NDP considering two peak-hour demands always performs no worse than other NDP schemes (including the NDPs considering one peak-hour demand) in terms of minimizing the total travel time of two commuting peaks. It is not difficult to verify this conclusion. By revisiting the design objective function in (1), it can be seen that, when $\alpha = \beta$, $\alpha, \beta > 0$, the design objective function is to directly optimize the network performance in terms of total travel time of two peaks, with no error due to introducing priority of any peak. The outcome in the numerical examples also indicates that NDP scheme will have the best effect when $\alpha = \beta = 0.5$.

5. Conclusions and Further Studies

In the previous NDP studies, it was always assumed that the NDP considering one peak-hour demand distribution is capable and effective to characterize real traffic situation on the network. However, we found that the NDP model considering only one peak-hour demand matrix might be unable to describe the real traffic patterns due to the asymmetric traffic characteristics in the morning and evening peaks, and the impacts of network structure and link sensitivity. The traditional NDP model considering one peak-hour demand matrix thus may lead to misleading outcomes for transportation management and planning. To address

these problems, we proposed an NDP model simultaneously considering morning and evening peak-hour demands. The proposed model can be used to some extent to avoid the impacts of asymmetric demand distributions in the morning and evening commute peaks, the network structure, and link sensitivity. The NDP problem is formulated as a bilevel programming model, in which the upper level is to minimize the weighted sum of total travel time for the network users travelling in both morning and evening commute peaks, and the lower level is to characterize the user equilibrium choice behaviors of the travelers in two peaks. Some extensions on elastic and uncertain peak-hour demands for two peaks are also discussed. The proposed NDP model is transformed into an equivalent MILP, which can be solved by optimization solvers (e.g., CPLEX). Through this transformation, a global solution can be guaranteed. Numerical examples are finally performed to demonstrate the effectiveness of the developed model. It was shown that the proposed NDP model can generate more promising design effect than the traditional NDP model considering one peak-hour demand and avoid the misleading decision. Meanwhile, it was found that how to choose a surveyed peak for data collection is very important for the traditional NDP model considering one peak-hour demand. That is, we should carefully determine which peak-hour demand is less likely to result in misleading outcomes. Furthermore, for weighted parameter setting, we recommend to set $\alpha = \beta$, $\alpha, \beta > 0$ which leads to the best network design effect.

Further studies could be carried out to extend the proposed model in the following aspects. First, although we introduce a conceptual framework of the NDP considering peak-hour uncertain demands in two peaks, a tractable and specific NDP model considering two peak-hour uncertain demands could be an interesting work. Second, the distribution of multiclass users with different value-of-times and mixed vehicles would also show a large diversity in the morning and evening commute peaks. To address such diversity in the NDP problem reveals important investigations.

Notations

Design Variables

$\mathbf{xv} = (\ldots, x_a, \ldots), a \in \overline{A}$: Design variables, namely, link capacity expansions.

Variables to Be Determined in Each Equilibrium

v_a^m, v_a^e: Flow on link $a \in A$ in the morning peak and evening peak, respectively

$t_a^m(v_a^m)$, $t_a^e(v_a^e)$: Travel time on link $a \in A$ in the morning peak and evening peak, respectively

$f_{r,w}^m$: Traffic flow on route $r \in R_w^m$, $w \in W^m$ in the morning peak

$f_{r,w}^e$: Traffic flow on route $r \in R_w^e$, $w \in W^e$ in the evening peak

$c_{r,w}^i$: Travel time on route r for O-D pair w in the morning if $i = m$ or in the evening peak if $i = e$

μ_w^m: Equilibrium minimal travel cost between O-D pair $w \in W^m$ in the morning peak

μ_w^e: Equilibrium minimal travel cost between O-D pair $w \in W^e$ in the evening peak

$b_a(x_a)$: Construction cost for each improved link $a \in \overline{A}$.

Parameter Given

a: Link $a \in A$

$\overline{A}$: Set of candidate links to be improved, $\overline{A} \subset A$

W^m, W^e: Sets of O-D pairs in morning and even peaks respectively

w: O-D pair $w \in W^m \bigcup W^e$

R_w^m: Set of routes connecting O-D pair $w \in W^m$ in morning commute peak

R_w^e: Set of routes connecting O-D pair $w \in W^e$ in evening commute peak

r: Route $r \in R_w^m \bigcup R_w^e$

$\delta_{ar,w}^m$: Link-route indicator $\delta_{ar,w}^m$ equals 1 if route r between O-D pair $w \in W^m$ uses link $a \in A$ at morning peak, and 0 otherwise

$\delta_{ar,w}^e$: Link-route indicator $\delta_{ar,w}^e$ equals 1 if route r between O-D pair $w \in W^e$ uses link $a \in A$ at evening peak, and 0 otherwise

q_w^m, q_w^e: Travel demand for O-D pair in the morning peak and evening peak respectively

α, β: Weighted parameter in objective function

t_a^0: Free flow travel time for link $a \in A$

c_a^0: Existing capacity for link $a \in A$

γ_a: Parameter in link capacity improvement construction cost.

Conflict of Interests

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

Acknowledgments

This study is supported by a program for new century excellent talents in university (Grant no. NCET-10-0637), a research Grant for 2013 Shanghai Postdoctoral Sustentation Fund (no. 13R21416400), and a research Grant for China Postdoctoral Science Foundation (no. 2013M531222).

References

[1] Q. Meng, H. Yang, and M. G. H. Bell, "An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem," *Transportation Research Part B*, vol. 35, no. 1, pp. 83–105, 2001.

[2] M. Abdulaal and L. J. LeBlanc, "Continuous equilibrium network design models," *Transportation Research Part B*, vol. 13, no. 1, pp. 19–32, 1979.

[3] T. L. Magnanti and R. T. Wong, "Network design and transportation-planning—models and algorithms," *Transportation Science*, vol. 18, no. 1, pp. 1–55, 1984.

[4] H. Yang and M. G. H. Bell, "Models and algorithms for road network design: a review and some new developments," *Transport Reviews*, vol. 18, no. 3, pp. 257–278, 1998.

[5] A. Karoonsoontawong and S. Travis Waller, "Comparison of system- and user-optimal stochastic dynamic network design models using Monte Carlo bounding techniques," *Transportation Research Record*, vol. 1923, pp. 91–102, 2005.

[6] S. V. Ukkusuri and S. T. Waller, "Linear programming models for the user and system optimal dynamic network design problem: formulations, comparisons and extensions," *Networks and Spatial Economics*, vol. 8, no. 4, pp. 383–406, 2008.

[7] A. Chen and X. Xu, "Goal programming approach to solving network design problem with multiple objectives and demand uncertainty," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4160–4170, 2012.

[8] A. Chen and C. Yang, "Stochastic transportation network design problem with spatial equity constraint," *Transportation Research Record*, vol. 1882, pp. 97–104, 2004.

[9] H. K. Lo, X. W. Luo, and B. W. Y. Siu, "Degradable transport network: travel time budget of travelers with heterogeneous risk aversion," *Transportation Research Part B*, vol. 40, no. 9, pp. 792–806, 2006.

[10] H. Shao, W. H. K. Lam, Q. Meng, and M. L. Tam, "Demand-driven traffic assignment problem based on travel time reliability," *Transportation Research Record*, vol. 1985, pp. 220–230, 2006.

[11] H. Wang, W. H. K. Lam, X. Zhang, and H. Shao, "Sustainable transportation network design with stochastic demands and chance constraints," *International Journal of Sustainable Transportation*, 2013.

[12] H. K. Lo and W. Y. Szeto, "Time-dependent transport network design: a study of budget sensitivity," *Journal of the Eastern Asia Society For Transportation Studies*, vol. 5, pp. 1124–1139, 2003.

[13] H. K. Lo and W. Y. Szeto, "Time-dependent transport network design under cost-recovery," *Transportation Research Part B*, vol. 43, no. 1, pp. 142–158, 2009.

[14] W. Y. Szeto and H. K. Lo, "Time-dependent transport network improvement and tolling strategies," *Transportation Research Part A*, vol. 42, no. 2, pp. 376–391, 2008.

[15] S. V. Ukkusuri and G. Patil, "Multi-period transportation network design under demand uncertainty," *Transportation Research Part B*, vol. 43, no. 6, pp. 625–642, 2009.

[16] S. T. Waller, K. C. Mouskos, D. Kamaryiannis, and A. K. Ziliaskopoulos, "A linear model for the continuous network design problem," *Computer-Aided Civil and Infrastructure Engineering*, vol. 21, no. 5, pp. 334–345, 2006.

[17] S. T. Waller and A. K. Ziliaskopoulos, "Stochastic dynamic network design problem," *Transportation Research Record*, vol. 1771, pp. 106–113, 2001.

[18] Q. Meng, H. Yang, and H. Wang, "The transportation network design problems," *Transportation Research Part B*, 2013.

[19] A. Chen, Z. Zhou, P. Chootinan, S. Ryu, C. Yang, and S. C. Wong, "Transport network design problem under uncertainty: a review and new developments," *Transport Reviews*, vol. 31, no. 6, pp. 743–768, 2011.

[20] Highway Capacity Manual, "Transportation Research Board," Washington, DC, USA, 2000.

[21] X. Zhang, H. Yang, H.-J. Huang, and H. M. Zhang, "Integrated scheduling of daily work activities and morning-evening commutes with bottleneck congestion," *Transportation Research Part A*, vol. 39, no. 1, pp. 41–60, 2005.

[22] X. Zhang, H.-J. Huang, and H. M. Zhang, "Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city," *Transportation Research Part B*, vol. 42, no. 1, pp. 38–56, 2008.

[23] X. Zhang and B. van Wee, "Efficiency comparison of various parking charge schemes considering daily travel cost in a linear city," *European Journal of Transport and Infrastructure Research*, vol. 11, no. 2, pp. 234–255, 2011.

[24] R. Venkatanarayana, B. L. Smith, and M. J. Demetsky, "Quantum-frequency algorithm for automated identification of traffic patterns," *Transportation Research Record*, vol. 2024, pp. 8–17, 2007.

[25] X. Zhang, "Effects of queue spillover in networks considering simultaneous departure time and route choices," *Transportation Planning and Technology*, vol. 36, no. 3, pp. 267–286, 2013.

[26] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice Hall, Englewood Cliff, NJ, USA, 1985.

[27] D. Z. W. Wang and H. K. Lo, "Global optimum of the linearized network design problem with equilibrium flows," *Transportation Research Part B*, vol. 44, no. 4, pp. 482–492, 2010.

[28] P. Luathep, A. Sumalee, W. H. K. Lam, Z.-C. Li, and H. K. Lo, "Global optimization method for mixed transportation network design problem: a mixed-integer linear programming approach," *Transportation Research Part B*, vol. 45, no. 5, pp. 808–827, 2011.

[29] X. Zhang and B. van Wee, "Enhancing transportation network capacity by congestion pricing with simultaneous toll location and toll level optimization," *Engineering Optimization*, vol. 44, no. 4, pp. 477–488, 2012.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 974398, 6 pages
http://dx.doi.org/10.1155/2014/974398

Research Article

Trafficability Analysis at Traffic Crossing and Parameters Optimization Based on Particle Swarm Optimization Method

Bin He[1] and Qiang Lu[1,2]

[1] *College of Electronics & Information Engineering, Tongji University, Shanghai 201804, China*
[2] *College of Information & Engineering, Taishan Medical University, Taian 271016, China*

Correspondence should be addressed to Bin He; hebin@tongji.edu.cn

Received 1 November 2013; Revised 26 December 2013; Accepted 3 January 2014; Published 3 March 2014

Academic Editor: Hu Shao

In city traffic, it is important to improve transportation efficiency and the spacing of platoon should be shortened when crossing the street. The best method to deal with this problem is automatic control of vehicles. In this paper, a mathematical model is established for the platoon's longitudinal movement. A systematic analysis of longitudinal control law is presented for the platoon of vehicles. However, the parameter calibration for the platoon model is relatively difficult because the platoon model is complex and the parameters are coupled with each other. In this paper, the particle swarm optimization method is introduced to effectively optimize the parameters of platoon. The proposed method effectively finds the optimal parameters based on simulations and makes the spacing of platoon shorter.

1. Introduction

With the rapid development of economy and growth of population, the number of vehicles becomes more and more in China and the increase in vehicle number is unprecedented. As a result, the traffic congestion has been becoming a serious problem. Therefore, it attracts many researchers' attention to improve the platoon's control which can shorten the spacing of platoon and increase traffic flow.

The best method to deal with this problem is automatic control of vehicles. One of the transportation automatic control systems is the Automated Highway System (AHS) which includes the longitudinal control, the lateral control, and the comprehensive control. An AHS is a proposed intelligent transportation system technology designed to provide driverless cars with specific rights-of-way. It is most often touted as a means of traffic congestion relief, as it would drastically reduce following distances and headway, thus allowing more cars to occupy a given stretch of road [1]. It belongs to longitudinal control to cross the traffic intersection at a fast speed. The aim of longitudinal control is to ensure a safe distance and maintain a relatively stable speed between vehicles.

The idea of longitudinal vehicle control has developed very quickly and become very attractive with the increasing issues of traffic congestion and road safety. The researchers showed many technical considerations in the design of longitudinal control systems, such as external forces, process and measurement noise, and sampling and quantization of measurements [2]. Many models which include linear model [3, 4] and nonlinear model [5, 6] had been established and the researchers had presented systematic analysis of longitudinal control for the platoon. The scholars applied many controllers, such as classical proportional integral controller [7, 8] and the fuzzy controller [9] in longitudinal control of platoon. Wang et al. [10] analyzed the local stability of platoon and studied the relationship between the safe headway and the global stability of platoon. However, there are many parameters to be set for the platoon. When the global optimization method is applied to find the optimum parameters of platoon, the method evaluates the fitness of application and numerous fitness evaluations are needed. As such, the convergence is also an important factor in the selection of a method for preventing a platoon from numerous iterations of the method.

The particle swarm optimization (PSO) method is a population based stochastic optimization method proposed by Kennedy and Eberhart in 1995 and is inspired by social behavior such as flocks of birds or schools of fish [11]. The main advantages of PSO method are simple to understand, easy to implement, and quick in convergence [12]. The PSO method has been widely studied. Pan et al. [13] analyzed the kinetic characteristic of three models of PSO method. Ren and Wang [14] proposed an accelerated convergence particle swarm optimization algorithm based on analyzing the convergence of basal PSO method. Coelho [15] presented a novel quantum-behaved PSO method using chaotic mutation operator. Chen and Zhao [16] proposed a particle swarm optimization method that uses an adaptive variable population size and periodic partial increasing or declining individuals in the form of ladder function. The PSO method has been successfully applied in the CPG parameter optimization [17], continuous nonlinear function optimization [11], reactive power and voltage control [18], and parameter tuning of controller for a power system [19].

The remainder of this paper is organized as follows. Section 2 builds a dynamic platoon model and presents a systematic analysis of a longitudinal control law for the platoon. Section 3 shows comprehensive and detailed analysis on parameters optimization based on the PSO method. Section 4 shows the simulation. The conclusion and some ideas about further research are discussed in Section 5.

2. Dynamic Model of Platoon

Various models for vehicles dynamics have been used in the study of longitudinal control of platoon. For a platoon travelling at a constant speed in a fixed direction, we adopt the following third-order model [2]. The state equation of the vehicle is given by

$$\begin{aligned} \dot{x}_i &= v_i, \\ \dot{v}_i &= a_i, \\ \dot{a} &= \frac{1}{\tau_i}\left(c_i - a_i\right), \end{aligned} \tag{1}$$

where x_i, v_i, and a_i denote the absolute position, velocity, and acceleration of the vehicle, respectively. The parameter τ_i is the time constant of the vehicle propulsion system and the parameter c_i is the control input. The assumed configuration for a platoon of five vehicles is shown in Figure 1.

The parameter δ_d is assigned position of vehicle along the road and the parameter δ_i is the deviation of the ith vehicle position from its assigned position. And it is described by

$$\delta_i = x_{i-1} - x_i - \delta_d. \tag{2}$$

Given the direction of platoon from right to left, the platoon variables are the velocity v_0 of the lead vehicle, the velocity v_i of the ith vehicle, the acceleration a_0 of the lead vehicle and the acceleration a_i of the ith vehicle.

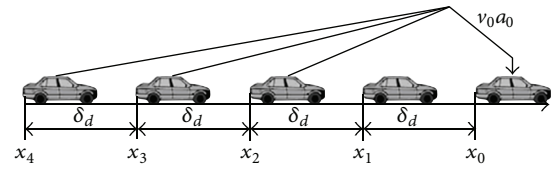


Figure 1: Platoon of five vehicles.

From the input/output point of view of the ith linearized model [3, 4], we can obtain

$$\ddot{x}_i = c_i. \tag{3}$$

The objective of longitudinal control is to maintain the spacing error below a predetermined level or, if possible, at zero. As to the first vehicle, the controller [6] can be written as

$$c_1 = c_{x1}\delta_1 + c_{v1}\dot{\delta}_1 + c_{a1}\ddot{\delta}_1 + c_{vL1}\left(v_0 - v_L\right) + c_{aL1}\left(a_0 - a_L\right), \tag{4}$$

where c_{x1}, c_{v1}, c_{a1}, c_{vL1}, and c_{aL1} are design constants [6]. The parameter v_L denotes the steady-state velocity of the lead vehicle and the parameter a_L is the steady-state acceleration of the lead vehicle.

When $w_0 = v_0 - v_L$ and $\dot{w}_0 = a_0 - a_L$, we can obtain

$$c_1 = c_{x1}\delta_1 + c_{v1}\dot{\delta}_1 + c_{a1}\ddot{\delta}_1 + c_{vL1}w_0 + c_{aL1}\dot{w}_0. \tag{5}$$

Differentiating both sides of (2) three times with respect to the time variable, we obtain

$$\dddot{\delta}_1 = \dddot{x}_0 - \dddot{x}_1. \tag{6}$$

Taking Laplace transforms, we obtain

$$h_{\delta_1 \omega_0}(s) = \frac{s^2 - c_{aL1}s - c_{vL1}}{s^3 + c_{a1}s^2 + c_{v1}s + c_{x1}}. \tag{7}$$

For the ith vehicle, we can also obtain

$$c_i = c_x\delta_i + c_v\dot{\delta}_i + c_a\ddot{\delta}_i + c_{vL}\left(v_0 - v_i\right) + c_{aL}\left(a_0 - a_i\right), \tag{8}$$

where c_x, c_v, c_a, c_{vL}, and c_{aL} are design constants [6]. Differentiating both sides of (2) three times with respect to the time variable, we obtain

$$\dddot{\delta}_i = \dddot{x}_{i-1} - \dddot{x}_i. \tag{9}$$

Taking Laplace transforms, we obtain

$$h_{\delta_i \delta_{i-1}}(s) = \frac{c_a s^2 + c_v s + c_x}{s^3 + \left(c_a + c_{aL}\right)s^2 + \left(c_v + c_{vL}\right)s + c_x}. \tag{10}$$

For the second vehicle, the transfer function is

$$h_{\delta_2 \delta_1}(s) = \frac{\left(c_{a1} - c_{aL}\right)s^2 + \left(c_{v1} - c_{vL}\right)s + c_{x1}}{s^3 + \left(c_a + c_{aL}\right)s^2 + \left(c_v + c_{vL}\right)s + c_x}. \tag{11}$$

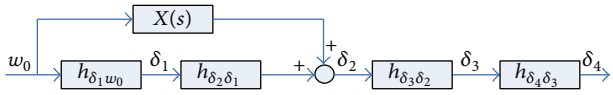


Figure 2: Block diagram for the platoon.

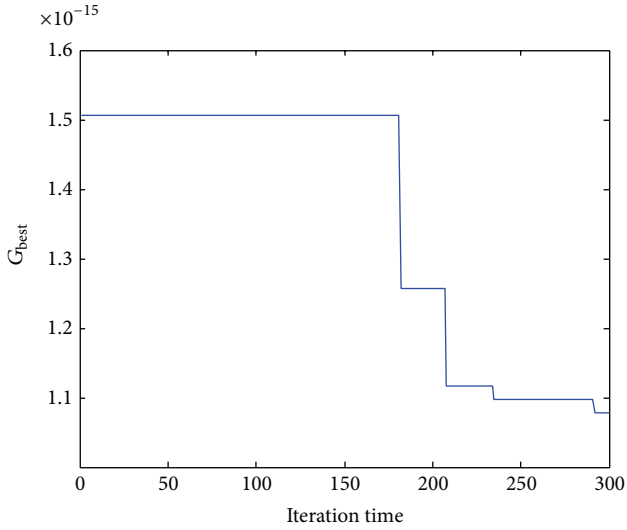


Figure 3: Curve of G_{best}.

Therefore, the main design objective for the longitudinal control law is shown in Figure 2, where $X(s) = (c_{aL1}s + c_{vL1})/(s^3 + (c_a + c_{aL})s^2 + (c_v + c_{vL})s + c_x)$.

We use the block diagram in Figure 2 to analyze the platoon. Some considerations suggest the main design objective for the longitudinal control law.

(1) In order to keep the platoon stable, all poles of the transfer function are required to be in the left half plane. According to Routh-Hurwitz stability criterion, we obtain $c_{a1} > 0$, $c_{x1} > 0$, $c_{a1}c_{v1} > c_{x1}$, and $(c_a + c_{aL})(c_v + c_{vL}) > c_x$.

(2) Since the perturbations in the lead vehicle's velocity from its steady-state value should not get magnified from one vehicle to the next as one goes down the platoon, we require that $|\delta_i(jw)/\delta_{i-1}(jw)| < 1$ for all $w > 0$.

Therefore, these parameters $c_{x1}, c_{v1}, c_{a1}, c_{vL1}, c_{aL1}, c_x, c_v, c_a, c_{vL}$, and c_{aL} are needed to be selected correctly.

3. Particle Swarm Optimization Method

The particle swarm optimization method uses the concept called particle and swarm [17, 20]. The particles correspond to an animal, bird, and insect in a herd, flock, and swarm, respectively. Let us consider a swarm including m particles which are seeking the optimum value of the objective function in an n-dimensional search space, each particle having a vector of position $Y_i = (y_{i1}, y_{i2}, \ldots, y_{in})$, $i = 1, 2, \ldots, m$ which is associated with a solution, a vector of velocity $U_i = (u_{i1}, u_{i2}, \ldots, u_{in})$, $i = 1, 2, \ldots, m$ which determines the movement value of a particle in each dimension to improve its current position, and a vector of particle best position $P_{\text{best}} = (p_{\text{best1}}, p_{\text{best2}}, \ldots, p_{\text{best}n})$ which is associated with most fitted positions of a particle from the first step of the algorithm. It is notable that the fitness of a position can easily be calculated considering the objective function of the optimization problem. A vector of the global best particle $G_{\text{best}} = (G_{\text{best1}}, G_{\text{best2}}, \ldots, G_{\text{best}n})$ is reserved for knowledge sharing among all particles of a swarm [21]. Using the aforementioned notations, each argument of the velocity and position vector for each particle in the swarm is updated through the multiple iterations of the algorithm using the following model:

$$U_i(t+1) = bU_i(t) + \varphi_1\left(P_{\text{best}} - Y_i(t)\right) + \varphi_2\left(G_{\text{best}} - Y_i(t)\right), \quad i = 1, 2, \ldots, m \tag{12}$$

$$Y_i(t+1) = Y_i(t) + U_i(t+1), \quad i = 1, 2, \ldots, m,$$

where $U_i(t)$ is a velocity and $Y_i(t)$ is a position of the particle at t iteration. P_{best} is a previous best position and G_{best} is a global best position of each particle obtained so far. φ_1 and φ_2 are determined as $\varphi_1 = \text{rand}(0, k_1)$ and $\varphi_2 = \text{rand}(0, k_2)$. k_1 is a cognition learning factor and k_2 is a social learning factor. b is the inertia weight which determines the particle speed prior to the current velocity and thus functions as a balancing algorithm between global search and local search capacities.

4. Simulations

4.1. Simulation Scheme. In this paper, the PSO method is used to optimize the parameters c_{vL1} and c_{aL1}. In order to reduce the particle dimension and avoid complicated analysis, we choose $c_{a1} = c_a + c_{aL}$, $c_{v1} = c_v + c_{vL}$, and $c_{x1} = c_x$. Then (7), (10), and (11) have the same poles. The values of parameters are set as $c_{a1} = 15$, $c_{v1} = 74$, $c_{x1} = 120$, $c_a = 5$, $c_v = 49$, $c_x = 120$, $c_{aL} = 10$, and $c_{vL} = 25$ [6]. The spacing between these vehicles is selected as fitness function. To examine the behavior of the platoon with optimized parameters, the simulation for platoon consisting of 5 vehicles is run in MATLAB. In the simulation, all the vehicles are assumed to be initially travelling at the velocity of 0 m/s. When $t = 0$ s, the lead vehicle's velocity increases from initial value and the acceleration is 2 m/s^2. Finally it reaches the value of 20 m/s. Taking inverse Laplace transforms for (7), the fitness function is obtained:

$$F(c_{aL1}, c_{vL1}, t) = \left(\frac{2c_{aL1} - c_{vL1}/2 + 8}{e^{4t}} + \frac{3c_{aL1} - c_{vL1}/2 + 18}{e^{6t}} - \frac{5c_{aL1} - c_{vL1} + 25}{e^{5t}}\right) \times 20, \tag{13}$$

where $t = 10$ s.

The size of the swarm is 50 and the number of iterations is 300. Parameters c_{vL1} and c_{aL1} are varied in the interval $[-3.1, 0]$ in step of 0.01. The parameters k_1 and k_2 are set as $k_1 = 0.5$ and $k_2 = 2.5$. The inertia weight [22] is described by

$$b = b_{\max} - \frac{b_{\max} - b_{\min}}{M} j, \tag{14}$$

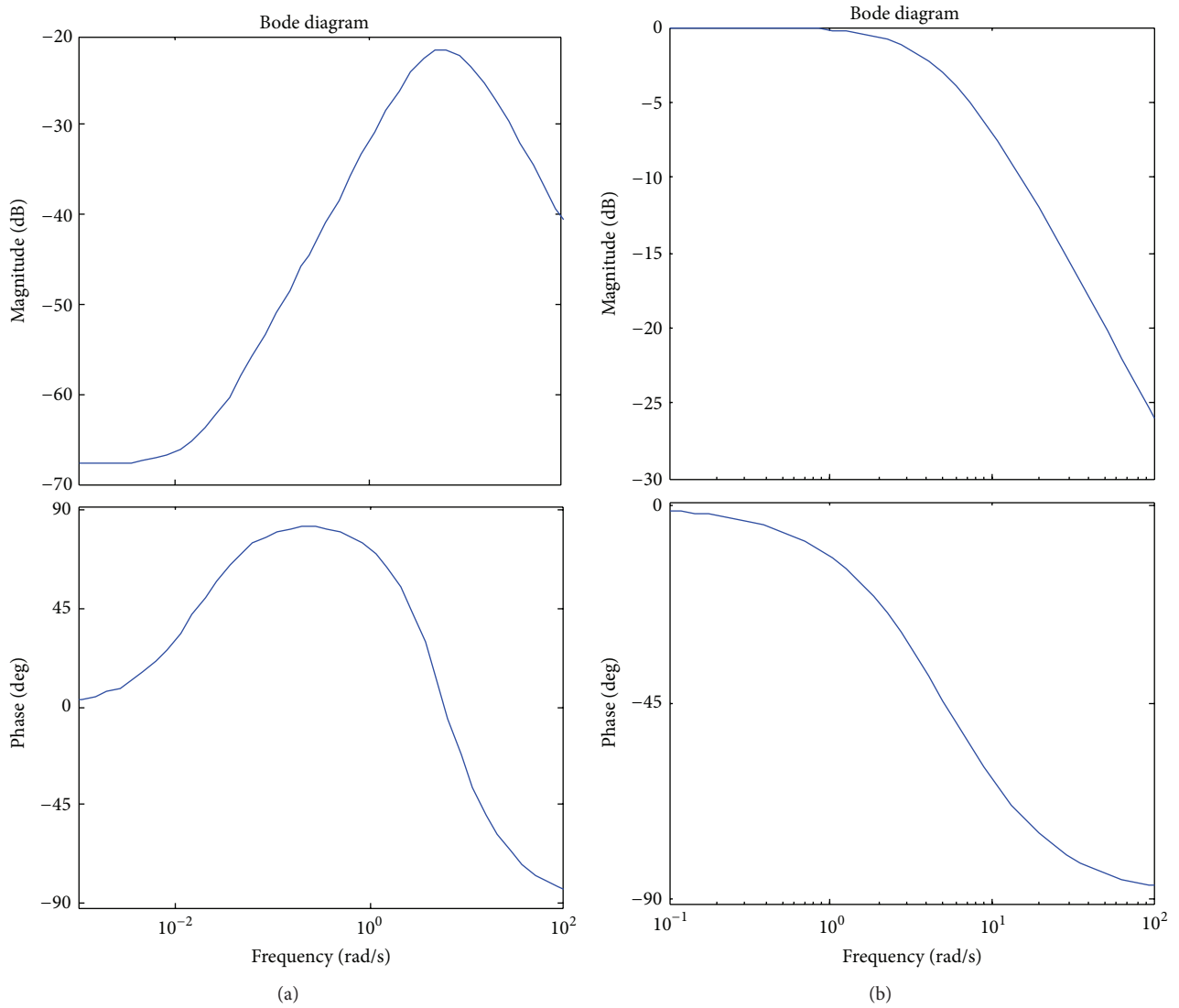


Figure 4: Bode diagrams.

where $b_{\max}$ and $b_{\min}$ are the maximal and minimum values of the inertia weight, and they are set as $b_{\max} = 1.2$ and $b_{\min} = 0.4$. M is the maximal iteration time and j is current iteration time.

The optimization procedure of the PSO method is given by the following steps.

Step 1. Initialize a population of particles with random positions and velocities.

Step 2. Evaluate each particle's fitness value.

Step 3. Compare each particle's fitness with the particle's P_{best}. If the current value is better than P_{best}, then set the P_{best} value equal to the current value and the P_{best} location equal to the current location.

Step 4. Compare the fitness with the population's overall previous best G_{best}. If the current value is better than G_{best}, then reset the G_{best} value to the current particle's array index and value.

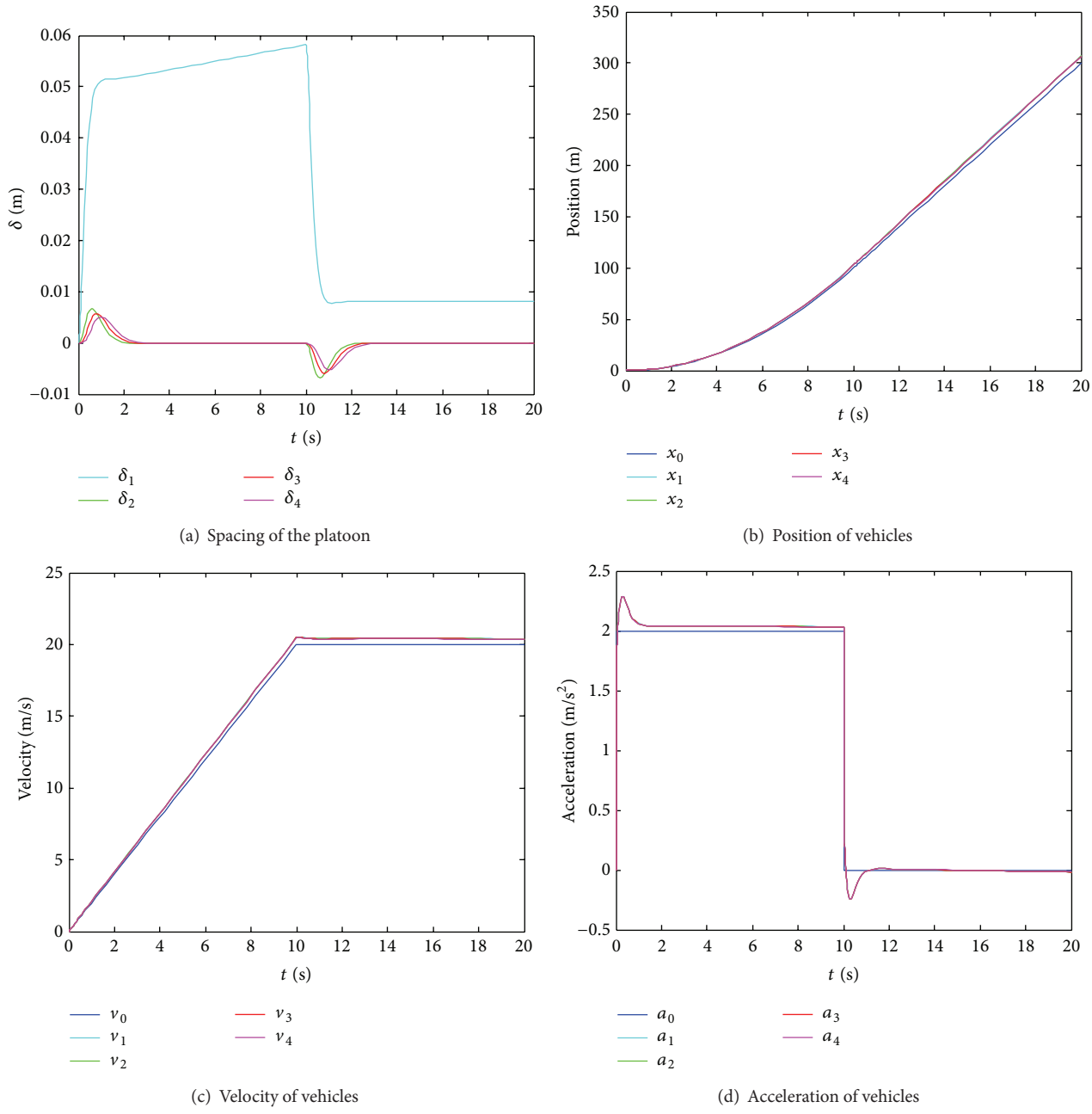
Step 5. Update each particle's velocity and position according to (12).

Step 6. Return to Step 2 until the maximum number of iterations is reached.

4.2. Simulations. Through the particle swarm optimization method, we obtain the values of parameters which are $c_{aL1} = -3.0315$ and $c_{vL1} = -0.0492$. The curve of G_{best} is obtained, as shown in Figure 3.

Then transfer functions are described by

$$h_{\delta_1\omega_0}(s) = \frac{s^2 + 3.0315s + 0.0492}{s^3 + 15s^2 + 74s + 120},$$



(a) Spacing of the platoon

(b) Position of vehicles

(c) Velocity of vehicles

(d) Acceleration of vehicles

Figure 5: Diagrams of the platoon.

$$h_{\delta_2\delta_1}(s) = h_{\delta_3\delta_2}(s) = h_{\delta_4\delta_3}(s) = \frac{5s^2 + 49s + 120}{s^3 + 15s^2 + 74s + 120}. \quad (15)$$

The Bode Diagrams of two transfer functions are shown in Figure 4. We can see that the frequence response meets the second condition of the platoon.

According to the longitudinal control law, the diagram of platoon is shown in Figure 5.

The simulation results show that the deviations of vehicles from their preassigned positions do not exceed 0.06 m. The accelerations of vehicles in the platoon are within the range of acceptable comfort limits.

5. Conclusion

In this paper, a mathematical model is built for the platoon's longitudinal movement and a longitudinal control law is analyzed in detail. It is well known that the parameter calibration of the platoon is a difficult problem. However, the PSO method effectively finds the parameters by storing the previous knowledge of particles and estimating the best positions and achieves the computation time of 2.7 s.

The Automated Highway System is seen as a better way to deal with the problem of the traffic congestion. In general, the road conditions are complex. A number of studies should be done to analyze the effects of disturbances and modeling errors which may need further investigation.

Conflict of Interests

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

Acknowledgments

The work was supported by the National Natural Science Foundation of China under Grants 61040056 and 61070127 and Shanghai Key Project of Foundation funding under Grant 09JC1414600, China.

References

[1] Wikipedia Contributors, "Platoon (automobile)," Wikipedia, The Free Encyclopedia, 2013, http://en.wikipedia.org/wiki/Platoon_(automobile).

[2] S. E. Shladover, "Longitudinal control of automotive vehicles in close-formation platoons," *Journal of Dynamic Systems, Measurement and Control*, vol. 113, no. 2, pp. 231–241, 1991.

[3] W. Yue and G. Guo, "Control of autonomous platoon under networked communication effect," *Control Theory and Applications*, vol. 28, no. 7, pp. 1041–1048, 2011.

[4] S. Sheikholeslam and C. A. Desoer, "Longitudinal control of a platoon of vehicles I: linear model," Path Research Report UCB-ITS-PRR-89-3, 1989.

[5] W. Yue and G. Guo, "Nonlinear modelling and control of autonomous platoon," *Control and Decision*, vol. 24, no. 7, pp. 996–1000, 2009.

[6] S. Sheikholeslam and C. A. Desoer, "Longitudinal control of a platoon of vehicles; III: nonlinear model," Path Research Report UCB-ITS-PRR-90-1, 1990.

[7] T. S. No, K. T. Chong, and D. H. Roh, "A lyapunov function approach to longitudinal control of vehicles in a platoon," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 1, pp. 116–124, 2000.

[8] V. Milanés, J. Villagrá, J. Pérez, and C. González, "Low-speed longitudinal controllers for mass-produced cars: a comparative study," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 1, pp. 620–628, 2012.

[9] Q. Wu, Z. He, X. Chu, and D. Lu, "An application of the adaptive fuzzy control in the longitudinal control of the platoon," in *Proceedings of the IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA '08)*, vol. 2, pp. 344–348, December 2008.

[10] F.-H. Wang, M. Yang, R.-Q. Yang, and C.-X. Wang, "Platoon stability based safe headway and its determination," *Journal of Shanghai Jiaotong University*, vol. 44, no. 6, pp. 787–791, 2010.

[11] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, December 1995.

[12] A. Abraham, H. Guo, and H. Liu, "Swarm intelligence: foundations, perspectives and applications," *Studies in Computational Intelligence*, vol. 26, pp. 3–25, 2006.

[13] F. Pan, J. Chen, M. G. Gan, T. Cai, and X.-Y. Tu, "Model analysis of particle swarm optimizer," *Acta Automatica Sinica*, vol. 32, no. 3, pp. 368–377, 2006.

[14] Z.-H. Ren and J. Wang, "Accelerate convergence particle swarm optimization algorithm," *Control and Decision*, vol. 26, no. 2, pp. 201–206, 2011.

[15] L. D. S. Coelho, "A quantum particle swarm optimizer with chaotic mutation operator," *Chaos, Solitons and Fractals*, vol. 37, no. 5, pp. 1409–1418, 2008.

[16] D. Chen and C. Zhao, "Particle swarm optimization with adaptive population size and its application," *Applied Soft Computing Journal*, vol. 9, no. 1, pp. 39–48, 2009.

[17] J. J. Kim, J. W. Lee, and J. J. Lee, "Central pattern generator parameter search for a biped walking robot using nonparametric estimation based particle swarm optimization," *International Journal of Control, Automation and Systems*, vol. 7, no. 3, pp. 447–457, 2009.

[18] Y. Fukuyama and H. Yoshida, "A particle swarm optimization for reactive power and voltage control in electric power systems," in *Proceedings of the IEEE Congress on Evolutionary Computation*, vol. 1, pp. 87–93, Seoul, Republic of Korea, May 2001.

[19] T. Okada, T. Watanabe, and K. Yasuda, "Parameter tuning of fixed structure controller for power system stability enhancement," in *Proceedings of the IEEE/PES Transmission and Distribution Conference and Exhibition*, vol. 1, pp. 162–167, October 2002.

[20] W. T. Sung and Y. C. Chiang, "Improved particle swarm optimization algorithm for android medical care IOT using modified parameters," *Journal of Medical Systems*, vol. 36, no. 6, pp. 3755–3763, 2012.

[21] K. Khalili-Damghani, A. R. Abtahi, and M. Tavana, "A new multi-objective particle swarm optimization method for solving reliability redundancy allocation problems," *Reliability Engineering and System Safety*, vol. 111, pp. 58–75, 2013.

[22] J. Yisu, J. Knowles, L. Hongmei, L. Yizeng, and D. B. Kell, "The landscape adaptive particle swarm optimizer," *Applied Soft Computing Journal*, vol. 8, no. 1, pp. 295–304, 2008.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 502708, 13 pages
http://dx.doi.org/10.1155/2014/502708

Research Article

Investigating the In-Vehicle Crowding Cost Functions for Public Transit Modes

Feifei Qin[1,2]

[1] *School of Urban Rail Transportation, Soochow University, No. 1 Shizi Street, Suzhou 215006, China*
[2] *Molde University College, P.O. Box 2110, 6402 Molde, Norway*

Correspondence should be addressed to Feifei Qin; merryqff@hotmail.com

Received 4 November 2013; Revised 12 December 2013; Accepted 13 December 2013; Published 23 February 2014

Academic Editor: X. Zhang

In the densely populated metropolitan area, empirical studies have found that overcrowding inside transit vehicles has become substantially worse and worse over recent years. Chronic in-vehicle crowding is not only caused by a lack of physical infrastructure, but also triggered by inadequate service provisions. Given the prevalence of overcrowded transit vehicles, this paper conducts both quantitative and qualitative studies, especially focusing on remodeling the in-vehicle crowding cost functions for different transit modes. Three numerical case studies show that applying distinct in-vehicle crowding cost functions to different transit modes has implications not only for the cost structure of transit systems and the magnitude of optimal service provisions but also for the presence of economies of scale in consumption.

1. Introduction

Residents in densely populated urban areas inevitably face travelling trauma. For example, during peak periods, the substantial growth of cars has quickly crammed road capacity, causing severe traffic congestion. Simultaneously, given the difficulties in providing additional rolling stocks, public transit facilities have not been able to cope with growing ridership. Thus, in rush hour, public transit passengers are often forced to tolerate the discomfort and stress associated with having to stand inside extremely overcrowded vehicles.

In the field of transport economics, a large number of theoretical researches and practical studies about road congestion have been conducted by many transport economists [1, 2]. However, the effect of in-vehicle crowding on efficient transit operations has not been carefully considered, even though some exploratory research attempted to address this problem as early as the 1970s [3, 4]. Specifically, the limited body of studies does not provide effective approaches to deal with the in-vehicle crowding effect, leaving a considerable gap between the theoretical research and the practical implementation. Reasons why the research on in-vehicle crowding is still in its infancy can be explored from three aspects. First, given the prevalence of overcrowding in transit, it is surprising that, until now, no definitive definition exists to describe the in-vehicle crowding issues. Second, owing to the diversity of transit modes, the existing research does not afford legible ways to measure the level of in-vehicle crowding, which in turn causes more confusion in terms of evaluating and modeling. Finally, there are many controversies about how to quantify and monetize the crowding costs and incorporate them into transit scheduling and pricing. Ambitiously aiming to solve the above problems in one shot, this paper comprehensively and systematically analyzes the in-vehicle crowding effect through qualitative and quantitative analysis.

The rest of paper is organized as follows. Section 2 reviews some relevant literature and seeks to develop a conceptual framework for precisely defining, measuring, and evaluating the in-vehicle crowding for different transit modes. In Section 3, after reexamining the extant crowding cost functions, three nonlinear function forms are set forth, which aim to capture the relationship between the average value of riding time and crowding levels. Additionally, to demonstrate the performances of the proposed nonlinear crowding cost functions, optimal frequencies are sought by minimizing the total system cost. In Section 4, we simulate the variation in optimal frequencies under a wide range of demand rate

for interurban bus, light rail, and heavy rail, respectively. Section 5 concludes with the main findings.

2. Conceptual Framework for the In-Vehicle Crowding Effect

Noting widespread dissatisfaction with overcrowding, a considerable number of studies have been conducted to specify the social effects of overcrowding from various disciplines, including sociology, psychology, and behavioral sciences. Unfortunately, in most transport economics studies, the in-vehicle crowding effect is incidental to the core research purpose. For example, Tian et al. [5] introduced an unspecified crowding cost function to study the equilibrium properties of many-to-one mass transit systems. de Palma et al. [6] optimized the location and pricing problems for metro stations through specifying one novel nonlinear crowding cost function. Only few scientific studies have been centered on addressing the in-vehicle crowding issues for public transit. To pave the way for the subsequent mathematical modeling and numerical experiments, different areas of research have been integrated into one conceptual framework, including passengers' crowding perceptions, quantitative and qualitative measurements, and the empirical estimates of crowding penalties.

As early as the 1970s, the complex nature of crowding problem attracted significant attention in the field of psychology [7, 8]. From a psychologist's point of view, the term "*crowding*" is made up of both objective and subjective elements [9]. Objective crowding is quantitatively measured by the number of people per unit of space, while the subjective crowding refers to a personal perceived state of mind that may occur when there is great disparity between expected interpersonal distance and actual one [10]. Thus, a central aspect of subjective crowding is the "*felt lack of behavioral freedom and privacy*" when the physical space becomes too limited. In the field of transport studies, "crowding" amounts to the unpleasant experiences of too many passengers fitting into a confined space, thus worsening passengers' well-being [11]. Accordingly, four key crowding effects have been specified in the literature, namely, in-vehicle crowding, platform crowding, excessive waiting time, and increased dwell time [12]. In this paper, only in-vehicle crowding effect is considered in the following analysis.

Since different transit modes are characterized by different interior layout with varying amounts of seats and standing space, passengers have different expectations of getting seats or travelling standing up. For example, because intercity buses and trains are merely designed for seating, passengers believe that purchasing one ticket entitles them to obtain one seat. Once they are forced to stand, they feel great stress and discomfort. In this case, crowding occurs when passengers are unable to obtain seats as they initially expected. On the contrary, the urban rail-based modes and some buses are designed to carry large number of standees rather than to provide as many seats as possible. Consequently, passengers do not view standing as crowding. As more passengers board, passengers start to perceive overcrowding when standing density exceeds one threshold, leading to an "*invasion of privacy*" [7]. In Table 1, the disparate perceptions of in-vehicle crowding are shown for different transit modes. Since transit modes are different in physical capacity design, we cannot apply one common definition for all modes to specify under what conditions passengers can perceive the in-vehicle crowding.

Table 1: The perception of in-vehicle crowding in different transit modes.

	Modes	Passengers' crowding feeling
Intercity (longer journeys)	Intercity rail	No space for seating
	Interurban bus	No space for seating
Commuter (shorter journeys)	Bus	Standees having less than 0.35 m^2 of space
	Light rail	Standees having less than 0.25 m^2 of space
	Heavy rail	Standees having less than 0.17 m^2 of space

Source: edited by author according to different vehicle design guidelines.

Since the crowding perceptions may differ among transit modes, the in-vehicle crowding is less easily measurable. The empirical research chooses either qualitative descriptions or quantitative measures to gauge its level. For example, to model the competition occurring on one rail-based route, Accent Marketing Research and Hague Consulting Group [13] merely described crowding levels from two ways: all seats occupied and easy to find a seat. Noting the limitations of this two-level measurement, Lam et al. [14], Accent [15], and Maunsell and MacDonald [16] all provided the multi-level qualitative descriptions to crudely specify whether the coach condition is crushed or not. Although these qualitative descriptions to some extent assist in specifying the crowding levels, the lack of accuracy and quantification prevents them from being widely used in transportation research. Consequently, to specify the crowding levels as precisely as possible, two quantitative metrics, namely, Load Factor and passenger standing density (pass/m^2), have been extensively adopted. As mentioned above, for interurban buses and trains where users feel overcrowding when they cannot find seats, it is appropriate to measure the crowding degree in terms of Load Factor (LF). However, for massive capacity transit modes, the Load Factor cannot be sensibly utilized, because a high standing capacity can easily make Load Factor exceed 300%. Thus, instead of consistently choosing Load Factor, current empirical studies [11, 17] have recommended adopting standing density (i.e., the number of standing passengers per square meter) as the proper measurement. Table 2 lists seat capacity, crush capacity, and proper crowding measurements for different transit modes.

As early as the 1970s, Goodwin [18] pointed out that overcrowding substantially affects the monetary values that passengers place on travelling time savings. As the crowding level increases, the disutility of travelling increases, so that passengers have to attach significantly high values of riding time to be equally well-off [19]. Similar to transfer penalty, the in-vehicle crowding can be evaluated by crowding penalties.

Table 2: Vehicle capacity and crowding measurements for transit modes.

Urban transport mode	Vehicle type	Seat capacity	Crush capacity[a]	Crowding perception	Crowding measurement
Intercity train	All types	65–95	100–170	Impossibility in finding seats	Load Factor
Interurban bus	All types	38–50	55–65	Impossibility in finding seats	Load Factor
Tram	All types	40–60	90–150	Standing in crush-loaded condition	Standing density
Bus	Single-decker	35–50	55–85	Impossibility in finding seats	Load Factor
	Double-decker	60–75	90–100	Impossibility in finding seats	Load Factor
	Articulated	60–80	100–160	Standing in crush-loaded condition	Load Factor
Light rail	All types	60–80	160–180	Standing in crush-loaded condition	Standing density
Subway	All types	60–80	200–300	Standing in crush-loaded condition	Standing density

[a]The crush capacity is the maximum number of people that it is physically possible to squeeze onto a transit vehicle.

In the published research on crowding valuation, three alternative ways are used to express crowding penalties, namely the time multiplier, the monetary value per time unit, and the monetary value per trip [20]. Since the time multiplier is inherently more transferable across different contexts compared with the monetary valuations, it has become the predominant approach.

Historically, the empirical estimates of crowding penalties have been reported in many unpublished consultancy repots. Only very recently has a small amount of research been published in academic journals. Of all these studies, the one conducted by Wardman and Whelan [21] is the most comprehensive in academic scope, as it estimated crowding penalties not only in terms of load factor but also in terms of standing density (see Table 3). The process can be described as follows: first, the levels of crowding inside transit vehicles are measured by the Load Factor and/or standing density. Then, a Stated Preference (SP) survey is conducted to investigate the passengers' trade-off between waiting time, fares, and the level of crowding. Finally, through Logit models (mainly multinomial Logit (MNL)), the crowding penalties are typically estimated by calculating the ratio of crowding coefficients to the noncrowding coefficient.

This paper largely builds on those values recommend by Whelan and Crockett [11] for the London area. The potential applications of the proposed crowding cost function in other regions are also possible if the city-specified crowding penalties can be obtained by SP/RP survey (see Li and Hensher [20] for Sydney region).

A close look at the magnitudes of crowding penalties given in Table 3 yields some interesting insights. First, the in-vehicle crowding degree, in terms of load factor or standing density, influences the value of riding time spent standing and seating. As the crowding degree increases to crush capacity (load factor >150% or standing density (pass/m^2) >6), the values of riding time get substantially higher. Second, the valuations of crowding penalties are 1.5 to 1.85 when few others are standing. When the vehicle is at crush capacity, crowding penalties exceed 2, which is very much in line with the recommended premium typically attached to walking time and waiting time. Finally, it is obvious that the variation of crowding penalties between successive Load Factors (from 1.5 at 100% to 2.37 at 200%) is much larger than the variation in terms of standing density (from 1.62 for one pass/m^2 to 2.04 for six pass/m^2).

Table 3: Crowding penalties for rail-based modes for London area.

Load Factor (LF)	Crowding penalties for standing	Stand density (pass./m^2)	Crowding penalties for standing
		0	1.53
100	1.5	1	1.62
120	1.67	2	1.70
140	1.85	3	1.79
160	2.02	4	1.87
180	2.2	5	1.96
200	2.37	6	2.04

Source: Whelan and Crockett (2009) [11].

3. Mathematical Modeling for In-Vehicle Crowding Cost

The current empirical studies suggest no matter what type of indicator is employed, crowding penalties monotonically increase with crowding levels. However, the appropriate function form employed to describe the mathematical relationship between the crowding penalties and crowding levels is still underresearched. Furthermore, since the perceptions of the in-vehicle crowding vary among transit modes, we cannot use one common crowding cost function for all transit modes. In what follows, after reexamining the classical function form proposed by Kraus (1991) [22] and other subsequent versions, three nonlinear functions are set out with some explanations.

3.1. Reinvestigating Existing In-Vehicle Crowding Cost Functions. In contrast to the conventional research that generally assigns the same time value regardless of in-vehicle conditions, Kraus (1991) [22] first investigated the effect of the in-vehicle crowding on optimal fare settings. Viewed as a seminal study, Kraus's work divided the value of riding time into two parts: the value of time for passengers who secure seats (P_{v0}) and the value of time in crowded situations where

passengers have to stand ($P_{v0} + \pi$). The generic form of the model is given as

$$P_v = \begin{cases} P_{v0} & \text{No people stand} \\ P_{v0} + \pi & \text{some people stand.} \end{cases} \quad (1)$$

In (1), π is the extra part perceived by standees. Since this simple function form provides an understandable way to capture the crowding externality, a large number of subsequent studies either directly adopted or slightly revised it based on their specific research purposes.

Jansson [23] and Tian et al. [5], respectively, developed their abstract crowding cost functions in which the value of riding time is a function of occupancy rate. But the exact relationship was not explicitly specified. Jara-Diaz and Gschwender [24] updated Kraus's piecewise-constant function by assuming that the values of riding time are no longer constant but rather a continuous on-decreasing function

$$P_v = P_{v0}\left(1 + \rho\phi\right), \quad (2)$$

where, as a markup on the value of riding time (P_{v0}), the crowding penalty (ρ) linearly varies with occupancy rate (ϕ). In contrast to this widely used linear function, a few studies have developed nonlinear ones as incidental to their core research [6, 25].

Although the traditional linear function reflects the additional discomfort and inconvenience associated with in-vehicle crowding, some shortcomings can be easily detected. First, the linear relationship between crowding penalties and crowding levels has been questioned by many empirical observations. A large amount of evidence suggests that the nonlinearity may be empirically supported, especially when psychological elements are factored into crowding perceptions. Second, the existing crowding cost functions only offer a deterministic form and do not specify the probability of having to stand or get a seat. Actually, introducing the probability of having to stand (or sit) as a weight may provide a more convincing way to present the average value of time. Third, a crucial assumption of the linear form is that the in-vehicle crowding effect is sufficiently smooth. Actually, the crowding perception would be sharply deteriorated at a particular crowding level. Finally, since transit modes are different in many dimensions, applying a common crowding cost function for all transit modes and disregarding the specific characteristics of each would lead to undesirable biases.

3.2. Developing Nonlinear Crowding Cost Functions for Different Transit Modes. Whelan and Crockett [11] not only provided a detailed estimation of crowding penalties but also examined possible function forms, such as linear, exponential, and Gompertz. The goodness of fit test for alternative forms indicated that the nonlinearity of crowding penalties with respect to crowding degrees is presented [21]. In addition, to deal with in-vehicle crowding problem for massive capacity modes, such as heavy rail, step function forms could be proper in that it allows discrete jumps among different standing densities [26].

Formula 1: One-Step Nonlinear Function for Crowding Costs. For some modes, such as interurban buses and rails, when the number of passengers inside the vehicle is lower than the seating capacity, the crowding penalty is not active. Thus, the value of riding time remains constant. However, as long as one passenger has to stand, the discomfort and stress resulting from standing would cause the value he attaches to riding time savings to be much higher than those seated passengers. To capture this, a one-step nonlinear function can be derived as follows:

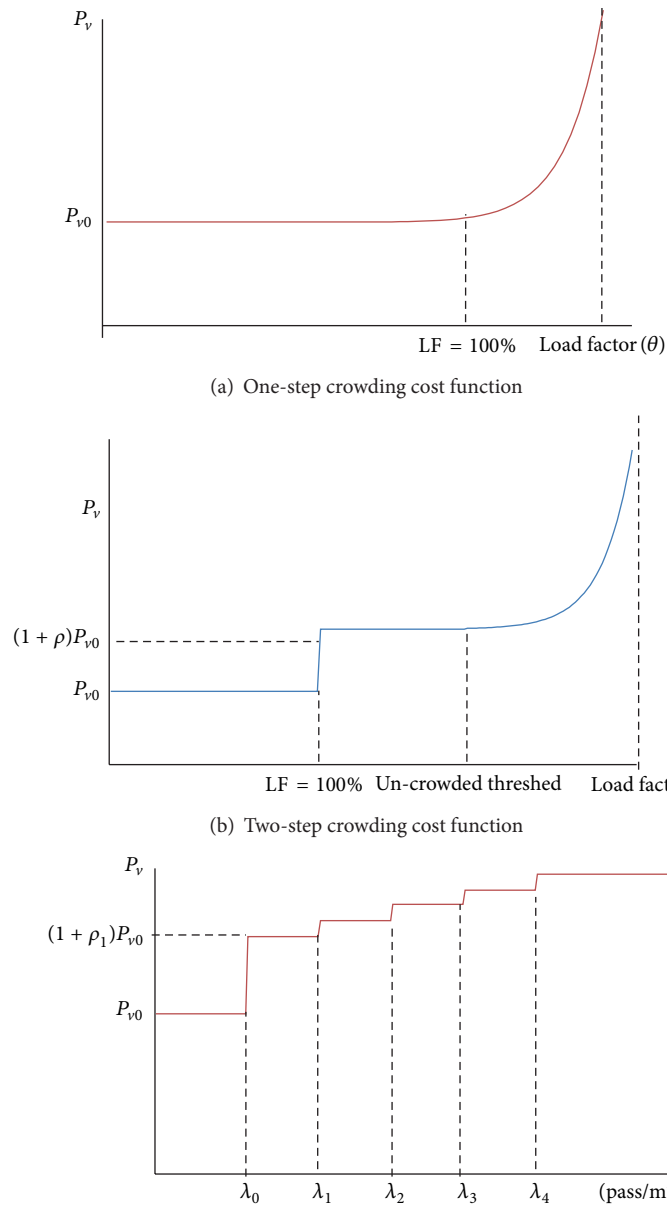
$$P_v = P_{v0}\frac{k}{N} + \left(1 - \frac{k}{N}\right)P_{v0}\left(1 + \beta e^{\alpha(\theta-1)}\right). \quad (3)$$

Equation (3) expresses the weighted value of riding time with a probabilistic form. If we denote k by seating capacity and N by number of passengers on board, the ratio of k/N shows the probability of getting a seat. To clarify the distinction between seating and standing, we assign a constant value of riding time (P_{v0}) to seated passengers and assume that crowding penalties for standees exponentially increase with load factor θ (i.e., N/k). In the exponential part, a certain value of α (more than five) can be arbitrarily given, which serves as an adjustment device to enlarge or shrink the exponential part. By contrast, another parameter β needs to be carefully calibrated from the empirical estimation of crowding penalties. For easy exploring, (3) can be re-arranged as

$$P_v = P_{v0} + \left(1 - \frac{k}{N}\right)P_{v0}\beta e^{\alpha(\theta-1)}. \quad (4)$$

It is envisaged that if the number of passengers is fewer than the number of seats ($N < k$), the item ($\theta - 1$) takes a negative sign. Through the scale parameter α, the exponential part $\beta e^{\alpha(\theta-1)}$ will get close to zero, causing the average value of riding time to be P_{v0}. If the number of passengers equals seating capacity ($N = k$), the item ($1-k/N$) is zero, indicating that the resulting average value of riding time is P_{v0}. As more passengers have to stand in the vehicle, a mark-up ($\beta e^{\alpha(\theta-1)}$) appears as the variable crowding penalty. In the most acute crowding condition, the value of riding time could approach infinity. To explain this function more clearly, we plot the weighted value of riding time against the load factor (θ) in Figure 1(a).

Formula 2: Two-Step Nonlinear Function for Crowding Costs. In contrast to interurban modes, most rail-based urban transit modes (such as tram and light rail) are designed to carry large numbers of standees. For those modes, in-vehicle crowding only takes place when the crowding level exceeds a certain threshold. Although until now no specific research has thus far focused on exploring this threshold, a general accepted observation is that in-vehicle crowding is active when load factor reaches 140% or standing density is above four pass/m^2. Observing this feature, only discriminating standing or sitting cannot describe the impact of crowding on



(a) One-step crowding cost function

(b) Two-step crowding cost function

(c) Step-wise crowding cost function

FIGURE 1: The graphical illustration of three non-linear crowding cost functions.

passengers' riding time savings. A new function form is thus called for to help differentiate the value of riding time among seating, standing in uncrowded conditions, and standing in crowded vehicles. The following two-step crowding cost function is actually borrowed from de Palma et al.'s work [6]. By slightly adjusting some parameters, the average value of time can be expressed as

$$P_v = P_{v0} + \frac{\rho P_{v0}}{1 + e^{\alpha(1-\theta)}} + \beta P_{v0} e^{\gamma(\theta-\delta)}. \quad (5)$$

Similarly, the first term on the right-hand side of (5) (P_{v0}) corresponds to the value of riding time in the base case where all passengers can find seats. The parameters α and γ can be given arbitrary values as they do not fundamentally affect the value of riding time. However, the parameters β and ρ should be carefully calibrated based on crowding penalties under different LFs. To illustrate the working mechanism, some explanations are made for three exclusive cases. If the load factor is less than one ($\theta < 1$), the denominator of the second term ($1 + e^{\alpha(1-\theta)}$) is scaled up to infinite value through any arbitrarily higher value of α, which in turn makes the second-term approaches zero. Meanwhile, since load factor is less than the threshold ($\theta < \delta$), the third term ($\beta P_{v0} e^{\gamma(\theta-\delta)}$) vanishes as it gets close to zero. Accordingly, the average value of riding time keeps the constant value—P_{v0}. As more passengers board, the limited seating capacity means that only a small proportion of them can find seats. The uncrowned threshold (δ) allows some passengers

Table 4: The standing density design guide for heavy rail.

Standing capacity	The crowding description	Japan	UK	Russia
0-1 pax/m^2	Fewer standing inside vehicle	Comfort	Comfort	Good
1-2 pax/m^2	Some standees	Comfort	Comfort	Good
2-3 pax/m^2	Standees can be free circulation	Comfort	Busy	Good
3-4 pax/m^2	0.25 m^2 of space among passengers Free circulation in the aisles	Little crowded (threshold)	Crowded	Normal
4-5 pax/m^2	0.20 m^2 of space among passengers Some restrictions in movement High probability of physical contact	Some crowded	Very crowded	Normal
5-6 pax/m^2	0.16 m^2 of space among passengers Very restricted circulation Frequent physical contact	Very crowded but tolerable	Maximal	Crowded
6-7 pax/m^2	Less 0.14 m^2 among passengers Impossible movement	Very crowded, crush loaded	Beyond capacity	Crowded
7-8 pax/m^2	Less 0.1 m^2 among passengers Physically squeeze into a car	Intolerable	Beyond capacity	unbearable

to travel by standing in uncrowded travelling conditions. Analogously, the scale parameter α causes the denominator of second-term $(1 + e^{\alpha(1-\theta)})$ not to appreciably differentiate from 1. Simultaneously, parameter γ scales down the third term $(\beta P_{v0} e^{\gamma(\theta-\delta)})$ to zero. Thus, in uncrowded condition, the impact of standing on the average value of time can be reflected by multiplying one mark-up to the value of time in the base case, which is $(1 + \rho)P_{v0}$. If load factor exceeds the threshold $(\theta > \delta)$, the vehicle gets more crowded and passengers have to stand in cramped conditions. In this case, the third term $(\beta P_{v0} e^{\gamma(\theta-\delta)})$ can no longer be ignored and it grows exponentially as each extra passenger boards the vehicle. The graphical representation of (5) is shown in Figure 1(b).

Formula 3: Stepwise Function for Crowding Costs. In contrast to bus-based modes, the physical capacity design guidelines for heavy rail usually adopt standing density to specify crowding levels instead of using load factor. In Table 4, eight levels of standing density are outlined for three representative countries. Furthermore, for each standing density, the travel conditions are described in terms of the available standing space and ability to move.

To this end, the average value of riding time can be modeled as a stepwise function. In this paper, the distinction of value of time is only made for six levels of standing density (i.e., from 0 pax/m^2 to 6 pax/m^2). The specific stepwise function for heavy rail is thus

$$P_v = P_{v0}\left(1 + \frac{\rho_1}{1 + e^{\gamma(\lambda_0 - N)}} + \frac{\rho_2}{1 + e^{\gamma(\lambda_1 - N)}} + \frac{\rho_3}{1 + e^{\gamma(\lambda_2 - N)}} + \frac{\rho_4}{1 + e^{\gamma(\lambda_3 - N)}} + \frac{\rho_5}{1 + e^{\gamma(\lambda_4 - N)}} + \frac{\rho_6}{1 + e^{\gamma(\lambda_5 - N)}}\right), \tag{6}$$

where λ_0 specifies the seating capacity. A new series of parameters λ_i $(i = 1, 2, \ldots, 5)$ denotes the number of passengers for the ith standing density level. The value of crowding penalty ρ_i $(i = 1, 2, \ldots, 6)$ is the incremental change between two consecutive levels. Through the scale-up parameter γ, the average value of time remains constant in each level but jumps discretely from one to another. The diagram of Figure 1(c) illustrates the shape of this function.

3.3. Total System Cost Minimization. In transit operation optimization, one of the approaches is the minimization of total system cost with respect to operating elements, such as fare, frequency, vehicle size, and routes. Since vehicle capacity and route density are assumed to be exogenously given, the only control variable here is frequency. As far as the modeling method is concerned, we adopt a very similar approach to Tirachini et al. [25].

As one of inputs supplied by users, waiting time cost makes up an appreciable part of total system costs. Since travelers can change their behaviors according to service type, we divide transit services into two types: frequency-based service and schedule-based service. For frequency-based transit services, passengers arrive at stations randomly so that the rule-of-thumb "*wait equals half headway*" is an approximation for waiting time. By contrast, for schedule-based services, passengers refer to the timetables before their departure. In this application, we assume that when frequency is greater than twelve vehicles/hour, the service can be viewed as a frequency-based type. When frequency is less than 12 vehicles/hour, the waiting behavior of passengers comprises two parts: passive waiting at other places and active waiting time at stations. It is interesting to note, since passive waiting time can be spent in a productive way, the value of time attached to passive waiting is less than the value of active waiting time at stations [24]. Thus, the general formula of average waiting time cost (w) for two service types is

$$w = P_w\left(t_w + \frac{2\mu}{f}\right), \tag{7}$$

where t_w is the fixed safety time that passengers spend waiting at stations. μ is the ratio of value of time for passive waiting to active waiting. P_w denotes the value of waiting time and f is frequency. Based on the above notations, total waiting time costs, denoted by C_{u-w}, are

$$C_{u-w} = P_w \left(t_0 + \frac{2\psi}{f} \right) Q, \tag{8}$$

where

$$t_0 = \begin{cases} 0, & \text{if } f \geq 12 \text{ veh/h}, \\ t_w, & \text{if } f < 12 \text{ veh/h}, \end{cases}$$
$$\psi = \begin{cases} 1, & \text{if } f \geq 12 \text{ veh/h}, \\ \mu, & \text{if } f < 12 \text{ veh/h}. \end{cases} \tag{9}$$

Another key component of costs—in-vehicle riding time costs—can be specified as the product of the average value of riding time (P_v) and the average riding time (T_v). If l is average trip length and L presents the route length, average riding time can be modeled as a fraction of the cycle time (t_c):

$$T_v = t_c \left(\frac{l}{2L} \right). \tag{10}$$

As passengers' value of riding time grows with crowding degree, the impact of crowding on riding time costs, can be dealt with by replacing P_v with the developed nonlinear crowding cost functions:

$$C_{u-v} = P_v t_c \left(\frac{l}{2L} \right) Q. \tag{11}$$

The so-called vehicle loaded (N) is the average number of passengers aboard, which can be calculated as the ratio of the total number of hourly passengers to the frequency:

$$N = \left(\frac{Q}{f} \right) * \left(\frac{l}{2L} \right). \tag{12}$$

Conventionally, the operating costs are divided into fixed and variable costs incurred in running services. Represented by C_{op}, the operating cost is

$$C_{\text{op}} = c_0 + c_1 B\eta + c_2 VB. \tag{13}$$

The first element is fixed costs per hour, c_o, which includes cost items that do not change with outputs. The second part (c_1) is unit cost per hour, which is determined by the fleet required in the peak period. The third one (c_2) is unit cost per vehicle kilometer. In practice, to avoid unexpected breakdowns, some vehicles remain unused at depots as back-ups. Consequently, a reserve rate of fleet, η, is introduced to reflect this. Denoting V by average running speed, the required fleet (B) can be formulated as

$$B = ft_c = f \left(\frac{2L}{V} \right). \tag{14}$$

Using (14) to eliminate B from (13), the operating cost function is

$$C_{\text{op}} = c_0 + c_1 \left(\frac{2L}{V} \right) \eta f + c_2 2Lf. \tag{15}$$

Ultimately, the objective is to minimize total system cost with respect to frequency (f):

$$\min_f \text{TSC} = C_{u-w} + C_{u-v} + C_{\text{op}}. \tag{16}$$

In practice, service frequency should be neither too low nor too high for operation and safety reasons. In inequity (17), $f_{\min}$ is the public desired minimum level of service and $f_{\max}$ is the maximum feasible frequency decided by the station capacity and safety considerations:

$$f_{\min} \leq f \leq f_{\max}. \tag{17}$$

Besides frequency constraints, the line capacity should be sufficient to accommodate demand:

$$\omega \left(\frac{Q}{2f} \right) \leq K, \tag{18}$$

where ω is the fraction of passengers traveling across the most loaded section and K denotes the physical capacity limit. Optimal frequency is sought by setting the first derivative of the total system cost function equal to zero and solving it subject to the above two constrains.

4. Numerical Experiments for Three Transit Modes

To illustrate the feasibility of incorporating the proposed nonlinear crowding cost functions into optimization, three cases are conducted for interurban bus, light rail, and metro. Two data sets are used. The first one is taken from Tirachini et al. [25]. The second one that has been used to calibrate crowding functions is taken from Whelan and Crockett's study. The values of parameters are summarized in Table 5.

To assess the relative merits of alternative crowding cost functions and the impact of crowding functions on system optimization, the following three scenarios are developed for each numerical case:

Scenario 1. Minimizing total system cost without considering the crowding effect.

Scenario 2. Minimizing total system cost with a classic linear crowding cost function.

Scenario 3. Minimizing total system cost with a nonlinear crowding cost function.

These simple numerical experiments could probably tell us at least three factors of general interests, which are also valid for more complex research. First, we want to know whether the proposed nonlinear functions are applicable for microeconomic modeling research. Second, we are interested

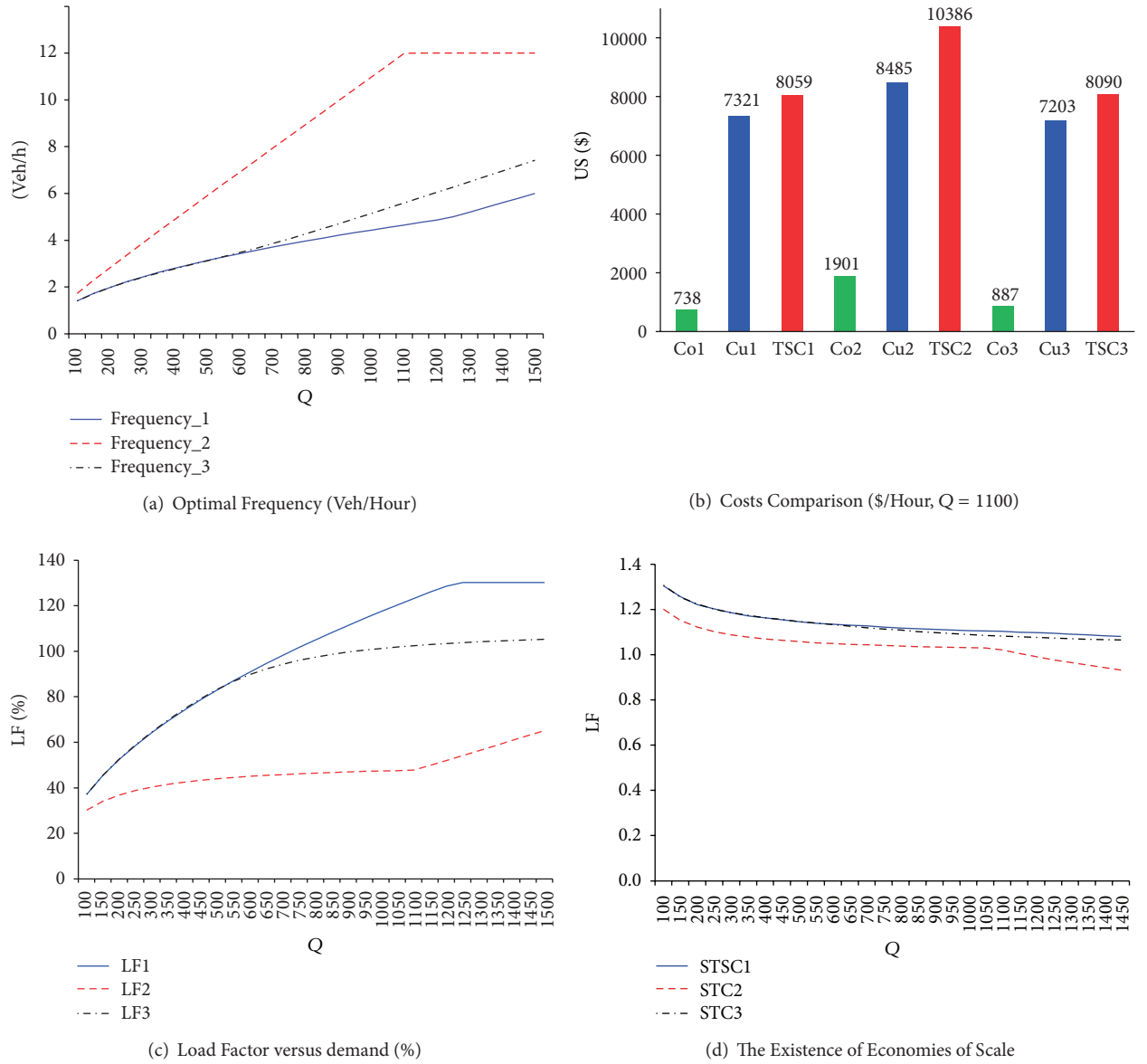
Table 5: Summaries of notations and parameters.

	Parameter	Description	Bus	Light rail	Heavy rail
Operating cost parameters	c_0	Fixed cost (\$/hour)	0	14,866	24,910
	c_1	Unit cost per vehicle-hour (\$/veh/h)	54	164	354
	c_2	Unit cost per vehicle-kilometer (\$/veh/km)	1.13	1.83	3.32
Mode operating parameters	V	Commercial speed (km/hour)	20	35	40
	t_c	Route travel trip time (hour)	2	1.14	1
	k	Seating capacity (seats/vehicle)	48	64	12
	K	Crush capacity (seats + standing/veh)	65	166 (5/m^2)	750 (6/m^2)
	$f_{\max}$	Maximum frequency allowed (veh/h)	200	80	40
Common parameters	P_v	Value of riding time (\$/hour)		9.45	
	P_w	Value of waiting time (\$/hour)		18.9	
	μ	Ratio of value of waiting time at home to at station		0. 33	
	t_w	Safety threshold time (min.)		4	
	L	Route length (km)		20	
	l	Average trip length (km)		10	
	η	Reserve rate for fleet		1.05	
	ω	Fraction of passengers across the loaded sections		30%	
	α	Scale parameter		5	
	γ	Scale parameter		20	

in finding out to what extent the optimal frequencies deviate among the three scenarios for different transit modes. In other words, compared with the linear crowding cost function, do the nonlinear crowding cost functions tend to overstate or understate optimal service frequency? Third, we want to know in which conditions the diseconomies of scale resulting from the "*Kraus effect*," offset the economies of scale arising from the "*Moring effect*," which finally leads to the diseconomies of scale on the consumption side.

4.1. Numerical Case One for Low Capacity Mode. In terms of low capacity transit modes, such as interurban buses and intercity rail, the design guidelines for internal standing areas are strict. For instance, comfortable loading for interurban buses, which allow standees on relatively short trips, should provide at least 0.45 m^2 for each standing passenger. Thus for these low capacity transit modes, the one-step function form for crowding costs is regarded as appropriate. Over a certain range of demand, we can examine the optimal outcomes for three scenarios.

First, embedding the linear crowding cost function into riding time costs generates the highest optimal frequencies across all demand levels. However, as depicted in Figure 2(a), there is no noticeable difference between Scenario 1 and Scenario 3 for low demand. However, the divergence is gradually evident when demand exceeds 700 pax/h. This divergent result can be explained by the fact that the optimal frequencies in Scenario 1 proportionally increase with the square root of demand. However, owning to the inclusion of the in-vehicle crowding externality, the optimal frequencies in Scenario 3 proportionally increase with demand rather than with the square root of demand for high patronage. Additionally, an in-depth comparison finds the linear function in Scenario 2 places more weight on riding time costs than the nonlinear function in Scenario 3 does. Thus, at any demand rate, the values of optimal frequencies in Scenario 2 are well above the optimal frequencies in Scenario 3. Second, Figure 2(b) depicts that when the demand rate is 1100 pax/h, Scenario 2 yields the highest operation costs (1901 \$/h) of the three scenarios, followed by Scenario 3 (887 \$/h) in second place. When the crowding cost is interpreted as a linear form, increased riding time cost dominate reduced waiting time cost, which finally yields the highest users costs in Scenario 2. By contrast, with a relatively smooth slope, the nonlinear function places less weight on riding time cost. Thus, in spite of its relatively higher frequency compared with Scenario 1, Scenario 3 provides lower consumer expenditures than Scenario 1 does. Third, in Figure 2(c), regarding optimal load factors, Scenario 2 intends to accommodate fewer passengers by providing more frequent services, resulting in the inefficient utilization of the vehicle facility. By contrast, by disregarding the crowding externality, Scenario 1 attempts to accommodate passengers as much as possible, which easily causes load factor to exceed 100%. As an intermediate case, Scenario 3 is likely to reduce the incidence of crowding with more efficient usage of transit vehicle. Finally, to assess the degree of scale economies, we introduce a new variable ζ, which is the ratio of the average costs to marginal costs. If $\zeta > 1$, we can confirm that there exist scale economies. The opposite case ($\zeta < 1$) denotes diseconomies of scale, while $\zeta = 1$ means neutral case [27]. From Figure 2(d), we can observe across all tested demand that Scenario 1 and Scenario 3 present the economies of scale. However, since the linear crowding cost curve is steeper than the nonlinear one, Scenario 2 puts more weight on the riding time cost. Thus, a considerable diseconomy of scale is present, particularly when ridership is high.



(a) Optimal Frequency (Veh/Hour)

(b) Costs Comparison ($/Hour, $Q = 1100$)

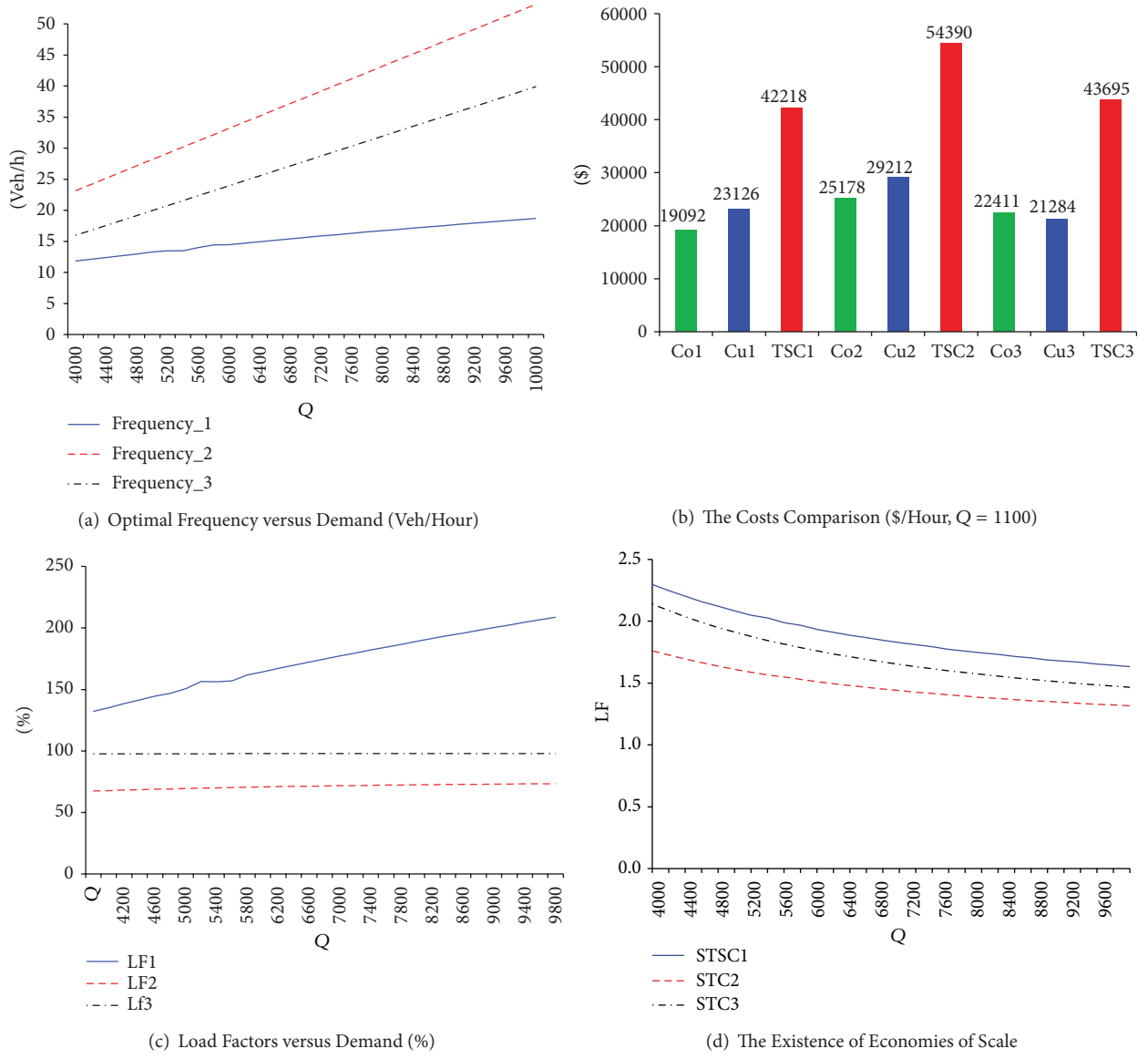
(c) Load Factor versus demand (%)

(d) The Existence of Economies of Scale

FIGURE 2: The Optimal Outputs for Numerical Case One (Here, C_o denotes the operating cost of the transit operator. C_u is the user cost and TSC means total system cost).

4.2. Numerical Case Two for High Capacity Modes. In the context of transit systems, most rail-based transit modes are designed to carry more standees than seated passengers so that passengers have high tolerance of in-vehicle crowding. Consequently, besides keeping the value of riding time for seating constant, it is reasonable to assume, when crowding level is below a certain threshold, the value of riding time also keeps constant but takes a higher value. Once crowding level is greater than threshold, the sharply increased discomfort makes the value of riding time present an exponential growth pattern. In this case, the threshold of uncrowded level is in line with the prevailing standards for physical vehicle capacity design (LF = 140%). The major results, in terms of optimal frequencies, cost components, and optimal LFs, are graphically represented in Figure 3.

Figure 3(a) shows the simulated optimal frequencies for three scenarios. Unlike Case 1, the frequency difference between Scenario 1 and Scenario 3 is apparent in this case. This divergence can be attributed to the fact that incorporating the two-step function into system optimization causes frequencies to vary proportionally with demand from the initial tested demand rate. As graphically illustrated by Figure 3(b), Scenario 2 implies the highest total system costs, user costs, and operating costs. However, the inclusion of the nonlinear function in total system cost minimization incurs slightly lower user costs than Scenario 1 does. This can be explained by the fact that although considering crowding effect generally increases the relative weight of riding time costs in the total cost function, the different function forms of crowding cost deliver different weights. A plot of the optimal



(a) Optimal Frequency versus Demand (Veh/Hour)

(b) The Costs Comparison ($/Hour, $Q = 1100$)

(c) Load Factors versus Demand (%)

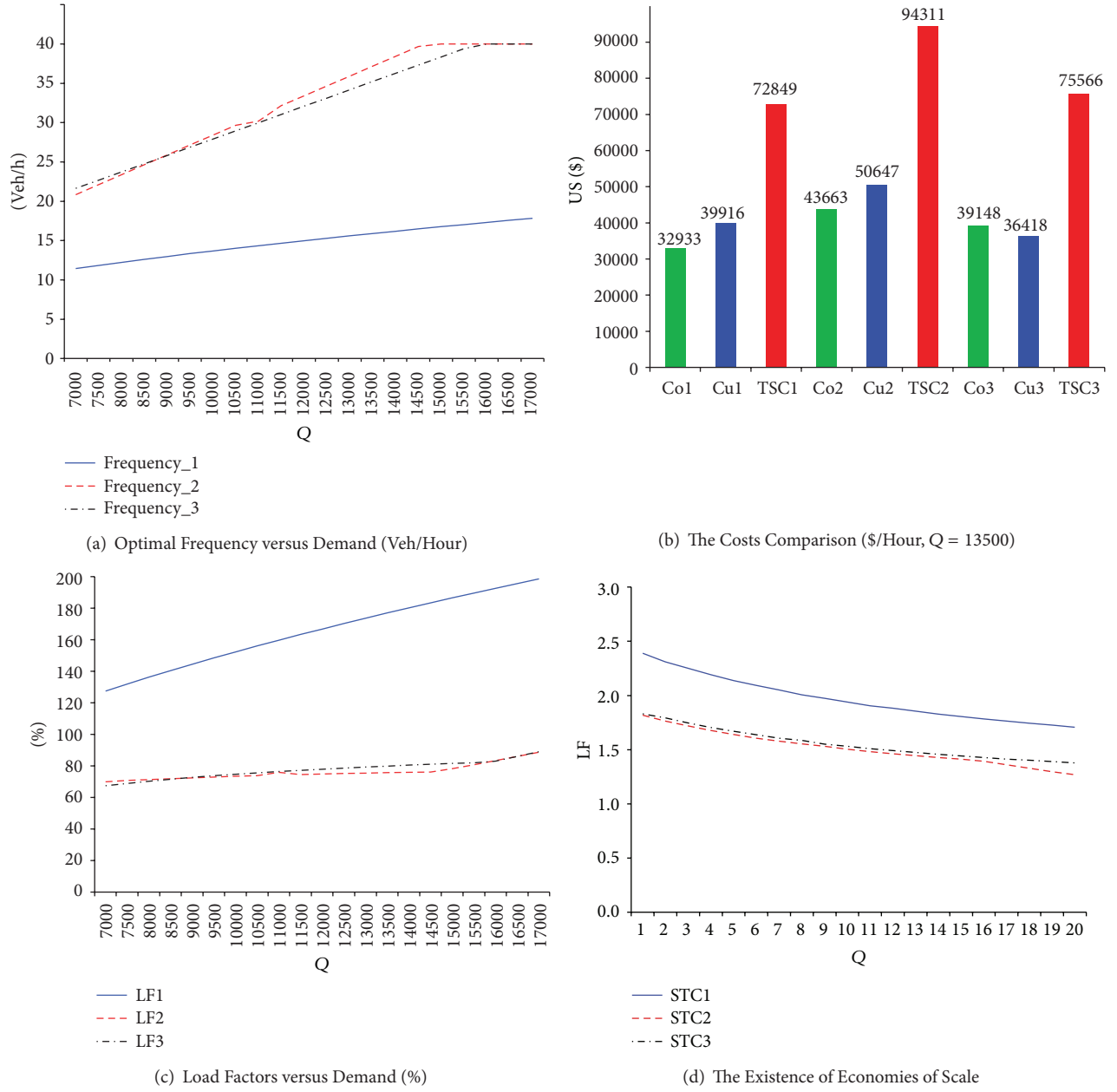
(d) The Existence of Economies of Scale

Figure 3: The Optimal Outputs for Numerical Case Two.

load factors versus demand is given in Figure 3(c). Due to the ignorance of crowding externality, Scenario 1 intends to fill vehicles with more passengers so that the optimal load factors range from 130% to 200%. In such a case, passengers consider discomfort to be impaired by high load factors. On the contrary, Scenario 2 has an incentive to run more frequent services with many empty seats, suggesting that the load factor is below 100%. The optimal load factors in Scenario 3 are around 100%. Thus, from a cost efficiency point of view, the most promising scenario would be Scenario 3. Concerning the degree of scale economies, Figure 3(d) depicts that all three scenarios involve pronounced scale economies. Furthermore, a closer look at Scenario 2 and Scenario 3 shows that no matter which type of crowding cost function is implemented the diseconomies of scale in users' riding costs resulting from introducing the crowding externality are dominated by the coexistence of economies in producer cost and waiting time cost, which finally generates the magnitude of scale economies in total system cost.

4.3. Numerical Case Three for Massive Capacity Modes. The latest version of PDFH 5 [17] provides six crowding penalties for a certain range of standing densities and recommends adopting these values for metro planning. Following the suggestion of PDFH, a stepwise function was developed for massive capacity transit modes, particularly for heavy rail. In this case, the stepwise function was examined for its applicability. The optimal solutions of three scenarios yield some insights.

First, due to disregarding the crowding effect, Scenario 1 intends to provide less frequent services along the entire range of demand, as shown in Figure 4(a). However, a comparison between Scenario 2 and Scenario 3 shows that the frequency difference is relatively small before the demand



(a) Optimal Frequency versus Demand (Veh/Hour)

(b) The Costs Comparison ($/Hour, $Q = 13500$)

(c) Load Factors versus Demand (%)

(d) The Existence of Economies of Scale

Figure 4: The Optimal Outputs for Numerical Case Three.

exceeds 11,500 pax/h. Beyond that point, Scenario 2 starts to adjust the frequency up in order to reduce the increasing crowding, since the slope of a linear function is steeper than the slope of stepwise function. Second, to show the effectiveness of vehicle utilization, we plot the load factor curves in Figure 4(c). In Scenario 1, the optimal load factors increase from 130% at 7,000 pax/h to 200% at 17,000 pax/h. On the contrary, in Scenario 2 and Scenario 3, the loaded passengers never exceed the vehicle capacity, implying that enough capacity is afforded to help relieve overcrowding. Finally, Figure 4(d) clearly depicts that Scenario 1 yields the scale economies, resulting from the combined scale economies on the consumption and production sides. By contrast, at high demand rates, Scenario 2 and Scenario 3 are prone to be diseconomies of scales. The likely explanation is that the diseconomy of scale from "*Kraus effect*" dominates the economy of scale from "*Mohring effect*," which finally leads to the diseconomy of scale on consumption side.

5. Summary

Recently, means of public transit often ply in extremely overcrowded conditions in large metropolitan cities. Although crowding can be mitigated through costly infrastructure improvement and network expansion, it can also be avoided through low-cost crowding relief measures if crowding cost can be correctly formulated. To provide insights into the in-vehicle crowding effect, this paper has presented three

nonlinear crowding cost functions and discussed the implementation in the system optimization.

A synthesis of the theoretical and numerical analysis generates valuable insights. First, with nice mathematical properties, the proposed nonlinear functions can be successfully implemented in system optimization. From a methodology point of view, the advantage of the nonlinear crowding cost function over the simple linear one is that nonlinear formulas provide a more realistic and rational representation. Second, numerical case studies show that, irrespective of the crowding costs function form (linear or nonlinear), the inclusion of crowding effect has complex impacts on the level of service provision, the cost of the system, and the degree of scale economies. Specifically, incorporating crowding costs functions into the system optimization usually generates frequent transit services, low users costs, and more efficient vehicle usage. With high demand rates, the diseconomies of scale from the *Kraus effect* may dominate the economies of scale from the *Mohring effect*, which probably leads to the presence of scale diseconomies in total system cost. Thus, we should treat the *Kraus effect* as importantly as the *Mohring effect*.

In most transit assignment applicants, in-vehicle crowding is not taken into account for modeling the mode/route choices. The prevailing software, such as EMME/2 and Tube, usually uses the fixed in-vehicle travel costs. In order to enable the transit assignment module to reflect crowding costs, all original fixed travel cost functions should be multiplied by the item $*(1 + \rho)$. In terms of the proposed crowding cost function, the equilibrium assignment function can be separated in a linear travel time cost part and a nonlinear crowding part. During the assignment steps, the nonlinear crowding part can be iteratively fed back to the generalized cost for each route/mode until the whole network reaches equilibrium.

The contributions of this paper to the literature are twofold. Firstly, after observing the disadvantages of conventional linear crowding cost functions, this paper proposed three nonlinear function forms for different transit modes to help decision-makers better examine the impacts of in-vehicle crowding on transit service planning. Secondly, for the first time, two separate strands of research work closely related to the in-vehicle crowding effect (i.e., the empirical estimation work on crowding penalties and theoretical in-vehicle crowding cost modeling) are integrated to fill the gap between theory and practice. It is hoped that the nonlinear crowding cost functions proposed here will help planners make better use of existing line capacities in order to relieve in-vehicle crowding before costly capital works are required.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The author is very grateful to two anonymous referees whose comments have led to a significant improvement of the paper. Financial support from the National Natural Science Foundation of China (Grant no. 71301112) is gratefully acknowledged.

References

[1] R. Arnott and K. Small, "The economics of traffic congestion," *American Scientist*, vol. 82, no. 5, pp. 446–455, 1994.

[2] O. Johansson, "Optimal road-pricing: simultaneous treatment of time losses, increased fuel consumption, and emissions," *Transportation Research D*, vol. 2, no. 2, pp. 77–87, 1997.

[3] H. Mohring, "Optimization and scale economies in urban bus transportation," *American Economic Review*, vol. 62, pp. 591–604, 1972.

[4] R. Turvey and H. Mohring, "Optimal bus fares," *Journal of Transport Economics and Policy*, vol. 9, pp. 280–286, 1975.

[5] Q. Tian, H.-J. Huang, and H. Yang, "Equilibrium properties of the morning peak-period commuting in a many-to-one mass transit system," *Transportation Research B*, vol. 41, no. 6, pp. 616–631, 2007.

[6] A. de Palma, M. Kilani, and S. Proost, "The location and pricing of mass transit stations," in *Proceedings of the 12th World Conference on Transport Research Society*, pp. 978–989, 2010.

[7] M. L. Fried and V. J. DeFazio, "Territoriality and boundary conflicts in the subway," *Psychiatry*, vol. 37, no. 1, pp. 47–59, 1974.

[8] E. S. Knowles, "The gravity model of crowding: application to social physics to the effects of others," in *Human Responses to Crowding*, A. Baum and Y. Epstein, Eds., Lawrence Erlbaum Associates, New York, NY, USA, 1978.

[9] D. Stokols, "On the distinction between density and crowding: some implications for future research," *Psychological Review*, vol. 79, no. 3, pp. 275–277, 1972.

[10] R. Baron and D. Richardson, *Human Aggression Edition*, Plenum, New York, NY, USA, 2nd edition, 2004.

[11] G. Whelan and J. Crockett, "An investigation of the willingness to pay to reduce rail overcrowding," in *Proceedings of the International Conference on Choice Modeling*, Harrogate, UK, 2009.

[12] Sinclair Knight Merz (SKM), "Critical Review of Transport Modelling Tools," 2011, http://www.bitre.gov.au/publications/34/Files.

[13] Accent Marketing and Research and Hague Consulting Group, "Rail Differential Pricing," Report Prepared for Rail Operational Research, 1996.

[14] W. H. K. Lam, C.-Y. Cheung, and C. F. Lam, "A study of crowding effects at the Hong Kong light rail transit stations," *Transportation Research A*, vol. 33, no. 5, pp. 401–415, 1999.

[15] Accent, "Rail Customer Valuations of Sitting, Standing and Crowding," Report Prepared for Transport for London, 2006.

[16] F. Maunsell and M. MacDonald, "Rail Overcrowding, Reliability and Frequency," Report Prepared for Demand Impacts Project for Centro, 2007.

[17] Association of Train Operating Companies (ATOC), "Passenger Demand Forecasting Handbook (Version 5)," Report prepared for British Railway Board, 2009.

[18] P. B. Goodwin, "Human effort and the value of travel time," *Journal of Transport Economics and Policy*, vol. 10, no. 1, pp. 3–15, 1976.

[19] A. Polydoropoulou and M. Ben-Akiva, "Combined revealed and stated preference nested logit access and mode choice model

for multiple mass transit technologies," *Transportation Research Record*, no. 1771, pp. 38–45, 2001.

[20] Z. Li and D. A. Hensher, "Crowding and public transport: a review of willingness to pay evidence and its relevance in project appraisal," *Transport Policy*, vol. 18, no. 6, pp. 880–887, 2011.

[21] M. Wardman and G. Whelan, "Twenty years of rail crowding valuation studies: evidence and lessons from British experience," *Transport Reviews*, vol. 31, no. 3, pp. 379–398, 2011.

[22] M. Kraus, "Discomfort externalities and marginal cost transit fares," *Journal of Urban Economics*, vol. 29, no. 2, pp. 249–259, 1991.

[23] K. Jansson, "Optimal public transport price and service frequency," *Journal of Transport Economics and Policy*, vol. 27, pp. 33–50, 1993.

[24] S. R. Jara-Diaz and A. Gschwender, "Towards a general microeconomic model for the operation of public transport," *Transport Reviews*, vol. 23, no. 4, pp. 453–469, 2003.

[25] A. Tirachini, D. A. Hensher, and S. R. Jara-Díaz, "Comparing operator and users costs of light rail, heavy rail and bus rapid transit over a radial public transport network," *Research in Transportation Economics*, vol. 29, no. 1, pp. 231–242, 2010.

[26] Australian Transport Council (ATC), *National Guidelines for Transport System Management in Australia, Canberra*, 2006.

[27] K. A. Small and E. T. Verhoef, *The Economics of Urban Transportation*, Routledge, London, UK, 2007.

Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 714267, 7 pages
http://dx.doi.org/10.1155/2014/714267

Research Article

Potential Field Cellular Automata Model for Pedestrian Evacuation in a Domain with a Ramp

Xiao-Xia Jian and Xiaoning Zhang

School of Economics and Management, Tongji University, Shanghai 200092, China

Correspondence should be addressed to Xiaoning Zhang; cexzhang@tongji.edu.cn

Received 14 November 2013; Accepted 23 December 2013; Published 11 February 2014

Academic Editor: Hu Shao

We propose a potential field cellular automata model with a pushing force field to simulate the pedestrian evacuation in a domain with a ramp. We construct a cost potential depending on the ramp angle and introduce a function to evaluate the pushing force, which is related to the cost and the desired direction of pedestrian. With increase of crowd density, there is no empty space for pedestrian moving forward; pedestrian will purposefully push another pedestrian on her or his desired location to arrive the destination quickly. We analyse the relationship between the slope of ramp and the pushing force and investigate the changing of injured situations with the changing of the slope of ramp. When the number of pedestrians and the ramp angle arrive at certain critical points, the Domino effect will be simulated by this proposed model.

1. Introduction

Crowd dynamics is the universal phenomenon and also the source of the catastrophe, for example, crowd disaster at the 2010 Love Parade electronic dance music festival in Duisburg, stampede ensued during Beckham's visit to Tongji University in Shanghai on June 20, 2013, and stampedes during large-scale sport and entertainment activities. Studying reasons of the catastrophe is quite critical to reduce the occurrence of disasters in the crowd. A plenty of research works, using video analysis [1–3], field study [1, 4], and modelling [5–8], focus on investigating the macroscopic crowd dynamics and microscopic variable characteristics exhibited by different individual pedestrians, analysing pedestrian dynamics in various scenarios. One of critical behavioral reactions for pedestrians is the pushing force among crowd under usual situation, which cannot be ignored in exploring the causes of crowd disaster.

As for our knowledge, the usual situation is that each pedestrian who prepares to attend the activities in stadiums or concerts is fully aware that he or she will experience high density crowd before entering the field; hence, his or her motion is stabile and the physical contact is the main factor to produce the pushing force. When people are condensed to some critical density, marginal physical forces of each individual add up to a lethal pressure and this is even true without the panic which inevitably bursts in such occasions. In particular, if the walking domain is even, for instance, a passageway with a ramp, the pushing force is affected by this external condition. In this case, pedestrians hold rational cogitations, so they desire to move with minimal travel cost or travel time, as described in [8] and references therein. In the potential field CA model, the cost potential [8] is the navigation of pedestrian movement, and the pedestrian moves along the direction of minimizing the cost, but this model is inadequate to simulate the pushing force among pedestrians.

A plenty of works focus on describing the avoidance behavior between bidirectional pedestrian flows [5, 7–13], others consider the friction [11, 14, 15] and repulsion [11, 15]. However, the contributions on investigating the pushing force [11, 16, 17] reproduced by physical contacts among pedestrians are scarce. Henein and White [16] introduce a force vector field to simulate the pushing force; Song et al. [17] investigate

the number of overlapping grids occupied by pedestrians to measure the pushing force. All these works did not discuss the pedestrians evacuation in a domain with the ramp.

In this paper, we describe the pedestrian evacuation in a domain with a ramp. This scenario can be used to simulate the domain around the entrance of stadiums or concert halls. Based on the basic rules of potential field CA model, we construct a new cost potential depending on the ramp angle and introduce a function to evaluate the pushing force, which is related to the cost and the desired direction of pedestrian; in other word, with increase of crowd density, there is no empty space for pedestrian moving forward; pedestrian will purposefully push another pedestrian on her or his desired location to reach the destination quickly; the magnitude of push force is determined by her or his potential value of current location. The former who is close to the exit produces small force and bears maximal pushing forces and the latter produces large force, and bear minimal pushing forces; when the pushing force is beyond a certain critical value, the pedestrian will fall onto the ground. What is more, the slope of a ramp is an important factor to affect pedestrian evacuation dynamics. When the slope of a ramp is deep, the critical value of pushing force is small and much more pedestrians could be injured; when the slope of ramp is small, the critical value of pushing force is large and fewer pedestrians may experience injury; the value of critical value is maximal when the angle of ramp is zero. When the number of pedestrians and the ramp angle are beyond certain critical points, the Domino effect will be simulated by this proposed model.

The remainder of this paper is organized as follows. In Section 2, we formulate a potential field cellular automata model with a pushing force field. In Section 3, the process of pedestrian evacuation in a domain with a ramp is simulated and the injured phenomenon that resulted from the pushing influences is reproduced to verify the practicality of this model; What is more, we analyse the sources of pushing disasters; one is the pushing force produced by local pedestrian group, which is dependent on the potential field; another is the deep ramp. Section 4 concludes the paper.

2. Potential Field CA Model with Pushing Force in Ramp

In this paper, we investigate the pedestrian evacuation in the walking domain with a ramp, where the walking domain is represented by $\widehat{\Omega}$, and the scenario is shown in Figure 1. We simulate pedestrians going across subdomain C, ramp B, and subdomain A and then leaving the walking domain from the exit at left side of subdomain A. Initially, pedestrians are assigned randomly to subdomains B and C with the initial density ρ_0.

The walking domain $\widehat{\Omega}$ is divided into cells with the space step h along x direction (y direction); each cell is empty or occupied by at most one pedestrian. After determining his or her movement direction, each pedestrian moves into his desired direction. The current position and possible movement direction of each pedestrian are shown in Figure 2(a).

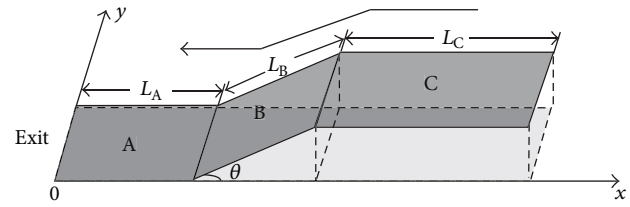


Figure 1: Illustration of the walking domain with a ramp. The walking domain $\widehat{\Omega}$ (gray) is constructed by three subdomains A, B, and C; level subdomains A and C are connected by ramp B. Here, L_A, L_B, and L_C are the lengths of subdomains, respectively; the black arrow indicates the direction of motion.

Here we assume that $h = 0.4$ m which produces an average area occupied by one pedestrian [10]; therefore, the dimensionless maximal density $\rho_{\max} = 1$ corresponds to actual density 6.25 ped/m^2. Besides, as the average density of real pedestrian is 1 m/s, we assume that the maximum speed of each pedestrian in our model is $v_{\max} = 1$ m/s and it implies that the time step is 0.4 s. The crucial step of the CA model is to determine the transition probabilities of Figure 2(b). In this proposed model, the transition probabilities are based on a potential field and a pushing force field.

2.1. Potential Field in a Domain with a Ramp. The potential field is a navigation of the pedestrian walking; here, the walking domain $\widehat{\Omega}$ is not on a two-dimensional plane. In order to compute potential field in the uneven walking domain $\widehat{\Omega}$, firstly, using (x, y) to represent the coordinate of the cell, we compute the virtual potential $\phi(x, y, t)$ on two-dimensional level domain Ω with the exit Γ_0 at each time t. Ω is set by leveling down $\widehat{\Omega}$; secondly, we reconstruct the potential $\phi(x, y, t)$ by introducing a ramp coefficient $\omega(x, y, \theta)$. The mathematical formulations of calculating the potential field $\widehat{\phi}(x, y, t, \theta)$ in $\widehat{\Omega}$ are as follows.

(a) The virtual potential $\phi(x, y, t)$ is determined by solving the following Eikonal equation:

$$\begin{aligned} \left\|\nabla\phi\left(x, y, t\right)\right\| &= \tau\left(x, y, t\right), \quad (x, y) \in \Omega, \\ \phi\left(x_0, y_0, t\right) &= 0, \quad (x_0, y_0) \in \Gamma_0, \end{aligned} \tag{1}$$

where $\|\nabla\phi(x, y, t)\| = \left(\phi_x(x, y, t)^2 + \phi_y(x, y, t)^2\right)^{1/2}$ and $\tau(x, y, t)$ is a cost distribution function which satisfies that $\tau(0) = 1$, $\tau'(\rho) > 0$, $0 \le \rho \le 1$ [8, 12, 13]. The existence and uniqueness of the solution of problem (1) are discussed by [8], and this problem can be solved by using the fast sweeping method [18].

(b) The potential $\widehat{\phi}(x, y, t, \theta)$ is reconstructed by $\phi(x, y, t)$ multiplying a ramp coefficient $\omega(x, y, \theta)$; that is,

$$\widehat{\phi}\left(x, y, t, \theta\right) = \omega\left(x, y, \theta\right)\phi\left(x, y, t\right), \tag{2}$$

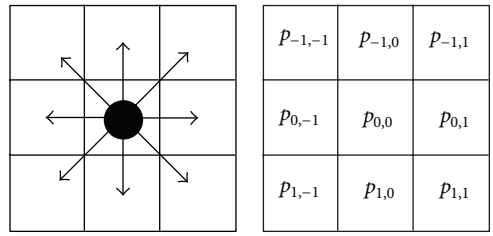


Figure 2: (a) An occupied cell and its eight neighboring cells, corresponding to (b) the nine probabilities for the pedestrian in the occupied cell to update his or her position.

where

$$\omega(x,y,\theta) = \begin{cases} 1, & x \in [0, L_A], \\ \exp\left(\alpha \dfrac{x - L_A}{h} \tan\theta\right), & x \in (L_A, L_A + L_B\cos\theta), \\ \exp\left(\alpha \dfrac{L_B}{h} \sin\theta\right), & x \in [L_A + L_B\cos\theta, L_A + L_B\cos\theta + L_C]. \end{cases} \quad (3)$$

α is a parameter that reflects the sensitivity of potential to the slope of ramp.

Obviously, we magnify the potential $\phi(x, y, t)$ by the coefficient $\omega(x, y, \theta) \geq 1$, which suggests a stronger effect with a deeper ramp angle; in other words, $\omega(x, y, \theta)$ is increasing with the ramp angle $\theta \in [0, 90^\circ)$. Here, after giving an angle θ of the ramp, the weight function $\omega(x, y, \theta)$ is given; although subdomains A and C are all flat, the height difference between two flat subdomains A and C is different with different θ; thus, the values of weight $\omega(x, y, \theta)$ in subdomain A are different than the ones in subdomain C, which implies that the potential values in subdomain C are influenced by the ramp angle θ. In addition, the final potential $\widehat{\phi}(x, y, t, \theta)$ is consistent with $\phi(x, y, t)$ if $\theta = 0$.

2.2. Aggregated Pushing Force Field. In the crowd, each pedestrian is surrounded by other pedestrians, hence, physical contacts among pedestrians occur, and pedestrians produce the pushing forces to desire for increasing the personal free-space and being close to the destinations. In the proposed model, we introduce the aggregated pushing force field to simulate the physical force effects.

While the density $\rho(x, y)$ is equal to the maximal density 1 (6.25 ped/m^2), which means that there is no empty neighbour cell for the pedestrian in (x, y) to arrive, the magnitude of the pushing force produced by the pedestrian in (x, y) at time t is

$$f(x,y,t) = \left(1 - \frac{\widehat{\phi}(x,y,t,\theta)}{\max_{(\widehat{x},\widehat{y})\in\widehat{\Omega}} \widehat{\phi}(\widehat{x},\widehat{y},t,\theta)}\right)^4. \quad (4)$$

This functional $f(x, y, t)$ reflects the sensitivity of pushing force to potential field, which is not a real pushing force. The functional $f(x, y, t)$ implies that the pushing force is decreasing with potential $\widehat{\phi}(x, y, t, \theta)$. In practice, while the pedestrian is near to the exit, his or her potential is much smaller than the maximal potential of all the domain and he or she holds stronger propulsion; on the contrary, the distance between the pedestrian's location and the exit is large; he or she is patient with small pushing force. The pedestrian considers that the pushing is helpful in arriving at the destination quickly; thus, he or she pushes along the direction of minimizing the cost potential $\widehat{\phi}(x, y, t, \theta)$; that is, the direction of $\vec{f}(x, y, t)$ is $\vec{l}(x, y, t) = (l_1, l_2) = -(\widehat{\phi}_x, \widehat{\phi}_y)/\|\nabla\widehat{\phi}\|$.

In fact, the pedestrian's pushing force is repeatedly retransmitted from person to person through interpersonal contacts within crowds [19]. In this proposed model, we model this process by adding up the pushing force in each time step. The incremental pushing force $\vec{f}^{\text{add}}$ in each cell (x, y) at each time step is produced by whether this pedestrian in (x, y) is directed by another pedestrian's desired direction. Here, using $\vec{f}^{\text{add}}(x, y, t)$ to represent the pushing force experienced by pedestrian in (x, y), if there exists a pedestrian in (x, y)'s neighbour $(\overline{x}, \overline{y})$ whose desired walking direction $(l_1, l_2) = (x - \overline{x}, y - \overline{y})$, pushing force $\vec{f}(\overline{x}, \overline{y}, t)$ is produced by pedestrian in $(\overline{x}, \overline{y})$ and experienced by the pedestrian in (x, y), where the magnitude $f^{\text{add}}(x, y, t)$ is equal to $f(\overline{x}, \overline{y}, t)$ and the direction is also along (l_1, l_2); otherwise, $\vec{f}^{\text{add}} = 0$.

Therefore, the aggregated pushing force (vector-wise) experienced by the pedestrian in the cell (x, y) is the vector sum of forces generated by the aggregated force from the beginning to time step $t-1$ and the incremental pushing force $\vec{f}^{\text{add}}(x, y, t)$. Obviously, $f^{\text{add}}(x, y, t) \geq 0$. Using function (4), we define the aggregated force vector field $\vec{F}(x, y, t)$ experienced by each pedestrian in (x, y) as

$$\begin{aligned} \vec{F}(x,y,t) &= \vec{F}(x,y,t-1) + f^{\text{add}}(\overline{x},\overline{y},t)\cdot\vec{l} \\ &= \left(F^1(x,y,t), F^2(x,y,t)\right), \end{aligned} \quad (5)$$

where $F^1(x, y, t)$ and $F^2(x, y, t)$ are two components of $\vec{F}(x, y, t)$ along axes x and y, respectively. For convenience, $\|\vec{F}(x, y, t)\| = ((F^1(x, y, t))^2 + (F^2(x, y, t))^2)^{1/2}$ represents the magnitude of $\vec{F}(x, y, t)$.

If $\|\vec{F}(x,y,t)\|$ exceeds a threshold, the pedestrian will become injured, and the injured pedestrians could not move again. Here, the threshold parameter, represented by $F^*(x,y,t)$, is the increasing function of pushing force $f(x,y,t)$ and is dependent on the angle of ramp. The mathematical formulation of $F^*(x,y,t)$ is defined as

$$F^*(x,y,t) = \begin{cases} \beta \cdot \dfrac{f(x,y,t)}{\cos\theta}, \\ \quad x \in (L_A, L_A + L_B\cos\theta), \\ \beta \cdot f(x,y,t), \\ \quad x \in [0, L_A] \cup [L_A + L_B\cos\theta, \\ \qquad\qquad L_A + L_B\cos\theta + L_C], \end{cases} \tag{6}$$

where β is a sensibility parameter and the formulation of β is defined as

$$\beta = \begin{cases} 10, & \theta = 0, \\ \csc\theta, & \theta \in (0, 90^\circ). \end{cases} \tag{7}$$

Here, β is a decreasing function related to the angle of ramp. While θ is zero, β is a large value and it implies that the pedestrian in this situation is not easily pushed onto fall the ground; with increase of θ, β decreases and it implies that the pedestrian is gradually easily pushed down. Obviously, another choice of critical threshold can be used to deeply investigate the different pushing processes of crowd dynamics.

2.3. Update Rules. The cost potential field $\widehat{\phi}(x,y,t,\theta)$ is calculated using (1) and (2) with current density distribution $\rho(x,y,t)$. The force field $\vec{F}(x,y,t)$ is calculated by formula (5). In this section, the transition probabilities and the injured probabilities are bounded by the comparisons between the $\|\vec{F}(x,y,t)\|$ and $F^*(x,y,t)$. Assign $(0,0)$ to cell occupied by uninjured pedestrian and then the update rules from time step t to $t+1$ are as follows.

(1) If $\|F(x,y,t)\| \le F^*(x,y,t)$, the pedestrian moves by the transition probabilities $p_{i,j}$ shown in Figure 2(b).

(1.1) $p_{i,j}$ is determined by

$$p_{ij} = \begin{cases} \dfrac{1}{|S_m|}, & (i,j) \in S_m, \\ 0, & (i,j) \notin S_m, \end{cases} \tag{8}$$

if $S_m \neq \emptyset$, $\delta\widehat{\phi}_{\min}(0) < 0$;

otherwise

$$p_{ij} = \begin{cases} 1, & (i,j) = (0,0), \\ 0, & (i,j) \neq (0,0), \end{cases} \tag{9}$$

where $\delta\widehat{\phi}_{ij}(0) \equiv (\widehat{\phi}(i,j) - \widehat{\phi}(0,0))/\iota_{ij}$ for $(i,j) \in S_0$ is the difference quotient, with $S_0 = \{(i,j) \mid (i,j)$ is empty$\}$, and ι_{ij} is the distance between the cells (i,j) and $(0,0)$. Besides, the set $S_m = \{(i,j) \mid \delta\widehat{\phi}_{ij}(0) = \delta\widehat{\phi}_{\min}(0)\} \subseteq S_0$, where $\delta\widehat{\phi}_{\min}(0) = \min_{(i,j)\in S_0}\delta\widehat{\phi}_{ij}(0)$, and $|S_m|$ is the number of elements in S_m.

(1.2) Under the parallel update, conflicts that resulted from the fact that an empty cell (i,j) could be a target cell of m $(0 \le m \le 8)$ pedestrians in the m neighboring cells $(0^{(k)}, 0^{(k)})$ if $m \ge 2$ can be resolved by the following probabilities:

$$p\left(0^{(k)}\right) = \begin{cases} \dfrac{1}{|S_{mm}|}, \\ \quad \text{if } m \ge 1, \ \left(0^{(k)}, 0^{(k)}\right) \in S_{mm}, \\ 0, \\ \quad \text{otherwise}, \end{cases} \tag{10}$$

where

$$S_{mm} = \left\{\left(0^{(k)}, 0^{(k)}\right) \mid \delta\widehat{\phi}_{\min}\left(0^{(k)}\right) = \delta\widehat{\phi}_{\min}, \ 1 \le k \le m\right\},$$
$$\delta\widehat{\phi}_{\min} = \min_{1\le k\le m} \Delta\widehat{\phi}_{\min}\left(0^{(k)}\right). \tag{11}$$

(2) If $\|F(x,y,t)\| > F^*(x,y,t)$, the pedestrian will be injured by the probabilities $\overline{p}_{i,j}$.

(2.1) The injured probabilities are defined by

$$\overline{p}_{i,j} = r^{\text{injured}} p_{i,j}, \tag{12}$$

where $r^{\text{injured}} \in [0, 0.5]$ is a random parameter. In this case, the pedestrian in $(0,0)$ is injured and falls in the location (i,j) with the probabilities $\overline{p}_{i,j}$; when $(i,j) \neq (0,0)$, this pedestrian will fall in an empty cell along the direction of his or her aggregated pushing force; otherwise, this pedestrian will fall in the current location.

(2.2) In this case, to avoid the situation that different neighbor pedestrians fall into the same empty cell simultaneously, we set equal falling-in possibility to different neighbor pedestrians; namely,

$$p\left(0^{(k)}\right) = \begin{cases} \dfrac{1}{m}, & \text{if } m \ge 1, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

One of injured pedestrians falls in the desired cell, and others fall in the current locations.

3. Simulation

In this simulation, the sizes of walking domain are 5×3 cells for subdomain A, 5×4 cells for ramp B, and 5×20 for subdomain C. The cost distribution $\tau(x,y,t)$ is given by

$$\tau(x,y,t) = 1 + g_0\rho^\gamma(x,y,t), \tag{14}$$

where we set $g_0 = 0.075$ and $\gamma = 2$.

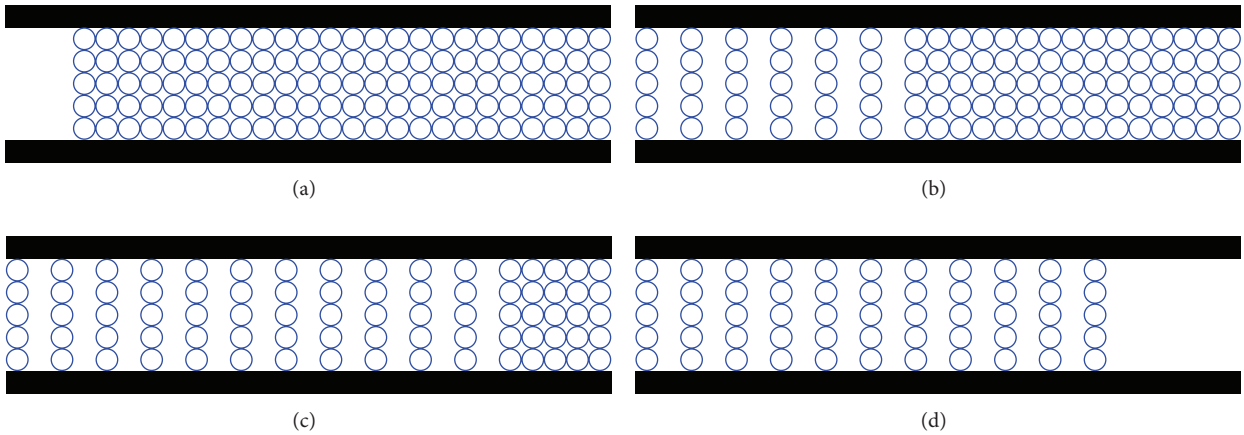


FIGURE 3: Evacuation process simulated by the proposed model, at $t = 0$, 10, 20, and 30 (color online). The initial density $\rho_0 = 1$ and the angle of ramp $\theta = 0$. The open circles represent pedestrians.

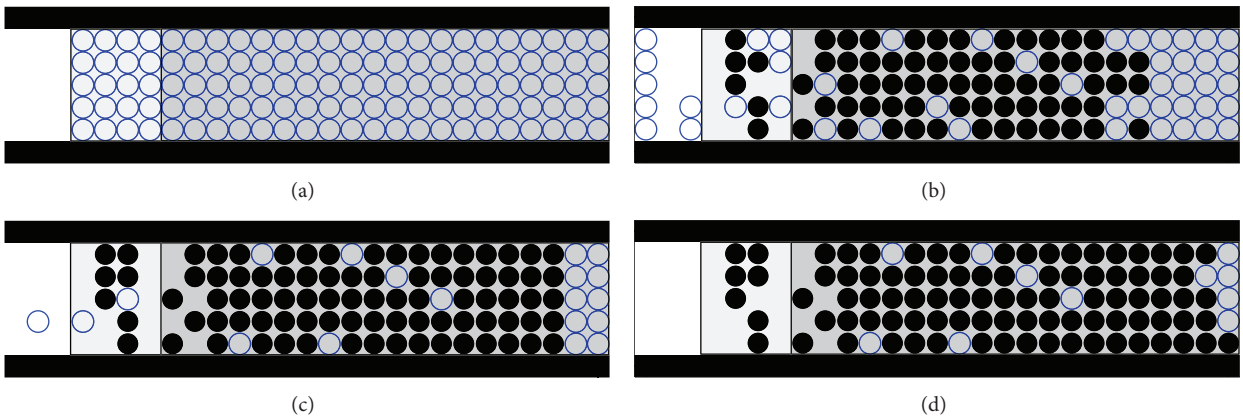


FIGURE 4: Evacuation process simulated by the proposed model, at $t = 0$, 6, 15, and 30 (color online). The initial density $\rho_0 = 1$ and the angle of ramp $\theta = 50^\circ$. Here, the light gray domain represents the projection of the ramp (subdomain B), and the gray is the subdomain C; the normal pedestrians and the injured pedestrians are indicated by open circles and the full circles, respectively.

Initially, N pedestrians ($N < 5 \times 4 + 5 \times 20$) are randomly assigned to the inner cell in subdomains B and C with the initial mean density $\rho_0 = N/(5\times4+5\times20)$. In the simulations, we simulate the injured phenomena and investigate relationships between the angle of the ramp and the crowd falling onto the ground under different initial densities ρ_0. Figure 3 shows the evacuation process simulated by the proposed model. In this simulation, the walking domain is without the ramp. Figures 3(a)–3(d) are the snapshots of an evacuation simulation at $t = 0$, 10, 20, and 30, respectively. This is in accordance with the underlying principle that the path-choice strategy in the propose model is somewhat optimal due to the pedestrians' awareness of the destination and the surroundings.

The other simulation is to investigate the evacuation dynamics in the walking domain with the ramp. Figure 4 shows the overlooked evacuation process in this situation here, angle of the ramp is $\theta = 50^\circ$. Obviously, Figures 4(b)–4(d) show the injured phenomena; several pedestrians were injured by aggregated pushing force. In addition, the pushing force can result in the Domino effect; from Figure 4(a) to Figure 4(b), after the front pedestrian falls onto the ground, the latter lacking the brace to fall down resulted from the large pushing force. And then, the injured numbers are not significantly increased from Figure 4(c) to Figure 4(d).

The functional relationship between the injured numbers and the angle of the ramp is shown in Figures 5(a) and 5(b), which are simulated by the proposed model. Figures 5(a) and 5(b) correspond the low initial density ($\rho_0 \in (0, 0.5)$) and the high initial density ($\rho_0 \in (0.5, 1]$), respectively. According to these curves, the injured numbers increase as the angle of the ramp increases. Moreover, one curve is always above the other, which corresponds to a higher average density, implying that the injured numbers increases with ρ_0 for a fixed ramp. These results concur with common sense, especially when the initial average density ρ_0 is increasing, which should give rise to many more pushing effects. Obviously, the curves in Figure 5(a) increase continuously as the angle of the ramp increases; with the high initial density, that is, $\rho_0 > 0.5$, while

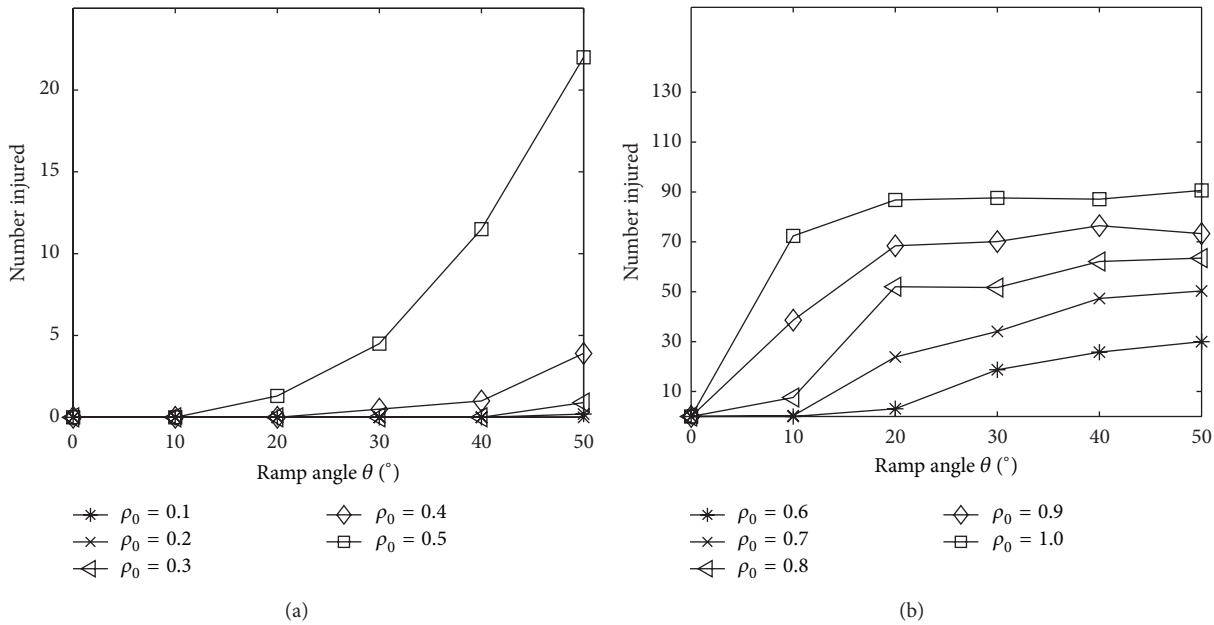


Figure 5: Functional relations between the injured number and the angle of the ramp for a variety of the average densities ρ_0, simulated by the proposed model. (a) $\rho_0 \leq 0.5$; (b) $\rho_0 > 0.5$. For each data, an average over 10 runs is plotted.

the ramp is more than 20°, the curves in Figure 5(b) reach a critical value and then the injured number will increase slowly and it implies that the Domino effect is obtained.

4. Conclusion

In this paper, a potential field CA model with an aggregated force field is proposed to reproduce pedestrian evacuation in a walking domain with ramp exit. And A potential field is defined to navigate pedestrian movement along minimizing the travel cost paths. In addition, an aggregated pushing force vector field is introduced to simulate the pushing force among pedestrians when pedestrian cannot move forward in high density domain. We investigate the effects of both the aggregated pushing force for pedestrian evacuation and the slope of ramp exit, the conclusion is that these two factors are the key causes of disasters. We will adjust the model parameters according to plentiful experiments and field studies in future work and will jointly simulate the pedestrian dynamics quantitatively and qualitatively using the proposed model.

Conflict of Interests

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

Acknowledgment

This study was supported by the National Natural Science Foundation of China (Grant no. 71125004).

References

[1] D. Helbing, A. Johansson, and H. Z. Al-Abideen, "Dynamics of crowd disasters: an empirical study," *Physical Review E*, vol. 75, no. 4, Article ID 046109, 2007.

[2] D. Helbing and P. Mukerji, "Crowd disasters as systemic failures: analysis of the love parade disaster," *EPJ Data Science*, vol. 1, no. 1, pp. 1–40, 2012.

[3] B. Krausz and C. Bauckhage, "Loveparade 2010: automatic video analysis of a crowd disaster," *Computer Vision and Image Understanding*, vol. 116, no. 3, pp. 307–319, 2012.

[4] S. C. Wong, W. L. Leung, S. H. Chan et al., "Bidirectional pedestrian stream model with oblique intersecting angle," *Journal of Transportation Engineering*, vol. 136, no. 3, pp. 234–242, 2010.

[5] H. Kuang, X. L. Li, T. Song, and S. Q. Dai, "Analysis of pedestrian dynamics in counter flow via an extended lattice gas model," *Physical Review E*, vol. 78, no. 6, Article ID 066117, 2008.

[6] H. Kuang, X. L. Li, Y. F. Wei, T. Song, and S. Q. Dai, "Effect of following strength on pedestrian counter flow," *Chinese Physics B*, vol. 19, no. 7, Article ID 070517, 2010.

[7] T. Xiong, P. Zhang, S. C. Wong, C. W. Shu, and M. P. Zhang, "A macroscopic approach to the lane formation phenomenon in pedestrian counterflow," *Chinese Physics Letters*, vol. 28, no. 10, Article ID 108901, 2011.

[8] P. Zhang, X. X. Jian, S. C. Wong, and K. Choi, "Potential field cellular automata model for pedestrian flow," *Physical Review E*, vol. 85, no. 2, Article ID 021119, 2012.

[9] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.

[10] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz, "Simulation of pedestrian dynamics using a two-dimensional cellular automaton," *Physica A*, vol. 295, no. 3-4, pp. 507–525, 2001.

[11] D. Helbing, "Traffic and related self-driven many-particle systems," *Reviews of Modern Physics*, vol. 73, no. 4, pp. 1067–1141, 2001.

[12] L. Huang, S. C. Wong, M. P. Zhang, C. W. Shu, and W. H. K. Lam, "Revisiting Hughes' dynamic continuum model for pedestrian flow and the development of an efficient solution algorithm," *Transportation Research B*, vol. 43, no. 1, pp. 127–141, 2009.

[13] Y. Q. Jiang, P. Zhang, S. C. Wong, and R. X. Liu, "A higher-order macroscopic model for pedestrian flows," *Physica A*, vol. 389, no. 21, pp. 4623–4635, 2010.

[14] A. Kirchner, K. Nishinari, and A. Schadschneider, "Friction effects and clogging in a cellular automaton model for pedestrian dynamics," *Physical Review E*, vol. 67, no. 5, Article ID 056122, 2003.

[15] W. G. Song, Y. F. Yu, W. C. Fan, and H. P. Zhang, "A cellular automata evacuation model considering friction and repulsion," *Science in China E*, vol. 48, no. 4, pp. 403–413, 2005.

[16] C. M. Henein and T. White, "Microscopic information processing and communication in crowd dynamics," *Physica A*, vol. 389, no. 21, pp. 4636–4653, 2010.

[17] W. Song, X. Xu, B.-H. Wang, and S. Ni, "Simulation of evacuation processes using a multi-grid model for pedestrian dynamics," *Physica A*, vol. 363, no. 2, pp. 492–500, 2006.

[18] H.-K. Zhao, "A fast sweeping method for Eikonal equations," *Mathematics of Computation*, vol. 74, no. 250, pp. 603–627, 2005.

[19] J. Fruin, "The causes and prevention of crowd disasters," in *Engineering for Crowd Safety*, R. A. Smith and J. F. Dickie, Eds., Elsevier, Amsterdam, The Netherlands, 1993.