

# Operations Research for Transportation and Sustainable Development

Lead Guest Editor: Jesica de Armas

Guest Editors: Helena Ramalhinho, Eduardo Lalla-Ruiz, Belen M. Batista, and  
Marcos M. Vega





---

# **Operations Research for Transportation and Sustainable Development**



Journal of Advanced Transportation

---

## **Operations Research for Transportation and Sustainable Development**

Lead Guest Editor: Jesica de Armas

Guest Editors: Helena Ramalhinho, Eduardo Lalla-Ruiz,  
Belen M. Batista, and Marcos M. Vega



Copyright © 2019 Hindawi Limited. All rights reserved.

This is a special issue published in "Journal of Advanced Transportation." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Francesco Bella, Italy  
Abdelaziz Bensrhair, France  
Cesar Briso-Rodriguez, Spain  
María Calderon, Spain  
Juan C. Cano, Spain  
Giulio E. Cantarella, Italy  
Maria Castro, Spain  
Anthony Chen, USA  
Nicolas Chiabaut, France  
Steven I. Chien, USA  
Antonio Comi, Italy  
Maria Vittoria Corazza, Italy  
Gonçalo Homem de Almeida Correia, The Netherlands  
Luca D'Acierno, Italy  
Andrea D'Ariano, Italy  
Alexandre De Barros, Canada  
Stefano de Luca, Italy  
Rocío de Oña, Spain  
Luigi Dell'Olio, Spain  
Cédric Demonceaux, France  
Sunder Lall Dhingra, India  
Roberta Di Pace, Italy  
Vinayak Dixit, Australia  
Yuchuan Du, China  
Nour-Eddin El-faouzi, France  
Juan-Antonio Escareno, France  
Francesco Galante, Italy  
Md. Mazharul Haque, Australia  
Jérôme Ha#rri, France  
Samiul Hasan, USA  
Serge P. Hoogendoorn, The Netherlands  
Hocine Imine, France  
Lina Kattan, Canada  
Victor L. Knoop, The Netherlands  
Alain Lambert, France  
Ludovic Leclercq, France  
Seungjae Lee, Republic of Korea  
Jaeyoung Lee, USA  
Zhi-Chun Li, China  
Yue Liu, USA  
Jose R. Martinez-De-Dios, Spain

Filomena Mauriello, Italy  
Rakesh Mishra, United Kingdom  
Andrea Monteriù, Italy  
Giuseppe Musolino, Italy  
Jose E. Naranjo, Spain  
Aboelmagd Noureldin, Canada  
Eneko Osaba, Spain  
Eleonora Papadimitriou, The Netherlands  
Dongjoo Park, Republic of Korea  
Paola Pellegrini, France  
Luca Pugi, Italy  
Hesham Rakha, USA  
Prianka N. Seneviratne, Philippines  
Fulvio Simonelli, Italy  
Richard S. Tay, Australia  
Pascal Vasseur, France  
Antonino Vitetta, Italy  
Francesco Viti, Luxembourg  
S. Travis Waller, Australia  
Shamsunnahar Yasmin, Australia  
Jacek Zak, Poland  
Guohui Zhang, USA

# Contents

## **Operations Research for Transportation and Sustainable Development**

Jesica de Armas, Helena Ramalhinho, Eduardo Lalla-Ruiz, Belen Melian-Batista, and Marcos Moreno-Vega

Editorial (2 pages), Article ID 4146362, Volume 2019 (2019)

## **The Lagrangean Relaxation for the Flow Shop Scheduling Problem with Precedence Constraints, Release Dates and Delivery Times**

Marcelus Fabri , Helena Ramalhinho , Mauricio C. de Souza, and Martin G. Ravetti

Research Article (10 pages), Article ID 3176074, Volume 2019 (2019)

## **Integrated Optimization of Tram Schedule and Signal Priority at Intersections to Minimize Person Delay**

Wen Zhou, Yun Bai , Jiajie Li , Yuhe Zhou , and Tang Li 

Research Article (18 pages), Article ID 4802967, Volume 2019 (2019)

## **Multiobjective Optimization Model for Sustainable Waste Management Network Design**

Sun Olapiriyakul , Warut Pannakkong , Warith Kachapanya, and Stefano Starita 

Research Article (15 pages), Article ID 3612809, Volume 2019 (2019)

## **Optimizing Transportation Network of Recovering End-of-Life Vehicles by Compromising Program in Polymorphic Uncertain Environment**

Jing Zhang, Jingjing Liu, and Zhong Wan 




Research Article (24 pages), Article ID 3894064, Volume 2019 (2019)

## **A Survey on the Electric Vehicle Routing Problem: Variants and Solution Approaches**

Tomislav Erdelić  and Tonči Carić 



Review Article (48 pages), Article ID 5075671, Volume 2019 (2019)

## **Adaptive Model Predictive Control for Cruise Control of High-Speed Trains with Time-Varying Parameters**

Xiaokang Xu, Jun Peng , Rui Zhang, Bin Chen, Feng Zhou, Yingze Yang , Kai Gao, and Zhiwu Huang 


Research Article (11 pages), Article ID 7261726, Volume 2019 (2019)

## **Optimal Location of Biogas Plants in Supply Chains under Carbon Effects: Insight from a Case Study on Animal Manure in North Dakota**

Yong Shin Park , Joseph Szmerekovsky , and Alan Dybing

Research Article (13 pages), Article ID 5978753, Volume 2019 (2019)


## **Determination of Bus Crowding Coefficient Based on Passenger Flow Forecasting**

Zhongyi Zuo , Wei Yin, Guangchuan Yang, Yunqi Zhang, Jiawen Yin, and Hongsheng Ge

Research Article (12 pages), Article ID 2751916, Volume 2019 (2019)



**A Two-Step Approach for Airborne Delay Minimization Using Pretactical Conflict Resolution in Free-Route Airspace**

Ramazan Kursat Cecen  and Cem Cetek



Research Article (17 pages), Article ID 4805613, Volume 2019 (2019)

**Equilibrium and Optimization in a Double-Ended Queueing System with Dynamic Control**

Yuejiao Wang  and Zaiming Liu


Research Article (13 pages), Article ID 6538265, Volume 2019 (2019)

**Evaluation of Transportation Network Reliability under Emergency Based on Reserve Capacity**

Xiongfei Zhang , Qi Zhong, and Qin Luo 

Research Article (13 pages), Article ID 1370159, Volume 2019 (2019)

**ST-LSTM: A Deep Learning Approach Combined Spatio-Temporal Features for Short-Term Forecast in Rail Transit**

Qicheng Tang, Mengning Yang , and Ying Yang

Research Article (8 pages), Article ID 8392592, Volume 2019 (2019)

## Editorial

# Operations Research for Transportation and Sustainable Development

**Jesica de Armas,<sup>1</sup> Helena Ramalhinho,<sup>1</sup> Eduardo Lalla-Ruiz,<sup>2</sup> Belen Melian-Batista,<sup>3</sup> and Marcos Moreno-Vega<sup>3</sup>**

<sup>1</sup>*Department of Economics and Business, Universitat Pompeu Fabra, Spain*

<sup>2</sup>*Department of Industrial Engineering and Business Information Systems, University of Twente, Netherlands*

<sup>3</sup>*Department of Computer Engineering and Systems, Universidad de La Laguna, Spain*

Correspondence should be addressed to Jesica de Armas; [jesica.dearmas@upf.edu](mailto:jesica.dearmas@upf.edu)

Received 21 July 2019; Accepted 25 July 2019; Published 31 December 2019

Copyright © 2019 Jesica de Armas et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sustainable development is the core principle for meeting human development goals while at the same time sustaining the ability of systems to produce, extract, process, or manufacture the natural resources and services upon which the economy and society depend.

Transportation produces water, noise, and air contamination, being a large contributor to greenhouse gas emissions. These environmental shortcomings can be mitigated by considering and integrating sustainable social, environmental, and economic goals that impact and accelerate green and sustainable development. Furthermore, sustainable transportation can be viewed and integrated as an essential ingredient in sustainable development strategies. Components for evaluating transport sustainability include particular transportation means used for road, water, or air transport; the way they are organized, the source of energy, and the infrastructure used to accommodate the transport. In this context, optimization plays a relevant role. Methods from Operations Research have been extensively used to successfully deal with many of these challenges.

This special issue aims at examining the current progress, challenges, and approaches in transportation to achieve a sustainable development through optimization. Thus, the scope of this issue covers high-quality original research and review articles focusing on the use of Operations Research techniques (mathematical programming, metaheuristics, simulation based approaches, etc.) to deal with transportation for promoting a sustainable development, mitigating

the negative environmental impacts, and identifying main opportunities to accelerate the transition to sustainability.

An overview of the topics in the selected articles composing this special issue is given below.

*Sustainable Waste Management Network Design.* Designing a cost-effective waste management supply chain, while considering sustainability issues such as land-use and public health impacts, is possible thanks to optimization models, and particularly to multiobjective approaches.

*Transportation Network of Recovering End-of-Life Vehicles.* From the viewpoints of environmental protection and resource utilization, it is valuable to investigate an effective strategy through optimization for recycling end-of-life vehicles. Fuzzy and stochastic parameters are keys to model these uncertain environments.

*Electric Vehicle Routing Problems.* Due to the challenges emerging from the integration of electric vehicles in logistics processes, a review of state-of-the-art Operations Research solution approaches and procedures for solving the Electric Vehicle Routing Problem (E-VRP) is useful.

*Cruise Control of High-Speed Trains.* High-speed train transportation has many more advantages than traditional railway traffic, such as high speed, large volume, and a safe and comfortable environment. Acceleration and deceleration also

have an impact on energy consumption and sustainability, so an optimized control method is important.

*Location of Biogas Plants in Supply Chains.* Biogas from animal manure not only provides energy efficiency but also minimizes carbon emissions compared to existing biomass products. Therefore, it is interesting the use of Operations Research techniques to consider biogas plants and minimize the total supply costs and carbon emissions.

*Bus Crowding and Passenger Flow.* Traffic congestion problems have seriously restricted economic development and affected people's daily lives. The improvement of public transportation should have priority, increasing the service quality and ameliorating the travel environment. One relief can be obtained minimizing the impact of the randomness of passenger flow on the determination of bus crowd coefficient.

*Airborne Delay Minimization in Free-Route Airspace.* Airborne transportation has a relevant economic impact for international business, trade, and tourism. It involves the management of a rising number aircrafts requiring safe separation distances among them. Thus, having decision support approaches for preventing collisions and managing conflicts at a pretactical level is necessary and crucial.

*Equilibrium and Optimization in a Double-Ended Queueing System.* Taxis service plays a relevant role in sustainable mobility. In order to make a more efficient use of this resource, optimal strategies which adjust the arrival rate of taxis to the passengers demand are necessities. This problem can be modelled as a double-ended queueing system.

*Transportation Network Reliability under Emergency.* Urban traffic networks are designed to address daily traffic demand. When an emergency occurs, the existing capacity of the infrastructure is usually not able to meet the demand of evacuation. To evaluate the maximum emergency evacuation demand that can be satisfied by of the network is necessary to consider its reliability under emergency.

*Short-Term Forecast in Rail Transit.* The short-term forecast of rail transit is one of the most essential issues in urban intelligent transportation system. Combining spatiotemporal features based on long short-term memory network allow to propose novel forecast models.

*Flow Shop Scheduling Problem.* This problem arise in many manufacturing companies. In the particular case of the automotive business, the idea is to schedule the jobs with precedence constraints, release dates and delivery times, in order to satisfy client demands as soon as possible, i.e., minimizing the latest date a client receives its products. This involves a better development of this industry by Operations Research techniques.

*Tram Schedule and Signal Priority at Intersections.* With the purpose of improving punctuality and reliability in trams, transit signal priority is used at intersections, which can extend or insert green phase to trams. Given the fact that

this may produce delays to crossing vehicles, an integrated optimization model on tram schedule and signal priority can balance the delay between trams and other vehicles in order to minimize person delay.

The guest editors of this special issue hope that the selected articles will be able to arise the interest of Operations Research for Transportation and Sustainable Development, especially in the aforementioned fields.

## Conflicts of Interest

The guest editors declare that there are no conflicts of interest regarding the special issue.

## Acknowledgments

The guest editors would like to thank the reviewers for the time and efforts they have devoted to provide detailed comments, advice, and suggestions for improvement of the technical and scientific level, as well as the presentation quality of the accepted papers.

Jesica de Armas  
Helena Ramalhinho  
Eduardo Lalla-Ruiz  
Belen Melian-Batista  
Marcos Moreno-Vega

## Research Article

# The Lagrangean Relaxation for the Flow Shop Scheduling Problem with Precedence Constraints, Release Dates and Delivery Times

Marcelus Fabri <sup>1</sup>, Helena Ramalhinho <sup>2</sup>, Mauricio C. de Souza,<sup>3</sup> and Martin G. Ravetti<sup>3</sup>

<sup>1</sup>Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain

<sup>2</sup>Department of Economics, Universitat Pompeu Fabra, Barcelona, Spain

<sup>3</sup>Department of Production Engineering, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Correspondence should be addressed to Marcelus Fabri; marcelusfl@gmail.com

Received 14 December 2018; Revised 14 May 2019; Accepted 28 June 2019; Published 23 December 2019

Academic Editor: Oded Cats

Copyright © 2019 Marcelus Fabri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work aims to present a methodology to support a company in the automotive business on scheduling the jobs on its final processes. These processes are: (i) checking the final product and (ii) loading the dispatch trucks. These activities are usually found in the outbound area of any manufacturing company. The problem faced is defined as the flow shop problem with precedence constraints, release dates, and delivery times. The major objective is to minimize the latest date a client receives its products. We present a time-indexed integer mathematical model to compute feasible solutions for the presented problem. Moreover, we take advantage of the Lagrangean Relaxation procedure to compute valid lower and upper bounds. The experiments were held based on the company's premises. As a conclusion, the results showed that the methodology proposed was able to compute feasible solutions for all the instances tested. Also, the Lagrangean Relaxation approach was able to calculate better bounds in a shorter computational time than the Mathematical problem for the more complicated instances.

## 1. Introduction

Many companies sort their processes in a sequential way. The sequential standard follows quite well the concept that each process has its suppliers and clients, which may be represented by the previous and the successor processes, respectively.

The problem defined above may be viewed as a flow shop one [1]. In the flow shop approach, each process may be viewed as a single or a set of machines. Each machine is responsible for executing a specific task. Moreover, all the jobs must follow the same order of machines. So, after a job completes a task in a machine, it must join the queue at the next machine. Furthermore, the tasks must be executed under some constraints such as release dates or resources availability. In this work, we approach a real-world problem that may be equivalent to a flow shop problem with precedence constraints, release dates, and delivery times. The company evaluated in this work, focuses its activities on the car-assembling business.

We concerned with a part of the company's processes to carry out the work. To be precise, we focus on the three final processes of the company's production flow, which are the checking, loading, and dispatching. In addition, we work under a set of premises. As a result, we state the studied problem as the flow shop problem with precedence constraints, release dates, and delivery times with the objective function of minimizing the makespan. That problem is defined as a strongly NP-hard problem [2]. The problem is depicted in Figure 1 and described as follows.

The observed company acts in the automotive business, and it is able to produce up to 2.000 products each day. We highlight that the products are not all the same, because there may exist differences between them. So, after the manufacturing phases, each product is assigned to a cluster. A cluster is represented by the letters A, B, and C in Figure 1. One cluster is set based on the similarities of the products. Also, each cluster must pass a quality control process to avoid sending poor-quality products to the final clients. The control consists



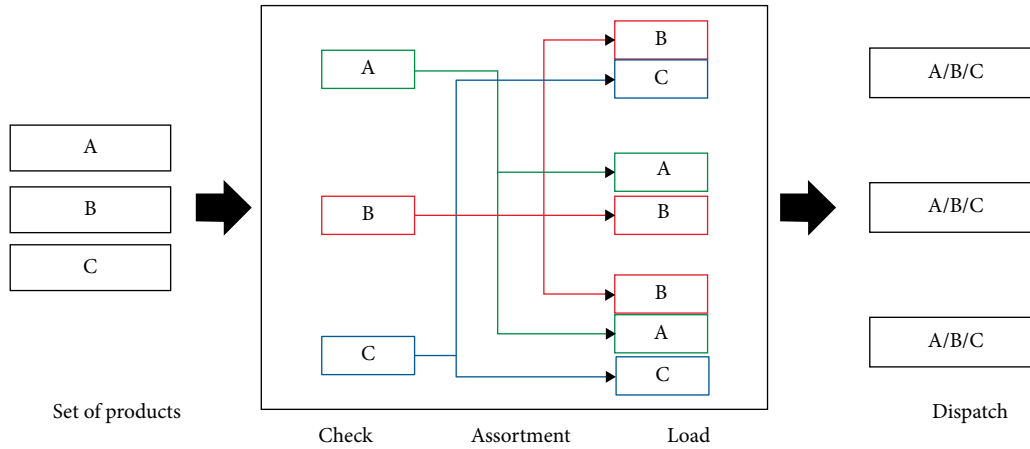


FIGURE 1: The processes' scheme.

of evaluating each product individually. As a result, the operator receives a list of clusters needed to be checked. Then, each cluster is schedule and checked one by one. It is noteworthy to state that the time spent at each cluster may vary according to the type of the products that compound the cluster, and the failures that may be found, as well. Usually, the cluster is compounded by a set of products that do not significantly differ from themselves.

In contrast, some products do differ from themselves, and that is the reason that justifies the clustering procedure. So, the company does not mix different products in the same cluster.

Afterward the checking phase, an established cluster is fragmented. In other words, the products that compound that cluster, are placed in a waiting zone and assorted according to its destinations. Usually, clients receive products that belong to a variety of clusters. So, after a truck's load has already been completely placed in the waiting area, the dispatching truck is authorized to get inside the marquee. Therefore, a truck is not allowed to get inside before all its cargo have already been placed in the waiting zone. So, we can assume that each outgoing truck has a precedent-jobs list. Then, after the truck arrival, the operators start to load the truck and conclude their task after an amount of time, which may vary according to the truck type, the cargo and the paperwork required. In this work, we assume that there is one team in charge of loading the trucks. As a result, we can not assume that more than one truck is loading at the same time.

Finally, the last process is the Dispatching one. Here the truck departs from the company towards the client's location. We highlight that we did not consider evaluating which route the truck driver should do, but the average duration to reach the clients' locations instead.

So, we considered modeling the checking and the loading process as machines. So, the checking process is defined as the Machine 1 ( $M_1$ ) and the loading process as the Machine 2 ( $M_2$ ). Moreover, we state the main objective is to minimize the maximum date when a client receives its products. As a result, we decided to model this problem as a two-machine flow shop scheduling problem. Therefore, we define this problem as the

two-machine flow shop scheduling problem with precedence constraints, release dates and delivery times.

Also, we modeled it through a time-indexed formulation which is based on the discretization of the time horizon. This kind of formulation is known to provide tighter linear relaxation bounds. However, the model presents a high number of variables.

In this work, we face two primary objectives. The first is to present a suitable model to cope with the flow shop problem with precedence constraints, release dates and delivery times. The second goal is to solve the proposed model problem through an appropriated decomposition method. Here, we propose an integer linear programming model on time-indexed variables and a Lagrangean Relaxation (LR) approach to obtain both upper and lower bounds. The sub-gradient method was chosen to carry out the convergence of the LR. See [3].

The paper is organized as follows. The mathematical model section presents the integer linear programming model. The Lagrangean Relaxation section describes the decomposition approach. The Computational Results section reports experiments results. Finally, the Conclusions section concludes the paper.

## 2. The Mathematical Model

The two-machine flow shop scheduling problem with precedence constraints, release dates and delivery times is set as  $(F2|prec, r_j, q_j|D_{max})$ . In other words, we aim to solve a scheduling problem, whose target is to minimize the date when the last product will be delivered to the client. Moreover, we deal with a set of constraints related to the company's machines. These constraints state that the machines are not allowed to work with more than one job at the same time. Also, all those jobs must be executed only once. Besides that, there is a precedence list in order to enable a job to be processed in machine 2. We present the mathematical model of this formulation next.

The notation was set based on [1]. The term  $F2$  means that there are two machines in sequence. The terms

( $prec, r_j, q_j$ ) mean that there are three classes of constraints applied, which are: the precedence, the release dates and the delivery times ones, respectively. Finally, the  $D_{\max}$  term refers to the objective function of minimizing the latest date a client receives its products. Moreover, the  $(F2|prec, r_j, q_j|D_{\max})$  can be described as follows. Consider a set  $J^1 = \{1, \dots, n\}$  of jobs to be processed on the first machine  $M_1$ , which correspond to the clusters to be processed at the first machine. Likewise, consider a set  $J^2 = \{1, \dots, m\}$  of jobs to be processed on the second machine  $M_2$ , which correspond to outgoing trucks to be loaded at the loading process. The processing time of a job  $j$  at a machine  $i = \{1, 2\}$  is denoted by  $p_{ij}$ , which corresponds to the time for processing a cluster at the  $M_1$ . The precedence constraints applied imply that to each job (truck)  $j \in J^2$  is associated a nonempty set  $S_j \subseteq J^1$  such that  $j$  can only be processed in  $M_2$  after all clusters belonging to  $S_j$  have been processed in  $M_1$ . Furthermore, we consider the delivery time  $d_j$  for each job  $j \in J^2$ . It corresponds to the time required to deliver the job  $j \in J^2$  to its final client. The delivery time data refers to the time spent on the road. As a reference, the reader may consider that the delivery time is related to the time spent from Spain to another European country and those in North Africa.

So, the problem consists of finding a sequence of clusters in  $J^1$  to be processed in  $M_1$  and a sequence of the jobs in  $J^2$  to be processed in  $M_2$  to minimize the maximum date when a client receives its job. In other words, we aim to minimize  $D_{\max}$ , which is equivalent to the latest time that one client receives its products.

To this aim, we propose a time-indexed model, which is based on a time-discretization of the planning horizon into a set  $T = \{0, \dots, h\}$  of periods. Time-indexed formulations have been shown in the literature to be likely to provide better LP-relaxation bounds than other formulations for scheduling problems, see [4–6]. Continuing, we define two binary variables:  $x_{jt}$  (resp.  $y_{jt}$ ) assumes value 1 if a job  $j \in J^1$  (resp.  $j \in J^2$ ) starts its processing in  $M_1$  (resp.  $M_2$ ) in period  $t \in T$ , and 0 otherwise. Also, the integer variable  $D_{\max}$ , which is the maximum time when the last client receives its products. The integer linear programming (ILP) formulation proposed for that version of the  $(F2|prec, r_j, q_j|D_{\max})$  is presented as follows:

$$\min D_{\max} \quad (1)$$

$$\sum_{t=0}^{h-p_{1j}} x_{jt} = 1 \quad \forall j \in J^1 \quad (2)$$

$$\sum_{t=0}^{h-p_{2j}} y_{jt} = 1 \quad \forall j \in J^2 \quad (3)$$

$$\sum_{t=0}^{h-p_{2j}} t y_{jt} - \sum_{t=0}^{h-p_{1k}} (t + p_{1k}) x_{kt} \geq 0 \quad \forall j \in J^2, \forall k \in S_j \quad (4)$$

$$\sum_{j \in J^1} \sum_{s=\max(0, t-p_{1j}+1)}^t x_{js} \leq 1 \quad \forall t \in T \quad (5)$$

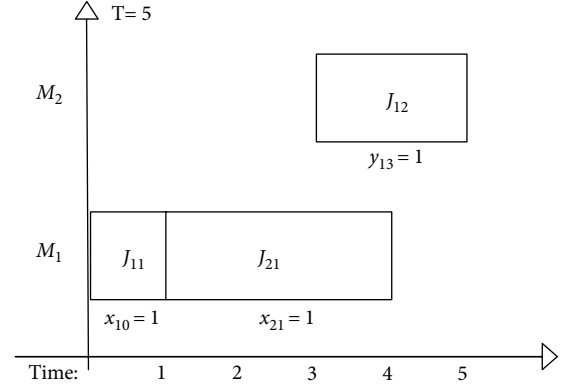


FIGURE 2: Scheduling violating the precedence constraints 4.

$$\sum_{j \in J^2} \sum_{s=\max(0, t-p_{2j}+1)}^t y_{js} \leq 1 \quad \forall t \in T \quad (6)$$

$$D_{\max} \geq \sum_{t=0}^{h-p_{2j}} (t + p_{2j} + d_j) y_{jt} \quad \forall j \in J^2 \quad (7)$$

$$x_{jt} \in \{0, 1\} \quad \forall j \in J^1, \forall t \in T, t \leq h - p_{1j} \quad (8)$$

$$y_{jt} \in \{0, 1\} \quad \forall j \in J^2, \forall t \in T, t \leq h - p_{2j} \quad (9)$$

$$D_{\max} \in \mathbb{Z}^+ \quad (10)$$

The objective function (1) minimizes the time of reception of the last job, which is equivalent to the biggest value possible of  $(t + p_{2j} + d_j) y_{jt}$ . Constraints (2) (resp. (3)) state that each job  $j \in J^1$  (resp.  $j \in J^2$ ) has to be started exactly once at  $M_1$  (resp.  $M_2$ ). In other words, one job must not be processed more than once by either the machines. The precedence constraints (4) ensure that a job  $j \in J^2$  cannot be processed at  $M_2$  before all jobs in  $S_j$  have been completed at  $M_1$ . Further explanation could be found in Figure 2. Constraints (5) (resp. (6)) state that machine  $M_1$  (resp.  $M_2$ ) can handle at most one job at any time period. Consequently, the  $M_1$  and  $M_2$  are not able to process more than one job at the same time. The  $D_{\max}$  value is established by the constraints (7). As a result, the period a final product the client is bigger or equal to the sum of its processing time in  $M_2$  ( $p_{2j}$ ), the delivery time ( $d_j$ ) from the manufacturer up to the client, and the period when it has been started to be processed in  $M_2(t)$ . Lastly, constraints (8)–(10) define the domain of the variables.

### 3. The Lagrangean Relaxation

The approach proposed to obtain lower and upper bounds for the  $(F2|prec, r_j, q_j|D_{\max})$  is based on applying a Lagrangean Relaxation (LR) procedure to the previous ILP model. The precedence constraints (4) that couple the scheduling on machines  $M_2$  and  $M_1$  are relaxed in the LR approach.

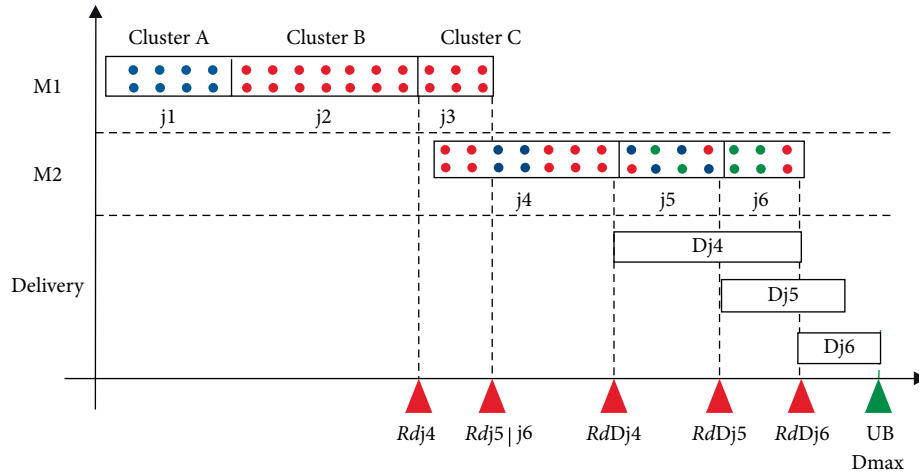


FIGURE 3: An upper bound solution example. The red triangles represent the release dates for each job. Furthermore, the green triangle represents the UB value.

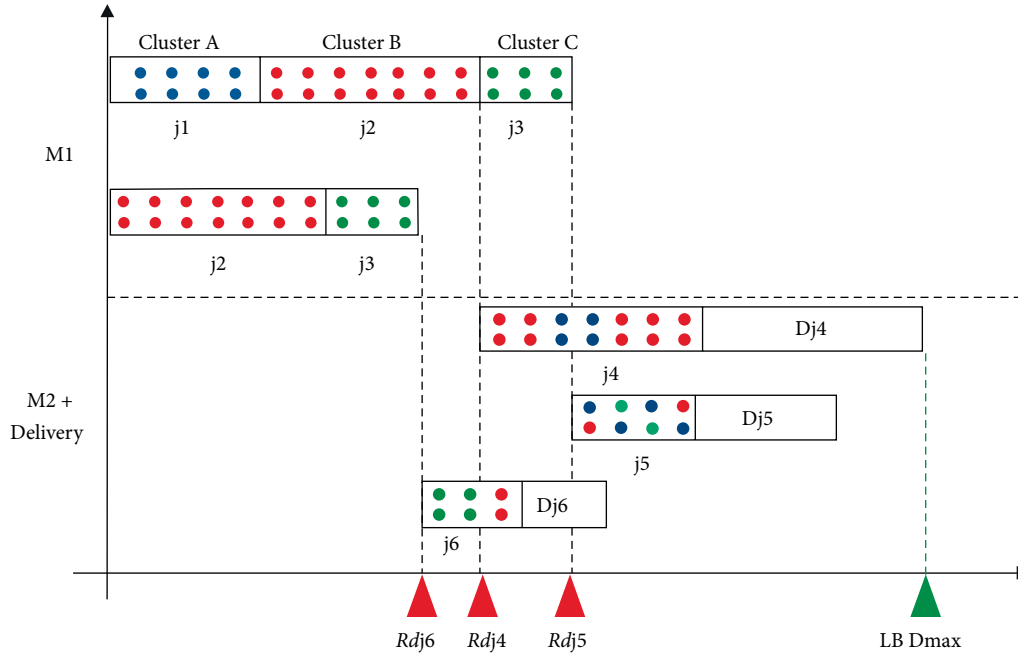


FIGURE 4: A lower bound solution example. The red triangles represent the release dates for each job. Furthermore, the green triangle represents the LB value. The figure depicts how the minimum release date ( $Rd$ ) is computed for each  $j \in J^2$ .

Figure 2 illustrates a scheduling which is not feasible to constraints (4). In this example, job 1 ( $J_{11}$ ) processed by  $M_1$  begins its processing at  $t = 0$ , so,  $x_{10} = 1$ , and job 2 ( $J_{21}$ ) processed by  $M_1$  begins its processing at  $t = 1$ , so,  $x_{21} = 1$ . Considering  $S_1 = \{J_{11}, J_{21}\}$  as the predecessors of job 1 ( $J_{12}$ ) processed by  $M_2$ , it can only start its processing at  $t = 4$ . However, in this example,  $J_{21}$  begins its processing at  $t = 3$ ,  $y_{13} = 1$ , and not at  $t = 4$  as expected, violating the precedence constraints.

Let  $\lambda_{jk} \geq 0, j \in J^2, k \in S_j$  be the Lagrangean multipliers associated with constraints (4). The Lagrangean multipliers are equivalent to the dual variables associated with each relaxed constraint placed in the objective function, see [3]. If a job is scheduled in  $M_2$  before one of its predecessor has been

completed, the objective function will be penalised in the Lagrangean subproblem  $L(\lambda)$  as follows:

$$L(\lambda) = \min D_{\max} + \sum_{j \in J^2} \sum_{k \in S_j} \lambda_{jk} \sum_{t=0}^{h-p_{1k}} (t + p_{1k}) x_{kt} - \sum_{t=0}^{h-p_{2j}} t y_{jt} \tag{11}$$

s.t. (2), (3), (5)–(10). We use the subgradient algorithm to solve the Lagrangean dual  $\max\{L(\lambda) : \lambda \geq 0\}$ . The resulting model allows us to decompose the Lagrangean subproblem into one smaller subproblem  $L(\lambda)_x$  in  $M_1$ .

First, the subproblem  $L(\lambda)_x$  is the total weighted completion time scheduling problem on one machine. The completion of a job  $k \in J^1$  is weighted by the sum of penalties applied

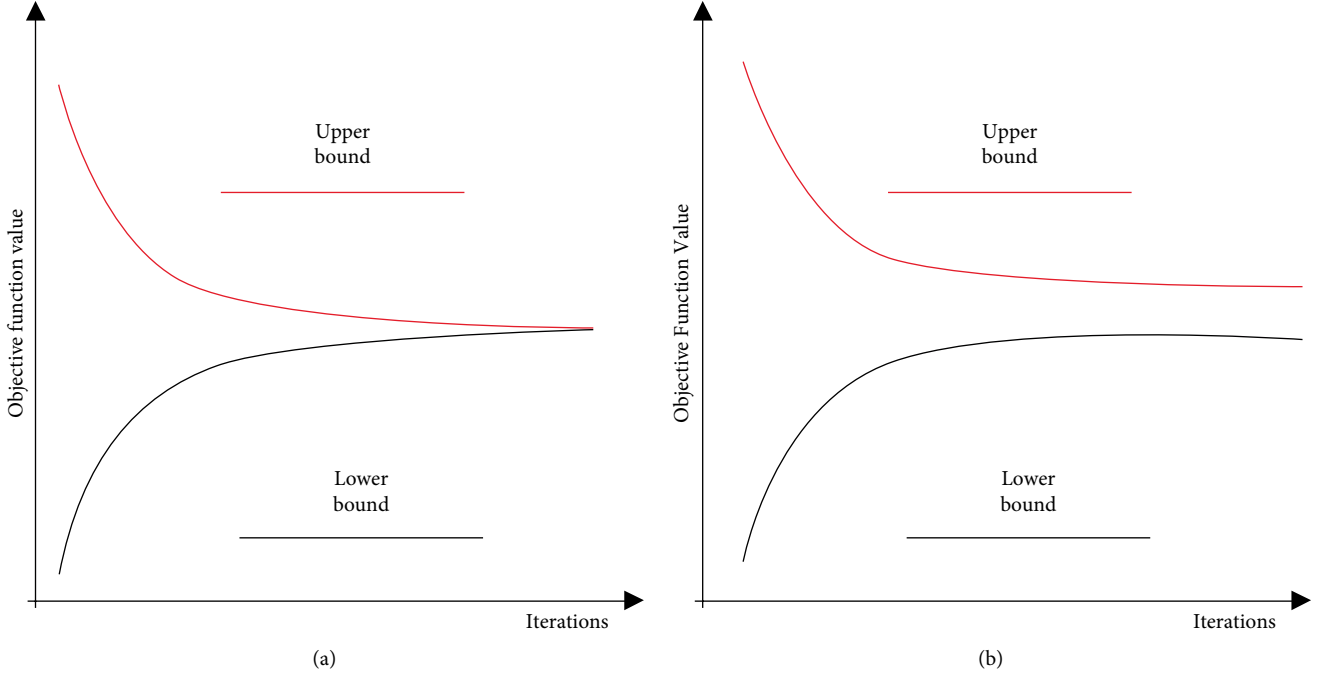


FIGURE 5: A scheme of the Lagrangean Relaxation method solved through the subgradient algorithm. The first figure represents a scenario that the optimal solution was found because the LB and UB achieved the same value. On the contrast, the second figure represents a scenario that the optimal solution was not found.

TABLE 1: Summary of the instances structure.

Group	$n$	$m$	$ S_j $	$p_j$	$d_j$	Identification
1	5	3-4-5-6-7	U(1, 4)	U(1, 10)	U(10 <sup>2</sup> , 10 <sup>3</sup> )	5-m-4-1
	10	6-8-10-12-14	U(1, 9)	U(1, 10)	U(10 <sup>2</sup> , 10 <sup>3</sup> )	10-m-9-1
	20	12-16-20-24-28	U(1, 19)	U(1, 10)	U(10 <sup>2</sup> , 10 <sup>3</sup> )	20-m-19-1
	40	24-32-40-48-56	U(1, 39)	U(1, 10)	U(10 <sup>2</sup> , 10 <sup>3</sup> )	40-m-39-1
	60	36-48-60-72-84	U(1, 59)	U(1, 10)	U(10 <sup>2</sup> , 10 <sup>3</sup> )	60-m-59-1
2	5	3-4-5-6-7	U(1, 4)	U(10, 10 <sup>2</sup> )	U(10 <sup>3</sup> , 5000)	5-m-4-2
	10	6-8-10-12-14	U(1, 9)	U(10, 10 <sup>2</sup> )	U(10 <sup>3</sup> , 5000)	10-m-9-2
	20	12-16-20-24-28	U(1, 19)	U(10, 10 <sup>2</sup> )	U(10 <sup>3</sup> , 5000)	20-m-19-2
	40	24-32-40-48-56	U(1, 39)	U(10, 10 <sup>2</sup> )	U(10 <sup>3</sup> , 5000)	40-m-39-2
	60	36-48-60-72-84	U(1, 59)	U(10, 10 <sup>2</sup> )	U(10 <sup>3</sup> , 5000)	60-m-59-2

on all jobs to be scheduled on  $M_2$  which have  $k$  among their predecessors. Setting  $\omega_k^1 = \sum_{j \in J^2: k \in S_j} \lambda_{jk}$ ,  $k \in J^1$ , subproblem  $L(\lambda)_x$  is written as follows:

$$L(\lambda)_x = \min \sum_{k \in J^1} \omega_k^1 \sum_{t=0}^{h-p_{1k}} (t + p_{1k}) x_{kt} \quad (12)$$

s.t. (2), (5), and (7). Subproblem  $L(\lambda)_x$  can be solved by the weighted shortest processing time first rule, in which jobs are sorted in decreasing order of  $\omega_k^1/p_{1k}$ . The proof is stated in [1]. So, we can infer an UB based on the jobs sequenced on the  $M_1$ . It will allow us to calculate the release dates of the jobs on the  $M_2$ . We will discuss the UB next.

In order to obtain a valid UB, let  $\bar{x}_{kt}$ ,  $k \in J^1$ ,  $t \in T$ , be an optimal solution of subproblem  $L(\lambda)_x$  for a given value of the Lagrangean multipliers. On the one hand, a schedule on  $M_2$  such

that the starting time  $r_j$  of each job  $j \in J^2$  satisfies  $r_j \geq \max_{k \in S_j} \left\{ \sum_{t=0}^{h-p_{1k}} (t + p_{1k}) x_{kt} \right\}$  is feasible and provides an upper bound for the  $(F2|prec, r_j, q_j|D_{\max})$  problem. On the other hand, the total weighted completion time scheduling problem on one machine with release dates is NP-Hard [7]. Thus, we propose a Lagrangean heuristic that consists of solving, for a given value of the Lagrangean multipliers, subproblem  $L(\lambda)_x$  in  $M_1$  and then applying the approximation algorithm proposed by Philipps et al. [8] to the resulting problem with release dates in  $M_2$ .

In the first step, the algorithm allows preemption to get an optimal schedule with the remaining weighted shortest processing time first rule. In the second step, jobs are nonpreemptively scheduled in the same order of their completion times. The algorithm produces, in  $O(n)$  time, a nonpreemptive schedule increasing the total weighted completion time by at most a factor of 2 to a preemptive schedule.



TABLE 2: Results for the ILP model running CPLEX with a time limit of 7,200 seconds, the GAP is defined as  $(UB - LB/UB)$ . Results depicted for the groups of instances 1 and 2. The bolded LB and UB values represent an instance's optimal solution.

Instance	Group 1 instances										Group 2 instances												
	Integer model CPLEX					Lagrangean Relaxation					Integer model CPLEX					Lagrangean Relaxation							
	LB	UB	GAP (%)	Time (sec)	Instance	LB	UB	GAP (%)	Time (sec)	Instance	LB	UB	GAP (%)	Time (sec)	LB	UB	GAP (%)	Time (sec)	Instance	LB	UB	GAP (%)	Time (sec)
5-3-4-1	794	794	0	0	5-3-4-2	4,630	4,630	0	1	4,630	4,660	1	0	5-3-4-2	4,630	4,660	1	0	5-3-4-2	4,630	4,660	1	0
5-4-4-1	671	671	0	0	5-4-4-2	3,222	3,222	0	12	3,222	3,222	0	0	5-4-4-2	3,222	3,222	0	0	5-4-4-2	3,222	3,222	0	0
5-5-4-1	842	842	0	0	5-5-4-2	3,222	3,222	0	26	3,222	3,305	2	0	5-5-4-2	3,222	3,305	2	0	5-5-4-2	3,222	3,305	2	0
5-6-4-1	801	801	0	0	5-6-4-2	4,855	4,855	0	26	4,840	4,959	3	0	5-6-4-2	4,855	4,959	3	0	5-6-4-2	4,855	4,959	3	0
5-7-4-1	901	901	0	0	5-7-4-2	4,954	4,954	0	21	4,954	5,090	3	0	5-7-4-2	4,954	5,090	3	0	5-7-4-2	4,954	5,090	3	0
10-6-9-1	939	939	0	1	10-6-9-2	5,271	5,272	0.001	7,200	5,221	5,317	2	0	10-6-9-2	5,271	5,317	2	0	10-6-9-2	5,271	5,317	2	0
10-8-9-1	983	983	0	5	10-8-9-2	4,289	4,289	0	52	4,289	4,289	0	0	10-8-9-2	4,289	4,289	0	0	10-8-9-2	4,289	4,289	0	0
10-10-9-1	1,002	1,002	0	10	10-10-9-2	5,080	5,080	0	3,138	5,080	5,123	1	0	10-10-9-2	5,080	5,123	1	0	10-10-9-2	5,080	5,123	1	0
10-12-9-1	1,001	1,001	0	2	10-12-9-2	4,802	4,802	0	70	4,802	4,865	1	0	10-12-9-2	4,802	4,865	1	0	10-12-9-2	4,802	4,865	1	0
10-14-9-1	906	906	0	0	10-14-9-2	5,151	5,151	0	157	5,151	5,220	2	0	10-14-9-2	5,151	5,220	2	0	10-14-9-2	5,151	5,220	2	0
20-12-19-1	926	926	0	28	20-12-19-2	5,398	6,124	12	7,200	5,921	6,003	1	1	20-12-19-2	5,398	6,124	12	7,200	20-12-19-2	5,398	6,124	12	7,200
20-16-19-1	962	962	0	24	20-16-19-2	5,248	5,824	10	7,200	5,697	5,963	4	1	20-16-19-2	5,248	5,963	10	7,200	20-16-19-2	5,248	5,963	10	7,200
20-20-19-1	1,048	1,048	0	43	20-20-19-2	5,394	5,726	6	7,200	5,726	6,124	6	2	20-20-19-2	5,394	5,726	6	7,200	20-20-19-2	5,394	5,726	6	7,200
20-24-19-1	1,082	1,082	0	51	20-24-19-2	5,313	7,033	24	7,200	5,581	6,201	1	2	20-24-19-2	5,313	7,033	24	7,200	20-24-19-2	5,313	7,033	24	7,200
20-28-19-1	1,100	1,100	0	189	20-28-19-2	5,647	7,328	23	7,200	6,067	6,617	8	3	20-28-19-2	5,647	7,328	23	7,200	20-28-19-2	5,647	7,328	23	7,200
40-24-39-1	1,012	1,012	0	2,332	40-24-39-2	5,928	7,941	25	7,200	6,281	7,296	14	6	40-24-39-2	5,928	7,941	25	7,200	40-24-39-2	5,928	7,941	25	7,200
40-32-39-1	1,096	1,159	5	7,200	40-32-39-2	-	-	-	7,200	6,025	7,132	15	7	40-32-39-2	-	-	-	7,200	40-32-39-2	-	-	-	7,200
40-40-39-1	1,084	1,128	4	7,200	40-40-39-2	-	-	-	7,200	7,296	7,989	9	12	40-40-39-2	-	-	-	7,200	40-40-39-2	-	-	-	7,200
40-48-39-1	1,074	1,146	6	7,200	40-48-39-2	-	-	-	7,200	6,562	8,227	20	15	40-48-39-2	-	-	-	7,200	40-48-39-2	-	-	-	7,200
40-56-39-1	1,051	1,126	7	7,200	40-56-39-2	-	-	-	7,200	7,111	7,976	11	17	40-56-39-2	-	-	-	7,200	40-56-39-2	-	-	-	7,200
60-36-59-1	1,184	1,288	8	7,200	60-36-59-2	-	-	-	7,200	7,349	8,144	10	16	60-36-59-2	-	-	-	7,200	60-36-59-2	-	-	-	7,200
60-48-59-1	1,160	1,505	23	7,200	60-48-59-2	-	-	-	7,200	8,101	9,839	18	34	60-48-59-2	-	-	-	7,200	60-48-59-2	-	-	-	7,200
60-60-59-1	1,194	1,580	24	7,200	60-60-59-2	-	-	-	7,200	7,572	9,609	21	43	60-60-59-2	-	-	-	7,200	60-60-59-2	-	-	-	7,200
60-72-59-1	1,195	1,618	26	7,200	60-72-59-2	-	-	-	7,200	7,150	9,905	28	45	60-72-59-2	-	-	-	7,200	60-72-59-2	-	-	-	7,200
60-84-59-1	1,118	1,100,843	99	7,200	60-84-59-2	-	-	-	7,200	7,388	10,302	28	61	60-84-59-2	-	-	-	7,200	60-84-59-2	-	-	-	7,200

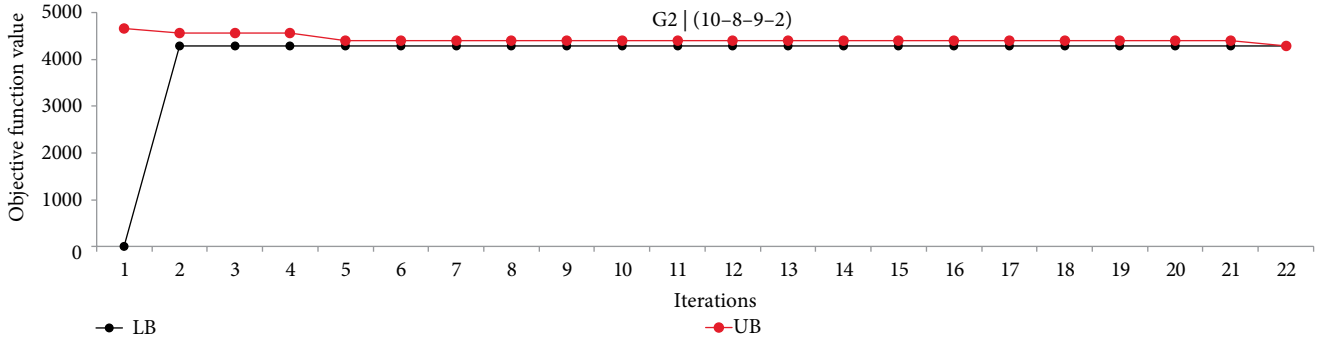


FIGURE 6: The instance G2 | (10-8-9-2) did converge its LB and UB values. As a result, both UB and the LB computed are considered optimal solutions.

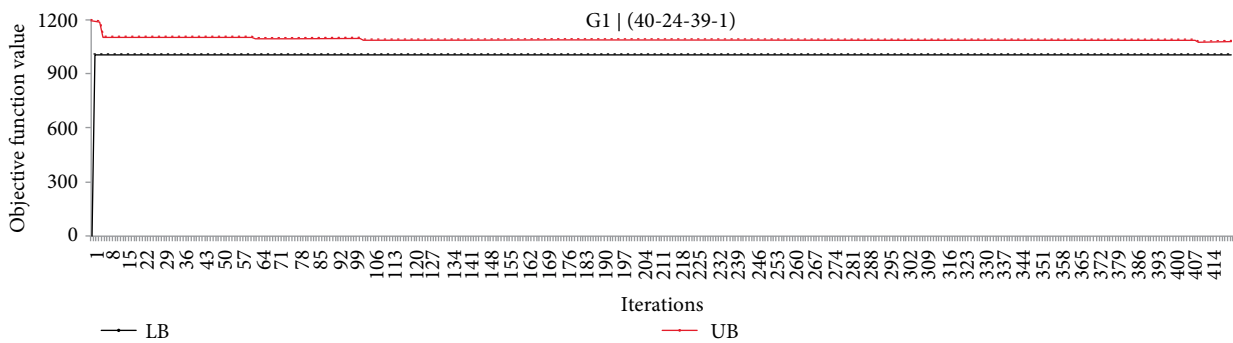


FIGURE 7: The instance G1 | (40-24-3 9-1) did not converge its LB and UB values. As a result, neither UB nor the LB computed is considered as optimal solutions.

Then, to complement the subgradient algorithm, we need to calculate a valid Lower Bound (LB) value. To do so, we aim to obtain the biggest possible LB value at each iteration of the subgradient algorithm. In this problem, a LB value may correspond to a relaxation of one of the constraints. In this sense, the constraints selected to be relaxed were the (6) ones. These constraints regard to the availability of the  $M_2$ . In other words, it means that two or more jobs cannot be processed at the same time by the  $M_2$ . Afterward, we need to compute a valid value of the  $D_{\max}$ . The  $D_{\max}$  can be evaluated as the biggest term of  $(t + p_{2j} + d_j)y_{jt}$ ,  $j \in J^2$ . It is computed following a rule we introduce next.

In order to calculate the date a client will receive its products, we need first to assign a minimum release date  $r_j$  value for each job  $j \in J^2$ , which is defined as  $r_j = \sum_{k \in S_j} p_{1k}$ . Then, we add to that value the respective processing time in the  $M_2$  ( $p_{2j}$ ). Next, we sum the delivery time required to deliver the products into the client's location ( $d_j$ ). Afterward, we sort the computed values in decreasing order. As a result, we assign the biggest value found, to the  $D_{\max}$  variable. Therefore, the reader may notice that it is equivalent to the relaxation of the constraints (6). In this sense, the value computed is a valid LB. The Figures 3 and 4 depict examples of valid UB and LB solutions, respectively.

Therefore, we did not use the complete formulation of the subproblem  $L(\lambda)$  (equation (11)) to execute the LR. We did not take into account the whole objective function of the Lagrangean dual to obtain the bounds. Instead, we considered the sequence calculated on M1. We took advantage of the LR

and the subgradient method structures to obtain valid UB values, as explained before. Moreover, the LB values were computed following a greedy rule based on the three components: (i) the minimum release date regarding each job  $j \in J^2$ , (ii) the processing time in the  $M_2$ , and (iii) the delivery time. Afterward, both UB and LB values were introduced to the subgradient algorithm to pursue its convergence. In this sense, we may refer to our approach as an adaptation of the LR method. The Figure 5 illustrates the conditions to the method that achieves the optimality.

As a result, the Lagrangean Relaxation achieves an optimal solution whenever the computed UB and the LB values are the same. Consequently, it requires that both bounds converge to an equal value to obtain the optimal solution.

#### 4. Computational Results

We report computational results on random instances, which are described as follows. The instances are divided into two groups: (i) short processing time jobs, and (ii) long processing time jobs. In the first (resp. second) group processing times are drawn from the uniform distribution between 1 and 10 (resp. 10 and 100). The number  $n$  of jobs in  $J^1$  is set to 5, 10, 20, 40, and 60, and the number  $m$  of jobs in  $J^2$  is set to  $0.6n$ ,  $0.8n$ ,  $n$ ,  $1.2n$ , and  $1.4n$ .

The set  $S_j$  of predecessors of a job  $j \in J^2$  is a random subset of  $J^1$  with cardinality  $|S_j|$  drawn from the uniform distribution

TABLE 3: Results obtained with the linear relaxation and the LR on instances of groups 1 and 2. The  $LB^*$  values refer to the lower bounds provided by the Lagrangean Relaxation. Likewise,  $LB$  values refer to the lower bounds computed by the linear relaxation.

Instance	Group 1 instances					Instance	Group 2 instances				
	Linear relaxation		Lagrangean Relaxation				Linear relaxation		Lagrangean Relaxation		
	$LB$	Time (sec)	$LB^*$	$(LB^*/LB)$	Time (sec)		$LB$	Time (sec)	$LB^*$	$(LB^*/LB)$	Time (sec)
5-3-4-1	788	0	794	1%	0	5-3-4-2	4,620	0	4,630	0%	0
5-4-4-1	666	0	671	1%	0	5-4-4-2	3,139	0	3,322	6%	0
5-5-4-1	840	0	842	0%	0	5-5-4-2	3,144	0	3,322	6%	0
5-6-4-1	795	0	801	1%	0	5-6-4-2	4,824	0	4,840	0%	0
5-7-4-1	895	0	901	1%	0	5-7-4-2	4,903	4	4,954	1%	0
10-6-9-1	932	0	939	1%	0	10-6-9-2	5,092	9	5,221	3%	0
10-8-9-1	966	0	983	2%	0	10-8-9-2	4,228	7	4,289	1%	0
10-10-9-1	985	0	996	1%	0	10-10-9-2	4,815	25	5,080	6%	0
10-12-9-1	995	0	1,001	1%	0	10-12-9-2	4,777	1	4,802	1%	0
10-14-9-1	905	0	906	0%	0	10-14-9-2	5,028	36	5,151	2%	0
20-12-19-1	909	1	926	2%	0	20-12-19-2	5,394	60	5,921	10%	1
20-16-19-1	941	2	954	1%	0	20-16-19-2	5,247	100	5,697	9%	1
20-20-19-1	1,009	4	1,048	4%	0	20-20-19-2	5,393	97	5,726	6%	2
20-24-19-1	1,044	6	1,082	4%	0	20-24-19-2	5,300	207	5,581	5%	2
20-28-19-1	1,046	5	1,100	5%	0	20-28-19-2	5,646	229	6,067	7%	3
40-24-39-1	978	10	1,006	1%	0	40-24-39-2	5,927	231	6,281	6%	6
40-32-39-1	1,066	10	1,144	7%	0	40-32-39-2	5,940	332	6,025	1%	7
40-40-39-1	1,060	15	1,092	3%	1	40-40-39-2	6,211	596	7,729	24%	12
40-48-39-1	1,064	15	1,116	5%	1	40-48-39-2	6,067	719	6,562	8%	15
40-56-39-1	1,045	26	1,069	2%	2	40-56-39-2	6,181	760	7,111	15%	17
60-36-59-1	1,135	28	1,233	9%	2	60-36-59-2	6,546	724	7,349	12%	16
60-48-59-1	1,118	40	1,225	10%	3	60-48-59-2	-	7,200	8,810	-	34
60-60-59-1	1,173	75	1,268	8%	4	60-60-59-2	-	7,200	7,757	-	43
60-72-59-1	1,174	100	1,289	10%	5	60-72-59-2	-	7,200	7,715	-	45
60-84-59-1	1,162	112	1,283	10%	7	60-84-59-2	-	7,200	7,738	-	61

between 1 and  $n - 1$ . In the Tables 1, 2, and 3, an instance is identified by  $n - m - np - g$ , where  $np$  is the maximum number of predecessors a job in  $J^2$  can have and  $g$  is the group. Regarding the delivery times, the delivery time will be calculated in terms of the uniform distribution between 100 up to 1000 and from 1000 up to 5000 for elements of the group 1 and 2, respectively. The reader may interpret each slot of time as 10 minutes in the real world. Moreover, we believe this set of data represents quite well the processes studied and its variability, as well.

Table 1 summarizes the structure of the generated instances. The first column shows the group. The second and third columns show the possible values for the number of jobs in  $J^1$  and  $J^2$ , respectively. The fourth and the fifth columns show the uniform distribution intervals from which  $|S_j|$  and the processing times are drawn, respectively. The sixth column refers to the time spent on the road to deliver the products to the final clients. The last column shows the identification, where one instance is generated for each value of  $m$  in the second column.

The experiments were carried out on the Operational System CentOS 7.x x86\_64 with 27 compute nodes, 720 cores and 7.4 TB of RAM as the maximum capacity. We highlight that we applied only one node to proceed with the calculations.

As a result, we did not use any parallel approach. Moreover, the programming languages C and C++ were used with compiler GNU GCC, and CPLEX 12.6.8. was used to solve the ILP models.

We first report the results obtained when trying to solve the ILP model running CPLEX with a time limit of 7,200 seconds. Table 2 shows results for the instances which CPLEX obtained lower and upper bounds within the time limit. Moreover, we introduced the same metrics for the Lagrangean Relaxation applied for the same instances. The results from instance group 1 are presented from the first to the ninth columns and for instance group 2 from the tenth to the eighteenth columns. For each group, we present the instance identification, the final upper (UB) and lower (LB) bounds obtained within the time limit, the percentage gap, and the time in seconds. The linear relaxation bounds are presented in the subsequent tables, along with results for the LR, as well. The percentage gap is computed as  $(UB - LB)/UB$ . The dash symbol “-” in the tables means that a method did not finish within the time limit.

Regarding the results, the mathematical model is able to obtain the optimal solution for 16 instances of group 1 and only 9 instances for group 2. Furthermore, the CPLEX was not able to provide neither a feasible UB nor a LB for the most

complicated instances of the Group 2. In contrast, the LR was able to compute valid bounds for all instances. Indeed, the LB values were verified to be the same as the optimal solutions in 21 opportunities. Moreover, the GAP values were smaller than 3% in 20 opportunities. To conclude, the maximum time spent by the LR to complete the method was 61 seconds and it was observed for the most complicated instance.

Moreover, we present two figures that illustrate the application of the subgradient algorithm when solving the Lagrangean Subproblem. The first figure represents a scenario that the optimal solution was achieved and the other figure represents a scenario that the convergence was not achieved, Figures 6 and 7, respectively.

Then, we report the results obtained with the linear relaxation of the ILP model. Also, we present the results based on the proposed LR. The time limit was set as 7,200 seconds for both experiments.

Table 3 shows the results for instance of the groups 1 and 2. Results obtained with Linear relaxation are shown from the second to the third columns and from the eighth to the ninth columns. Moreover, the LR results are shown from the fourth to the sixth columns and from tenth to the twelfth columns. The first column presents the instance, the second column the linear relaxation solution, and the third column presents the computational time in seconds. Then, we present the LR lower bound in the next column, followed by the variance between both lower bounds found, and finally the computational time in the LR in seconds. Afterward, the same structure repeats for instances of the Group 2. The dash symbol '-' in the tables means that a method did not finish within the time limit.

For both groups of instances, the results show a common pattern. The LR provided either equal or better LB result than linear relaxation for all out of the 50 instances. Also, the CPLEX was not able to compute a valid LB for the most complicated set of instances. As a result, LR outperformed the linear relaxation for all out of 46 instances. Furthermore, the LR provided valid LB for the four most complicated instances. The linear relaxation was not able to compute valid LB for those instances within 7,200 seconds computing time.

As previously mentioned, instances of group 2 were generated with longer processing times than those of group 1, c.f., Table 1. As a result, those instances present a much larger time-horizon, increasing the number of variables drastically. This fact has a significant impact on the performance of the methods.

## 5. Conclusions

In this work, we considered a two-machine flow shop scheduling problem with precedence constraints, release dates and delivery time. Moreover, the problem's objective is minimizing the time a client receives the last job. This problem is the usual case for many manufacturers' production procedures, which must check each product as soon as it is done. By doing so, the company avoids sending poor-quality products to the final client.

We also propose an adaptation of the Lagrangean Relaxation (LR) approach, which presented the best overall

results. On the one hand, the LR has obtained the optimal solution only in 3 cases out of 50. On the other hand, the LR outperformed the CPLEX for the most complicated instances. The LR was able to compute feasible solutions for all instances within 61 seconds of computing time. Even though the ILP provided the optimal solution for 26 instances, those optimal solutions were achieved only for the easier instances.

Therefore, the work presents an alternative way for companies that must schedule its activities in a flow shop fashion. Besides, the activities described in that work may be adapted for a range of others scenarios. As a result, the methodology presented is a contribution to the companies that must schedule their processes, in particular in the outbound area.

As future works, we consider developing a metaheuristic to provide better UB and LB in order to achieve better solutions for the large instances.

## Data Availability

The instances data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work has been partially supported by Doctorats Industrials, Agència de Gestió d'Ajuts Universitaria I de Recerca, Generalitat de Catalunya [2016 DI 022] (Marcelus Fabri), the Spanish Ministry of Economy and Competitiveness [TRA2013-48180-C3-2-P] (Helena Ramalhinho), and CAPES, CNPq, and FAPEMIG, Brazil (Martín Gómez Ravetti, Mauricio C. de Souza).

## References

- [1] M. L. Pinedo, "Scheduling: theory, algorithms, and systems," in *Operations Research Proceedings 1991*, W. Gaul, A. Bachem, W. Habenicht, W. Runge, and W. W. Stahl, Eds., pp. 35–42, Springer, Berlin, Heidelberg, 2016.
- [2] H. Ramalhinho, "Sevast'yanov's algorithm for the flow-shop scheduling problem," *European Journal of Operational Research*, vol. 91, no. 1, pp. 176–189, 1996.
- [3] F. Vanderbeck and L. A. Wolsey, "Reformulation and decomposition of integer programs," in *50 Years of Integer Programming 1958–2008*, M. Jünger, Th. M. Lieblich, D. Naddef et al., pp. 431–502, Springer, Berlin, Heidelberg, 2010.
- [4] P. M. Cota, B. M. R. Gimenez, D. P. M. Araújo, T. H. Nogueira, M. C. de Souza, and M. G. Ravetti, "Time-indexed formulation and polynomial time heuristic for a multi-dock truck scheduling problem in a cross-docking centre," *Computers and Industrial Engineering*, vol. 95, pp. 135–143, 2016.
- [5] J. P. Sousa and L. A. Wolsey, "A time indexed formulation of non-preemptive single machine scheduling problems," *Mathematical Programming*, vol. 54, no. 1–3, pp. 353–367, 1992.



- [6] J. M. Van Den Akker, C. A. J. Hurkens, and M. W. P. Savelsbergh, "Time-indexed formulations for machine scheduling problems: column generation," *Informs Journal on Computing*, vol. 12, no. 2, pp. 111–124, 2000.
- [7] J. K. Lenstra, A. H. G. Rinnooy Kan, and P. Brucker, "Complexity of machine scheduling problems," *Annals of Discrete Mathematics*, vol. 1, pp. 343–362, 1977.
- [8] C. Phillips, C. Stein, and J. Wein, "Minimizing average completion time in the presence of release dates," *Mathematical Programming*, vol. 82, no. 1–2, pp. 199–223, 1998.

## Research Article

# Integrated Optimization of Tram Schedule and Signal Priority at Intersections to Minimize Person Delay

Wen Zhou,<sup>1</sup> Yun Bai ,<sup>1</sup> Jiajie Li ,<sup>1</sup> Yuhe Zhou ,<sup>1</sup> and Tang Li <sup>2</sup>

<sup>1</sup>Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>Department of Civil and Environmental Engineering, Imperial College London, London, SW7 2AZ, UK

Correspondence should be addressed to Yun Bai; [yunbai@bjtu.edu.cn](mailto:yunbai@bjtu.edu.cn)

Received 14 December 2018; Revised 29 April 2019; Accepted 2 July 2019; Published 18 July 2019

Guest Editor: Helena Ramalhinho

Copyright © 2019 Wen Zhou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Modern trams, as a rapidly developed high-volume transport model, have strict requirements on schedule, because the delay will reduce the attractiveness of public transportation to passengers. To improve punctuality and reliability, Transit Signal Priority (TSP) has been employed at intersections, which can extend or insert green phase to trams. However, extending or inserting the green phase for every tram might lead to heavy delays to crossing vehicles. To address this problem, this study developed an integrated optimization model on tram schedule and signal priority which can balance the delay between trams and other vehicles to minimize person delay. Three conditional strategies named early green, green extension, and phase insertion are proposed for the signal priority. Simultaneously, arrival time, departure time of trams at stations, and stop line are optimized as well. The proposed model is tested with a numerical case and a real-world case at Ningbo tramline in China. The results indicate that the integrated optimization can reduce the average delay of all passengers on trams and other vehicles, compared to timetable optimization only and TSP only. It is also found that the proposed model is able to adapt to the fluctuation in the ratio of tram passenger to auto vehicle user, compared with only minimizing tram passenger delay or auto vehicle user delay.

## 1. Introduction

Modern trams have been developed rapidly in many cities across the world. In China, for example, tram services were provided in 14 cities by the end of 2017, which is 7 times that in 2010. In most cities, Semi-exclusive rights-of-way (ROW) have been adopted where trams have an exclusive lane in the road section and share ROW with other vehicles at the intersection. Apparently, trams might be interrupted by other vehicles or by a red signal at intersections.

To reduce the stop of trams at intersections, Transit Signal Priority (TSP) has been applied, which allows preferential treatment of transit vehicles at signalized intersections. In general, TSP can be divided into two main types: passive priority and active priority. Passive priority is to create a green wave band for transit vehicles by adjusting the sequence and shift of signaling phases among successive intersections. Ideally, trams travelling at the recommended speed can pass through all intersections without a stop. However, the

practical performance of passive priority might deteriorate due to interruptions at stations and intersections. For example, longer boarding time than the scheduled one will cause the pre-set green wave to fail, which means that the tram might hit a red signal if it still travels at the recommended speed.

Different from the passive priority, active priority can detect trams and give green phase to trams whenever they are approaching the intersection. This makes the active priority less sensitive to the disturbance of tram service. The strategies of active TSP can be further classified into unconditional and conditional. Unconditional active TSP, which gives an absolute priority to trams at intersections, is able to minimize tram travel time at the expense of imposing heavy delays to non-priority vehicles at intersections. Although unconditional TSP always provides transit vehicles with green, it inevitably causes disturbances to non-priority vehicles. On the opposite, conditional active TSP only grants priority based on a set of conditions and rules, such as whether

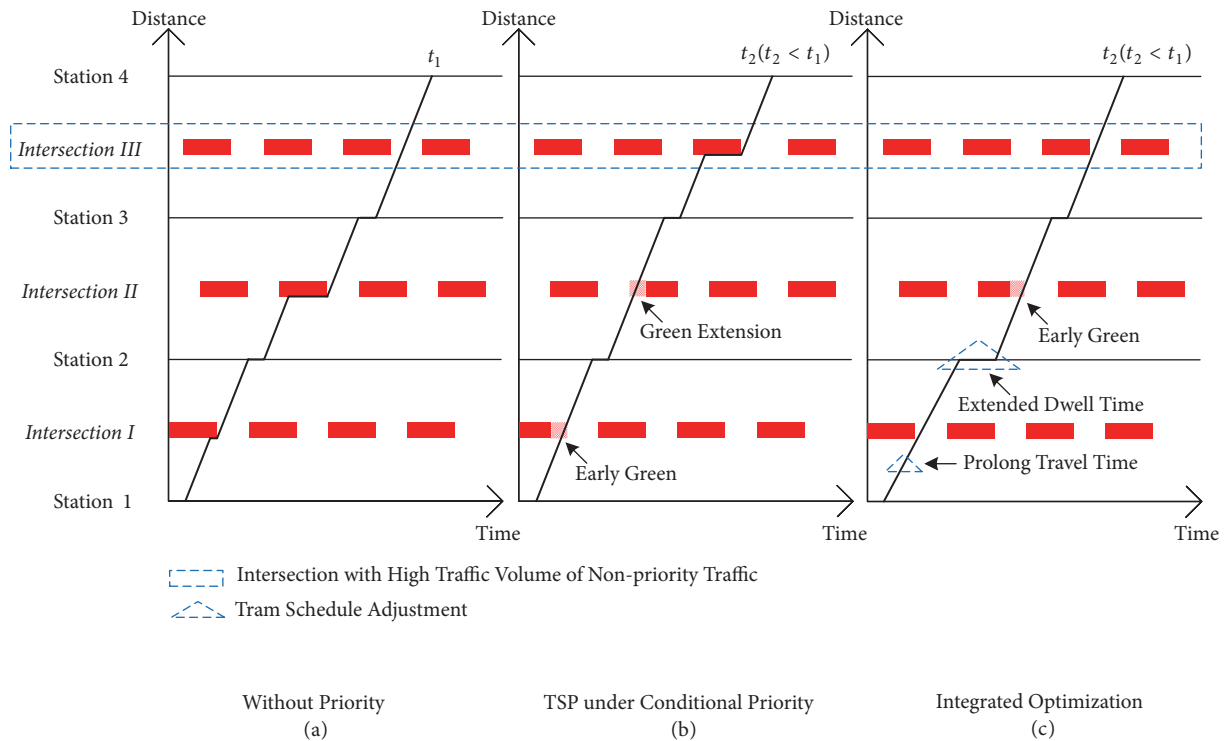


FIGURE 1: Benefits of integrating conditional active TSP strategy and tram schedule adjustment.

or not the disturbance to non-priority vehicles exceeds the threshold value and the transit vehicle is already delayed. In other words, conditional active TSP considers the trade-off between trams and other vehicles at intersections. It has been proposed to minimize the average travel time of each person on trams and other vehicles at intersections. Compared to unconditional active TSP, conditional active TSP only has minor impacts on other vehicles.

In addition to TSP, tram timetable optimization is another way to reduce the potential stop of trams at intersections. By adjusting timetable, e.g., the run-time on inter-station section and the dwell time at stations, trams can be scheduled to arrive at intersections during a green time. However, in practice the adjustment range on the timetable is limited. This is because the minimal run-times and dwell times of trams are constrained by the speed limits and the numbers of boarding and alighting passengers. The maximum run-times and dwell times are also specified to keep a reasonable service level. Therefore, tram timetable optimization is unable to ensure all the trams meeting green signal at intersections. Compared with only optimizing tram timetable or sole TSP, the integrated optimization on tram timetable and conditional active TSP might further reduce the average travel time of passengers on both trams and other vehicles. Figure 1 demonstrates the possible benefits of integrated optimization. Without TSP, as shown in Figure 1(a), the tram, operating with original timetable, meets red signals at intersections I and II, which prolongs tram travel time.

With conditional active TSP, as shown in Figure 1(b), early green and green extension are employed at intersections I and II respectively, to allow tram meeting green signals

at these two intersections where the traffic volumes of non-priority vehicles are relatively lower. At intersection III, the active TSP strategy would not be triggered due to the higher traffic volume of non-priority vehicles, which indicates that the tram with original timetable might hit red signals. Compared to the case (a) with no TSP, the average travel time of passengers on both trams and other vehicles could be reduced in case (b) where the conditional active TSP is applied.

With integrated optimization on conditional active TSP strategy and tram timetable, only early green is activated at intersection II as shown in Figure 1(c). By prolonging travel time during station 1 and intersection I, the tram can avoid red signal at intersection I without TSP, which means that other non-priority vehicles will not be interrupted. Meanwhile, the tram does not experience red signal at intersection III, which is the combined effect of prolonging the travel time during station 1 and station 2 and extending dwell time at station 2. In other words, the duration time of tram waiting at a red light at intersection III is transferred to the prolonged time during section and the extended dwell time at station 2. Thus no TSP strategy is required to activate at intersection III. Therefore, the delays imposed on non-priority vehicles are smaller than those in case (b) while the tram travel time remains, which indicates less average travel time of passengers on both trams and other vehicles.

As discussed above, this integrated optimization model of conditional active TSP and tram schedule can further reduce tram stops at intersections than previous studies which only focus on TSP or tram timetabling. At the same time, interruptions to other non-priority vehicles can be

reduced, which means that all passengers on trams and auto vehicles are considered in the proposed optimization model to achieve higher operational efficiency of the whole system.

The remainder of this paper is organized as follows. The literature review of the problem is elaborated in Section 2. Then the problem has been described in Section 3, which is followed by the formulation of the integrated optimization model on conditional active TSP and tram timetable in Section 4. The methodology to solve the proposed model is developed in Section 5. Case studies are conducted to investigate the performance of the proposed method in Section 6 followed by summarizing the conclusions of this study in Section 7.

## 2. Literature Review

TSP allows preferential treatment of transit vehicles at signalized intersections. It gives transit vehicles a little extra green time or a little less red time at traffic signals to reduce the time wasted on stop or deceleration [1]. The implementation of TSP strategies can contribute to lower delays for priority transit vehicles and auto vehicles that travel in the same directions [2, 3]. Other potential benefits include improved transit schedule reliability, increased passenger comfort and ultimately increased the attractiveness of the transit service [4–6].

Passive priority strategy and active priority are the two main types of the TSP strategies [7]. Passive priority is to create a green wave band for transit vehicles along their trips by adjusting the sequence and shift of signaling phases among different intersections [8]. MAXBAND, MULTIBAND and AM-BAND are the three classical models for signal control with passive priority. MAXBAND aims to find the optimal signaling parameters and to maximize the bandwidth of green wave [9–11]. MULTIBAND is able to expand the bandwidth of green wave attained by MAXBAND, by relaxing the constraint of equal width on green wave band at different sections [12–14]. AM-BAND further releases the symmetrical constraint on green wave band, which improves the utilization of green time and decreases stops and delays of vehicles at intersections [15]. The practical performance of passive priority might be deteriorated due to disturbance of tram service, which is not uncommon in practice.

Different from passive priority, active priority is widely used because of its flexibility. Active priority strategy grants green signal to trams by adjusting the start time and/or end time of the stage or phase giving transit vehicles right away in each signal cycle [16, 17]. The common strategies to carry out the active TSP include green extension, early green and phase insertion [18–20]. Green extension prolongs the green time when a tram arrives at intersections just behind the end of a green phase. Early green shortens the time of the phase ahead of green when a tram arrives at intersections just before the start of a green phase. Phase insertion denotes that a special green phase is inserted within the red phase. The phase insertion is generally adopted when a tram arrives at intersections at the middle of the red phase [21, 22].

The active priority can be further classified into unconditional and conditional [23]. Unconditional priority strategies

grant an absolute priority to transit vehicles once they are detected at the upstream of the intersections. There are no differential treatments between the type of transit vehicle and the state of non-priority vehicles [24–27]. Although unconditional TSP always provides transit vehicles with green phase and speeds up transit vehicles by guaranteeing no red signals at intersections, it inevitably causes disturbances to non-priority vehicles, particularly when the volume of non-priority vehicles is high [28–30]. Wahlstedt [31] indicated that TSP results in shorter travel times for buses and longer travel times for crossing traffic and traffic following the prioritized buses in one direction. Skabardonis and Christofa [32] confirmed that, under high traffic volume conditions, the provision of transit signal priority could deteriorate the Level of Service (LOS) on cross-streets by up to two levels.

The conditional priority strategies based on a set of conditions and rules have also been developed to reduce disturbances to non-priority vehicles [23]. Conditional TSP balances the benefits to trams and the disruptions to non-priority traffic and is only granted for trams if the disruption is beyond a certain range. Altun and Furth [33] indicated that, compared to unconditional active TSP, conditional active TSP resulted in only minor impacts on non-priority traffic. Furth and Muller [34] proposed a conditional priority method for buses and the results show that, compared to no priority, absolute priority caused severe increases in delay, while conditional priority had almost no impact. Similarly, Kenny and Amer [35] showed that the proposed conditional control policy could effectively reduce the transit headway deviation and causes smaller disturbance to cross street traffic compared with the existing unconditional transit signal priority algorithm.

Additionally, tram schedule optimization can also decrease stops of trams at intersections. The traditional scheduling problem is to define a series of trips with prescribed starting and ending times, with the objective of minimizing costs, including capital costs for required buses and operating costs based on traveling distance and idle time [36]. The majority of existing studies focus on railway and subway system, while scheduling problems in the tram system are quite limited in the literature. Nachtigall and Voget [37] developed a timetabling model for the railway system, which is to minimize the passenger waiting time at stations. Wong and Ho [38] presented the application of a dynamic programming approach, with the aid of an event-based model, to devise an optimal set of dwell times and run times for trains under given operational constraints over a regional level. Niu and Zhou [39] applied a timetable optimization approach for the urban rail system. The objective was to minimize passengers' waiting time at stops and also reduce the waiting time passengers who were not able to board their desired service suffered because of congestions. Robenek et al. [40] highlighted the consideration of passenger satisfaction in the design of train timetables. Zhou et al. [41] developed an integrated optimization model on train control and timetable to minimize the net energy consumption. However, tram scheduling problems cannot be solved by these present models. The reason is that the operating environment of the

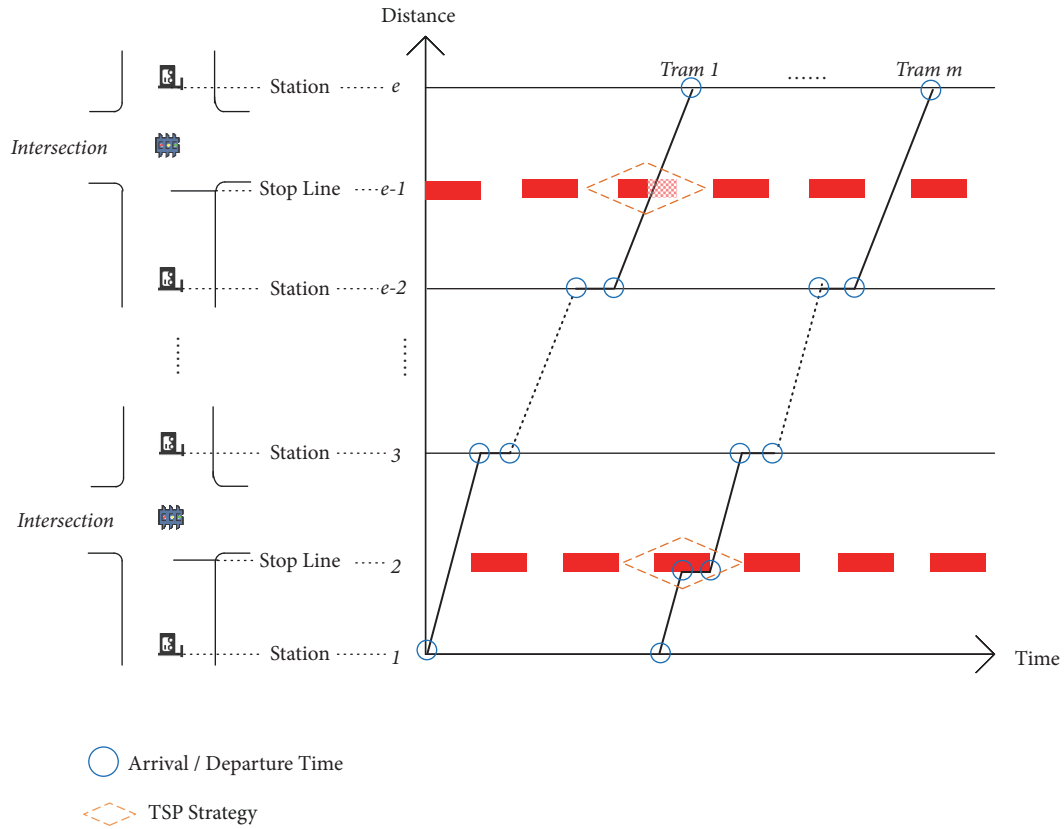


FIGURE 2: Integrated optimization on signal priority and tram scheme.

tram is different from that of the railway and subway system. Train operation would not be influenced by other traffic along its rail track, while trams share intersections with other traffic and tram operations are inevitably affected at the intersection.

Conditional active TSP and tram schedule adjustments still have their limitations when they are applied separately. Therefore by integrating conditional active TSP and tram schedule adjustment, fewer stops of tram at intersections than they are applied separately. Recently, Shi et al. [42] have explored at the planning level the benefits of coordinating tram movements and signal timings at controlled intersections. Its objective is to minimize the weighted sum of the total tram travel time and TSP's negative impacts on other traffic. However, Shi et al. [42] applied unconditional TSP and give trams absolute priority, which means that the negative impacts on other traffic are not the minimum. The disturbances to other traffic can be reduced without sacrificing the total tram travel time if conditional TSP is applied. Moreover, Shi et al. [42] only considered tram operations and ignored the operations of auto vehicles, which cannot maximize the operational efficiency of the whole intersections.

This paper aims to propose an integrated optimization model on conditional active TSP and tram schedule adjustments to minimize person delay of the whole system referring to all trams and non-priority vehicles along intersections. By applying conditional active TSP, minor impacts are imposed on other traffic compared to the unconditional one.

Considering each person on trams and auto vehicles rather than only one transportation mode can contribute to the higher operational efficiency of the whole intersections. TSP disturbance is measured as auto vehicle delay caused by TSP activation, which is more adaptable to changes in signal timing and traffic volume.

### 3. Problem Description

With semi-exclusive ROW, trams will not be disturbed at any road section between successive intersections. They only stop at the station for passenger boarding and alighting or stop at intersection stop line to wait for a red signal. Therefore, stop lines at intersections and stations along the tramline are considered as the critical nodes ( $1 \dots e$ ) for TSP strategy and timetable optimization, as shown in Figure 2. This paper integrally optimizes TSP strategy at each intersection and tram schedule along the tramline to minimize average person delay. The conditional active TSP strategies at intersections, the signal offsets among different intersections, and tram schedule are the decision variables to be optimized in the discussed problem.

The proposed conditional active TSP strategies consist of three parts: (1) whether TSP measures are activated for the incoming tram, (2) which specific TSP measure is applied for the tram, and (3) how long the TSP measure will continue. These three parts are defined by three variables which are  $\mu_{sk}$ ,  $\mu_{sk}^w$  and  $t_{skNj}^w$ . The first part is to estimate whether the

TABLE 1: Decision variables.

$t_{me}^A$	Arrival time of tram $m$ at node $e$
$t_{me}^D$	Departure time of tram $m$ from node $e$
$t_{ik}^{GS}$	The original green start of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$t_{ik}^{GE}$	The original green end of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$\mu_{ik}^T$	1, if a TSP measure is taken in cycle $k$ at intersection $i$ ; 0, otherwise
$\mu_{ik}^W$	1, if green extension/phase insertion/early green is taken in cycle $k$ at intersection $i$ ; 0, otherwise
$t_{ik}^W$	Extra green time for a TSP action in cycle $k$ at intersection $i$
$t_{ik\alpha}^E$	The end time of green extension $\alpha$ in cycle $k$ at intersection $i$
$t_{ik\beta}^S$	The start time of phase insertion $\beta$ in cycle $k$ at intersection $i$
$t_{ik\beta}^E$	The end time of phase insertion $\beta$ in cycle $k$ at intersection $i$
$t_{iky}^S$	The start time of early green $\gamma$ in cycle $k$ at intersection $i$

arrival time of the tram is during green phase. If tram arrives at intersections during green phase, no TSP is activated. Otherwise, TSP is activated depending on the efficiency of the whole system. In other words, no TSP is activated even trams hit red phase at intersections if the activation causes too much delay to auto vehicles and results in lower efficiency of the whole system. The second part is to decide whether to implement green extension, phase insertion, or early green. The decision is made according to the arrival time of the tram at intersection. If the tram will arrive at intersection during the beginning of the red phase, green extension is applied. If the tram will arrive during the end of the red phase, early green is applied. If the tram will arrive at the middle of red phase, phase insertion is applied. In the third part, the duration of the specific TSP measure is specified based on the minimal person delay.

The decision variables with respect to tram schedules could be denoted by tram arrival and departure time at each node denoted as  $t_{me}^A$  and  $t_{me}^D$ . If a tram arrives at an intersection during a green phase or TSP measures are activated when a tram hits the red signal at intersections, the arrival time is equal to the departure time. As for tram 1 in Figure 2, its arrival time is equal to departure time at node 2 and node  $e-1$ . If trams arrive at intersections during red phase with no TSP measure activation, trams depart from intersections at the beginning of the next green phase.

The average person delay of the whole system involves both auto vehicle user delay and tram passenger delay. Auto vehicle delay is calculated based on a classical vehicle delay model proposed by the Australian Road Research Board (ARRB) [43]. The ARRB model is applicable to both under-saturated and over-saturated conditions [44]. Tram delay is defined as the difference between the actual travel time and the minimum travel time, which is the time that trams run at the highest speed and do not experience any red phase at intersections.

In order to formulate the integrated optimization of signal priority and tram schedule, the following assumptions are made throughout this paper. Tram is the only public transit mode along the road. All intersections are equipped with TSP control devices. Similar to the metro timetable optimization in Niu et al. [39] and Yang et al. [45], only schedules for trams heading in one direction are considered. For tramlines,

tram headway is rather long and there is little chance that two trams may meet at the same intersections. Thus trams in the other direction can be optimized in the same way.

## 4. Mathematical Model

4.1. *Notations.* See Tables 1 and 2.

4.2. *Objective Function.* To improve punctuality and reliability of trams, TSP strategies have been employed at intersections, which might lead to heavy delays to auto vehicles. To address this problem, tram schedules and TSP strategies are integrally optimized to balance the delay between trams and other vehicles. The proposed model aims to increase the efficiency of the whole system considering both tram passengers and auto vehicle users. Therefore, the weighted average delay of both auto vehicle users and tram passengers is chosen as the objective of the proposed model to be minimized, as shown in Eq. (1). In this study, a higher weight is assigned to tram passengers. The reason is that public transit should be granted more priorities to increase the attractiveness to passengers, thereby reducing traffic congestion.

$$\min \frac{\sum_{i,k} \xi_a Q_i \bar{D}_{iknj} + \eta \cdot \sum_{i,s} \xi_t \sigma_{s,s+1}^m |t_{mi}^A - t_{mi}^{A'}|}{\sum_{i,k} \xi_a Q_i + \sum_{m,s} \xi_t \bar{\sigma}^m} \quad (1)$$

The weighted factor  $\eta$  decides the degree of priority assigned to tram passengers, which should be greater than one.

4.3. *Auto Vehicle Delay.* Auto vehicle delay model is the primary tool to analyze vehicle delay for signalized intersections. Four auto vehicle delay models are widely used to analyze vehicle delay for signalized intersections, including the Webster model, the ARRB model, the HCM1985 model, and the HCM2000 model. Yao et al. [46] analyze the applicability of four vehicle delay models for signalized intersections and design experimental environments on the basis of certain traffic demands and signal control. The results shown that, for isolated signalized intersections, the ARRB model displays the best performance in vehicle delay estimation. Thus the ARRB model is applied to calculate auto vehicle delay in this paper.



TABLE 2: Notations and parameters.

General notations	
$e$	Index of nodes, including stations and intersections
$i$	Index of intersections
$s$	Index of stations
$m$	Index of trams
$w$	Index of TSP actions, where green extension, phase insertion and early green are denoted as $\alpha, \beta, \gamma$ respectively
$j$	Index of approach lane direction, where through, left-turning and right-turning are denoted as $z, l, r$ , respectively.
Tram parameters	
$t_{me}^A$	Arrival time of tram $m$ at node $e$ based on the minimal travel time schedule
$t_{me}^D$	Departure time of tram $m$ from node $e$ based on the minimal travel time schedule
$\rho_{e,e+1}^{\min}$	Minimum running time for each tram from node $e$ to node $e+1$
$\rho_{e,e+1}^{\max}$	Maximum running time for each tram from node $e$ to node $e+1$
$\theta_s^{\min}$	Minimum dwell time for each tram at station $s$
$\theta_s^{\max}$	Maximum dwell time for each tram at station $s$
$h_a$	Minimum headway between two consecutive trams arrivals at the same station
$h_d$	Minimum headway between two consecutive trams departures at the same station
$h_a^d$	Minimum headway between a tram departure and another tram arrival at the same station
$\sigma_{s,s+1}^m$	Load factor of tram $m$ between two consecutive stations
$\bar{\sigma}^i$	Average load factor of tram $m$ along the tramline
$\xi_m$	Fixed passenger occupancy of tram $m$
$\eta$	Weight factor determining tradeoff of person delay between tram and auto vehicles
Signal parameters	
$k$	Index of signal cycle at intersection $i$
$n$	Index of phase in cycle $k$ for intersection $i$
$N$	The phase serving tram, which is also called critical phase
$\tau_{n=N}$	1, if phase $n$ is the critical phase $N$ ; 0, otherwise
$\tau_{n=N-1}$	1, if phase $n$ is the previous phase before critical phase $N$ ; 0, otherwise
$C_{ik}^l$	The unadjusted cycle length in cycle $k$ at intersection $i$
$g_{iknj}^l$	The unadjusted green duration of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$\lambda_{iknj}^W$	Green ratio of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$\delta_{\min}^W$	Minimum green duration for TSP actions
$\delta_{\max}^W$	Maximum green duration for TSP actions
$\varphi_{\min}$	Minimum red time between two green phases in a signal cycle
$\mu_{ik}$	1, if a TSP action is taken in cycle $k$ at intersection $i$ ; 0, otherwise
Traffic parameters	
$D_{iknj}$	Average vehicle delay of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$\bar{D}_{ik}$	Average vehicle delay in cycle $k$ at intersection $i$
$q_{iknj}^l$	Original traffic volume of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$\Delta q_{iknj}^W$	Variation of traffic volume caused by applying TSP strategy
$x_{iknj}$	Saturation of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$F_{iknj}$	Saturation volume of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$c_{iknj}$	Capacity of lane group $j$ during phase $n$ in cycle $k$ at intersection $i$
$\omega_{iknj}^{mA}$	1, if tram $m$ arrives at intersection $i$ during phase $n$ in cycle $k$ ; 0, otherwise
$Q_{ik}$	Auto traffic volume of intersection $i$ in cycle $k$
$\xi_a$	Passenger occupancy of an auto vehicle

The ARRB model is formulated from Eq. (2) to Eq. (4). Eq. (2) defines auto vehicle delay of a lane or a lane group. Eq. (3) specifies that flow ratio is calculated based on traffic

volume and saturation flow. Eq. (4) describes the saturation calculation when average overflow queue is approximately zero.

$$D_{iknj} = \begin{cases} \frac{C_k (1 - \lambda_{iknj})^2}{2(1 - y_{iknj})} + \frac{c_{iknj} x_{iknj} T}{4q_{iknj}} \left[ (x_{iknj} - 1) + \sqrt{(x_{iknj} - 1)^2 + \frac{12(x_{iknj} - x_{0iknj})}{c_{iknj} T}} \right], & x_{iknj} > x_{0iknj} \\ \frac{C_{ik} (1 - \lambda_{iknj})^2}{2(1 - y_{iknj})}, & x_{iknj} \leq x_{0iknj} \end{cases} \quad (2)$$

$$y_{iknj} = \frac{q_{iknj}}{F_{iknj}} \quad (3)$$

$$x_{0iknj} = 0.67 + \frac{F_{iknj} g_{iknj}}{600} \quad (4)$$

Vehicle delay at one intersection can be calculated by the aggregated vehicle delay in each lane or lane group, as shown in Eq. (5).

$$\bar{D}_{iknj} = \frac{\sum_{n,j} D_{iknj} q_{iknj}}{\sum_{n,j} q_{iknj}} \quad (5)$$

Green ratio, capacity, and saturation are all calculated from green duration, cycle length, traffic volume, and saturation flow. Especially, the value of saturation flow only depends on intersection infrastructure, which means that the value of saturation flow in one specific lane is constant. Therefore, the ARRB model depends on three variables, which are green duration, cycle length, and traffic volume. Thus auto vehicle delay of an intersection is simplified as Eq. (6).

$$\bar{D}_{iknj} = d_a^r (g_{iknj}, C_{ik}, q_{iknj}) \quad (6)$$

With given green duration, cycle length, and traffic volume at different intersections, auto vehicle delay of each intersection can be calculated.

*Scenario 1: Green Extension Activation.* Eq. (7) to Eq. (10) describe the change of green duration, cycle length, and traffic volume when the green extension is activated. Each lane group  $j$  during phase  $n$  in cycle  $k$  at intersection  $i$  is the calculation object in this paper. The phase served tram is considered as critical phase, which is denoted as  $N$ . Green extension strategy is applied only when the tram arrives at intersections during phase  $N+1$ , which is the next phase after the critical phase  $N$ .

$$\tau_{n=N} = \begin{cases} 1, & \text{if } n = N \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$g_{sknj} = g'_{sknj} + \tau_{n=1} t_{skNj}^\alpha \quad (8)$$

$$C_{sk} = C'_{sk} + t_{skNj}^\alpha \quad (9)$$

$$q_{sknj} = q'_{sknj} + \tau_{n=1} \Delta q_{skNj}^\alpha \quad (10)$$

Eq. (7) identifies whether the current phase is in the critical phase  $N$  or not.

Eq. (8) describes the change of green duration when the green extension is activated. Once the green extension strategy is activated, green duration of the critical phase  $N$  is extended by  $t_{skNj}^\alpha$  and green duration of other phases is still the same.

Eq. (9) describes the change of cycle length when the green extension is activated. Because the green duration of critical phase  $N$  is increased by  $t_{skNj}^\alpha$  and other phases keep the same, the cycle length of cycle  $k$  is increased by  $t_{skNj}^\alpha$  as well.

Eq. (10) describes the change of traffic volume when the green extension is activated. Green extension strategy leads to the green duration of phase  $N$  increasing. Meanwhile, traffic volume increases by  $\Delta q_{skNj}^\alpha$ .

Above all, auto vehicle delay of one intersection when the green extension is activated is simplified as Eq. (13).

$$\bar{D}_{sknj}^\alpha = d_a^r (g'_{sknj} + \tau_{n=1} t_{skNj}^\alpha, C'_{sk} + t_{skNj}^\alpha, q'_{sknj} + \tau_{n=1} \Delta q_{skNj}^\alpha) \quad (11)$$

*Scenario 2: Early Green Activation.* Eq. (12) to Eq. (15) describe the change of green duration, cycle length, and traffic volume when the early green strategy is activated. Early green strategy is applied only when the tram arrives at intersections during phase  $N-1$ , which is the last phase before the critical phase  $N$ .

$$\tau_{n=N-1} = \begin{cases} 1, & \text{if } n = N - 1 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$g_{sknj} = g'_{sknj} + (\tau_{n=N} - \tau_{n=N-1}) t_{skNj}^\alpha \quad (13)$$

$$C_{sk} = C'_{sk} \quad (14)$$

$$q_{sknj} = q'_{sknj} + (\tau_{n=N} - \tau_{n=N-1}) \Delta q'_{sknj} \quad (15)$$

Eq. (12) identifies whether the current phase is in the phase  $N-1$  or not.

Eq. (13) describes the change of green duration when the early green strategy is activated. Early green is also called red truncation, which means the red phase is reduced. Once the



early green strategy is activated, the green duration of phase  $N-1$  is shortened by  $t_{skNj}^y$  and green duration of other phases is still the same.

Eq. (14) describes that cycle length does not change when the early green strategy is activated. The extra green duration for phase  $N$  is at the cost of reducing the green duration of phase  $N-1$ . Thus the cycle length is the same as the original one.

Eq. (15) describes the change of traffic volume when the green extension is activated. The green duration of phase  $N-1$  is shortened by  $t_{skNj}^y$  and traffic volume is reduced by  $\Delta q_{sk,N-1,j}^y$ . In terms of the residual phases, traffic volume is kept the same.

Above all, auto vehicle delay of one intersection when the early green is activated is simplified as Eq. (16).

$$\begin{aligned} \bar{D}_{sknj}^y &= d_a^r \left( g'_{sknj} + (\tau_{n=N} - \tau_{n=N-1}) t_{skNj}^y, C'_{sk}, q'_{sknj} \right. \\ &\quad \left. + (\tau_{n=N} - \tau_{n=N-1}) \Delta q_{sknj}^y \right) \end{aligned} \quad (16)$$

*Scenario 3: Phase Insertion Activation.* Eq. (17) to Eq. (25) describe the change of green duration, cycle length, and traffic volume when the phase insertion strategy is activated. Once phase insertion is active, an extra phase is granted for trams denoted as  $N^1$ . The extra phase  $N^1$  is inserted to another phase  $n^T$  according to tram arrival time at intersections. Phase  $n^T$  is interrupted and thus divided into two phases  $n^1$  and  $n^2$ . The residual phases do not get any influence which are denoted as  $n^0$ .

$$\tau_{n=n^0} = \begin{cases} 1, & \text{if } n = n^0 \\ 0, & \text{otherwise} \end{cases} \quad (17)$$

$$\tau_{n=N^1} = \begin{cases} 1, & \text{if } n = N^1 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

$$\tau_{n=n^1} = \begin{cases} 1, & \text{if } n = n^1 \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

$$\tau_{n=n^2} = \begin{cases} 1, & \text{if } n = n^2 \\ 0, & \text{otherwise} \end{cases} \quad (20)$$

$$\begin{aligned} g_{sknj} &= \tau_{n=n^0} g'_{skn^0j} + \tau_{n=N^1} t_{skNj}^\beta \\ &\quad + \tau_{n=n^1} (t_{sknj}^{GE} - t_{skNj}^S) \\ &\quad + \tau_{n=n^2} (t_{sknj}^{GE} - t_{skNj}^S) \end{aligned} \quad (21)$$

$$C_{sk} = C'_{sk} + t_{skNj}^\beta \quad (22)$$

$$\begin{aligned} q_{sknj} &= \tau_{n=n^0} q'_{skn^0j} + \tau_{n=N^1} \Delta q_{skN^1j}^\beta \\ &\quad + \tau_{n=n^1} q_{skn^1j}^\beta + \tau_{n=n^2} q_{skn^2j}^\beta \end{aligned} \quad (23)$$

$$q_{skn^1j}^\beta + q_{skn^2j}^\beta < q'_{sknj} \quad (24)$$

Eq. (17) - (19) identify the category of the current phase.

Eq. (20) describes the change of green duration when the phase insertion strategy is activated. For unaffected phases, the green duration is the same as the original ones. For the extra phase  $N^1$ , the starting time and the end time are denoted as  $t_{skNj}^S$  and  $t_{skNj}^E$ . Obviously  $t_{skNj}^\beta = t_{skNj}^E - t_{skNj}^S$  is the green duration of the extra phase  $N^1$ . For the two divided phases, the total green duration of the two phases is still the same.

Eq. (21) describes the change of cycle length when the phase insertion strategy is activated. The cycle length is increased by the green duration of the extra phase  $N^1$ , which is  $t_{skNj}^\beta$ .

Eq. (22) describes the change of traffic volume when green extension is activated. For unaffected phases, the traffic volume is the same as the original one. For extra phase  $N^1$ , the traffic volume is based on the green duration of phase  $N^1$ . The traffic volume during phase  $N^1$  is denoted as  $\Delta q_{skN^1j}^\beta$ . For the two divided phases  $n^1$  and  $n^2$ , traffic volume is not the same as before.

As described in Eq. (24), the sum of traffic volume during phases  $n^1$  and  $n^2$  is smaller than the traffic volume during the former intact phase  $n^T$ . Because the continuous traffic flow is disrupted by the inserted phase, traffic flow is shortened.

Above all, auto vehicle delay of one intersection when the phase insertion strategy is activated is simplified as Eq. (25).

$$\bar{D}_{sknj}^\beta = d_a^r (g_{sknj}, C_{sk}, q_{sknj}) \quad (25)$$

#### 4.4. Constraints

*4.4.1. Signal Constraints.* Signal constraints deal with signal variables to address the TSP strategy and the signal timing of each intersection.

(1) *Constraint of Green Extension.* Once the green extension action is activated in cycle  $k$  at intersection  $s$ , the extended green time is between  $\delta_{\min}^\alpha$  and  $\delta_{\max}^\alpha$ .

$$\mu_{ik}^\alpha \delta_{\min}^\alpha \leq t_{ikNj}^\alpha \leq \mu_{ik}^\alpha \delta_{\max}^\alpha \quad (26)$$

Green extension action is applied only when trams arrive during phase  $N+1$ .

$$\mu_{ikNj}^\alpha = \mu_{ikNj} \cdot \omega_{ik,N+1,j}^{mA} \quad (27)$$

(2) *Constraint of Phase Insertion.* Once the phase insertion action is applied in cycle  $k$  at intersection  $s$ , the green duration of the insertion phase is between  $\delta_{\min}^\beta$  and  $\delta_{\max}^\beta$ .

$$\mu_{ik}^\beta \delta_{\min}^\beta \leq t_{ikNj}^\beta \leq \mu_{ik}^\beta \delta_{\max}^\beta \quad (28)$$

TABLE 3: Numbers of variables and constraints in the model.

Type	Name	Theoretical dimension
Continuous variable	$t_{me}^A, t_{me}^D$	$ M  \cdot  E $
	$t_{ik}^W$	$ W  \cdot  I  \cdot  K $
Binary variable	$t_{ik}^{GS}, t_{ik}^{GE}$	$ I  \cdot  K $
	$t_{ik\alpha}^E, t_{ik\beta}^S, t_{ik\beta}^E, t_{iky}^S$	$ I  \cdot  K $
	$\mu_{ik}^T, \mu_{ik}^W$	$ W  \cdot  I  \cdot  K $
	Constraint (26)-(33)	$ I  \cdot  K $
Constraints	Constraint (34)	$ M  \cdot  S $
	Constraint (35)	$ M  \cdot  E  - 1$
	Constraint (36)-(38)	$ M  - 1 \cdot  E $

(3) *Constraint of Early Green.* Once early green action is activated in cycle  $k$  at intersection  $s$ , the advanced green time is between  $\delta_{\min}^y$  and  $\delta_{\max}^y$ .

$$\mu_{ik}^y \delta_{\min}^y \leq t_{ikNj}^y \leq \mu_{ik}^y \delta_{\max}^y \quad (29)$$

Early green action is applied only when trams arrive during phase  $N-1$ .

$$\mu_{ikNj}^y = \mu_{ikNj} \cdot \omega_{ik,N-1,j}^{mA} \quad (30)$$

(4) *The Minimum Time Gap between Two Green Phases.* The starting time of the inserted phase  $\beta$  is at least  $\varphi_{\min}$  later than the ending of the previous green phase.

$$t_{ikNj\beta}^S - t_{ikNj}^{GE} \geq \varphi_{\min} \quad (31)$$

The end time of the inserted phase  $\beta$  is at least  $\varphi_{\min}$  earlier than the start of the next green phase.

$$t_{ikNj\beta}^E - t_{ik+1,Nj}^{GS} \geq \varphi_{\min} \quad (32)$$

(5) *Activate a TSP Action One Time in a Cycle.* The following relation holds in practice, which means no more than one TSP action can be applied in one signal cycle at an intersection.

$$\sum_w \mu_{ik}^w \leq 1 \quad (33)$$

4.4.2. *Tram Schedule Constraints.* Tram schedule constraints deal with tram variables to address the arrival and departure time of tram at each station.

(1) *Dwell Time.* A feasible range for dwell time is enforced as follows:

$$\theta_s^{\min} \leq t_{ms}^D - t_{ms}^A \leq \theta_s^{\max} \quad (34)$$

(2) *Link Travel Time.* The travel time between two nodes is also adjustable:

$$\rho_{e,e+1}^{\max} \leq t_{m,e+1}^A - t_{me}^D \leq \rho_{e,e+1}^{\min} \quad (35)$$

(3) *Headway.* The minimum headway between arrivals of trams  $m$  and  $m+1$  at station  $s$  is described as follows.

$$t_{m+1,s}^A - t_{ms}^A \geq h_a \quad (36)$$

The minimum headway between departures of trams  $m$  and  $m+1$  at station  $s$  is described as follows.

$$t_{m+1,s}^D - t_{ms}^D \geq h_a \quad (37)$$

The minimum headway between the departure of trams  $m$  and the arrival of tram  $m+1$  is described as follows.

$$t_{m+1,e}^A - t_{me}^D \geq h_a^d \quad (38)$$

## 5. Methodology

5.1. *Complexity of Formulation.* According to the optimization model formulated in Section 4, all the variables can be classified into two types. The first type refers to continuous decision variables including arrival time and departure time of trams, the duration of TSP strategies, and the signal offsets among different intersections. The second type is binary decision variables including whether or not TSP measures are activated for the incoming tram and which specific TSP measure is applied for the tram. It is obvious that the formulated model in Section 4 is a non-linear programming model. The numbers of variables and critical constraints of the problem are given in Table 3.

The complexity of the formulated model is mainly affected by the number of trams — $M$ —, the number of intersections — $I$ —, and the number of stations — $S$ —. According to Table 3, the quantity of model variables is 89 when one tram passes by one intersection and one station. However, the total number of variables becomes  $89^2$  when another intersection and another station are taken into account. As the quantity of variables exponentially increases with the scale of tramlines, enumeration solution might be unable to find the optimal solution within limited computing time. Therefore, GA is adopted to solve the proposed model for its high efficiency, because the formulated problem turns out to be a large scale problem in practice.

5.2. *The Solution Method.* As stated above, the formulated problem turns out to be a large scale problem that the

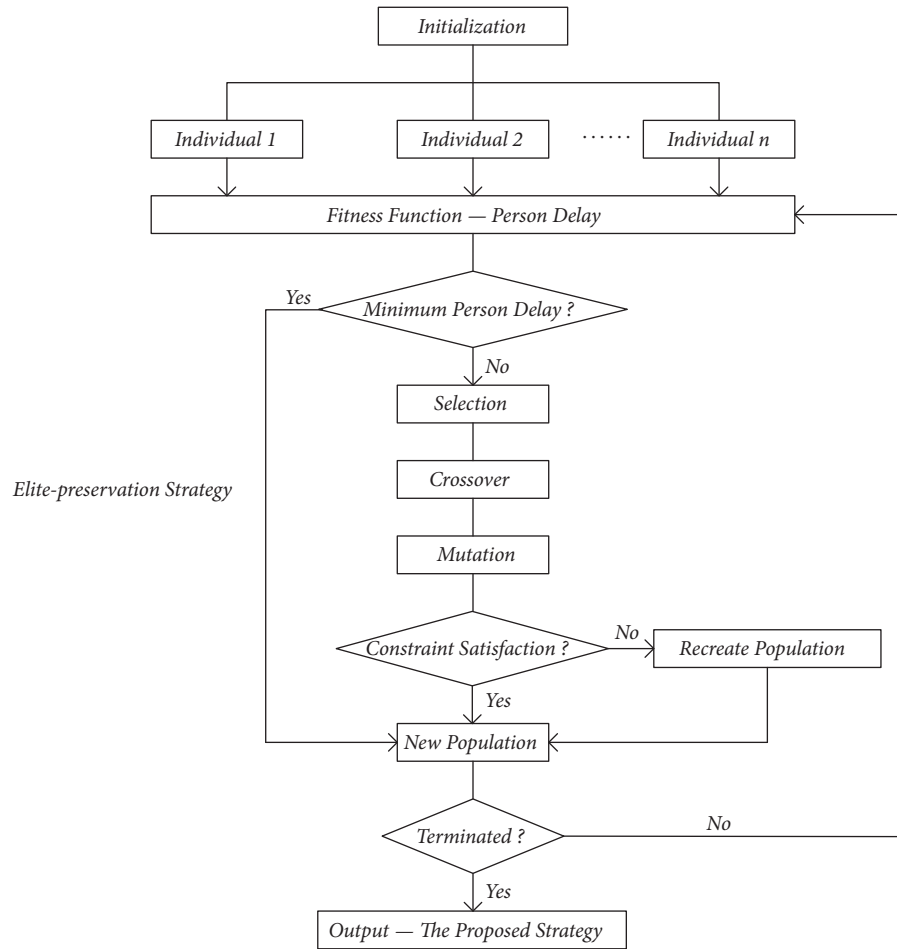


FIGURE 3: Flowchart of GA algorithm.

enumeration solution cannot address it effectively [47]. Therefore an intelligent search algorithm, Genetic Algorithm (GA), is applied to solve the model. The GA is a powerful multi-objective evolutionary algorithm which is competent to quickly find satisfied solutions of large scale problems. Figure 3 gives the flowchart of employing the GA to solve the formulated model.

The pseudocode of the GA is presented in Algorithm 1. The inputs include the number of nodes, the distance between each node and constraints. The output is the solution to the proposed problem, which consists of arrival and departure time of trams, unadjusted signal timing, and TSP strategies. The first step of the proposed GA is chromosome encoding, as shown in Figure 4. Real-number encoding is used to represent the chromosomes. The second step is to initialize populations. Random initialization within a certain range has been applied, where each chromosome is initialized based on the constraints of the proposed model to make sure it is a feasible solution. Thirdly, taking person delay as an objective, the delay of each chromosome will be calculated. Then the chromosome with the minimal person delay is selected into the new population directly, which is called the elite-preservation strategy. After that, the remaining chromosomes

are input to the process of crossover and mutation. Each individual is checked again whether they are satisfied with constraints after crossover and mutation. The individual with the minimal average person delay is the last generation and is the attained solution of the proposed model. The evolution of GA populations stops when the pre-determined maximum generation is reached.

**5.2.1. Chromosome Encoding and Initialization.** At the beginning stage of GA, an initial parent population is generated based on the practical timetable configurations. Real number coding method is adopted to represent the chromosomes of individuals. The tramline covers  $|I|$  intersections and  $|S|$  stations, which involves  $|E|$  critical nodes. There are  $|M|$  trams scheduled in the given time  $|K|$ . Each chromosome can be mainly divided into two parts: tram schedule and signal timing, as shown in Figure 4. The length of each gene part is equal to the quantity of its corresponding variables.

**5.2.2. Selection, Crossover, and Mutation.** Selection, crossover, and mutation are the main genetic operators to generate offspring. In this study, the solutions in the parent population are selected by spinning the roulette wheel to produce

```

(1) Input: Tramline Distance ( $TD$ )
(2)   Quantity of Trams, Intersections and Stations ( $QM, QI, QS$ )
(3)   Quantity of Tram Passengers and Auto Vehicle Users ( $QTP, QAV$ )
(4)   Constraints ( $C$ )
(5) Output: Tram Arrival Time ( $AT$ )
(6)   Tram Departure Time ( $DT$ )
(7)   Unadjusted Signal Timings ( $ST$ )
(8)   TSP strategies ( $TSP$ )
(3)  $initialPopulation \leftarrow$  Initialization ( $TD, QM, QI, QS, QTP, QAV, C$ )
(4)  $newPopulation \leftarrow \emptyset$ 
(5) repeat
(6)   for each  $individual \in initialPopulation$  do
(7)     if  $fitnessCalculation (individual) =$  minimum person delay;
(8)        $newPopulation \leftarrow newPopulation \cup individual$ ;
(9)     else
(10)       $currentPopulation \leftarrow initialPopulation$ ;
(11)    end
(12)  end
(13)   $selectedCouples \leftarrow$  selection ( $currentPopulation$ );
(15)   $crossedCouples \leftarrow$  crossover ( $selectedCouples$ );
(17)   $offspring \leftarrow$  mutation ( $crossedCouples$ )
(18)   $newPopulation \leftarrow newPopulation \cup offspring$ ;
(19) until termination criterion

```

ALGORITHM 1: A Genetic Algorithm to solve the proposed model.

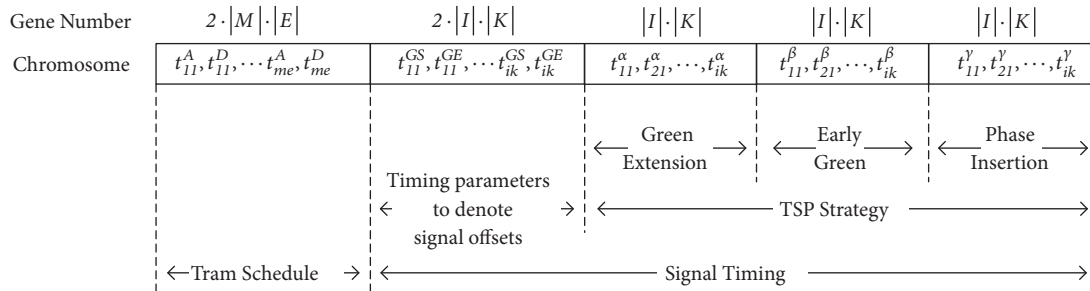


FIGURE 4: Chromosome encoding.

offspring. The crossover operator employed in this study only occurs amongst the same group of genes as illustrated by the Figure 5. The mutation operator modifies the value of a randomly designated gene in a chromosome with a predetermined probability.

## 6. Case Study

**6.1. Numerical Case.** Four trams running on a tramline covering two stations and two intersections, as shown in Figure 6, are employed here to demonstrate the performance of the proposed approach. Traffic volume, signal timings, lane direction, and the number of the lanes of the two intersections are specified in Table 4.

**6.1.1. Performance Comparison with and without Integration.** Figure 7 shows the schedules of four trams during one hour in detail. When the second tram arrives at intersection  $I$ , green

extension strategy is triggered to grant the tram with green signal. Shown by the third and fourth trams, trams can arrive at the intersection during green signal, by increasing the travel time between the stations and the intersections. Thus it reduces the possibility of triggering TSP strategy, which causes delay to auto vehicles.

In order to further demonstrate the performance of the integrated model, the computational results with and without the integration of conditional active TSP and tram schedule are compared. In detail, without integration can be divided into only optimizing tram schedule and only applying TSP strategy which is subdivided into unconditional TSP and conditional TSP.

Compared with the integrated model, as shown in Table 5, only optimizing tram schedule leads to higher tram passenger delay and person delay. Auto vehicle user delay is reduced because no TSP is activated. The reduction of auto vehicle user delay is at the cost of increasing tram passenger

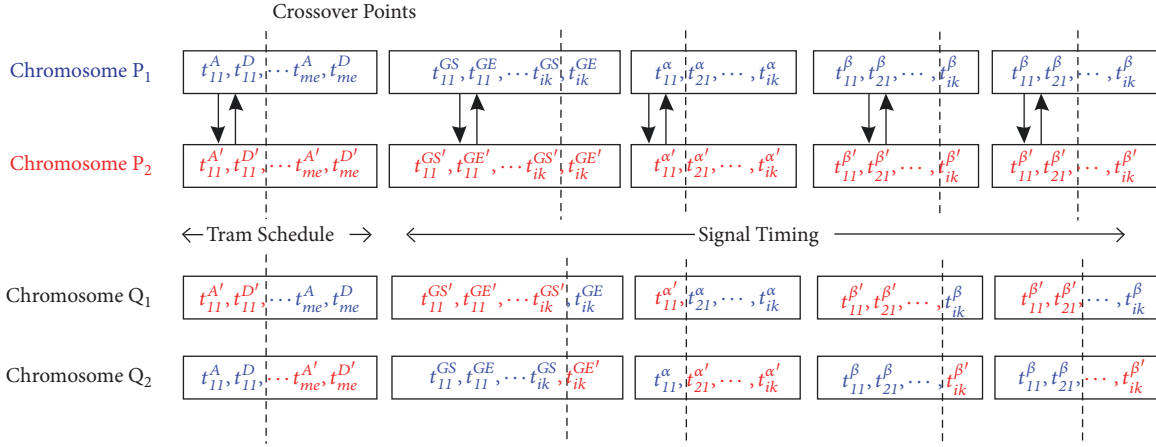


FIGURE 5: Crossover.

TABLE 4: Traffic volume and signal timings.

Intersection	Traffic flow (Number of Lane)			Split (Cycle 180 sec)		Intersection	Traffic flow (Number of Lane)			Split (Cycle 180 sec)		
	Left	Through	Right	Left	Through		Left	Through	Right	Left	Through	
S1	South	152(1)	1020(2)	88(1)	25	90	South	148(1)	1084(2)	96(1)	22	90
	North	147(1)	1016(2)	44(1)	25	90	North	153(1)	1096(2)	2(1)	22	90
	East	144(1)	512(2)	372(2)	23	42	East	148(1)	527(2)	212(2)	23	45
	West	142(1)	492(2)	16(1)	23	42	West	146(1)	518(2)	16(1)	23	45

TABLE 5: Performance comparison with and without the integration of the numerical case.

Delay/s (Variation rate/%)	Schedule only	TSP only		Integrated model
		unconditional	conditional	
Tram passenger	60.9 (+95.34%)	0	39.2(+25.58%)	31.2
Auto vehicle user	39.1 (-8.45%)	89.2 (+108.45%)	48.2 (+12.68%)	42.8
Average person	40.7 (+6.25%)	61.1 (+59.38%)	43.7 (+14.06%)	38.3
Weighted average person	45.3 (+28.57%)	59.7 (+69.64%)	39.0 (+10.71%)	35.2

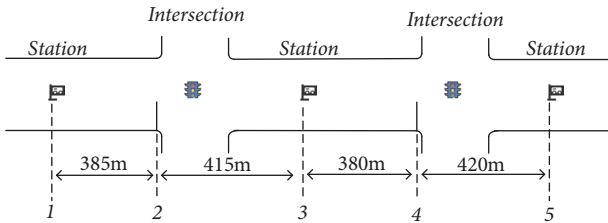


FIGURE 6: The layout of the tramline in the numerical case.

delay. Due to travel time limits between each link, only adjusting tram schedule is not able to reduce delay for tram passengers and each person of the whole system.

When unconditional TSP is applied, tram passenger delay is zero. Although tram passengers do not suffer from any delay, auto vehicle user delay is extremely high, which due to unconditional TSP does not take into account the operation of auto vehicles. When the number of auto vehicles is larger

than that of tram passenger at intersection, higher auto vehicle user delay will lead to higher person delay of the whole system.

When only conditional TSP is applied, in order to trade off auto vehicle user and tram passengers, trams are not always granted priority to pass intersections without stops. As a result, tram passenger delay increases. The integrated model can further reduce auto vehicle user delay than conditional TSP because tram can be scheduled to arrive at intersections during green signals. Thus no TSP is required to activate and no delay is imposed to auto vehicle. Compared with unconditional TSP, conditional TSP can greatly reduce person delay of the whole system, although it increases tram passenger delay. Thus, conditional active priority is more suitable for system optimization.

6.1.2. Comparison between the Enumeration Solution and GA Solution. In this paper, genetic algorithm (GA) is adopted to solve the problem. Since GA is an intelligent search algorithm, the global optimal solution cannot be achieved.

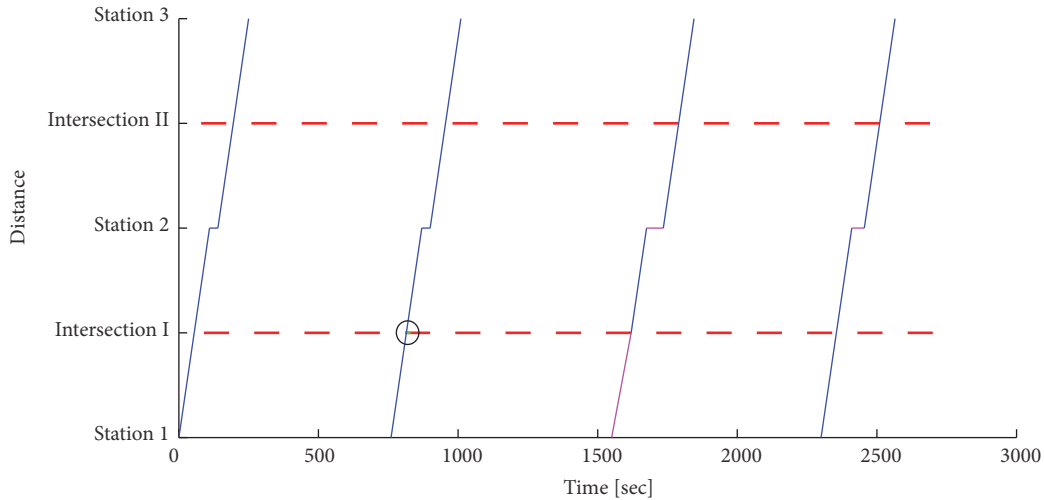


FIGURE 7: Tram schedule of four trams.

TABLE 6: Performance comparison of GA and enumeration.

Quantity of trams	Quantity of nodes		Method	CPU time (sec)	Objectives (sec)	Gap (%)
	Intersections	Stations				
4	2	3	Enumeration	3954	34.7	-
			GA	2361	35.2	1.4%
	4	5	Enumeration	12814	42.3	-
			GA	1600	43.0	1.6%
8	5	7	Enumeration	>21600	54.8*	-
			GA	2473	54.6	-0.4%
	8	9	Enumeration	>21600	70.9*	-
			GA	3521	70.3	-0.8%
12	10	11	Enumeration	>21600	82.1*	-
			GA	3864	80.4	-2.1%
	12	13	Enumeration	>21600	89.6*	-
			GA	4057	88.5	-1.2%

(The value with \* is the current optimal solution in a given CPU time)

In order to testify the accuracy of the proposed GA, compared the proposed GA with enumeration solution. The pseudocode of the enumeration procedure is presented in Algorithm 2.

In the enumeration procedure, all possible combinations of decision variables of the proposed model are listed and evaluated first, followed by choosing the group of decision variables which leads to the minimal average person delay as the optimal solution. In the resubmitted manuscript, the pseudocode of the enumeration procedure has been supplemented to explain how it finds the global optimal solution. In this study, the decision variables of the proposed model include the signal offsets among all intersections, tram schedules and TSP strategies. The outer layer of the enumeration procedure lists all signal offsets among different intersections without changing the signal cycle length and green duration. The inner layer of the enumeration procedure then lists all possible groups of tram schedules and TSP strategies for each signal offset produced in the first step.

It should be noted that TSP strategies might affect the departure time of trams at intersections, which is a component of tram scheme. As a result, within the inner layer of the enumeration procedure, the departure time of the first tram from the first station is enumerated at first, followed by attaining all possible groups of arrival and departure time of the first tram at the other stations, considering each possible TSP strategy at each intersection, each possible inter-station run-time and station dwell time in their valid ranges. Once all possible schedules of the first tram are attained, the possible schedules of the other trams are enumerated one by one taking into account the valid ranges of tram headways, inter-station run-time and station dwell time, as well as the TSP strategies.

As shown in Table 6, when the problem is getting large, the enumeration is not capable of finding the optimal solution in a given time. The minor gap between enumeration solution and GA shows that the proposed GA has a high accuracy. In the case of 5 nodes including 2 intersections and 3 stations with 4 trams scheduled, the CPU time of enumeration and



```

(1) Input: Quantity of trams ( $m$ )
(2)   Nodes / Intersection / Station ( $N / I / S$ )
(3)   Arrival / Departure time constraints ( $AT / DT$  constraint)
(4)   Dwell / Link travel time constraints ( $DT / LT$  constraint)
(5)   Signal timing constraint ( $ST$  constraint)
(6)   TSP strategy constraint ( $TSP$  constraint)
(7) Output: Solutions; The best solution
(8) repeat
(9)   for each intersection do
(10)    Unadjusted signal timing  $\leftarrow$  Initialization ( $ST$  constraint)
(11)  end
(12)  for each unadjusted signal timing do
(13)   for each tram do
(14)    Arrival time  $\leftarrow$  Initialization ( $AT$  constraint)
(15)    for each node  $N$  do
(16)     if node  $N \in$  Station
(17)      Departure time  $\leftarrow$  Initialization ( $DT$  constraint)
(18)    else
(19)     if Arrival time during Green phase
(20)      Departure time = Arrival time
(21)    else
(22)     TSP strategy  $\leftarrow$  Initialization ( $TSP$  constraint)
(23)     for each TSP strategy do
(24)      if TSP measure is activated
(25)       Departure time = Arrival time
(26)     else
(27)      Departure time = Arrival time + Waiting time
(28)     end
(29)    end
(30)   end
(31)  end
(32)  end
(33)  end
(34)  end
(35)  Solutions  $\leftarrow$  {Unadjusted signal timing; Arrival time; Departure time; TSP strategy}
(36)  Person delay  $\leftarrow$  Calculation (Solutions)
(37)  The best solution  $\leftarrow$  Minimum (Person delay)
(38) until finding The best solution

```

ALGORITHM 2: The enumeration procedure for the numerical case.

GA both is acceptable, and the gap is 1.4%. When the quantity of nodes and trams becomes larger, especially in the case of 17 nodes including 8 intersections and 9 stations with 8 trams to schedule, enumeration solution cannot find the global optimal solution in 6 hours. Obviously the length of CPU time depends on the combinations of the quantity of trams and the quantity of nodes. Thus either the quantity of trams or the quantity of nodes becomes larger and the CPU time prolongs.

**6.2. Case Study in Ningbo Tramline.** A section of a tramline in Ningbo, a city of China, is employed to evaluate the proposed model. The tramline section consists of 5 stations and 7 intersections. The layout of intersections and stations is given in Figure 8. The maximum carrying capacity on one tram is 368 passengers. Taking peak-hours, for example, traffic volume, original signal timings at different intersections where the cycle time is 180 s, are presented in Table 7. The

range of dwell time at different stations and the range of travel time as well as the loading factor of each link are specified in Table 8.

**6.2.1. Performance Comparison by Different Objectives.** The objective of the proposed model is to minimize person delay of the whole system including tram passengers and auto vehicle users. In order to further demonstrate the benefit of minimizing person delay, the computational results of three different objectives are compared. Beside the proposed objective, other two objectives are to minimize tram passenger delay and to minimize auto vehicle user delay.

Compared with the proposed objective, as shown in Table 9, only minimizing tram passenger delay reduces tram passenger delay to zero and results in a significant increase in auto vehicle user delay. When the number of auto vehicles is larger than that of trams at intersections, auto vehicle user delay has significant influence on person delay of the

TABLE 7: Traffic volume and signal timings at intersections along the Ningbo tramline.

Intersection	Traffic flow (Number of Lane)			Split (Cycle 180 sec)		Intersection	Traffic flow (Number of Lane)			Split (Cycle 180 sec)			
	Left	Through	Right	Left	Through		Left	Through	Right	Left	Through		
S1	South	152(1)	1020(2)	88(1)	25	90	S5	South	140(1)	1008(2)	60(1)	27	90
	North	147(1)	1016(2)	44(1)	25	90		North	142(1)	1024(3)	168(2)	27	90
	East	144(1)	512(2)	372(2)	23	42		East	157(1)	470(2)	214(2)	23	40
	West	142(1)	492(2)	16(1)	23	42		West	212(2)	468(2)	74(1)	23	40
S2	South	148(1)	1084(2)	96(1)	22	90	S6	South	144(1)	1178(2)	362(2)	25	86
	North	153(1)	1096(2)	2(1)	22	90		North	137(1)	1032(2)	58(2)	25	86
	East	148(1)	527(2)	212(2)	23	45		East	240(2)	408(2)	410(2)	34	35
	West	146(1)	518(2)	16(1)	23	45		West	144(1)	412(3)	72(1)	34	35
S3	South	134(1)	1064(2)	24(1)	27	90	S7	South	146(1)	1052(2)	258(1)	26	77
	North	146(1)	1004(2)	20(1)	27	90		North	141(1)	1003(2)	82(1)	26	77
	East	132(1)	510(2)	120(2)	23	40		East	142(1)	522(2)	120(1)	36	41
	West	138(1)	512(2)	108(1)	23	40		West	137(1)	514(2)	132(1)	36	41
S4	South	148(1)	1056(2)	60(1)	21	90							
	North	139(1)	1096(2)	176(1)	21	90							
	East	148(1)	482(2)	132(1)	26	43							
	West	144(1)	512(2)	80(1)	26	43							

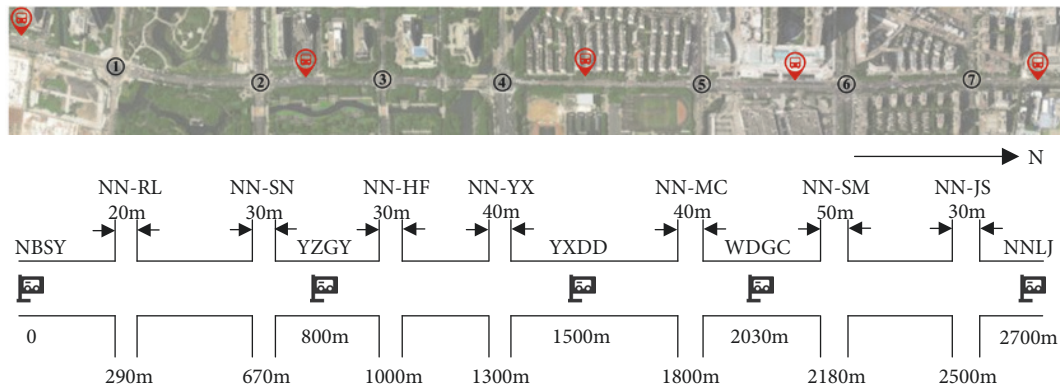


FIGURE 8: Layout of the Ningbo tramline.

TABLE 8: Tram timetable formulation parameters and loading factors.

Station	Range of dwell time [sec]	Link	Range of travel time [sec]	Loading factor
NBSY	(10,20)	NBSY- YZGY	(105,125)	0.7
YZGY	(15,25)	YZGY- YXDD	(101,121)	0.9
YXDD	(20,30)	YXDD- WDG	(85,105)	0.8
WDG	(20,30)	WDG- NNLJ	(99,119)	0.6
NNLJ	(10,20)			

TABLE 9: Performance comparison by different objectives.

Delay/s (Variation rate/%)	Minimizing tram passenger delay	Minimizing auto vehicle user delay	The proposed model
Tram passenger	0	83 (+93.02%)	43
Auto vehicle user	148 (+108.45%)	48 (-32.39%)	71
Average person	102 (+59.38%)	67 (+4.69%)	64
Weighted average person	95 (+69.64%)	75 (+33.39%)	56



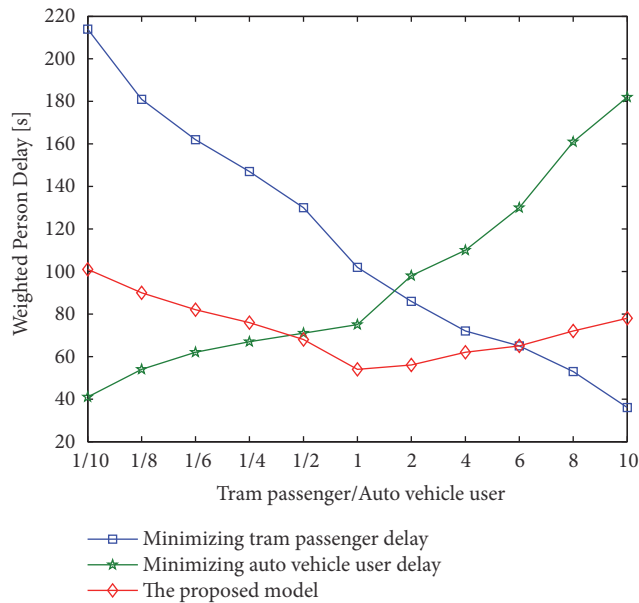


FIGURE 9: Performance on different ratio of tram passenger to auto vehicle user with different objectives.

whole system. This leads to person delay of the whole system which is relatively high due to the high auto vehicle user delay. When only minimizing auto vehicle user delay, the interests of auto vehicle passengers will be ensured by less TSP activated for trams. Thus trams passenger delay increase, which reduced the attraction of public transportation. Considering optimizing tram passengers and auto vehicle users that separately reduced the efficiency of the whole system, these two stakeholders should be optimized simultaneously.

**6.2.2. Performance on Different Ratio of Tram Passenger to Auto Vehicle User.** Figure 9 shows the changes in different ratio of tram passengers to auto vehicle users with different objectives. When the ratio varies, weighted person delay fluctuates greatly with the objectives of only minimizing tram passengers or only minimizing auto vehicle passengers. Neither of them can adapt to the situation when the ratio changes. However, by taking tram passengers and auto vehicle users as a whole, it can stabilize the weighted person delay and adapt to changes in different ratio.

In practice, the ratio of tram passenger to auto vehicle user varies from day to day, especially the changes between weekdays and weekends, peak hour and non-peak hour. For example, the number of auto vehicles is larger than that of trams on weekdays and the number of trams becomes larger on weekends. If a single type of person is considered for optimization, such a change cannot be adapted. Therefore, the proposed model has better applicability.

## 7. Conclusions

Operating in semi-exclusive ROW, trams may stop at intersections to wait red signals, which results in extra intersection run-time or even delays. To improve the tram service

quality, Transit Signal Priority (TSP) has been applied at intersections by offering extra green phase for trams. However, providing extra green phase for every tram might lead to heavy delays to crossing vehicles. To address this problem, this study developed an integrated optimization model of tram schedule and signal priority which can balance the delay between trams and other vehicles to minimize the average person delay.

Due to the complexity of the problem, the GA is employed to solve the proposed model. A numerical case is conducted to testify the computing time and solution optimality of the proposed GA and the enumeration procedure. The results show that the enumeration is unable to find the optimal solution or even a near-optimal solution when the tramline covers several stations and intersections due to the huge solution space, while the GA outperforms the enumeration procedure within the restrained computing time. The case studies on the Ningbo tramline indicate that the integrated optimization reduces the average delay of passengers on both trams and auto vehicles, in comparison with only optimizing tram timetable or only applying TSP. It is also found that minimizing the average person delay of the whole system is able to adapt to the possible fluctuations in the ratio of tram passengers to auto vehicle users, compared with only minimizing tram passenger delay or auto vehicle user delay.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research is funded by the Fundamental Research Funds for the Central Universities (2019YJS096, 2018JBM025) and the National Natural Science Foundation of China (71571016).

## References

- [1] H. R. Smith, P. B. Hemily, and M. Ivanovic, *Transit Signal Priority (TSP): A Planning and Implementation Handbook*, ITS America, Washington, DC, USA, May 2005.
- [2] G.-L. Chang, M. Vasudevan, and C.-C. Su, "Bus-preemption under adaptive signal control environments," *Transportation Research Record*, no. 1494, pp. 146–154, 1995.
- [3] J. Collura, H. Rakha, and J. Guifford, *Guidelines for the Planning and Deployment of Emergency Vehicle Preemption and Transit Priority Strategies*, Research Report, Virginia Tech Transportation Institute, 2003.
- [4] M. A. Turnquist, "Strategies for improving reliability of bus transit service," *Transportation Research Record*, vol. 818, pp. 7–13, 1978.
- [5] M. D. Abkowitz and M. Lepofsky, "Implementing headway-based reliability control on transit routes," *Journal of Transportation Engineering-ASCE*, vol. 116, no. 1, pp. 49–63, 1989.

- [6] R. L. Jackson and D. Ibarra, "Service reliability program of the Southern California Rapid Transit District: a test of the short-term impacts on Line 26-51," *Transportation Research Record*, vol. 1338, pp. 3-13, 1992.
- [7] W. J. Ma, K. L. Head, and Y. H. Feng, "Integrated optimization of transit priority operation at isolated intersections: a person-capacity-based approach," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 49-62, 2014.
- [8] Y. Bai, J. Li, T. Li, L. Yang, and C. Lyu, "Traffic signal coordination for tramlines with passive priority strategy," *Mathematical Problems in Engineering*, vol. 2018, Article ID 6062878, 14 pages, 2018.
- [9] J. T. Morgan and J. D. C. Little, "Synchronizing traffic signals for maximal bandwidth," *Operations Research*, vol. 12, no. 6, pp. 896-912, 1964.
- [10] J. D. C. Little, "The synchronization of traffic signals by mixed-integer linear programming," *Operations Research*, vol. 14, no. 4, pp. 568-594, 1966.
- [11] J. D. C. Little, M. D. Kelson, and N. H. Gartner, "MAXBAND: a program for setting signal on arteries and triangular network," *Transportation Research Record*, no. 795, pp. 40-46, 1981.
- [12] N. H. Gartner, S. F. Assman, F. Lasaga, and D. L. Hou, "A multi-band approach to arterial traffic signal optimization," *Transportation Research Part B: Methodological*, vol. 25, no. 1, pp. 55-74, 1991.
- [13] C. Stamatiadis and N. H. Gartner, "MULTIBAND-96: a program for variable-bandwidth progression optimization of multiarterial traffic networks," *Transportation Research Record*, vol. 1554, no. 1, pp. 9-17, 1996.
- [14] N. H. Gartner and C. Stamatiadis, "Arterial-based control of traffic flow in urban grid networks," *Mathematical and Computer Modelling*, vol. 35, no. 5-6, pp. 657-671, 2002.
- [15] C. Zhang, Y. Xie, N. H. Gartner, C. Stamatiadis, and T. Arsava, "AM-band: an asymmetrical multi-band model for arterial traffic signal coordination," *Transportation Research, Part C: Emerging Technologies*, vol. 58, pp. 515-531, 2015.
- [16] N. B. Hounsell, B. P. Shrestha, and J. R. Head, "The way ahead for London's bus priority at traffic signals," *IET Intelligent Transport Systems*, vol. 2, no. 3, pp. 193-200, 2008.
- [17] A. H. F. Chow, S. Li, and R. Zhong, "Multi-objective optimal control formulations for bus service reliability with traffic signals," *Transportation Research Part B: Methodological*, vol. 103, pp. 248-268, 2017.
- [18] A. Skabardonis, "Control strategies for transit priority," *Transportation Research Record*, vol. 1727, no. 1, pp. 20-26, 2000.
- [19] G. Currie and A. Shalaby, "Active transit signal priority for streetcars: experience in Melbourne and Toronto," in *Proceedings of the Transportation Research Board Meeting*, 2008.
- [20] M. T. Islam, J. Tiwana, A. Bhowmick, and T. Z. Qiu, "Design of LRT signal priority to improve arterial traffic mobility," *Journal of Transportation Engineering*, vol. 142, no. 9, Article ID 04016034, 2016.
- [21] A. H. F. Chow, S. Li, and R. Zhong, "Modelling urban traffic dynamics based upon the variational formulation of kinematic waves," *Transportmetrica B: Transport Dynamics*, vol. 3, no. 3, pp. 169-191, 2015.
- [22] W. Ma, X. Yang, and Y. Liu, "Development and evaluation of a coordinated and conditional bus priority approach," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2145, no. 2145, pp. 49-58, 2010.
- [23] F. Ahmed and Y. E. Hawas, "An integrated real-time traffic signal system for transit signal priority, incident detection and congestion management," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 52-76, 2015.
- [24] S. Yagar and B. Han, "A procedure for real-time signal control that considers transit interference and priority," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 315-331, 1994.
- [25] R. J. Baker, J. Collura, J. J. Dale, L. Head, B. Hemily, and M. Ivanovic, "An overview of transit signal priority," *Computers*, 2002.
- [26] D. François and R. Hesham, "Integration transit signal priority with adaptive traffic signal control systems," in *Proceedings of the Transportation Research Board 84th Annual Meeting*, pp. 1-18, 2005.
- [27] G. Zhou and A. Gan, "Design of transit signal priority at signalized intersections with queue jumper lanes," *Journal of Public Transportation*, vol. 12, no. 4, pp. 117-132, 2009.
- [28] J. Viegas and B. Lu, "Widening the scope for bus priority with intermittent bus lanes," *Transportation Planning and Technology*, vol. 24, no. 2, pp. 87-110, 2001.
- [29] J. Viegas and B. Lu, "The intermittent bus lane signals setting within an area," *Transportation Research Part C: Emerging Technologies*, vol. 12, no. 6, pp. 453-469, 2004.
- [30] M. Eichler and C. F. Daganzo, "Bus lanes with intermittent priority: strategy formulae and an evaluation," *Transportation Research, Part B (Methodological)*, vol. 40, no. 9, pp. 731-744, 2006.
- [31] J. Wahlstedt, "Impacts of bus priority in coordinated traffic signals," *Procedia - Social and Behavioral Sciences*, vol. 16, pp. 578-587, 2011.
- [32] A. Skabardonis and E. Christofa, "Impact of transit signal priority on level of service at signalized intersections," *Procedia - Social and Behavioral Sciences*, vol. 16, pp. 612-619, 2011.
- [33] S. Z. Altun and P. G. Furth, "Scheduling buses to take advantage of transit signal priority," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2111, no. 7, pp. 50-59, 2009.
- [34] P. G. Furth and T. H. J. Muller, "Conditional bus priority at signalized intersections: better service with less traffic disruption," *Transportation Research Record*, vol. 1731, no. 1, pp. 23-30, 2000.
- [35] K. Ling and A. Shalaby, "Automated transit headway control via adaptive signal priority," *Journal of Advanced Transportation*, vol. 38, no. 1, pp. 45-67, 2010.
- [36] J. Li and K. L. Head, "Sustainability provisions in the bus-scheduling problem," *Transportation Research Part D: Transport and Environment*, vol. 14, no. 1, pp. 50-60, 2009.
- [37] K. Nachtigall and S. Voget, "Minimizing waiting times in integrated fixed interval timetables by upgrading railway tracks," *European Journal of Operational Research*, vol. 103, no. 3, pp. 610-627, 1997.
- [38] K. K. Wong and T. K. Ho, "Dwell-time and run-time control for DC mass rapid transit railways," *IET Electric Power Applications*, vol. 1, no. 6, p. 956, 2007.
- [39] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117-135, 2015.
- [40] T. Robenek, Y. Maknoon, S. S. Azadeh, J. Chen, and M. Bierlaire, "Passenger centric train timetabling problem," *Transportation Research Part B: Methodological*, vol. 89, pp. 107-126, 2016.

- [41] Z. Yuhe, B. Yun, L. Jiajie, M. Baohua, and L. Tang, "Integrated optimization on train control and timetable to minimize net energy consumption of metro lines," *Journal of Advanced Transportation*, vol. 2018, Article ID 7905820, 19 pages, 2018.
- [42] J. Shi, Y. Sun, P. Schonfeld, and J. Qi, "Joint optimization of tram timetables and signal timing adjustments at intersections," *Transportation Research Part C: Emerging Technologies*, vol. 83, pp. 104–119, 2017.
- [43] F. V. Webster, "Traffic signal settings," Road Research Laboratory Technical 39, HMSO, London, UK, 1958.
- [44] A. Rahmi, "Traffic signals: capacity and timing analysis," Research Report, 1981.
- [45] L. Yang, J. Qi, S. Li, and Y. Gao, "Collaborative optimization for train scheduling and train stop planning on high-speed railways," *OMEGA - The International Journal of Management Science*, vol. 64, pp. 57–76, 2016.
- [46] R.-H. Yao, M.-N. Liu, and H.-F. Xu, "Applicability analysis of vehicle delay models for isolated signalized intersection," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 46, no. 02, pp. 390–398, 2016.
- [47] H. R. Lourenço, "Sevast'yanov's algorithm for the flow-shop scheduling problem," *European Journal of Operational Research*, vol. 91, no. 1, pp. 176–189, 1996.

## Research Article

# Multiobjective Optimization Model for Sustainable Waste Management Network Design

Sun Olapiriyakul , Warut Pannakkong , Warith Kachapanya, and Stefano Starita 

*School of Manufacturing Systems and Mechanical Engineering, Sirindhorn International Institute of Technology, Thammasat University, Pathum Thani 12121, Thailand*

Correspondence should be addressed to Stefano Starita; [stefano@siit.tu.ac.th](mailto:stefano@siit.tu.ac.th)

Received 14 December 2018; Revised 11 March 2019; Accepted 14 April 2019; Published 13 May 2019

Guest Editor: Jesica de Armas

Copyright © 2019 Sun Olapiriyakul et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inefficient or poorly planned waste management systems are a burden to society and economy. For example, excessively long waste transportation routes can have a negative impact on a large share of the population. This is exacerbated by the rapid urbanization happening worldwide and in developing countries. Sustainability issues should be accounted for at every stage of decision making, from strategic to daily operations. In this paper, we propose a multiobjective optimization model to design a cost-effective waste management supply chain, while considering sustainability issues such as land-use and public health impacts. The model is applied to a case study in Pathum Thani (Thailand) to provide managerial insights.

## 1. Introduction

An unsanitary and inefficient municipal solid waste management (MSWM) system has long been a challenging issue to overcome. An inadequate waste management budget and the lack of public participation in waste segregation at the source are among the leading causes for the long-term accumulation of uncollected and improperly disposed of solid wastes in developing countries. In urban areas, unsanitary disposal sites are generally more problematic, considering the limited land availability and the need to attain adequate waste disposal capacity to serve a rapidly growing population. The risk of exposure to contaminants emitted from solid waste disposal is also more pronounced for urban residents. First, this is due to the problem of waste disposal capacity shortage, caused by a lack of land suitable for solid waste receptacles or a delay in building sufficient capacity. Waste flows beyond manageable capacity are among the causes of the uncontrolled release of leachates and gases, which can lead to the formation of strong unpleasant odors and spontaneous fires. Second, in an urban context, expanding the existing disposal facilities or locating new ones has to be made in the vicinity of densely populated areas. The release of hazardous constituents from disposal and transport operations can impose substantial environmental stress on the

surrounding communities, especially for centralized MSWM systems [1]. The neighboring population has to deal with the immediate impacts, such as uncleanliness, odors, and inadequate air quality. The long-term impacts of unsanitary MSWM practices on the local population include various health risks [2] and decreased environmental quality [3].

Previous studies highlight the fact that in developing countries MSWM results in decreasing property values [4] and imposes an additional financial burden on the local population in terms of higher sanitation fees [5]. Consequently, a disproportionate environmental impact is an increasingly important issue in MSWM planning. To resolve this issue, environmental justice must be added as one of the strategic goals to be achieved. The principles of environmental justice, in this case, refer to the idea that all communities must be protected from excessive or disproportionate environmental stressors. The level of environmental stressors experienced by communities needs to be evaluated. Whenever possible, the capacity of each community to tolerate unfavorable environmental impacts should be considered.

Aside from disproportionate impacts and justice concerns, the negative externalities of MSWM can also impede the development of desired social aspects such as strong community social cohesion and efficient youth development [6]. The waste management literature [7] also reveals that



environmental problems of waste management generate a number of social issues, such as those related to employment opportunities [8], unsanitary working conditions [9], and community satisfaction [10]. Attempts to offset these issues usually require significant spending on improving sanitary infrastructure and services. Such financial needs can ultimately leave communities with an insufficient sanitary budget and inability to properly collect and dispose of municipal solid wastes, leading to a vicious cycle of decline in local public health and environmental conditions. Therefore, a sustainable MSWN requires that significant economic, environmental, and social issues be integrated into the strategic decision-making processes.

The successful establishment of sustainable MSWM is dependent on the network design and transportation planning stage. Environmentally benign site selection requires a careful land suitability analysis. The potential interactions of disposal sites with local communities and existing geographical, hydrological, and socioeconomic parameters need to be taken into account. For example, in Thailand, there is a rigorous protocol for solid waste disposal site selection. However, a large number of disposal sites are still located in environmentally sensitive areas. Due to the rapid and unplanned urban growth in many cities, some disposal sites, which were originally in suitable vacant land areas, are now in close proximity to rapidly urbanizing areas. Under the current MSWM situation, the number of affected communities is only expected to rise in the near future. Local governments will also expend a great deal of effort to cope with strong public opposition, should any future solid waste disposal siting or expansion decision be taken. To sum up, a sustainable MSWM requires the development of network design approaches to help in resolving the following MSWM challenges:

- (i) finding suitable land for waste disposal,
- (ii) public health impacts caused by the operation of MSWM facilities and waste transport,
- (iii) lack of land-use planning for MSWM,
- (iv) environmental justice and disproportionate environmental impacts on nearby communities.

The remainder of the paper is organized as follows: Section 2 provides a review of the relevant literature. Section 3 introduces a mixed-integer multiobjective program to design a sustainable supply chain for waste management. A case study in Pathum Thani (Thailand) province is analyzed in Section 4. Finally, the conclusion and future research directions are discussed in Section 5.

## 2. Background

This section provides a literature review to gain an insight into the research trends in sustainable MSWM design. We focus on three broad categories: sustainable supply chain network design (SCND), solid waste management, and environmental justice. Results are summarized in a Venn diagram in Figure 1. The diagram reveals that each research field has received

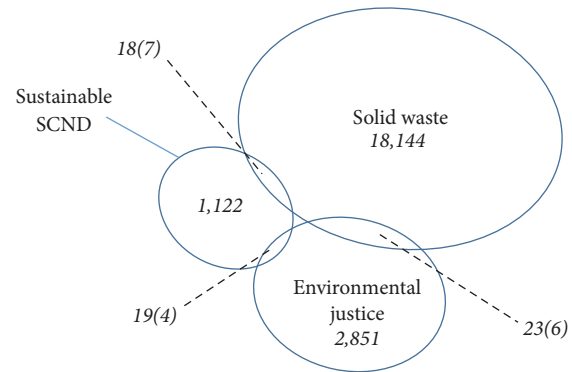


FIGURE 1: Relevant research areas.

considerable attention over the years. There are some pairwise overlaps. However, to the best of our knowledge, no research paper has studied the problem incorporating the three issues at the same time.

Despite a growing research interest in sustainable supply chain management and environmental justice, there is still limited consideration of environmental justice issues in the sustainable SCND literature. Past studies combine equality and justice issues with the environmental and social dimensions. Moura et al. [11] incorporate environmental justice aspects into a biobjective transportation network design model while addressing the (three pillars of) sustainability. Their model uses a restrictive constraint that protects the surrounding communities from being overly burdened with noise and air pollution. The motorists cost of increased travel time due to the congestion effects of the transport network is regarded as the social impact metric. Constraints are imposed to achieve a network design with desired environmental justice levels. There are also previous sustainable SCND studies that investigate social justice issues. These papers generally aim to minimize the inequality in accessing public services and in the quality-of-life improvement opportunities. An SCND study by Ferguson et al. [12] shows how to incorporate social equality into a transit service design problem. They aim to minimize the variation in the level of access to basic amenities provided by the designed transit system. Jafari et al. [13] propose a sustainable SCND model for textile industries, with the aim of promoting social justice by maximizing the employment level in different geographical areas. Despite the increased research efforts, there are still social justice issues that need to be translated into a well-defined optimization problem. Manaugh et al. [14] point out social justice issues and measures that can potentially be incorporated into urban transport planning. These issues are related to the disproportionate accessibility of transport service among different groups of people. Oswald Beiler and Mohammed [15] address many demographic, social-economic, and location-based factors that can be considered in developing transportation justice metrics and frameworks.

The topic of solid waste management has been studied in depth due to the steadily growing urban population and consequent waste generation. Life-cycle environmental impacts of various MSWM systems have been explored

extensively as reviewed by Bernstad Saraiva et al., 2018 [16]. Also, a number of economic, environmental, and social performance indicators for MSWM have been proposed by researchers as addressed by Rodrigues et al., 2018 [17]. However, most of sustainable SCND studies are related to industrial applications. The strategic planning of MSWM infrastructures and transport network has been confined to the analysis of infrastructure investment, operating expenses, and economic viability. For instance, Zhang et al. [18] develop an optimization model to minimize the costs of inventory, transportation, and disposal of MSWM. In their study, an interval programming approach is employed to deal with uncertainties of MSWM planning parameters. Toso and Alem [19] propose a deterministic and stochastic capacitated facility location model to determine the optimal location planning design for recycling urban solid wastes. Their model solely focuses on minimizing the overall costs having budget constraints. Just recently more attention is being given to MSWM planning problems looking beyond economic feasibility to add the sustainability perspective. A network design study by Inghels et al. [20] evaluates the financial viability of using multimodal transportation to reduce the carbon emissions and social impact of MSWM. Their model evaluates the societal cost burden associated with different transportation modes, measured as the sum of disturbance effects on nearby residents. The effects include accidents, noise, air pollution, congestion, and construction of transport infrastructure. Xu et al. [21] propose a SCND model for the reverse logistic supply chain of solid wastes. The amount of carbon emissions created during the transportation of recyclable e-wastes is used as an environmental metric.

The number of affected people is used to estimate the negative effects of MSWM facilities on local communities. According to the research trend, all aspects of sustainability are currently being addressed in the context of MSWM. However, as pointed out by Eskandarpour et al. [22], there is a research need for a broader consideration of social and environmental impact matrices. Aside from the global warming potential, the use of other LCA-based environmental impact indicators needs to be explored. Indicators that have been used by LCA studies to evaluate the environmental performance of MSWM systems are summarized in the review papers by Khandelwal et al. [23] and Yadav and Samadder [24].

As already pointed out, there is a lack of studies about how justice affects the sustainability performances of an MSWM system. The development of MSWM infrastructure that positively contributes to environmental and social justice must be based on a careful evaluation of the current stage of urban development and ongoing injustice issues [25–27]. There is a difference between social and environmental justice problems in developed and developing countries. Developed countries generally experience inequalities for side effects of urban growth such as air pollution, waste water, and traffic congestion. In developing countries and low-density cities, the problems are more related to spatial inequalities of economic development activities, sanitation services, and allocation of public resources. As shown in the case of the Perth Metropolitan area, people living in

the outer suburbs have the least accessibility provided by car and public transport to job, education, shopping, and healthcare opportunities [28]. The case of plastic bag waste in Nairobi, Kenya, is a typical example of environmental justice issues associated with the disproportionate share of sanitary services and environmental protection policy [29]. The case study describes how political influences on local businesses result in unsustainable patterns of production and consumption of plastic bags, and, thereby, there is a vast accumulation of solid and plastic wastes in communities with low socioeconomic status.

In addition to pollution problems, land-use issues are also critical issues in MSWM planning. Agyeman and Evans [30] explore urban development initiatives that demonstrate inherent links between environmental justice and sustainability issues. They point out that solid waste and land-use planning are common concerns for both environmental justice and sustainability. A land-use policy should focus on preventing disproportionate land-use impacts within or among communities. Without adequate land-use and environmental justice policies, solid waste facilities are likely to be located in poor and minority communities as discovered by Norton et al. [31] when focusing on North Carolina waste infrastructure.

This study addresses the three fundamental sustainability dimensions: environmental, social, and economic, in the context of MSWM in rapidly urbanized regions. To bridge the research gap and contribute to the field of sustainable MSWM, this develops a social impact metric based on environmental justice and incorporate it into a sustainable SCND model for MSWM system. Specifically, the issue of environmental justice related to land-use stress caused by MSWM facility establishment is considered. The inclusion of a land-use equality objective is used to obtain a balanced network design where land-use stress is fairly distributed across an area. Furthermore, the damage of MSWM to human health is also introduced as a measure for the environmental impact of an MSWM system. We use the disability-adjusted life years (DALYs) metric, according to the life-cycle impact assessment (LCIA) method. Despite being used by WHO as a measure of the global burden of disease for many years, there have been very limited applications of DALYs in both the supply chain and the waste management literature. Lastly, facility and transportation costs are taken into account to evaluate the economic aspects of sustainability.

### 3. Methodology

**3.1. Public Health Impact Assessment.** In this study, the public health impact is defined as the overall mortality and disease burden of nearby residents caused by MSWM facilities and waste transport activities. The public health impact is estimated in units of disability-adjusted life years (DALYs), which is one of the well-established endpoint LCIA metrics. DALYs represent the number of years of life lost due to premature mortality and healthy years of life lost due to disability [32]. We translate the impact of waste management operations and waste transportation into DALYs per person

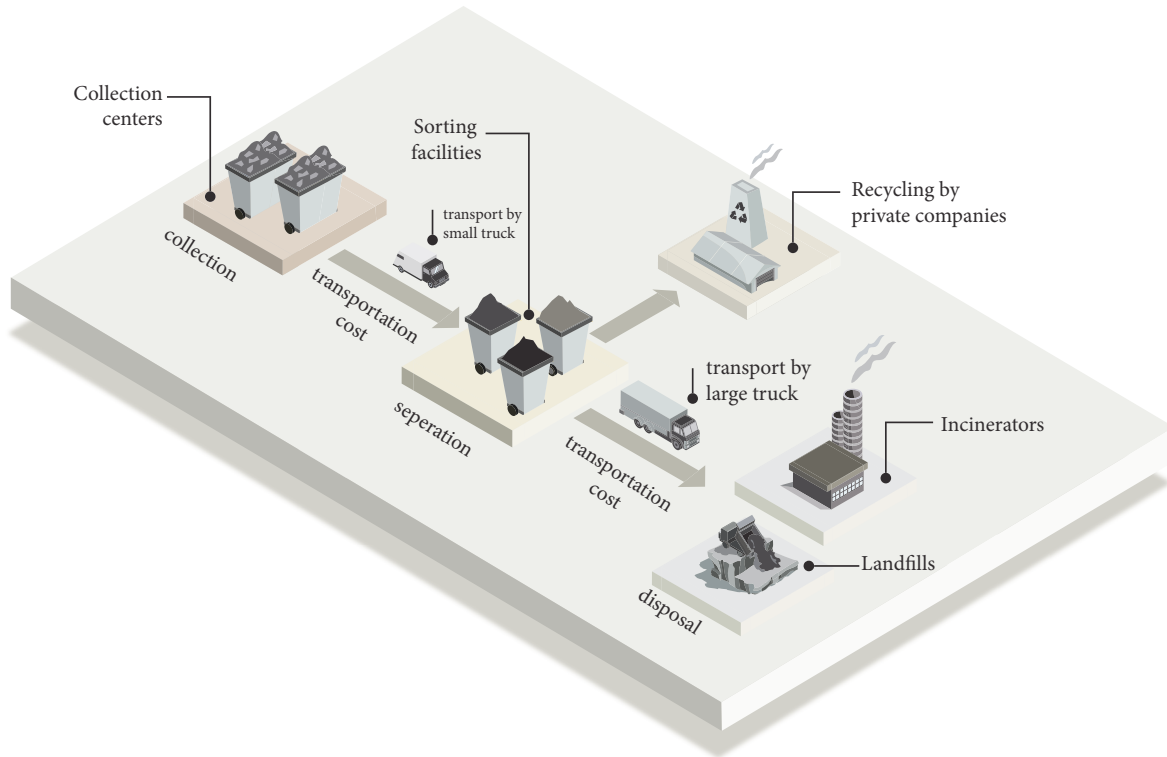


FIGURE 2: Waste management supply chain.

using the ReCiPe 2008 Endpoint LCIA method [33]. Then, the total DALYs are obtained by multiplying the individual impact of exposure by the total number of people living within the affected area.

**3.2. Land-Use Impact Assessment.** Processes in MSWM facilities including construction, operation, and closure normally take place over long timescales. Under traditional centralized waste systems, the life expectancy of MSWM facilities is longer, due to the need for larger-sized facilities to cope with increased waste generation in cities. Communities surrounding waste facilities have to deal with the long-term negative external effects of municipal solid waste. Therefore, the issue of disproportionate environmental burden among population in certain areas is a pressing concern, especially for rapidly urbanized cities. From a strategic point of view, it is important that each administrative area in a city is not overly burdened by land-use impact or other important environmental stressors caused by MSWM facilities. The principle of environmental justice must be adopted at the early phase of MSWM planning.

In this study, the spatial planning of MSWM infrastructure uses a land-use equality strategy to mitigate the impact on local land-use in areas with substantial land-use stress. The proposed planning approach involves a two-step process. The first step is to evaluate the current land availability in each geographical or administrative area within a city. This

step generally requires knowledge of land-use policies and the use of GIS tools to screen out portions of land which are not suitable for development. The second step is to calculate the land-use stress of each administrative area, which is the ratio of land-use impact caused by MSWM facilities to the available land. In our study, the level of land-use impact is calculated based on the total amount of direct and indirect land-use. The direct land-use is the actual land area used for facility establishment. The indirect land-use is estimated by land occupation LCIA methodology, which assesses the impact on land quality over a given period. The ReCiPe Midpoint (H) V1.07/Europe ReCiPe H method is used in our study to account for the land occupation impact based on the type and capacity of facilities. Previous attempts to integrate direct and indirect land-use impacts have been made in LCA studies to account for the relevant land-use impacts [34, 35].

**3.3. Optimization Model for Sustainable MSWM.** In this section, we introduce a mathematical formulation for the sustainable MSWM problem. The multiobjective mixed-integer model is a customization of the popular facility location model. We consider a 3-echelon supply chain, where solid wastes are gathered in collection centers, then moved to sorting facilities, and finally sent to either incinerators or landfills. Figure 2 shows an example of this supply chain.

We assume that decisions can be made on both locations and sizes of tier 2 and tier 3 of the supply chain (i.e., sorting

facilities and landfills/incinerators). This directly affects the capacity of each facility, its land-use, and its impact on public health. The objective is to identify locations, sizes, and routes to minimize costs, land-use, and public health impact. Superscripts  $S$ ,  $I$ , and  $L$  are used throughout the mathematical formulation to refer to sorting, incinerator, and landfill facilities, respectively.

The model's notation is given below:

#### Sets and Indices

- (i)  $I$  is the set of collection centers, indexed by  $i$ ;
- (ii)  $J$  is the set of sorting facility locations, indexed by  $j$ ;
- (iii)  $K$  is the set of incinerator locations, indexed by  $k$ ;
- (iv)  $K'$  is the set of landfill locations, indexed by  $k'$ ;
- (v)  $L(j), L(k), L(k')$  are the set of available sizes at locations  $j, k$ , and  $k'$ , indexed by  $l$ .

#### Parameters

- (i)  $t_{ij}, t_{jk}, t_{jk'}$  are the transportation costs on links  $(i, j), (j, k)$ , and  $(j, k')$ , respectively;
- (ii)  $c_{jl}^S, c_{kl}^I, c_{k'l}^L$  are the fixed costs to open facilities of size  $l$ ;
- (iii)  $C_{jl}^S, C_{kl}^I, C_{k'l}^L$  are the storage capacities of facilities of size  $l$ ;
- (iv)  $\bar{C}_{ij}, \bar{C}_{jk}, \bar{C}_{jk'}$  define the maximum amount of waste that a single trip can carry over links  $(i, j), (j, k)$ , and  $(j, k')$ , respectively;
- (v)  $o_j^S, o_k^I, o_{k'}^L$  are unit operations costs to manage the flow of solid waste, for each facility;
- (vi)  $s_{jl}^S, s_{kl}^I, s_{k'l}^L$  are land-use stress ratios for facilities of size  $l$ ;
- (vii)  $D_i$  is the amount of solid waste available at collection center  $i$ ;
- (viii)  $p_{ij}, p_{jk}, p_{jk'}$  are the number of people living nearby links  $(i, j), (j, k)$ , and  $(j, k')$ , respectively;
- (ix)  $p_{jl}^S, p_{kl}^I, p_{k'l}^L$  are the number of people living near a facility of size  $l$ ;
- (x)  $d_{ij}, d_{jk}, d_{jk'}$  are the DALYs per person, due to transportation activities on links  $(i, j), (j, k)$ , and  $(j, k')$ , respectively;
- (xi)  $d_{jl}^S, d_{kl}^I, d_{k'l}^L$  are the DALYs per person, due to size  $l$  facility operations.

#### Decision Variables

- (i)  $y_{jl}^S, y_{kl}^I, y_{k'l}^L$  are binary location variables, equal to 1 when sorting facilities, incinerators, and landfills of size  $l$  are open at their respective locations  $j, k$ , and  $k'$ ;
- (ii)  $x_{ij}, x_{jk}, x_{jk'}$  are the number of trips on links  $(i, j), (j, k)$ , and  $(j, k')$ , respectively;

- (iii)  $f_{ij}, f_{jk}, f_{jk'}$  denote the amount of solid waste transported on links  $(i, j), (j, k)$ , and  $(j, k')$ , respectively.

The cost function  $F_c$  is computed as follows.

$$F_c = \sum_{j \in J} \sum_{l \in L(j)} c_{jl}^S y_{jl}^S + \sum_{k \in K} \sum_{l \in L(k)} c_{kl}^I y_{kl}^I + \sum_{k' \in K'} \sum_{l \in L(k')} c_{k'l}^L y_{k'l}^L + \sum_{i \in I} \sum_{j \in J} (t_{ij} + o_j^S) x_{ij} + \sum_{j \in J} \sum_{k \in K} (t_{jk} + o_k^I) x_{jk} + \sum_{j \in J} \sum_{k' \in K'} (t_{jk'} + o_{k'}^L) x_{jk'} \quad (1)$$

The overall cost is the sum of the fixed costs to open sorting facilities, incinerators, and landfills plus the operational costs of transporting and managing the solid waste flow across the network.

A second function  $F_u$  is introduced to measure the average land-use stress. The function is formally defined as follows.

$$F_u = \sum_{j \in J} \sum_{l \in L(j)} s_{jl}^S y_{jl}^S + \sum_{k \in K} \sum_{l \in L(k)} s_{kl}^I y_{kl}^I + \sum_{k' \in K'} \sum_{l \in L(k')} s_{k'l}^L y_{k'l}^L \quad (2)$$

The function is the sum of all the land-use ratio across all candidate locations. Parameters  $s_{jl}, s_{kl}, s_{k'l}$  represent the ratios for land used and land available. This ratio can be computed by looking at the entire network or by narrowing down the focus to smaller districts, to compute the impact of land-use at a local level.

Finally, a third function  $F_h$  is used to evaluate the impact of transportation and facilities on the population's health. The function is defined as follows.

$$F_h = \sum_{j \in J} \sum_{l \in L(j)} p_{jl}^S d_{jl}^S y_{jl}^S + \sum_{k \in K} \sum_{l \in L(k)} p_{kl}^I d_{kl}^I y_{kl}^I + \sum_{k' \in K'} \sum_{l \in L(k')} p_{k'l}^L d_{k'l}^L y_{k'l}^L + \sum_{i \in I} \sum_{j \in J} p_{ij} d_{ij} x_{ij} + \sum_{j \in J} \sum_{k \in K} p_{jk} d_{jk} x_{jk} + \sum_{j \in J} \sum_{k' \in K'} p_{jk'} d_{jk'} x_{jk'} \quad (3)$$

The Sustainable Waste Management Network (SWMN) design model can be formulated as a multiobjective mixed-integer program.

$$[\text{SWMN}]: \min \quad \{F_c, F_u, F_h\} \quad (4)$$

$$\text{s.t.} \quad \sum_{j \in J} f_{ij} = D_i \quad \forall i \in I \quad (5)$$

$$\sum_{i \in I} f_{ij} = \sum_{k \in K} f_{jk} \quad \forall j \in J \quad (6)$$



$$\sum_{i \in I} f_{ij} \leq \sum_{l \in L(j)} C_{jl} y_{jl}^S \quad \forall j \in J \quad (7)$$

$$\sum_{j \in J} f_{jk} \leq \sum_{l \in L(k)} C_{kl} y_{kl}^I \quad \forall k \in K \quad (8)$$

$$\sum_{j \in J} f_{jk'} \leq \sum_{l \in L(k')} C_{k'l} y_{k'l}^L \quad \forall k' \in K' \quad (9)$$

$$f_{ij} \leq \bar{C}_{jl} x_{ij} \quad \forall i \in I, j \in J \quad (10)$$

$$f_{jk} \leq \bar{C}_{kl} x_{jk} \quad \forall j \in J, k \in K \quad (11)$$

$$f_{jk'} \leq \bar{C}_{k'l} x_{jk'} \quad \forall j \in J, k' \in K' \quad (12)$$

$$\sum_{l \in L(j)} y_{jl}^S \leq 1 \quad \forall j \in J \quad (13)$$

$$\sum_{l \in L(k)} y_{kl}^I \leq 1 \quad \forall k \in K \quad (14)$$

$$\sum_{l \in L(k')} y_{k'l}^L \leq 1 \quad \forall k' \in K' \quad (15)$$

$$f_{ij}, f_{jk}, f_{jk'} \geq 0 \quad \forall i, j, k, k' \quad (16)$$

$$y_{jl}^S, y_{kl}^I, y_{k'l}^L \in \{0, 1\} \quad \forall j, k, k', l \quad (17)$$

$$x_{ij}, x_{jk}, x_{jk'} \in \mathbb{Z}^+ \quad \forall i, j, k, k' \quad (18)$$

The objective (4) is to minimize the costs, land-use ratios, and impact of transportation. Constraints (5) state that the outflow of waste from any given collection center  $i$  must be equal to  $D_i$ . Equations (6) are flow balance constraints, enforcing that the inflow at a sorting facility  $j$  is entirely forwarded to incinerators  $k$  and landfills  $k'$ . Constraints (7)-(9) are the size-dependent capacity constraints at any facility of the network. Constraints (10)-(12) are the transportation capacity limitations. Inequalities (13)-(15) enforce that only one size can be selected if a facility is open at a given location. Constraints (16)-(18) define the decision variables' domains.

**3.4. Solution Algorithms.** The computational results, displayed in Section 4, are obtained from various optimization models solved under single-objective functions and multiobjective functions. We refer to these models using the notation SWMN(), where we define inside the brackets the objectives being optimized together. For example, SWMN( $F_c, F_h$ ) is the multiobjective problem optimizing both costs and health impact whereas SWMN( $F_u$ ) is the single-objective model minimizing only the land-use impact. Single-objective models can be solved with numerical methods such as branch and bound method. As for multiobjective models, a min-max approach is implemented to identify solutions that minimize the deviations from the ideal results. Formally, let  $F_c^*$ ,  $F_u^*$ , and  $F_h^*$  be the optimal value obtained by solving the respective single-objective problems. Furthermore, let  $F_c^{max}$ ,  $F_u^{max}$ , and  $F_h^{max}$  upper bounds be set equal to the worst value obtained by each function across the single-objective problems. We can

now define the deviation levels between each objective and its ideal target by normalizing the functions as follows.

$$\sigma_c = \frac{F_c - F_c^*}{F_c^{max} - F_c^*} \quad (19)$$

$$\sigma_u = \frac{F_u - F_u^*}{F_u^{max} - F_u^*} \quad (20)$$

$$\sigma_h = \frac{F_h - F_h^*}{F_h^{max} - F_h^*} \quad (21)$$

By adding a new continuous decision variable  $z$ , we can formulate a min-max goal attainment on the deviations of three objectives as follows.

$$[\text{SWMN}(F_c, F_u, F_h)] : \min \quad z \quad (22)$$

$$\text{s.t.} \quad \sigma_c \leq z \quad (23)$$

$$\sigma_u \leq z \quad (24)$$

$$\sigma_h \leq z \quad (25)$$

$$(5) - (18) \quad (26)$$

The objective of min-max SWMN is to minimize the largest deviation from optimal targets across the three functions considered. The model can be easily modified to target only two objectives.

## 4. Case Study

In this section, we apply the proposed SWMN model to a case study in Pathum Thani province in Thailand. The waste management system in Thailand is of interest as very little planning has been used in the past to locate waste facilities. Nowadays, among a total of 2,490 municipal solid waste sites, only 499 (20%) adopt safety standards, such as sanitary/engineer landfill, controlled dumping, and incineration with air pollution-controlled system [36]. The rest of the sites implement unhealthy practices such as open dumping and incineration without air pollution-controlled system. As a consequence, the uncontrolled release of leachates and gases from these dumps is frequent. This is taking a toll on both environment and society. For example, in 2010 a fire broke out at the Phraeksa dump site causing hundreds of residents to flee the area. This incident prompted local governments across Thailand to reexamine their regulatory approaches in health and environmental safety.

Pathum Thani is located in the central region of Thailand, within the Bangkok metropolitan area (Figure 3). It covers a total area of 1,525.9 km<sup>2</sup>, organized into 7 districts, 60 subdistricts, and 529 villages. Its population has steadily increased in the past decade to about 1.2 million. As a result, the amount of waste generated increased to 0.612 million tonnes/year [36]. As the province is quite vast, we narrow the focus to Muang Pathum Thani, Sam Khok, and Lat Lum Kaeo districts for our case study (Figure 4). Currently, solid wastes from communities are gathered at collection centers across

TABLE 1: Safety distance from waste facilities.

Criteria	Sorting facility (km)	Incinerator (km)	Landfill (km)
Residential area	1	2	1
Archaeological heritage site	1	1	1
River	1	1	1
Pond	-	0.3	0.3
Main road	-	0.3	0.3



FIGURE 3: Central Thailand.

subdistricts. Wastes at collection centers are transported to sorting facilities and subsequently sent to either incinerators or landfills. Recycling is purposely not included in this study as it is currently done by private companies. The validation of the proposed model is done by determining the location and size of sorting and disposal facilities.

#### 4.1. Parameters

**Land Availability Assessment and Candidate Locations.** In order to quantify the land available, an initial screening is necessary to exclude from the analysis places such as rivers, ponds, main roads, archaeological heritage sites, and residential areas. We further include a buffering area to guarantee safe distances between waste sites. As shown in Table 1, the size of these buffers is set according to the regulation and guideline of MSWM developed by the Pollution Control Department (PCD).

A number of polygons are identified by combining available land with subdistrict boundaries. Based on their sizes, these polygons are further divided and their centroid is selected as a facility location. Available land for landfill siting is shown in Figure 4. Since candidate locations of the three types of facilities overlap, we further constrain the mathematical model to avoid colocations.

To ensure environmental justice across the three districts, we perform the land-use assessment at the subdistrict level. For any location within a subdistrict, we compute parameters  $s_{jl}^S$ ,  $s_{kl}^I$  and  $s_{k'l}^L$  as the following ratio.

$$\frac{\text{Direct + Indirect land use, with facility of size } l}{\text{Total land available in the sub-district}} \quad (27)$$

For each facility, 3 capacity levels are considered: small, medium, and large. Their direct and indirect land-use are shown in Table 2. The values are estimated by land occupation LCIA methodology as previously described in our land-use impact assessment.

**Traveling Distances.** All wastes are transported from the collection centers to the sorting facilities using 16-tonne (light) trucks. From the sorting facilities to both the incinerator and landfill sites, 32-tonne (heavy) trucks are used. Traveling distances are estimated using ArcGIS and a digital map of Pathum Thani. Specifically, having defined the candidate locations, we use a network analysis tool to determine the shortest routes.

**Public Health Impact Assessment.** To measure the public health impact, this study computes the DALYs for people affected by the MSWM system. The first step is to estimate the number of people living near the supply chain. No data is currently available showing the population distribution at the household level. ArcGIS is used to count the total number of residential buildings surrounding the waste transportation routes and MSWM facilities.

To estimate the damage to public health caused by an MSWM system, this study adopts the emission-to-exposure model used by Gouge et al. [37] and Greco et al. [38]. They use several buffer distances to estimate the health impact caused by transport pollution. In our study, we set the buffer distance for transportation routes to 100 m. For MSWM facilities, the affected areas are assumed to increase with the waste disposal dimension. Moreover, due to a long history of unsanitary practices in Thailand, landfills are assumed to have a larger impact than other MSWM facilities (Table 3).

The estimation of the number of affected people is made based on the information of the total population living in the area and the number of residential buildings. The proposed estimation approach is expected to give a more accurate estimate of affected people than the previous approach [39]. Varying impact distances, corresponding to different types and sizes of facilities, are used instead of one single impact distance. The second step is to multiply the number of affected

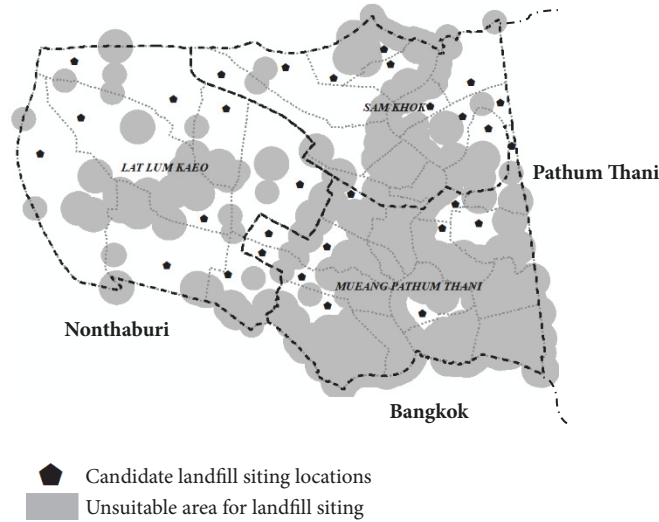


FIGURE 4: Landfill locations.

TABLE 2: Direct and indirect land-use impacts.

Waste facility	Direct land-use ( $m^2$ )			Indirect land-use ( $m^2$ )		
	Small size	Medium size	Large size	Small size	Medium size	Large size
Sorting facility	4,800	8,000	16,000	6,191	10,780	21,559
Incinerator	8,000	16,000	24,000	11,971	23,941	35,911
Landfill	80,000	160,000	192,000	127,770	255,525	335,295

TABLE 3: Facilities impact on public health.

Facilities	Sized	Capacity (tonne/day)	Affected areas (km)	DALYs
Sorting Facility	Small	50	0.3	0.07
	Medium	100	0.6	0.14
	Large	300	1.8	0.28
Incinerator	Small	50	0.5	5.95
	Medium	100	1	11.9
	Large	150	1.5	17.85
Landfill	Small	50	1	3.89
	Medium	100	2	7.78
	Large	150	3	11.66

people near the transportation routes and MSWM facilities by the DALYs coefficients, which are summarized in Tables 3 and 4. The number of DALYs is estimated based on a 30-year operational period, using the ReCiPe 2008 Endpoint LCIA method [33]. Only local-scale environmental impacts are taken into account, including human toxicity, photochemical oxidant formation, PM formation, and ionizing radiation. The damage due to climate change and ozone depletion is neglected. Finally, the capacity of waste sites and trucks is selected from those available in SimaPro 7.3 (LCA software) which covers a wide range of typical facilities and operations. The entire dataset can be found in Kachapanya [40].

**4.2. Results and Discussion.** This section presents the results from the computational analysis which is carried out on a Windows 10 machine using an Intel i7-6700HQ processor with 8 GB of RAM. CPLEX 12.6 optimization studio is used to solve the mathematical models. The analysis is organized in two subsections: single-objective and multiobjective optimizations.

The results are interpreted in terms of satisfaction levels of each objective. A satisfaction level of 100%, indicates that the objective is equal to its optimal level. Conversely, smaller satisfaction levels indicate that there is a gap between the objective and its best achievable target. Formally, these levels

TABLE 4: Transportation impact on public health.

Fleet types	Capacity (ton)	Affected areas (km)	DALYs Empty load (per km)	DALYs full load (per tkm)
Light truck	16	0.1	$5.82 \cdot 10^{-7}$	$5.62 \cdot 10^{-8}$
Heavy truck	32	0.1	$1.16 \cdot 10^{-6}$	$1.12 \cdot 10^{-7}$

TABLE 5: Results of single-objective optimization.

		Minimizing Cost	Minimizing Avg. Land-Use Stress	Minimizing Public Health Impact
Total Cost (USD)	Total transportation cost	11,973,205	23,519,346	9,658,807
	Total land cost	7,753,691	1,854,007	21,035,012
	Total construction cost	17,114,396	19,456,136	27,692,869
	Total operation cost	14,103,023	19,823,517	16,354,068
	<i>Total cost</i>	<i>50,944,315</i>	<i>64,653,006</i>	<i>74,740,757</i>
	<i>Satisfaction Level</i>	<i>100%</i>	<i>42%</i>	<i>0%</i>
Land-Use Impact	<i>Avg. Land-use Stress</i>	<i>0.573</i>	<i>0.028</i>	<i>0.985</i>
	<i>Satisfaction Level</i>	<i>43%</i>	<i>100%</i>	<i>0%</i>
Public Health Impact (DALY)	Transportation	13,042	47,868	6,214
	Waste Facilities	31,639	10,072	1,010
	<i>Total Public Health Impact</i>	<i>44,681</i>	<i>57,940</i>	<i>7,224</i>
	<i>Satisfaction Level</i>	<i>26%</i>	<i>0%</i>	<i>100%</i>
Average Satisfaction Level		56%	47%	33%
Computational Time (Sec.)		24	10	13

are obtained from the deviation values introduced in the methodology section (i.e.,  $1 - \sigma_c$ ,  $1 - \sigma_u$ , and  $1 - \sigma_h$ ).

#### 4.2.1. Single-Objective Optimization

**Total Cost Optimization (SWMN( $F_c$ )).** Figure 5(a) shows the optimal layout of the MSWM system obtained by solving SWMN( $F_c$ ). Results show that when cost is the main driver, the number of facilities is low. Out of 37 candidate locations for sorting facilities, only 6 are selected. Similarly, only 1 incinerator and 3 landfills are selected. While these facilities provide sufficient capacity for all subdistrict, they also generate routes longer than 30 km.

The summary of costs and the average land-use stress associated with the solution are shown in Table 5 under the column named “Minimizing Cost”. The construction cost is the largest cost. As a consequence, landfills are chosen over incinerators. The cost-optimum of the MSWM layout is about 50,944,315 US dollars composed of the transportation (11,973,205 US dollars), land (7,753,691 US dollars), construction (17,114,396 US dollars), and operational costs (14,103,023 US dollars). The average land-use stress and public health impact for this solution are estimated at 0.573 and 44,681 DALYs, respectively. The public health impact is from transportation (13,042 DALY) and waste facilities (31,639 DALYs). This suggests that although SWMN( $F_c$ ) achieves minimum cost, it comes at the expense of public health and land-use. This is mostly due to disposal sites located close

to the urban areas where waste is generated, resulting in excessive land-use stress and high public health impact.

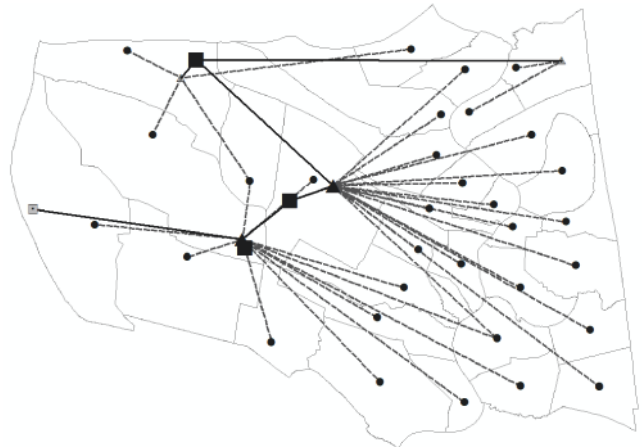
**Average Land-Use Stress Optimization (SWMN( $F_u$ )).** Number and location of facilities obtained by solving SWMN( $F_u$ ) are shown in Figure 5(b). Again, the number of open facilities is relatively small. Four sorting facilities and four incinerators are selected. The locations are different, compared to the SWMN( $F_c$ ) case. Sorting facilities and incinerators are located in rural areas, reducing the excessive land-use stress returned by the cost minimization model. No landfill is selected. However, displacing facilities away from urban areas leads to high transportation cost and high public health impact.

The results of average land-use stress are shown in Table 5 under the column named “Minimizing Avg. Land-Use Stress”. The results show that the total cost is 64,653,006 US dollars consisting of transportation cost (24,053,902 US dollars), land cost (1,896,146 US dollars), construction cost (19,898,341 US dollars), and disposal cost (20,274,072 US dollars). The average land-use stress of the MSWM system drops to 0.028. The public health impact is 57,940 DALYs (47,868 DALYs from transportation and 10,072 DALYs from waste facilities).

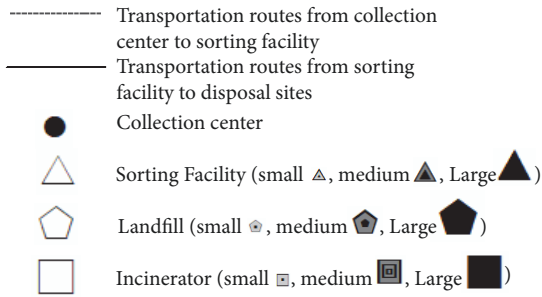
**Public Health Impact Optimization (SWMN( $F_h$ )).** Solving SWMN( $F_h$ ) leads to a layout that is quite different. Most of the facilities are sparsely located across the region (Figure 5(c)).



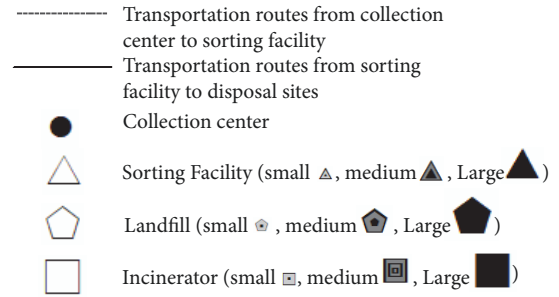
Optimal layout of MSWM system under cost objective



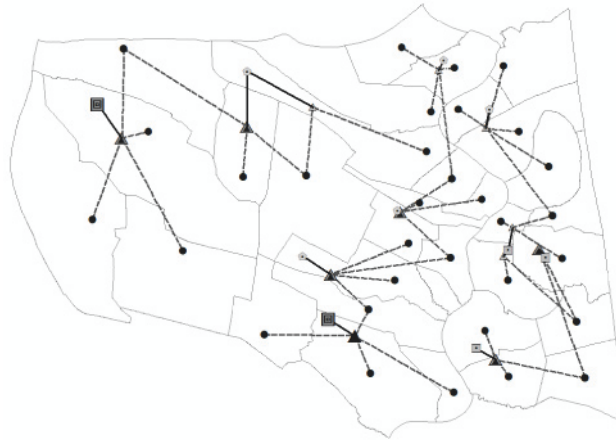
Optimal layout of MSWM system under Avg. land use stress objective



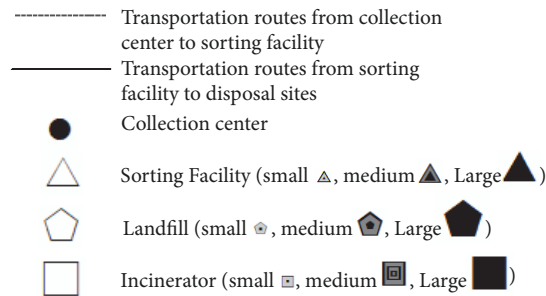
(a)



(b)



Optimal layout of MSWM system under public health impact objective



(c)

FIGURE 5: Optimal layouts of MSWM system under single-objective functions.



TABLE 6: Results of multiobjective optimization.

		Bi-objective			Tri-objective
		Minimizing Cost and Avg. Land Use Stress	Minimizing Cost and Public Health Impact	Minimizing Avg. Land Use Stress and Public Health	Minimizing Cost, Avg. Land Use Stress, and Public Health
Total Cost (USD)	Total transportation cost	13,123,748	12,309,316	12,128,780	10,313,519
	Total land cost	6,692,561	10,973,973	12,124,521	10,511,253
	Total construction cost	22,465,381	21,349,224	23,941,083	22,864,008
	Total operation cost	17,448,975	16,729,273	19,823,509	17,268,910
	<i>Total cost</i>	<i>59,730,664</i>	<i>61,361,787</i>	<i>68,017,894</i>	<i>60,957,690</i>
<i>Satisfaction Level</i>		<i>63%</i>	<i>56%</i>	<i>28%</i>	<i>58%</i>
Land Use Impact	<i>Avg. Land-use Stress</i>	<i>0.159</i>	<i>0.857</i>	<i>0.242</i>	<i>0.466</i>
	<i>Satisfaction Level</i>	<i>86%</i>	<i>13%</i>	<i>78%</i>	<i>54%</i>
Public Health Impact (DALY)	Transportation	24,483	17,704	8,997	8,384
	Waste Facilities	31,117	5,050	4,018	1,342
	<i>Total Public Health Impact</i>	<i>55,600</i>	<i>22,754</i>	<i>13,015</i>	<i>9,726</i>
	<i>Satisfaction Level</i>	<i>5%</i>	<i>69%</i>	<i>89%</i>	<i>95%</i>
<i>Average Satisfaction Level</i>		<i>51%</i>	<i>46%</i>	<i>65%</i>	<i>69%</i>
Computational Time (Sec.)		37	35	33	82

A total of 22 facilities are opened (12 sorting facilities, 5 landfills, and 5 incinerators). Results in Table 5 show that the optimal public health drops to 7,224 DALYs. The average land-use stress and total cost increase to 0.985 and 74,740,757 US dollars, respectively. As expected, the land cost (21,035,012 US dollars) and construction cost (27,692,869 US dollars) are mostly responsible for the cost increment. Clearly, reducing public health is mostly a consequence of reducing the transportation routes. This is also evident by looking at the transportation cost (9,658,807 US dollars) which decreases considerably. This is achieved by locating a large number of facilities sparsely across the area. Consequently, the land-use ratio is bound to worsen dramatically.

To sum up, the transportation cost reaches its minimum when  $SWMN(F_h)$  is solved. Its value is 20% smaller than  $SWMN(F_c)$ . However, the land cost from  $SWMN(F_u)$  is lower than  $SWMN(F_c)$  by 76%, and it is lower than  $SWMN(F_h)$  by 91%. For the total construction cost,  $SWMN(F_c)$  is 12% lower than  $SWMN(F_u)$  because landfills have lower construction cost than incinerators, and it is 38% lower than  $SWMN(F_h)$  because the number of sorting facilities is smaller. The reason is that incinerators reduce the amount of land required, and the cost of land in urban areas is typically more expensive. The lowest operational cost is from  $SWMN(F_c)$ . Moreover, focusing on the land-use stress impact,  $SWMN(F_u)$  results in lower values compared to  $SWMN(F_c)$  and  $SWMN(F_h)$  (95% and 97%, respectively). Finally, regarding the public health impact, the  $SWMN(F_h)$  result is lower than  $SWMN(F_c)$  by

approximately 84% and lower than  $SWMN(F_u)$  by approximately 88%.

**4.2.2. Multiobjective Optimization.** The results and the layouts of MSWM networks are shown in Tables 6 and 7 and Figure 6.

**Total Cost and Average Land-Use Stress Optimization ( $SWMN(F_c, F_u)$ ).** When the problem is solved minimizing costs and land-use, sorting facilities and incinerators are evenly distributed between urban and rural areas. The landfills are located in rural areas due to larger land requirements (Figure 6(a)). This scenario shows the tradeoff between costs and land-use stress. Looking at  $SWMN(F_c)$ , land-use stress satisfaction increases from 43% to 86% but it deteriorates the total cost satisfaction level by 37% (Table 6). Due to the higher potential impact of landfill facilities on the overall land-use stress,  $SWMN(F_c, F_u)$  reduces the number of facilities to only one large facility (Table 7). However, to satisfy the total demand, three large incinerators are built. Furthermore, the number of sorting facilities increases to 11 (7 small, 1 medium, and 3 large). As expected, this scenario has high satisfaction levels of cost (63%) and land-use stress (86%). Nevertheless, the satisfaction level of the public health objective is extremely low (5%).

**Total Cost and the Public Health Impact ( $SWMN(F_c, F_h)$ ).** The results of optimizing costs and public health show that

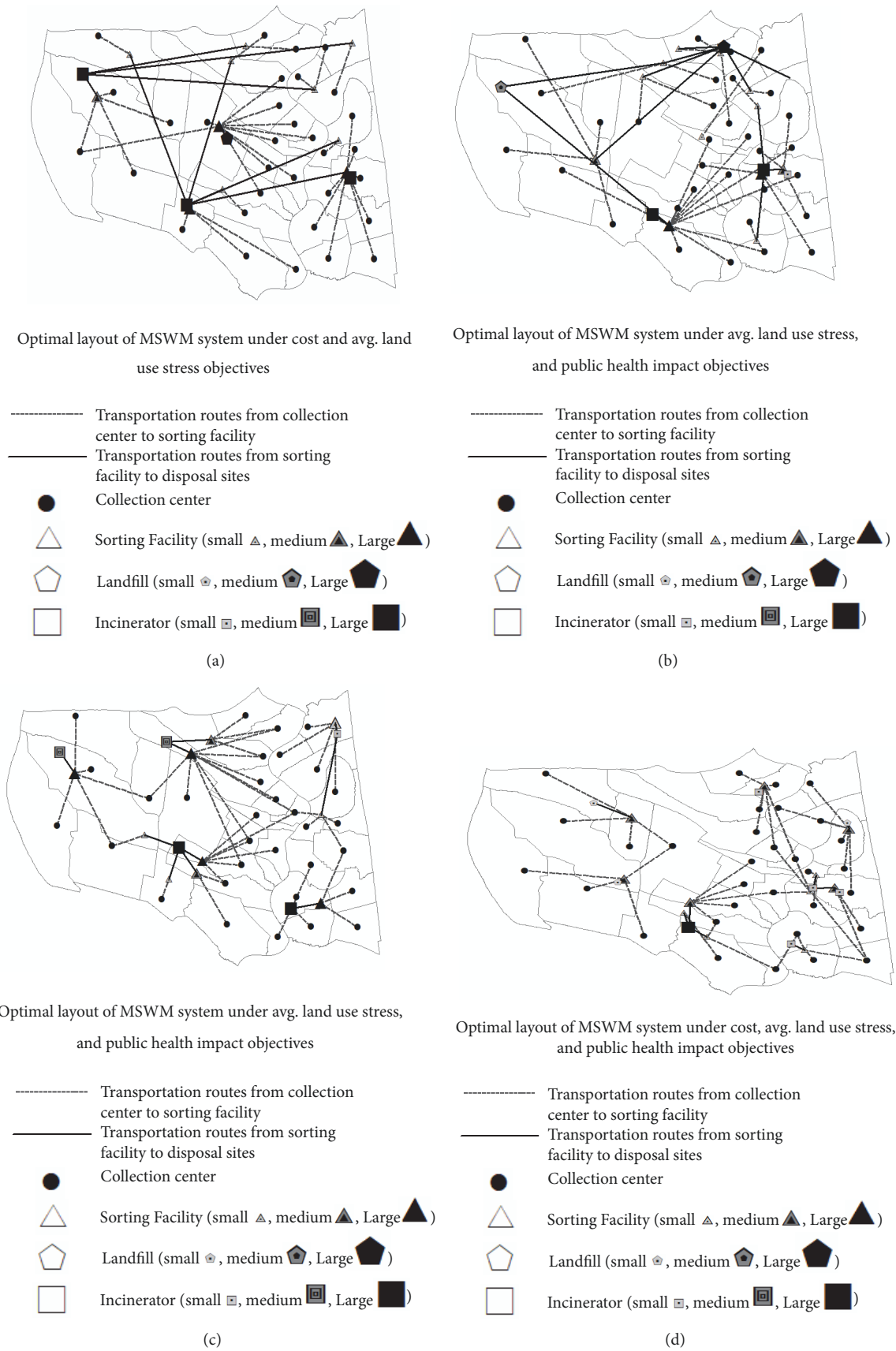


FIGURE 6: Optimal layouts of MSWM system under multiobjective functions.

TABLE 7: Number of facilities in the optimization scenarios.

Scenario	Sorting Facility			Incinerator			Landfill			Avg. satisfaction level
	Small size	Medium size	Large size	Small size	Medium size	Large size	Small size	Medium size	Large size	
Minimizing cost	1	3	2	0	0	1	1	0	2	56%
Minimizing avg. land use stress	2	0	2	1	0	3	0	0	0	47%
Minimizing public health impact	5	6	1	3	2	0	5	0	0	33%
Minimizing cost and avg. land use stress	7	1	3	0	0	3	0	0	1	51%
Minimizing cost and public health impact	3	3	3	0	0	2	2	0	1	46%
Minimizing avg. land use stress and public health	4	2	5	1	2	2	0	0	0	65%
Minimizing cost, avg. land use stress, and public health	6	6	0	4	0	1	3	0	0	69%

all facilities are sparsely distributed in urban and rural areas (Figure 6(b)). This is because the two objectives share the common goal of reducing the transportation routes so that both health and costs are reduced. From Table 7, the sorting facilities are now three for each size. Two large incinerators are built and three landfills locations are selected (2 small 1 large). The average satisfaction is 46%. The satisfaction level of cost is 56% while the satisfaction level of public health is 69% (Table 6). This result suggests a clear conflict between these two objectives, despite the common goal of reducing transportation routes.

*Average Land-Use Stress and Public Health Impact* ( $SWMN(F_w, F_h)$ ). Optimizing land-use and health impact disregarding costs results in selecting only incinerators (1 small, 2 medium, and 2 large) due to their lower land-use as opposed to landfills (Figure 6(c)). The number of sorting facilities increases to 11, 5 of which are large, 2 medium, and 4 small. The average satisfaction level is 65% as land-use stress (78%) and public health impact (89%) simultaneously reach high levels (Table 6). However, the satisfaction level of the total cost is reduced dramatically to 28%, suggesting that this solution is not practical.

*Total Cost, Average Land-Use Stress and Public Health Impact* ( $SWMN(F_c, F_w, F_h)$ ). Previous results have shown that both single- and biobjective models fail to reach a good compromise. Therefore, the problem is solved minimizing the three objectives at the same time. This leads to an MSWM system layout that is well balanced across the region (Figure 6(d)) as this model mainly chooses small and medium-sized facilities (Table 7). Normally, it is difficult to obtain solutions with high satisfaction levels across all conflicting objectives. However, this scenario gives the highest average satisfaction level (69%) and offers an effective compromise between the three objectives as each satisfaction level is above 50%.

## 5. Concluding Remarks

In this paper, we propose a novel network design optimization model for MSWM which accounts for sustainability in a comprehensive way. Specifically, we incorporate environmental and social impact indicators with the economic objective. A formal methodology is introduced to model public health and land-use impacts. The first is measured in terms of DALYs imposed by the waste operations on



the population living close to the supply chain. To enforce a fair use across a city's subdistricts, the land-use metric is computed as the ratio between used and available land. The multiobjective formulation is translated into a single-objective model aiming to minimize the maximum gap of each objective from its optimal target.

A case study in Pathum Thani (Thailand) is developed to validate the model while also providing results that can be of public interest, to increase awareness and engage with local stakeholders. The single-objective analysis highlights the fact that focusing only on cost generates a supply chain which imposes a serious burden on society. In fact, the resulting land-use and public health metrics are far from sustainable. However, single-objective models focusing on public health or land-use are inefficient and they deteriorate the metrics outside the objective. This further motivates the multiobjective approach studied in this paper, where a model is proposed to minimize the deviation of each criterion from its optimal target. The best tradeoff between the metrics is indeed achieved when all dimensions are considered simultaneously.

The scope of this work, together with an increasing push for a wide-range research focus on sustainability, provides several interesting further extensions. Integrated modeling approaches should be developed to simultaneously consider interrelated supply chain decisions, as demonstrated by Mota and et al. [41]. To this aim, optimization models can be developed to incorporate facility location decisions with other decisions, such as waste collection schemes, transport modes, and disposal technologies. This will require developing and solving complex multiobjective location-routing problems. Another line of research should focus on incorporating uncertainty into the problem. It is clearly of interest to investigate the extent to which the problem's features such as the amount of waste, the transportation costs, and their inherent uncertainties can impact sustainability metrics and costs. Finally, given the complexity of the current model and its possible extensions, a further research direction should focus on the development of efficient solution algorithms to obtain good solutions on large realistic networks.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] L. Bastin and D. M. Longden, "Comparing transport emissions and impacts for energy recovery from domestic waste (EfW): Centralised and distributed disposal options for two UK counties," *Computers, Environment and Urban Systems*, vol. 33, no. 6, pp. 492–503, 2009.
- [2] F. Ncube, E. J. Ncube, and K. Voyi, "A systematic critical review of epidemiological studies on public health concerns of municipal solid waste handling," *Perspectives in Public Health*, vol. 137, no. 2, pp. 102–108, 2016.
- [3] L. A. Nwaogu, C. O. Ujowundu, C. I. Iheme, T. N. Ezejiyor, and D. C. Belonwu, "Effect of sublethal concentration of heavy metal contamination on soil physicochemical properties, catalase and dehydrogenase activities," *International Journal of Biochemistry Research & Review*, vol. 4, no. 2, pp. 141–149, 2014.
- [4] G. Owusu, E. Nketiah-Amponsah, S. N. A. Codjoe, and R. L. Afutu-Kotey, "How do Ghanas landfills affect residential property values? A case study of two sites in Accra," *Urban Geography*, vol. 35, no. 8, pp. 1140–1155, 2014.
- [5] V. Spoann, T. Fujiwara, B. Seng, and C. Lay, "Municipal solid waste management: Constraints and opportunities to improve capacity of local government authorities of Phnom Penh Capital," *Waste Management & Research*, vol. 36, no. 10, pp. 985–992, 2018.
- [6] G. Owusu, "Social effects of poor sanitation and waste management on poor urban communities: a neighborhood-specific study of Sabon Zongo, Accra," *Journal of Urbanism*, vol. 3, no. 2, pp. 145–160, 2010.
- [7] V. Ibáñez-Forés, M. D. Bovea, C. Coutinho-Nóbrega, and H. R. de Medeiros, "Assessing the social performance of municipal solid waste management systems in developing countries: Proposal of indicators and a case study," *Ecological Indicators*, vol. 98, pp. 164–178, 2019.
- [8] P. Ferrão, P. Ribeiro, J. Rodrigues et al., "Environmental, economic and social costs and benefits of a packaging waste management system: a portuguese case study," *Resources, Conservation & Recycling*, vol. 85, pp. 67–78, 2014.
- [9] H. Yuan, "A model for evaluating the social performance of construction waste management," *Waste Management*, vol. 32, no. 6, pp. 1218–1228, 2012.
- [10] H. Ak and W. Braidia, "Sustainable municipal solid waste management decision making: Development and implementation of a single score sustainability index," *Management of Environmental Quality: An International Journal*, vol. 26, no. 6, pp. 909–928, 2015.
- [11] J. L. Moura, A. Ibeas, and L. dell'Olio, "Optimization-simulation model for planning supply transport to large infrastructure public works located in congested urban areas," *Networks and Spatial Economics*, vol. 10, no. 4, pp. 487–507, 2010.
- [12] E. M. Ferguson, J. Duthie, A. Unnikrishnan, and S. T. Waller, "Incorporating equity into the transit frequency-setting problem," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 1, pp. 190–199, 2012.
- [13] H. R. Jafari, M. Seifbarghy, and M. Omidvari, "Sustainable supply chain design with water environmental impacts and justice-oriented employment considerations: a case study in textile industry," *Scientia Iranica*, vol. 24, no. 4, pp. 2119–2137, 2017.
- [14] K. Manaugh, M. G. Badami, and A. M. El-Geneidy, "Integrating social equity into urban transportation planning: a critical evaluation of equity objectives and measures in transportation plans in north america," *Transport Policy*, vol. 37, pp. 167–176, 2015.
- [15] M. O. Beiler and M. Mohammed, "Exploring transportation equity: Development and application of a transportation justice framework," *Transportation Research Part D: Transport and Environment*, vol. 47, pp. 285–298, 2016.

- [16] A. Bernstad Saraiva, R. G. Souza, C. F. Mahler, and R. A. B. Valle, "Consequential lifecycle modelling of solid waste management systems – Reviewing choices and exploring their consequences," *Journal of Cleaner Production*, vol. 202, pp. 488–496, 2018.
- [17] A. P. Rodrigues, M. L. Fernandes, M. F. F. Rodrigues, S. C. Bortoluzzi, S. E. Gouvea da Costa, and E. Pinheiro de Lima, "Developing criteria for performance assessment in municipal solid waste management," *Journal of Cleaner Production*, vol. 186, pp. 748–757, 2018.
- [18] Y. M. Zhang, G. H. Huang, and L. He, "An inexact reverse logistics model for municipal solid waste management systems," *Journal of Environmental Management*, vol. 92, no. 3, pp. 522–530, 2011.
- [19] E. A. Toso and D. Alem, "Effective location models for sorting recyclables in public management," *European Journal of Operational Research*, vol. 234, no. 3, pp. 839–860, 2014.
- [20] D. Inghels, W. Dullaert, and D. Vigo, "A service network design model for multimodal municipal solid waste transport," *European Journal of Operational Research*, vol. 254, no. 1, pp. 68–79, 2016.
- [21] Z. Xu, A. Elomri, S. Pokharel, Q. Zhang, X. G. Ming, and W. Liu, "Global reverse supply chain design for solid waste recycling under uncertainties and carbon emission constraint," *Waste Management*, vol. 64, pp. 358–370, 2017.
- [22] M. Eskandarpour, P. Dejax, J. Miemczyk, and O. Péton, "Sustainable supply chain network design: An optimization-oriented review," *OMEGA - The International Journal of Management Science*, vol. 54, pp. 11–32, 2015.
- [23] H. Khandelwal, H. Dhar, A. K. Thalla, and S. Kumar, "Application of life cycle assessment in municipal solid waste management: a worldwide critical review," *Journal of Cleaner Production*, vol. 209, pp. 630–654, 2019.
- [24] P. Yadav and S. R. Samadder, "A critical review of the life cycle assessment studies on solid waste management in Asian countries," *Journal of Cleaner Production*, vol. 185, pp. 492–515, 2018.
- [25] N. S. Kubanza, D. K. Das, and D. Simatele, "Some happy, others sad: exploring environmental justice in solid waste management in kinshasa, the democratic republic of congo," *Local Environment*, vol. 22, no. 5, pp. 595–620, 2017.
- [26] N. S. Kubanza and D. Simatele, "Social and environmental injustices in solid waste management in sub-saharan africa: a study of kinshasa, the democratic republic of congo," *Local Environment*, vol. 21, no. 7, pp. 866–882, 2016.
- [27] N. S. Kubanza and D. Simatele, "Sustainable solid waste management in sub-Saharan African cities: application of system thinking and system dynamic as methodological imperatives in kinshasa, the democratic republic of congo," *Local Environment*, vol. 23, no. 2, pp. 220–238, 2018.
- [28] K. Kelobonye, G. McCarney, J. C. Xia, M. S. H. Swapan, F. Mao, and H. Zhou, "Relative accessibility analysis for key land uses: a spatial equity perspective," *Journal of Transport Geography*, vol. 75, pp. 82–93, 2019.
- [29] J. Njeru, "The urban political ecology of plastic bag waste problem in Nairobi, Kenya," *Geoforum*, vol. 37, no. 6, pp. 1046–1058, 2006.
- [30] J. Agyeman and T. Evans, "Toward just sustainability in urban communities: building equity rights with sustainable solutions," *Annals of the American Academy of Political and Social Science*, vol. 590, no. 1, pp. 35–53, 2003.
- [31] J. M. Norton, S. Wing, H. J. Lipscomb, J. S. Kaufman, S. W. Marshall, and A. J. Cravey, "Race, wealth, and solid waste facilities in North Carolina," *Environmental Health Perspectives*, vol. 115, no. 9, pp. 1344–1350, 2007.
- [32] C. J. L. Murray, "Quantifying the burden of disease: the technical basis for disability-adjusted life years," *Bulletin of the World Health Organization*, vol. 72, no. 3, pp. 429–445, 1994.
- [33] M. Goedkoop, R. Heijungs, M. Huijbregts, A. De Schryver, J. V. Z. R. Struijs, and R. V. Zelm, "ReCiPe 2008, A life cycle impact assessment method which comprises harmonised category indicators at the midpoint and the endpoint level," First edition Report I: Characterisation, 2009, <http://www.lcia-recipe.net>.
- [34] L. M. I. Canals, C. Bauer, J. Depestele et al., "Key elements in a framework for land use impact assessment within LCA," *The International Journal of Life Cycle Assessment*, vol. 12, no. 1, pp. 5–15, 2007.
- [35] T. Koellner, L. De Baan, T. Beck et al., "UNEP-SETAC guideline on global land use impact assessment on biodiversity and ecosystem services in LCA," *The International Journal of Life Cycle Assessment*, vol. 18, no. 6, pp. 1188–1202, 2013.
- [36] PCD, "Solid waste of community management in 2016," Pollution Control Department Report, Pollution Control Department Thailand, 2016.
- [37] B. Gouge, H. Dowlatabadi, and F. J. Ries, "Minimizing the health and climate impacts of emissions from heavy-duty public transportation bus fleets through operational optimization," *Environmental Science & Technology*, vol. 47, no. 8, pp. 3734–3742, 2013.
- [38] S. L. Greco, A. M. Wilson, J. D. Spengler, and J. I. Levy, "Spatial patterns of mobile source particulate matter emissions-to-exposure relationships across the United States," *Atmospheric Environment*, vol. 41, no. 5, pp. 1011–1025, 2007.
- [39] S. Olapiriyakul, "Designing a sustainable municipal solid waste management system in Pathum Thani, Thailand," *International Journal of Environmental Technology and Management*, vol. 20, no. 1-2, pp. 37–59, 2017.
- [40] W. Kachapanya, *Sustainable solid waste management system in urban areas of Pathum Thani Thailand. (Master of Science; Management Mathematics)*, Sirindhorn International Institute of Technology, Thammasat University, 2018, <https://drive.google.com/drive/folders/14Okc0vUmjCHbgkBCdqLbPR1Gy7IDVaVg>.
- [41] B. Mota, M. I. Gomes, A. Carvalho, and A. P. Barbosa-Povoa, "Sustainable supply chains: An integrated modeling approach under uncertainty," *OMEGA - The International Journal of Management Science*, vol. 77, pp. 32–57, 2018.

## Research Article

# Optimizing Transportation Network of Recovering End-of-Life Vehicles by Compromising Program in Polymorphic Uncertain Environment

Jing Zhang,<sup>1,2</sup> Jingjing Liu,<sup>2</sup> and Zhong Wan <sup>2</sup>

<sup>1</sup>Experimental Teaching Center, Guangdong University of Foreign Studies, Guangdong, Guangzhou, China

<sup>2</sup>School of Mathematics and Statistics, Central South University, Changsha, China

Correspondence should be addressed to Zhong Wan; wanmath@163.com

Received 13 November 2018; Revised 10 February 2019; Accepted 14 March 2019; Published 12 May 2019

Guest Editor: Helena Ramalhinho

Copyright © 2019 Jing Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With rapid development of technology and improvement of living standards, the per capita holding of automobiles greatly increases, and the amount of end-of-life vehicles (ELVs) becomes larger and larger such that it is valuable to investigate an effective strategy for recycling ELVs from the viewpoints of environmental protection and resource utilization. In this paper, an optimization model with fuzzy and stochastic parameters is built to formulate the transportation planning problems of recycling ELVs in polymorphic uncertain environment, where the unit processing and transportation costs, the selling price of reused items, and the fixed cost are all fuzzy, while the demand in secondary market and the production capacity are random owing to features underlying the practical data. For this complicated polymorphic uncertain optimization model, a unified compromising approach is proposed to hedge the uncertainty of this model such that some powerful optimization algorithms can be applied to make an optimal recycling plan. Then, an interactive algorithm is developed to find a compromising solution of the uncertain model. Numerical results show efficiency of the algorithm and a number of important managerial insights are revealed from the proposed model by scenario analysis and sensitivity analysis.

## 1. Introduction

**1.1. Background.** With rapid development of technology and improvement of living standards, the per capita holding of automobiles greatly increases. In China, as the largest developing country with a population of around 1.3 billion, huge amount of end-of-life vehicles (ELVs) is bringing enormous pressure on its environment and human life. Actually, the civilian car ownership in China has reached 137 million in 2013, and has been almost doubling every four years. If the average lifespan of a car is 8-9 years, then the number of ELVs will exceed 14 million in 2020 [1]. In the world, it is estimated that the number of vehicles will rise to 1.85 billion by 2030, and the scrap generated from the ELVs will be 3.71 billion tonnes [2]. In an era of resources shortage and environmental deterioration, recycling the ELVs can give birth to new-style industry as a typical low-carbon and sustainable production approach [3, 4].

It is well known that the ELVs contain a great amount of reusable components and materials such as steel, copper, rubber, etc. Therefore, recycling the ELVs offer considerable economic and environmental benefits [5]. This fact has been paid great attention either by governments, by industry or by academia. Actually, the European Union (EU) has established legal regulation that manufacturers are responsible for take-back of ELVs from end-users, dismantling, shredding, and recycling of ELVs [6]. Directive 2000/53/EC required that, no later than 1 January 2015, for all the ELVs, the reuse and recovery rate shall be increased to a minimum of 95% by an average weight per vehicle and year. Within the same time limit, the reuse and recycling rate shall be increased to a minimum of 85% by an average weight per vehicle and year. Japan had an ELV recovery rate of 85% in 2002, its attained target reached 95% by 2015 [7].

In China, a series of relevant policies and regulations have been issued to improve the management mechanism

for the ELV recycling industry since 2001. The “automotive products recycling technology policy” was implemented in 2006, which specifies utilization rate targets of the recyclable products in China [8]. The established China automotive material data system (CAMDS) in 2009 has played an important role in implementing the automotive products recyclable rate and the managing the ELVs. Chinese government has legislated that “yellow label cars” (heavily polluting vehicles) must be eliminated by 2017 [3]. Automotive components remanufacturing, as an essential part in automotive life-cycle development, has become a prominent direction to promote sustainable planning of automobile industry in China [9]. Actually, these policies are bringing a great of economic and environmental benefits to China.

*1.2. Literature Review.* In recent years, recovery of used products has become increasingly important owing to economic reasons and growing environmental or legislative concern [10]. Particularly, the ELV recycling plays an important role for sustainable development. For example, each remanufactured engine could save 68-83% of the energy required to manufacture a new engine, and decrease carbon dioxide emissions by 73-87%. The coolants and batteries in ELVs can also be recycled, and it reduces emissions of greenhouse gases and gases that lead to acidification. The main ingredient of coolant is Solid CO<sub>2</sub>, and the electrolyte of lead accumulator is lead accumulator [1]. Recycling metals from the ELVs could decrease the amounts of resources consumed building new cars. If all the vehicle materials can be recycled to produce new vehicles, about 30% of the energy consumption can be saved [11].

Summarily, the ELVs contain a great quantity of reusable components and materials such as steel, copper, rubber, plastics, etc. which can be reused or remanufactured. Thus, the remanufacturing industry of ELVs necessarily has strategic significance as it better utilizes resources and creates higher values. Especially, recycling ELVs can play an important role in realizing the country’s sustainable development goals. In 2003, China has introduced extended producer responsibility (EPR), which requires that any manufacturer should participate in ELV take-back, dismantling, remanufacturing, and so on [12]. The “automotive product recycling technology policy” in China requires carmakers to improve the design of vehicles, spare parts, and raw materials, as well as reduce the use of lead and other environmentally hazardous substances. Actually, this policy is an encouragement to the carmakers such that more recycled materials from the ELVs are used [13].

It is easy to see that recycling the ELVs depends on establishment of an efficient ELV recycling network, which not only can reduce the impact on the environment during the recycling process, but also can facilitate the effective reuse of recycled resources [14]. Furthermore, construction of optimization models for production planning problems of recycling the ELVs is helpful to provide the decision-makers an optimal plan for the practical operation of the recycling system [15].

In this connection, Cruz-Rivera et al. developed a reverse logistics network design for collection of the ELVs in Mexico

[16]. Demirel et al. proposed a deterministic mixed integer linear programming (MILP) model for the ELV recycling network design, where all of the end-users, collection centers, dismantlers, shredders, landfills, recycling facilities, and secondary markets are included [6]. On the basis of [6], Demirel et al. in 2017 developed a closed-loop supply chain for the ELVs recycling, where some reusable components after processing are sold to suppliers for remanufacturing [17]. Finally, new vehicles with the remanufactured components will flow to consumers. Ene et al. considered refurbishment in ELVs recycling network; the reusable parts must be refurbished before they could be sold to secondary markets [18]. Phuc et al. designed inspection centers in the ELVs recycling system; those ELVs passing inspection is repaired in the repair centers and then is sold to the used vehicle markets [19]. Additionally, in [2], municipal solid waste incinerator and advanced thermal treatment measures are applied to dispose autoshrredder residue (ASR) besides land-filling.

In some of the existing models for recycling the ELVs, various kinds of uncertainties have also been considered. It suggests in [20] that uncertainty seems to be the key factor influencing the management of ELVs. Özkır et al. stated that the selling price of products can be described by trapezoidal fuzzy sets such that both seller’s and buyer’s satisfaction levels are reflected [21]. In [19], the fixed cost, the transportation cost and the processing cost were also regarded as trapezoidal fuzzy sets in a reverse ELV recovery network. In [22], the capacities of sorting entities for recycling the ELVs were observed as random parameters, while the procurement cost, the transportation cost, the processing cost, and the storage cost were assumed to be interval parameters. In a global supply chain management model proposed by Wan et al., the demand of products for retailers is assumed to be stochastic and depends on the price of products [23].

Very recently, a polymorphic uncertain equilibrium model (PUEM) was developed by Wan et al. for the problem of decentralized supply chain management, where the demand of consumers was regarded as a continuous random variable, and the holding cost of the retailer and the transaction cost between the manufacturer and retailer were described by fuzzy sets [24]. Then, for the PUEM, a deterministic equivalent formulation (DEF) was first derived by compromise programming approach such that the existing powerful algorithms in the standard smooth optimization were employed to find an approximate equilibrium point for the uncertain problem. It is also noted that there are many different approaches to removing uncertainty in the uncertain model. For example, expectation method was applied in [19] to deal with the fuzzy objective such that the fuzzy objective can be converted into a deterministic one. Chance-constrained programming method was adopted in [22] to deal with the random constraints.

However, in the existing results for optimizing the system of recycling the ELVs, there are still some deficiencies, which can be briefly summarized as follows.

(1) Polymorphic uncertainty is rarely considered in this special reverse logistics network. Especially, randomness of



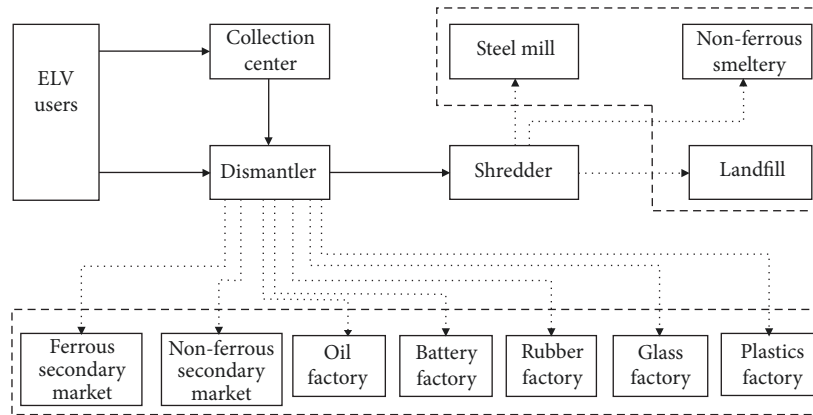


FIGURE 1: Structure of the ELV recovery network.

the demand in the secondary markets of reusable components has not been taken account into design of an optimal ELV recovery network.

(2) To hedge fuzziness of the objective function, the expectation method is often applied to transform the fuzzy objective into a deterministic one. Clearly, this method can not address the feature of variance in a fuzzy set. A more reasonable method should capture all information in a fuzzy objective function, which includes the lower and upper variances of a fuzzy set, as well as its center value.

*1.3. Motivation of This Research.* From the above literature review, it is necessary to build a new polymorphic uncertain optimization model for a more efficient system of the ELV recovery management. In particular, this model should simultaneously capture fuzziness and randomness of model parameters in a ELV recovery network design. Then, as done in [24], a unified compromising programming approach should be presented to convert the PUOM into a deterministic one such that the existing powerful optimization algorithms can be applied to find an approximately optimal strategy for recycling the ELVs.

In this paper, just like the mentioned reasons in [19, 21, 23, 24], we suppose that all of the fixed cost, the unit transportation cost, the unit processing cost, and the unit selling price of reused parts are fuzzy model parameters, and both of the capacity and the demand are regarded to be random variables. Then, our investigation proceeds along the following three subsequent steps.

*Step 1.* In a polymorphic uncertain environment, we construct a new optimization model to formulate the production planning problems of ELV recovery system.

*Step 2.* To hedge uncertainty of the model, a unified compromising programming approach will be proposed, which is associated with the following two phases: (1) in the first phase, the original problem is converted into an auxiliary crisp multiple-objective mixed integer linear programming problem; (2) in the second phase, a novel interactive fuzzy programming approach is proposed to find a preferred compromising solution through an interaction between the

decision-maker with preference and the rational model [25].

*Step 3.* To answer what is the practical significance of the new model and the developed algorithm in this paper, we will reveal some important managerial insights from the proposed model by scenario analysis and sensitivity analysis.

The rest of the paper is organized as follows. Next section is devoted to the description of problem and construction of model. In Section 3, an interactive algorithm is developed. In Section 4, numerical results of case study are reported. In Section 5, sensitivity analysis is conducted, and some practical managerial implications are revealed from the constructed model. Some conclusions and suggestions on future research are presented in the last section.

## 2. Problem Description and Formulation

*2.1. Problem Description.* Similar to the setting in [6], the network structure of ELV recovery system to be addressed in this paper is shown in Figure 1.

As shown in Figure 1, the network nodes basically consist of the ELV sources such as the last owners, insurance companies and abandoned vehicles, the collection centers, the dismantlers, the shredders, the recycling facilities, the secondary markets, and the landfills.

Specifically, the process flow of the recycling network can be stated as follows. All of the ELVs must be treated in formal dismantling companies. The collection centers or dismantlers first procure the ELVs from the ELV sources. Then, the ELVs in the collection centers will all be transported to the dismantlers. In the dismantlers, it is first required to remove and store separately the fuel, the motor oil, the oil from transmission system, the hydraulic oil, the cooling liquid, the liquid from the brake system, and other liquids and hazardous substances if any. Subsequently, the components or materials removed from the scrap car are considered for reuse and recycling. Reusable ferrous and nonferrous components are sold to the secondary market, while the recycling materials, such as batteries, tyres, glass, plastics, and waste oil, are sold to the recycling factories. The remaining

hulks are shipped to the shredders for further recycling. In the shredders, some materials can be mechanically recycled by shredder, air suction, magnetic sorters, and eddy current sorters. Finally, the hulks are divided into ferrous material, nonferrous material, and autoshredder residue (ASR). The sorted metals will be allocated to steel mills or nonferrous smelteries for further recycling, while ASR will be directly transported to the landfill.

In the recycling network in [6], all of the end-users, collection centers, dismantlers, shredders, landfills, recycling facilities, and secondary markets are the nodes of this network, and it is assumed that the last owners must return their vehicles to one of the collection centers or dismantlers. Different from the network in [6], Figure 1 indicates that (1) only ferrous and nonferrous components for reusing are sold to the secondary markets in our network; (2) there is no need to build the recycling facilities for processing the battery, tyre, glass, and plastics; instead, all of them are separately sold to the existent factories which have equipment for recycling; (3) our network is more in accordance with the suggested recycling system in China by [26, 27].

The goal of this paper is to formulate the above recycling network such that the total recycling cost is minimized, which is associated with the costs of transportation, processing, disposal, and the fixed opening costs in a multistage production plan. In order to build an optimization model that is more realistic than those available in the literature, uncertainty in recycling system must be incorporated into construction of model. Actually, due to incompleteness and unavailability of desired data, it is inappropriate to assume that the fixed opening cost, the unit transportation cost, the unit processing cost, the unit selling price of reconditioned parts, the capacity, and demand are all fixed constants. An acceptable approach is to describe these parameters by uncertain mathematical concepts from the theory of stochastic or fuzzy mathematics. Especially, in this paper, we assume that the fixed opening cost, the unit transportation cost, the unit processing cost, and the unit selling price of reconditioned parts are fuzzy sets, and the capacity levels of dismantlers, shredders, landfills, and the demand of secondary markets are random variables.

For simplicity, we suppose that the membership function of the relevant fuzzy sets is subject to possibility distribution (see Figure 2). Mathematically, any trapezoidal fuzzy number  $\tilde{c}$  is given by a membership function  $\mu_{\tilde{c}} : R \rightarrow [0, 1]$ , where for any  $c$ ,

$$\mu_{\tilde{c}}(c) = \begin{cases} 0, & c \leq c^p, \\ \frac{c - c^p}{c^{m_1} - c^p}, & c^p \leq c \leq c^{m_1}, \\ 1, & c^{m_1} \leq c \leq c^{m_2}, \\ \frac{c^0 - c}{c^0 - c^{m_2}}, & c^{m_2} \leq c \leq c^0, \\ 0, & c \geq c^0, \end{cases} \quad (1)$$

$c^p$ ,  $c^{m_1}$ ,  $c^{m_2}$ , and  $c^0$  are given constants. Different values of  $c^p$ ,  $c^{m_1}$ ,  $c^{m_2}$ , and  $c^0$  define various fuzzy sets. From this viewpoint, any fuzzy number can be denoted by a quaternion  $\tilde{c} = (c^p, c^{m_1}, c^{m_2}, c^0)$ . For fuzzy number  $\tilde{c}$ , we call  $(c^{m_1} - c^p)$

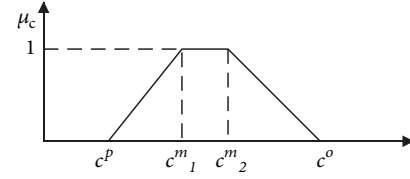


FIGURE 2: The trapezoidal possibility distribution of fuzzy number  $\tilde{c}$ .

its the upper variance, call  $(c^0 - c^{m_2})$  its lower variance, and  $(c^{m_1} + c^{m_2})$  the center value.

2.2. Notations. Before construction of model, we first introduce the following notations for readability.

#### Indexes

- $i$ : the labels of ELV sources,  $i = 1, 2, \dots, I$ .
- $j$ : the labels of collection centers,  $j = 1, 2, \dots, J$ .
- $k$ : the labels of dismantlers,  $k = 1, 2, \dots, K$ .
- $l$ : the labels of shredders,  $l = 1, 2, \dots, L$ .
- $s$ : the labels of secondary markets,  $s = 1, 2, \dots, S$ .
- $m$ : the labels of steel mills,  $m = 1, 2, \dots, M$ .
- $n$ : the labels of nonferrous smelteries,  $n = 1, 2, \dots, N$ .
- $p$ : the labels of oil recycling factories,  $p = 1, 2, \dots, P$ .
- $q$ : the labels of battery recycling factories,  $q = 1, 2, \dots, Q$ .
- $r$ : the labels of rubber recycling factories,  $r = 1, 2, \dots, R$ .
- $v$ : the labels of glass recycling factories,  $v = 1, 2, \dots, V$ .
- $w$ : the labels of plastics recycling factories,  $w = 1, 2, \dots, W$ .
- $u$ : the labels of landfills,  $u = 1, 2, \dots, U$ .
- $t$ : the processing periods,  $t = 1, 2, \dots, T$ .

#### Parameters

- $R_{it}$ : the amount of ELVs returned from ELV source  $i$  in period  $t$  (ton).
- $\tilde{f}_{kt}$ : the fixed opening cost for dismantler  $k$  in period  $t$  (yuan).
- $\tilde{f}_{lt}$ : the fixed opening cost for shredder  $l$  in period  $t$  (yuan).
- $\tilde{p}_{kt}$ : the unit cost of dismantling at dismantler  $k$  in period  $t$  (yuan/ton).
- $\tilde{p}_{lt}$ : the unit cost of shredding at shredder  $l$  in period  $t$  (yuan/ton).
- $\tilde{p}_{ut}$ : the unit cost of disposal at landfill  $u$  in period  $t$  (yuan/ton).
- $\tilde{s}_{1t}$ : the unit price of selling of dismantler for ferrous components for reusing in period  $t$  (yuan/ton).

$\bar{s}_{2t}$ : the unit price of selling of dismantler for nonferrous components for reusing in period  $t$  (yuan/ton).

$\bar{s}_{3t}$ : the unit price of selling of dismantler for oil for recycling in period  $t$  (yuan/ton).

$\bar{s}_{4t}$ : the unit price of selling of dismantler for battery for recycling in period  $t$  (yuan/ton).

$\bar{s}_{5t}$ : the unit price of selling of dismantler for tyre for recycling in period  $t$  (yuan/ton).

$\bar{s}_{6t}$ : the unit price of selling of dismantler for glass for recycling in period  $t$  (yuan/ton).

$\bar{s}_{7t}$ : the unit price of selling of dismantler for plastics for recycling in period  $t$  (yuan/ton).

$\bar{z}_{1t}$ : the unit price of selling of shredder for ferrous material for recycling in period  $t$  (yuan/ton).

$\bar{z}_{2t}$ : the unit price of selling of shredder for nonferrous material for recycling in period  $t$  (yuan/ton).

$\bar{t}c_{ijt}$ : the unit cost of transportation between ELV source  $i$  and collection center  $j$  for ELV in period  $t$  (yuan/ton·km).

$\bar{t}c_{ikt}$ : the unit cost of transportation between ELV source  $i$  and dismantler  $k$  for ELV in period  $t$  (yuan/ton·km).

$\bar{t}c_{jkt}$ : the unit cost of transportation between collection center  $j$  and dismantler  $k$  for ELV in period  $t$  (yuan/ton·km).

$\bar{t}c_{klt}$ : the unit cost of transportation between dismantler  $k$  and shredder  $l$  for hulk in period  $t$  (yuan/ton·km).

$\bar{t}c_{lut}$ : the unit cost of transportation between shredder  $l$  and landfill  $u$  for ASR in period  $t$  (yuan/ton·km).

$d_{ij}$ : the distance between ELV source  $i$  and collection center  $j$  (km).

$d_{ik}$ : the distance between ELV source  $i$  and dismantler  $k$  (km).

$d_{jk}$ : the distance between collection center  $j$  and dismantler  $k$  (km).

$d_{kl}$ : the distance between dismantler  $k$  and shredder  $l$  (km).

$d_{lu}$ : the distance between shredder  $l$  and landfill  $u$  (km).

$\bar{c}a_{jt}$ : the capacity of collection center  $j$  in period  $t$  (ton).

$\bar{c}a_{kt}$ : the capacity of dismantler  $k$  in period  $t$  (ton).

$\bar{c}a_{lt}$ : the capacity of shredder  $l$  in period  $t$  (ton).

$\bar{c}a_{ut}$ : the capacity of landfill  $u$  in period  $t$  (ton).

$\bar{d}e_{1st}$ : the demand of secondary market  $s$  for reusable ferrous components in period  $t$  (ton).

$\bar{d}e_{1st}$ : the demand of secondary market  $s$  for reusable nonferrous components in period  $t$  (ton).

$\alpha$ : the weight percentage of hulk in ELV.

$\beta_1$ : the weight percentage of reusable ferrous components in ELV.

$\beta_2$ : the weight percentage of reusable nonferrous components in ELV.

$\beta_3$ : the weight percentage of oil in ELV.

$\beta_4$ : the weight percentage of batteries in ELV.

$\beta_5$ : the weight percentage of tyres in ELV.

$\beta_6$ : the weight percentage of glass in ELV.

$\beta_7$ : the weight percentage of plastics in ELV.

$\eta$ : the weight percentage of ASR in hulk.

$\eta_1$ : the weight percentage of ferrous material in hulk.

$\eta_2$ : the weight percentage of nonferrous material in hulk.

### Decision Variables

$B_{ijt}$ : the amount of ELVs transported from ELV source  $i$  to collection center  $j$  in period  $t$ .

$C_{ikt}$ : the amount of ELVs transported from ELV source  $i$  to dismantler  $k$  in period  $t$ .

$D_{jkt}$ : the amount of ELVs transported from collection center  $j$  to dismantler  $k$  in period  $t$ .

$E_{klt}$ : the amount of hulk transported from dismantler  $k$  to shredder  $l$  in period  $t$ .

$F_{lut}$ : the amount of ASR transported from shredder  $l$  to landfill  $u$  in period  $t$ .

$Q1_{kst}$ : the amount of ferrous components transported from dismantler  $k$  to secondary market  $s$  in period  $t$ .

$Q2_{kst}$ : the amount of nonferrous components transported from dismantler  $k$  to secondary market  $s$  in period  $t$ .

$Q3_{kpt}$ : the amount of oil transported from dismantler  $k$  to oil recycling factory  $p$  in period  $t$ .

$Q4_{kqt}$ : the amount of batteries transported from dismantler  $k$  to battery recycling factory  $q$  in period  $t$ .

$Q5_{krt}$ : the amount of tyres transported from dismantler  $k$  to rubber recycling factory  $r$  in period  $t$ .

$Q6_{kvt}$ : the amount of glass transported from dismantler  $k$  to glass recycling factory  $v$  in period  $t$ .

$Q7_{kwt}$ : the amount of plastics transported from dismantler  $k$  to plastics recycling factory  $w$  in period  $t$ .

$Q8_{lmt}$ : the amount of ferrous material transported from shredder  $l$  to steel mill  $m$  in period  $t$ .

$Q9_{lnt}$ : the amount of nonferrous material transported from shredder  $l$  to nonferrous smeltery  $n$  in period  $t$ .

$O_{kt}$ : if dismantler  $k$  is opened in period  $t$ ,  $O_{kt} = 1$ ; otherwise,  $O_{kt} = 0$ .

$O_{lt}$ : if shredder  $l$  is opened in period  $t$ ,  $O_{lt} = 1$ ; otherwise,  $O_{lt} = 0$ .



2.3. *Construction of Model for the ELV Recycling.* We now formulate the objective function of the ELV recovery system in a centralized decision-making mode, being referred as to  $\tilde{\pi}$ . The objective is to minimize the total cost of recycling system. It is associated with the fixed opening cost, the transportation cost, the processing cost, and the revenue from the sale income of the recycled resources. Therefore, the total cost function is written as

$$\tilde{\pi} = \widetilde{FC} + \widetilde{PC} + \widetilde{TC} - \widetilde{RE}, \quad (2)$$

where the total fixed opening cost is

$$\widetilde{FC} = \sum_k \sum_t \tilde{f}_{kt} O_{kt} + \sum_l \sum_t \tilde{f}_{lt} O_{lt}, \quad (3)$$

the total transportation cost on each arc of the network reads

$$\begin{aligned} \widetilde{TC} = & \sum_i \sum_j \sum_t \tilde{t}c_{ijt} B_{ijt} d_{ij} + \sum_i \sum_k \sum_t \tilde{t}c_{ikt} C_{ikt} d_{ik} \\ & + \sum_j \sum_k \sum_t \tilde{t}c_{jkt} D_{jkt} d_{jk} + \sum_k \sum_l \sum_t \tilde{t}c_{klt} E_{klt} d_{kl} \\ & + \sum_l \sum_u \sum_t \tilde{t}c_{lut} F_{lut} d_{lu} \\ & + \sum_k \sum_s \sum_t \tilde{t}c_{kst} d_{ks} (Q1_{kst} + Q2_{kst}) \\ & + \sum_k \sum_p \sum_t \tilde{t}c_{kpt} Q3_{kpt} d_{kp} \\ & + \sum_k \sum_q \sum_t \tilde{t}c_{kqt} Q4_{kqt} d_{kq} \\ & + \sum_k \sum_r \sum_t \tilde{t}c_{krt} Q5_{krt} d_{kr} \\ & + \sum_k \sum_v \sum_t \tilde{t}c_{kvt} Q6_{kvt} d_{kv} \\ & + \sum_k \sum_w \sum_t \tilde{t}c_{kwt} Q7_{kwt} d_{kw} \\ & + \sum_l \sum_m \sum_t \tilde{t}c_{lmt} Q8_{lmt} d_{lm} \\ & + \sum_l \sum_n \sum_t \tilde{t}c_{lnt} Q9_{lnt} d_{ln}, \end{aligned} \quad (4)$$

the total processing cost of dismantlers, shredders, and landfills is

$$\begin{aligned} \widetilde{PC} = & \sum_i \sum_k \sum_t \tilde{p}c_{kt} C_{ikt} + \sum_j \sum_k \sum_t \tilde{p}c_{kt} D_{jkt} \\ & + \sum_k \sum_l \sum_t \tilde{p}c_{lt} E_{klt} + \sum_l \sum_u \sum_t \tilde{p}c_{ut} F_{lut}, \end{aligned} \quad (5)$$

and the income from the sale of isolated materials is

$$\begin{aligned} \widetilde{RE} = & \sum_k \sum_s \sum_t (\tilde{s}_{1t} Q1_{kst} + \tilde{s}_{2t} Q2_{kst}) \\ & + \sum_k \sum_p \sum_t \tilde{s}_{3t} Q3_{kpt} + \sum_k \sum_q \sum_t \tilde{s}_{4t} Q4_{kqt} \\ & + \sum_k \sum_r \sum_t \tilde{s}_{5t} Q5_{krt} + \sum_k \sum_v \sum_t \tilde{s}_{6t} Q6_{kvt} \\ & + \sum_k \sum_w \sum_t \tilde{s}_{7t} Q7_{kwt} + \sum_l \sum_m \sum_t \tilde{z}_{1t} Q8_{lmt} \\ & + \sum_l \sum_n \sum_t \tilde{z}_{2t} Q9_{lnt}. \end{aligned} \quad (6)$$

*Remark 1.* Note that, in (3), (4), (5), and (6), the opening cost, the unit transportation cost, the unit sale prices of all the reused components from the ELVs, and the unit processing cost for dismantling, shredding, and disposal at landfill are all supposed to be fuzzy model parameters. The advantage of this assumption is that these fuzzy parameters can reflect uncertain information of cost or revenue which can not be precisely calculated by statistical (financial) data. Actually, it is often that depreciation degree of the production tools can only be described as higher or lower, and market reputation of the reused products can only be said to be more credible. A fuzzy set is the most appropriate mathematical concept to characterize these uncertain language evaluations such as high, low, and credible.

Next, we present some practical constraints in minimization of the total recycling expense.

The first type of constraints is on material flow balance of network. It reads

$$\sum_j B_{ijt} + \sum_k C_{ikt} = R_{it}, \quad \forall i \in I, t \in T, \quad (7)$$

$$\sum_i B_{ijt} = \sum_k D_{jkt}, \quad \forall j \in J, t \in T, \quad (8)$$

$$\sum_l E_{klt} = \alpha \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (9)$$

$$\forall k \in K, t \in T,$$

$$\sum_s Q1_{kst} = \beta_1 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (10)$$

$$\forall k \in K, t \in T,$$

$$\sum_s Q2_{kst} = \beta_2 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (11)$$

$$\forall k \in K, t \in T,$$

$$\sum_p Q3_{kpt} = \beta_3 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (12)$$

$$\forall k \in K, t \in T,$$

$$\sum_q Q4_{kqt} = \beta_4 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (13)$$

$$\forall k \in K, t \in T,$$

$$\sum_r Q5_{krt} = \beta_5 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (14)$$

$$\forall k \in K, t \in T,$$

$$\sum_v Q6_{kvt} = \beta_6 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (15)$$

$$\forall k \in K, t \in T,$$

$$\sum_w Q7_{kwt} = \beta_7 \left( \sum_i C_{ikt} + \sum_j D_{jkt} \right), \quad (16)$$

$$\forall k \in K, t \in T,$$

$$\sum_u F_{lut} = \eta \sum_k E_{klt}, \quad \forall l \in L, t \in T, \quad (17)$$

$$\sum_m Q8_{lmt} = \eta_1 \sum_k E_{klt}, \quad \forall l \in L, t \in T, \quad (18)$$

$$\sum_u Q9_{lmt} = \eta_2 \sum_k E_{klt}, \quad \forall l \in L, t \in T. \quad (19)$$

The second type of constraints is on capacities of collection centers, dismantlers, shredders, and landfills. Owing to randomness of capacities, it says that the following stochastic inequalities should be satisfied:

$$\sum_i B_{ijt} \leq \widehat{c}a_{jt}, \quad \forall j \in J, t \in T, \quad (20)$$

$$\sum_i C_{ikt} + \sum_j D_{jkt} \leq \widehat{c}a_{kt} \cdot O_{kt}, \quad \forall k \in K, t \in T, \quad (21)$$

$$\sum_k E_{klt} \leq \widehat{c}a_{lt} \cdot O_{lt}, \quad \forall l \in L, t \in T, \quad (22)$$

$$\sum_l F_{lut} \leq \widehat{c}a_{ut}, \quad \forall u \in U, t \in T, \quad (23)$$

The third type of constraints is on the limited demands in secondary markets for the used components. Since the demands are random, the following stochastic inequalities hold:

$$\sum_k Q1_{kst} \leq \widehat{d}e_{1st}, \quad \forall s \in S, t \in T, \quad (24)$$

$$\sum_k Q2_{kst} \leq \widehat{d}e_{2st}, \quad \forall s \in S, t \in T. \quad (25)$$

The last type of constraints is on nonnegativity of decision variables and binary variables:

$$B_{ijt}, C_{ikt}, D_{jkt}, E_{klt}, F_{lut}, Q1_{kst}, Q2_{kst}, Q3_{kpt}, Q4_{kqt}, Q5_{krt}, \\ Q6_{kvt}, Q7_{kwt}, Q8_{lmt}, Q9_{lmt} \geq 0, \quad (26)$$

$$\forall i, j, k, l, s, m, n, p, q, r, v, w, u, t,$$

$$O_{kt}, O_{lt} \in \{0, 1\}, \quad \forall k \in K, l \in L, t \in T. \quad (27)$$

Consequently, we obtain a polymorphic uncertain optimization model (PUOM) for the management problem of the ELV recovery system:

$$\begin{aligned} \min \quad & \bar{\pi} \\ \text{subject to} \quad & (7) - (27). \end{aligned} \quad (28)$$

*Remark 2.* Note that in (20)-(23), the capacities for the collection centers, dismantlers, shredders, secondary markets, and landfill are supposed to be random. This assumption is based on the given probability distribution inferred from statistical data of capacity, especially after long-term practical production. In [20], the capacity of sorting entities for recycling the ELVs was also assumed to be random.

*Remark 3.* Compared with the deterministic model built in [18], PUOM (28) takes into account the fuzziness of cost and selling price, as well as the randomness of capacity and demand. Instead of simply describing the unit cost and selling price by interval parameters as done in [22], we characterize them by fuzzy sets which can contain more uncertain information, especially those perceptual evaluations of decision-makers in practice. Consequently, the proposed PUOM (28) in this paper is a polymorphic uncertain programming problem.

*Remark 4.* Since any uncertain optimization problem has no any optimal solution from the viewpoint of standard optimization theory, we have to make an optimal decision by a compromising programming approach, as done in [24]. Specifically, chance-constrained and multiobjective optimization approach will be first proposed to transform the polymorphic uncertain optimization problem into a deterministic equivalent formulation in Section 3.1. Then, an interactive algorithm will be developed to find a compromising solution for the original uncertain optimization model on the basis of analytic tools and the existing powerful algorithms in the classical optimization theory.

### 3. Unified Compromising Programming Approach and an Interactive Algorithm

In this section, we intend to develop a unified compromising programming approach to treat the PUOM (28).

Our basic idea can be stated as follows. By chance-constrained programming approach, we first hedge randomness of constraints in PUOM (28) such that the obtained model is only involved with fuzzy parameters. Then, we construct an auxiliary multiple-objective optimization problem

(MOOP) to convert the fuzzy model into a mixed integer linear programming model. Finally, we develop an interactive algorithm to find the compromising solution of the MOOP.

**3.1. Unified Compromising Programming Approach.** For simplicity of statement, we now only consider a linear stochastic constraint

$$a^T x \leq \hat{b}, \quad (29)$$

where  $x \in R^e$  is a vector of decision variables,  $a \in R^e$  is given crisp vector of coefficients, and  $\hat{b}$  is a random variable with probability distribution function  $F_{\hat{b}}(\cdot)$ .

Based on the idea of chance-constrained programming approach, equivalent deterministic formulation of (29) can be obtained (see, for example, [22–24, 28, 29]). Specifically, for a given confidence level  $1 - \delta$  ( $\delta \in [0, 1]$ ), (29) is equivalent to that the following inequality:

$$Pr(a^T x \leq \hat{b}) \geq 1 - \delta \quad (30)$$

always holds for any  $\delta \in [0, 1]$ , where  $Pr(\cdot)$  represents the probability that a stochastic inequality holds. The constraint (30) can be replaced by

$$F_{\hat{b}}(a^T x) \leq \delta. \quad (31)$$

Thus, by definition of probability distribution function, (31) can be approximated by an ordinary crisp constraint:

$$a^T x \leq F_{\hat{b}}^{-1}(\delta) \quad (32)$$

for a given violation degree  $\delta$ . Particularly, the random parameter  $\hat{b}$  is subject to normal distribution, i.e.,  $\hat{b} \sim N(\mu_{\hat{b}}, \sigma_{\hat{b}}^2)$ , where  $\mu_{\hat{b}}$  and  $\sigma_{\hat{b}}$  are the mean and standard deviation, respectively. Let  $\Phi(\cdot)$  express the standard normal distribution function. Then, (32) reads

$$a^T x \leq \mu_{\hat{b}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{b}}. \quad (33)$$

Let  $F_{\hat{c}_{ajt}}(\cdot)$ ,  $F_{\hat{c}_{akt}}(\cdot)$ ,  $F_{\hat{c}_{alt}}(\cdot)$ ,  $F_{\hat{c}_{aut}}(\cdot)$ ,  $F_{\hat{d}_{e_{1st}}}(\cdot)$ , and  $F_{\hat{d}_{e_{2st}}}(\cdot)$  be the probability distribution functions of the random variables  $\hat{c}_{ajt}$ ,  $\hat{c}_{akt}$ ,  $\hat{c}_{alt}$ ,  $\hat{c}_{aut}$ ,  $\hat{d}_{e_{1st}}$ , and  $\hat{d}_{e_{2st}}$ , respectively. From (32), it follows that, for a given violation degree  $\delta$ , the constraints (20)–(25) are replaced by

$$\sum_i B_{ijt} \leq F_{\hat{c}_{ajt}}^{-1}(\delta), \quad \forall j \in J, t \in T, \quad (34)$$

$$\sum_i C_{ikt} + \sum_j D_{jkt} \leq F_{\hat{c}_{akt}}^{-1}(\delta) \cdot O_{kt}, \quad \forall k \in K, t \in T, \quad (35)$$

$$\sum_k E_{klt} \leq F_{\hat{c}_{alt}}^{-1}(\delta) \cdot O_{lt}, \quad \forall l \in L, t \in T, \quad (36)$$

$$\sum_l F_{lut} \leq F_{\hat{c}_{aut}}^{-1}(\delta), \quad \forall u \in U, t \in T, \quad (37)$$

$$\sum_k Q3_{kst} \leq F_{\hat{d}_{e_{1st}}}^{-1}(\delta), \quad \forall s \in S, t \in T, \quad (38)$$

$$\sum_k Q4_{kst} \leq F_{\hat{d}_{e_{2st}}}^{-1}(\delta), \quad \forall s \in S, t \in T. \quad (39)$$

If all the probability distributions are normal, then from (33), we know that (20)–(25) can be further rewritten as

$$\sum_i B_{ijt} \leq \mu_{\hat{c}_{ajt}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{c}_{ajt}}, \quad (40)$$

$$\forall j \in J, t \in T,$$

$$\sum_i C_{ikt} + \sum_j D_{jkt} \leq (\mu_{\hat{c}_{akt}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{c}_{akt}}) \cdot O_{kt}, \quad (41)$$

$$\forall k \in K, t \in T,$$

$$\sum_k E_{klt} \leq (\mu_{\hat{c}_{alt}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{c}_{alt}}) \cdot O_{lt}, \quad (42)$$

$$\forall l \in L, t \in T,$$

$$\sum_l F_{lut} \leq \mu_{\hat{c}_{aut}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{c}_{aut}}, \quad (43)$$

$$\forall u \in U, t \in T,$$

$$\sum_k Q3_{kst} \leq \mu_{\hat{d}_{e_{1st}}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{d}_{e_{1st}}}, \quad (44)$$

$$\forall s \in S, t \in T,$$

$$\sum_k Q4_{kst} \leq \mu_{\hat{d}_{e_{2st}}} + \Phi^{-1}(\delta) \cdot \sigma_{\hat{d}_{e_{2st}}}, \quad (45)$$

$$\forall s \in S, t \in T.$$

We are now in a position to treat the fuzzy objective function in PUOM (28).

Since it is assumed that all the fuzzy parameters in  $\tilde{\pi}$  are subject to trapezoidal membership functions,  $\tilde{\pi}$  is a trapezoidal fuzzy set as a sum of trapezoidal fuzzy ones (see [25, 30, 31]). Denote  $\tilde{\pi} = (\pi^p, \pi^{m_1}, \pi^{m_2}, \pi^o)$ . Instead of minimizing  $(\pi^p + \pi^{m_1} + \pi^{m_2} + \pi^o)/4$  as in [19, 32, 33], we replace minimization of a fuzzy objective  $\tilde{\pi}$  by minimizing an integrated crisp objective function, defined by the membership functions of fuzzified  $(\pi^{m_1} + \pi^{m_2})$ ,  $(\pi^o - \pi^{m_2})$ , and  $(\pi^{m_1} - \pi^p)$  (see model (57)). To minimize the total cost and capture the uncertain (fuzzy) information in the total cost function, we first need to construct an auxiliary multiobjective optimization problem as follows:

$$\begin{aligned} \min \quad & Z_1 = \pi^{m_1} + \pi^{m_2} \\ \max \quad & Z_2 = \pi^{m_1} - \pi^p \\ \min \quad & Z_3 = \pi^o - \pi^{m_2} \\ \text{s.t.} \quad & (7) - (19), (26) - (27), (34) - (39), \end{aligned} \quad (46)$$

where

$$\begin{aligned} Z_1 = & \sum_k \sum_t (f_{kt}^{m_1} + f_{kt}^{m_2}) O_{kt} + \sum_l \sum_t (f_{lt}^{m_1} + f_{lt}^{m_2}) O_{lt} \\ & + \sum_i \sum_j \sum_t (tc_{ijt}^{m_1} \end{aligned}$$

$$\begin{aligned}
& + t c_{ijt}^{m_2} B_{ijt} d_{ij} + \sum_i \sum_k \sum_t (t c_{ikt}^{m_1} + t c_{ikt}^{m_2}) C_{ikt} d_{ik} \\
& + \sum_j \sum_k \sum_t (t c_{jkt}^{m_1} \\
& + t c_{jkt}^{m_2}) D_{jkt} d_{jk} + \sum_k \sum_l \sum_t (t c_{klt}^{m_1} + t c_{klt}^{m_2}) E_{klt} d_{kl} \\
& + \sum_l \sum_u \sum_t (t c_{lut}^{m_1} \\
& + t c_{lut}^{m_2}) F_{lut} d_{lu} + \sum_i \sum_k \sum_t (p c_{kt}^{m_1} + p c_{kt}^{m_2}) C_{ikt} \\
& + \sum_j \sum_k \sum_t (p c_{kt}^{m_1} \\
& + p c_{kt}^{m_2}) D_{jkt} + \sum_k \sum_l \sum_t (p c_{st}^{m_1} + p c_{st}^{m_2}) E_{klt} \\
& + \sum_l \sum_u \sum_t (p c_{ut}^{m_1} \\
& + p c_{ut}^{m_2}) F_{lut} \\
& + \sum_k \sum_s \sum_t ((t c_{kst}^{m_1} + t c_{kst}^{m_2}) d_{ks} \\
& - (s_{1t}^{m_1} + s_{1t}^{m_2})) Q1_{kst} \\
& + \sum_k \sum_s \sum_t ((t c_{kst}^{m_1} + t c_{kst}^{m_2}) d_{ks} \\
& - (s_{2t}^{m_1} + s_{2t}^{m_2})) Q2_{kst} \\
& + \sum_k \sum_p \sum_t ((t c_{kpt}^{m_1} + t c_{kpt}^{m_2}) d_{kp} \\
& - (s_{3t}^{m_1} + s_{3t}^{m_2})) Q3_{kpt} \\
& + \sum_k \sum_q \sum_t ((t c_{kqt}^{m_1} + t c_{kqt}^{m_2}) d_{kq} \\
& - (s_{4t}^{m_1} + s_{4t}^{m_2})) Q4_{kqt} \\
& + \sum_k \sum_r \sum_t ((t c_{krt}^{m_1} + t c_{krt}^{m_2}) d_{kr} \\
& - (s_{5t}^{m_1} + s_{5t}^{m_2})) Q5_{krt} \\
& + \sum_k \sum_v \sum_t ((t c_{kvt}^{m_1} + t c_{kvt}^{m_2}) d_{kv} \\
& - (s_{6t}^{m_1} + s_{6t}^{m_2})) Q6_{kvt} \\
& + \sum_k \sum_w \sum_t ((t c_{kwt}^{m_1} \\
& + t c_{kwt}^{m_2}) d_{kw} - (s_{7t}^{m_1} + s_{7t}^{m_2})) Q7_{kwt} \\
& + \sum_l \sum_m \sum_t ((t c_{lmt}^{m_1} + t c_{lmt}^{m_2}) d_{lm}
\end{aligned}$$

$$\begin{aligned}
& - (z_{1t}^{m_1} + z_{1t}^{m_2}) Q8_{lmt} \\
& + \sum_l \sum_n \sum_t ((t c_{lnt}^{m_1} + t c_{lnt}^{m_2}) d_{ln} \\
& - (z_{2t}^{m_1} + z_{2t}^{m_2})) Q9_{lnt}, \tag{47}
\end{aligned}$$

$$\begin{aligned}
Z_2 = & \sum_k \sum_t (f_{kt}^{m_1} - f_{kt}^p) O_{kt} + \sum_l \sum_t (f_{lt}^{m_1} - f_{lt}^p) O_{lt} \\
& + \sum_i \sum_j \sum_t (t c_{ijt}^{m_1} \\
& - t c_{ijt}^p) B_{ijt} d_{ij} + \sum_i \sum_k \sum_t (t c_{ikt}^{m_1} - t c_{ikt}^p) C_{ikt} d_{ik} \\
& + \sum_j \sum_k \sum_t (t c_{jkt}^{m_1} \\
& - t c_{jkt}^p) D_{jkt} d_{jk} + \sum_k \sum_l \sum_t (t c_{klt}^{m_1} - t c_{klt}^p) E_{klt} d_{kl} \\
& + \sum_l \sum_u \sum_t (t c_{lut}^{m_1} \\
& - t c_{lut}^p) F_{lut} d_{lu} + \sum_i \sum_k \sum_t (p c_{kt}^{m_1} - p c_{kt}^p) C_{ikt} \\
& + \sum_j \sum_k \sum_t (p c_{kt}^{m_1} \\
& - p c_{kt}^p) D_{jkt} + \sum_k \sum_l \sum_t (p c_{st}^{m_1} - p c_{st}^p) E_{klt} \\
& + \sum_l \sum_u \sum_t (p c_{ut}^{m_1} \\
& - p c_{ut}^p) F_{lut} \\
& + \sum_k \sum_s \sum_t ((t c_{kst}^{m_1} - t c_{kst}^p) d_{ks} \\
& - (s_{1t}^{m_1} - s_{1t}^p)) Q1_{kst} \\
& + \sum_k \sum_s \sum_t (t c_{kst}^{m_1} - t c_{kst}^p) d_{ks} - (s_{2t}^{m_1} - s_{2t}^p) Q2_{kst} \\
& + \sum_k \sum_p \sum_t ((t c_{kpt}^{m_1} - t c_{kpt}^p) d_{kp} \\
& - (s_{3t}^{m_1} - s_{3t}^p)) Q3_{kpt} \\
& + \sum_k \sum_q \sum_t ((t c_{kqt}^{m_1} - t c_{kqt}^p) d_{kq} \\
& - (s_{4t}^{m_1} - s_{4t}^p)) Q4_{kqt} \\
& + \sum_k \sum_r \sum_t ((t c_{krt}^{m_1} - t c_{krt}^p) d_{kr} \\
& - (s_{5t}^{m_1} - s_{5t}^p)) Q5_{krt}
\end{aligned}$$

$$\begin{aligned}
& + \sum_k \sum_v \sum_t ((tc_{kvt}^{m_1} - tc_{kvt}^p) d_{kv} \\
& - (s_{6t}^{m_1} - s_{6t}^p)) Q6_{kvt} \\
& + \sum_k \sum_w \sum_t ((tc_{kwt}^{m_1} - tc_{kwt}^p) d_{kw} \\
& - (s_{7t}^{m_1} - s_{7t}^p)) Q7_{kwt} \\
& + \sum_l \sum_m \sum_t ((tc_{lmt}^{m_1} - tc_{lmt}^p) d_{lm} \\
& - (z_{1t}^{m_1} - z_{1t}^p)) Q8_{lmt} \\
& + \sum_l \sum_n \sum_t ((tc_{lnt}^{m_1} - tc_{lnt}^p) d_{ln} \\
& - (z_{2t}^{m_1} - z_{2t}^p)) Q9_{lnt} \\
& + \sum_k \sum_q \sum_t ((tc_{kqt}^o - tc_{kqt}^{m_2}) d_{kq} - (s_{4t}^o - s_{4t}^{m_2})) Q4_{kqt} \\
& + \sum_k \sum_r \sum_t ((tc_{krt}^o - tc_{krt}^{m_2}) d_{kr} - (s_{5t}^o - s_{5t}^{m_2})) Q5_{krt} \\
& + \sum_k \sum_v \sum_t ((tc_{kvt}^o - tc_{kvt}^{m_2}) d_{kv} - (s_{6t}^o - s_{6t}^{m_2})) Q6_{kvt} \\
& + \sum_k \sum_w \sum_t ((tc_{kwt}^o - tc_{kwt}^{m_2}) d_{kw} - (s_{7t}^o - s_{7t}^{m_2})) \\
& \cdot Q7_{kwt} \\
& + \sum_l \sum_m \sum_t ((tc_{lmt}^o - tc_{lmt}^{m_2}) d_{lm} - (z_{1t}^o - z_{1t}^{m_2})) Q8_{lmt} \\
& + \sum_l \sum_n \sum_t ((tc_{lnt}^o - tc_{lnt}^{m_2}) d_{ln} - (z_{2t}^o - z_{2t}^{m_2})) Q9_{lnt}.
\end{aligned} \tag{48}$$

and

$$\begin{aligned}
Z_3 = & \sum_k \sum_t (f_{kt}^o - f_{kt}^{m_2}) O_{kt} + \sum_l \sum_t (f_{lt}^o - f_{lt}^{m_2}) O_{lt} \\
& + \sum_i \sum_j \sum_t (tc_{ijt}^o - tc_{ijt}^{m_2}) B_{ijt} d_{ij} \\
& + \sum_i \sum_k \sum_t (tc_{ikt}^o - tc_{ikt}^{m_2}) C_{ikt} d_{ik} \\
& + \sum_j \sum_k \sum_t (tc_{jkt}^o - tc_{jkt}^{m_2}) D_{jkt} d_{jk} \\
& + \sum_k \sum_l \sum_t (tc_{klt}^o - tc_{klt}^{m_2}) E_{klt} d_{kl} \\
& + \sum_l \sum_u \sum_t (tc_{lut}^o - tc_{lut}^{m_2}) F_{lut} d_{lu} \\
& + \sum_i \sum_k \sum_t (pc_{kt}^o - pc_{kt}^{m_2}) C_{ikt} \\
& + \sum_j \sum_k \sum_t (pc_{kt}^o - pc_{kt}^{m_2}) D_{jkt} \\
& + \sum_k \sum_l \sum_t (pc_{st}^o - pc_{st}^{m_2}) E_{klt} \\
& + \sum_l \sum_u \sum_t (pc_{ut}^o - pc_{ut}^{m_2}) F_{lut} \\
& + \sum_k \sum_s \sum_t ((tc_{kst}^o - tc_{kst}^{m_2}) d_{ks} - (s_{1t}^o - s_{1t}^{m_2})) Q1_{kst} \\
& + \sum_k \sum_s \sum_t ((tc_{kst}^o - tc_{kst}^{m_2}) d_{ks} - (s_{2t}^o - s_{2t}^{m_2})) Q2_{kst} \\
& + \sum_k \sum_p \sum_t ((tc_{kpt}^o - tc_{kpt}^{m_2}) d_{kp} - (s_{3t}^o - s_{3t}^{m_2})) \\
& \cdot Q3_{kpt}
\end{aligned}$$

Clearly, the first objective in (46) is to minimize the total cost and the last two objectives are used to capture the uncertain information in the total cost as much as possible. In addition, if (34)-(39) in model (46) are replaced by (40)-(45), then we get a special case of model (46) under assumption of normal distribution.

*Remark 5.* With the proposed unified compromising programming approach, the randomness of constraints in PUOM (28) first vanishes by chance-constrained programming method. It is possible that the feasible region of PUOM (28) is empty if the choice of violation degree  $\delta$  is very small [24]. In other words, a suitable choice of violation degree is important to the unified compromising programming approach, which will be further addressed in Section 4. It is also noted that the constraint of PUOM (28) is involved with the inverse of distribution function. So, there does not exist an explicit expression of such a constraint in the case that it is difficult to calculate the inverse of distribution function. Consequently, many powerful algorithms in standard optimization theory can not be applied to solve PUOM (28). Under assumption of normal distribution, explicit expressions of the constraints can be obtained (see the case study in Section 4).

*Remark 6.* Similar to [25, 31], we construct an auxiliary multiobjective optimization model (46) to hedge the fuzziness of the objective function in (28). Intuitively, the reason why we transform a fuzzy objective function into a three-objective one is that it can push the trapezoidal possibility distribution of the fuzzy total cost to the left as far as possible (see Figure 3).

Actually, the first objective in (46) implies the total cost, especially the part with degree of membership 1, is to be minimized by minimizing the center value. The third objective in (46) can minimize the area of the region (II) in Figure 3 so as to reduce the risk of higher cost, while the second objective in (46) can maximize the area of region (I) in Figure 3 so as to increase the possibility of obtaining lower cost. Clearly, the crisp model (46) is a deterministic

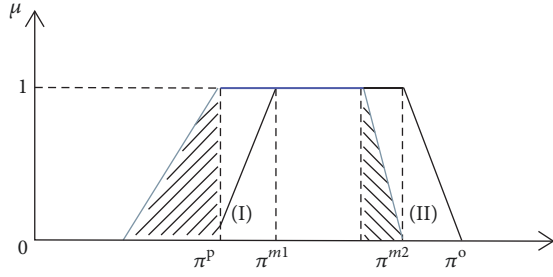


FIGURE 3: Principle of three-objective optimization approach.

equivalent formulation of the PUOM (28). To the best of our knowledge, it is the first time that the management problem of ELV recovery is treated by this approach in a polymorphic uncertain environment. For example, an expectation method was recently presented to convert the fuzzy model into a crisp one in [19].

3.2. *Interactive Algorithm.* With the preparation in Section 3.1, we now develop an interactive algorithm to find a compromising solution of PUOM (28), different from the hybrid heuristic algorithm developed in [24].

*Algorithm 1.*

*Step 1.* Choose an acceptable violation degree  $\delta$  for the random constraints. Solve the following three mixed integer linear programming problems (MILP):

$$\begin{aligned} \min \quad & Z_1 \\ \text{s.t.} \quad & (7) - (19), (26) - (27), (34) - (39), \end{aligned} \quad (50)$$

$$\begin{aligned} \max \quad & Z_2 \\ \text{s.t.} \quad & (7) - (19), (26) - (27), (34) - (39), \end{aligned} \quad (51)$$

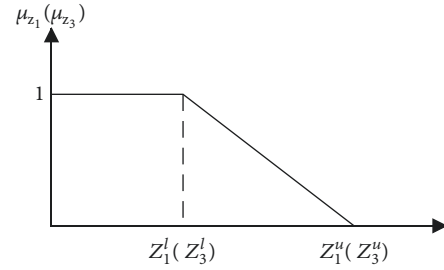
$$\begin{aligned} \min \quad & Z_3 \\ \text{s.t.} \quad & (7) - (19), (26) - (27), (34) - (39). \end{aligned} \quad (52)$$

Denote  $x^{1*}$ ,  $x^{2*}$ , and  $x^{3*}$  by the optimal solutions of (50), (51), and (52), respectively. The corresponding optimal values of the three objective functions are referred to as  $Z_1^l$ ,  $Z_2^u$ , and  $Z_3^l$ , respectively. Compute

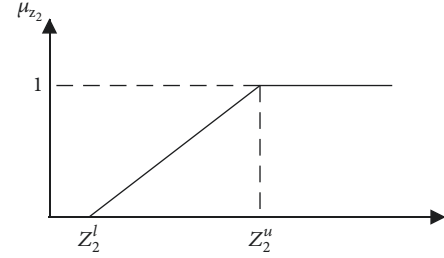
$$\begin{aligned} Z_1^u &= \max \{Z_1(x^{2*}), Z_1(x^{3*})\}, \\ Z_2^l &= \min \{Z_2(x^{1*}), Z_2(x^{3*})\}, \\ Z_3^u &= \max \{Z_3(x^{1*}), Z_3(x^{2*})\}. \end{aligned} \quad (53)$$

Clearly,  $Z_i^l$  and  $Z_i^u$  are the upper and lower bounds of the  $i$ -th objective function,  $i = 1, 2, 3$ .

*Step 2.* With  $Z_i^l$  and  $Z_i^u$ , formulate a trapezoidal membership function for the objective function  $Z_i$ . Specifically, we fuzzify



(a) Membership function for  $Z_1(Z_3)$



(b) Membership function for  $Z_2$

FIGURE 4: Fuzzification of the three objective functions.

$Z_1$  and  $Z_2$  and  $Z_3$  by specifying their membership functions (also see Figures 4(a) and 4(b)):

$$\mu_{\bar{Z}_1}(Z_1) = \begin{cases} 1, & Z_1 \leq Z_1^l \\ \frac{Z_1^u - Z_1}{Z_1^u - Z_1^l}, & Z_1^l \leq Z_1 \leq Z_1^u \\ 0, & Z_1 \geq Z_1^u, \end{cases} \quad (54)$$

$$\mu_{\bar{Z}_2}(Z_2) = \begin{cases} 0, & Z_2 \leq Z_2^l \\ \frac{Z_2 - Z_2^l}{Z_2^u - Z_2^l}, & Z_2^l \leq Z_2 \leq Z_2^u \\ 1, & Z_2 \geq Z_2^u, \end{cases} \quad (55)$$

and

$$\mu_{\bar{Z}_3}(Z_3) = \begin{cases} 1, & Z_3 \leq Z_3^l \\ \frac{Z_3^u - Z_3}{Z_3^u - Z_3^l}, & Z_3^l \leq Z_3 \leq Z_3^u \\ 0, & Z_3 \geq Z_3^u, \end{cases} \quad (56)$$

respectively.

*Step 3.* The decision-maker decides a compensation coefficient  $\gamma$  and a weight  $\theta_h$  for the  $h$ -th objective function,  $h = 1, 2, 3$ . Then, an integrated model is defined by

$$\begin{aligned} \max \quad & \lambda(x) = \gamma\lambda_0 + (1 - \gamma) \sum_{h=1}^3 \theta_h \mu_{\bar{Z}_h}(x) \\ \text{s.t.} \quad & \lambda_0 \leq \mu_{\bar{Z}_h}(x), \quad h = 1, 2, 3, \end{aligned} \quad (57)$$

$$(7) - (19), (26) - (27), (34) - (39),$$

where  $\lambda_0$  is an auxiliary variable.



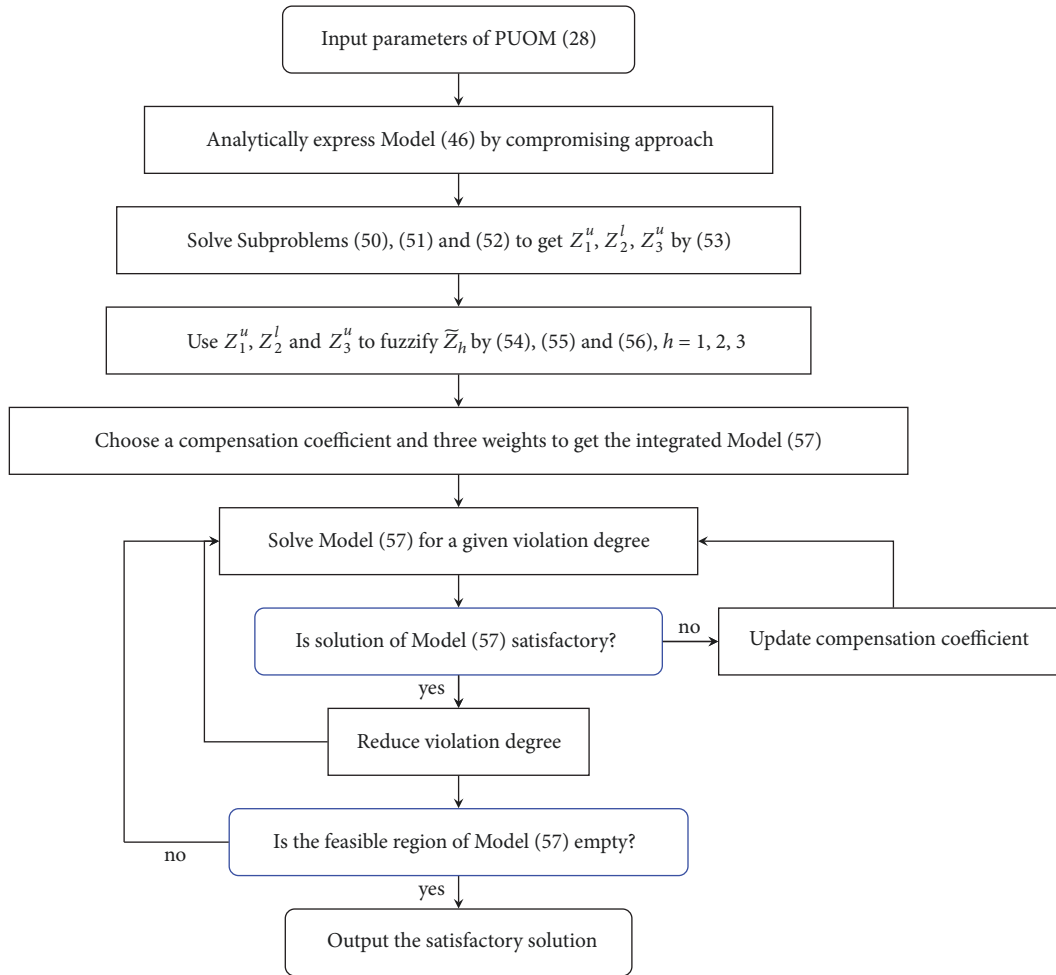


FIGURE 5: Logic chat of Algorithm 1.

*Step 4.* Select a violation degree  $\delta^*$  as small as possible to ensure solution existence of the deterministic model (57).

*Step 5.* Solve model (57). Denote  $x^*$  and  $\lambda_0^*$  the optimal solution of (57). Clearly, corresponding to  $\delta^*$ , the satisfaction degree of the three objective functions at least attains  $\lambda_0^*$ .

*Step 6.* If the decision-maker is satisfied with this current compromising solution, go to Step 7. Otherwise, update the value of compensation coefficient by  $\gamma =: \gamma + \Delta\gamma$ , where  $1 - \gamma > \Delta\gamma > 0$ . Return to Step 5.

*Step 7.* Choose a smaller  $\delta^*$ . If the feasible region of (57) is empty for  $\delta^*$ , then the algorithm stops and the decision-maker determines the most satisfactory  $\delta^*$ ,  $x^*$  and  $\lambda_0^*$  amongst all of the previous numerical results. Otherwise, return to Step 5.

The logic of Algorithm 1 can be demonstrated by the flowchart in Figure 5.

*Remark 7.* In Step 2, the three membership functions defined by (54), (55), and (56) are degenerate trapezoidal functions.

Since our aim is to minimize  $Z_1$ , smaller values of  $Z_1$  than  $Z_1^l$  are always preferable. Thus, it is nature to make the degree of membership equal to 1 for any value of  $Z_1$  less than  $Z_1^l$ . In this case, the left side of the membership function  $\mu_{\bar{Z}_1}(Z_1)$  is degenerate. The similar reason can be used to explain the definitions of  $\mu_{\bar{Z}_2}(Z_2)$  and  $\mu_{\bar{Z}_3}(Z_3)$ .

*Remark 8.* In Step 3, as done in [34], a fuzzy programming approach is employed to convert a multiobjective optimization problem into a single-objective problem such that an interactive algorithm is developed to find a compromising solution of the PUOM (28). Clearly, by maximizing the integrated degree of membership  $\lambda$  in model (57), we obtain a compromising solution such that both the total cost and its variance (the decision-making risk) become as smaller as possible. Actually, the first term in the objective function of model (57) describes the lower bound of the membership degree of all the three fuzzy objectives; greater lower bound means higher degree of overall satisfaction, with which the three objectives are simultaneously improved. The second term in the objective function of model (57) makes a trade-off between the overall satisfaction and the individual one.



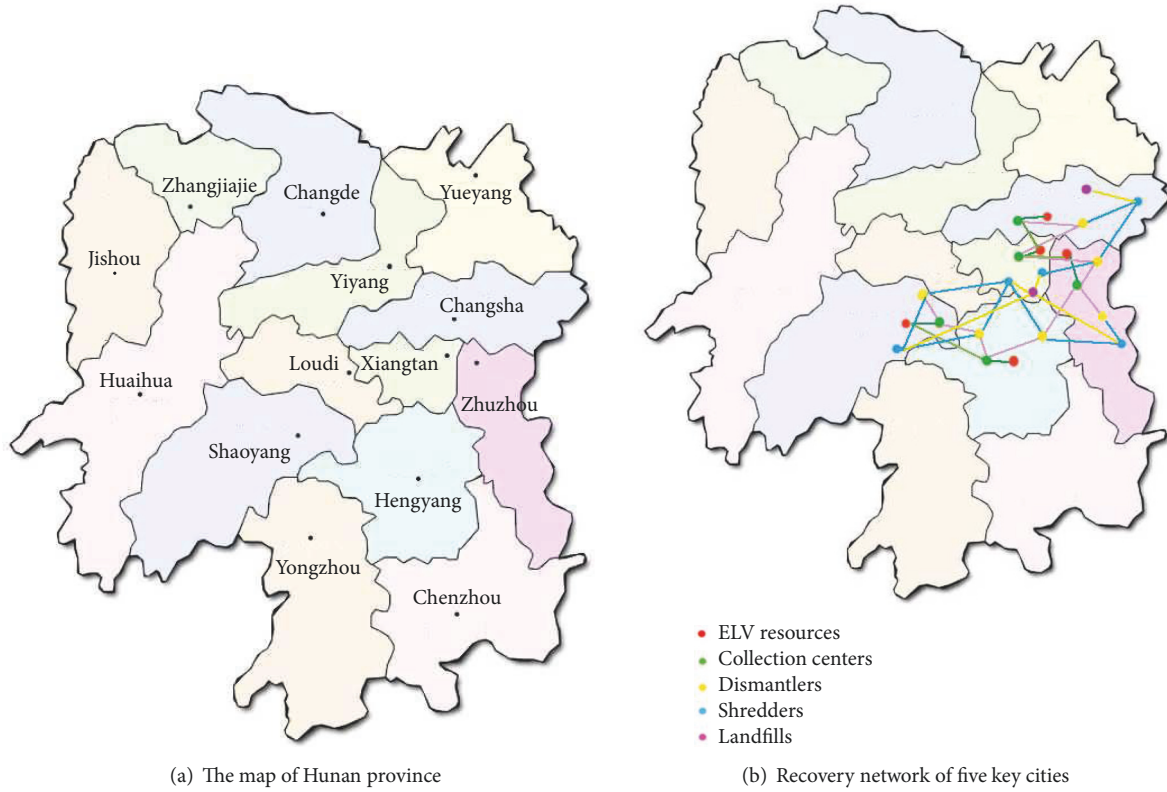


FIGURE 6: Recovery network in Hunan province.

*Remark 9.* In [34], an interactive algorithm was developed to solve a problem of purchasing, production, and distribution, which is involved in multiple suppliers, one manufacturer, and multiple distribution centers. It is shown that this algorithm always generates unbalanced and balanced solutions based on the decision-maker’s preferences. Similar to [34], we develop Algorithm 1 to find a compromising solution of the polymorphic uncertain recycling problems of ELVs in this paper. Numerical experiments in Sections 4 and 5 will show that Algorithm 1 can reveal more valuable managerial insights from the proposed model (2) than the solution methods available in the literature. Especially, by our algorithm, we can show how to choose an optimal compensation coefficient and a violation degree (see Section 4).

#### 4. Case Study with Different Algorithmic Parameters

In this section, we will apply the presented model and algorithm in Sections 2 and 3 to solve practical ELV recovery management problems in Hunan, China.

*4.1. Case Description.* According to [3], China has built a number of ELV recycling enterprises. In this case study, we attempt to deal with the end-of-life car recovery management problem of the five key cities (Changsha, Zhuzhou, Xiangtan, Hengyang, and Shaoyang) in Hunan (see Figure 6), considering that the data can be collected completely.

For convenience, the centers of the five cities are regarded as the ELV sources, and all of the relevant collecting enters, dismantlers, shredders, landfills, secondary markets, and recycling factories are distributed in these cities with given locations (see Figure 6(b) and Tables 1–6). It is noted that, for some cities, there are more than one dismantler, secondary market, oil, or glass factory, or there is no any landfill, secondary market, rubber, or plastics factory in practice.

A part of available data on unit costs of transportation and processing, selling price and material weight percentages of ELVs are deregistered from [6] (see Tables 7–11). For simplifying this case study, we only consider a single-period recycling problem. Referring to the government data on scrapped vehicles [12], the amount of ELVs from the five sources are estimated by

$$\begin{aligned}
 R_1 &= 1073, \\
 R_2 &= 587, \\
 R_3 &= 418, \\
 R_4 &= 1087, \\
 R_5 &= 1077.
 \end{aligned} \tag{58}$$

The aim of this case study is to answer the following questions:

- (1) How to determine an optimal transportation plan for the ELV recycling network?

TABLE 1: Size of the decision variables.

<i>I</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>U</i>	<i>S</i>	<i>P</i>	<i>Q</i>	<i>R</i>	<i>V</i>	<i>W</i>	<i>M</i>	<i>N</i>
5	5	6	5	2	2	2	2	2	2	2	2	2

TABLE 2: Sites of all items.

	Changsha	Zhuzhou	Xiangtan	Hengyang	Shaoyang
Resource	1	2	3	4	5
Collection center	1	2	3	4	5
Dismantler	1	2,3	4	6	5
Shredder	2	1	3	5	4
Landfill	2	-	1	-	-
Secondary market	1,2	-	-	-	-
Steel mill	1,2	-	-	-	-
Non-ferrous smeltery	1	-	-	2	-
Oil factory	1,2	-	-	-	-
Battery factory	1	-	-	2	-
Rubber factory	-	1	-	-	2
Glass factory	1,2	-	-	-	-
Plastics factory	-	-	1	-	2

TABLE 3: Distance between areas (km).

	Resources					Collection centers				
	1	2	3	4	5	1	2	3	4	5
Dismantlers										
1	12.8	39.7	40	150.1	186.1	21.8	59.8	46.4	150.9	153.1
2	56.9	11.1	23.7	114.1	178.4	60.2	13.8	36.3	116.4	138.7
3	46.3	5.7	18.5	120.3	177.6	49.8	22.9	31.9	122.2	139.2
4	126.3	82.9	82.4	44	148.4	124.9	64.8	81.8	48.8	105.1
5	157.3	146	134.1	90.1	28.2	148.8	143.3	121.6	84.3	15.5
6	149.6	118.9	111.5	22.5	93.4	144.2	107.3	103.6	16.6	53.3
Collection centers										
1	10	49.8	43.7	147.6	170.2	-	-	-	-	-
2	67.4	20.2	27.2	100.7	170.5	-	-	-	-	-
3	46.9	26.9	13.4	104.8	146.8	-	-	-	-	-
4	152	117	111.4	6.1	109.9	-	-	-	-	-
5	148.3	133.6	122.1	75.5	43.5	-	-	-	-	-

(2) What are the impacts of violation degree  $\delta$  on the three objectives  $Z_1, Z_2,$  and  $Z_3$  for a fixed compensation coefficient  $\gamma$ ?

(3) What are the influences of compensation coefficient  $\gamma$  on the least satisfaction degree  $\lambda_0$ ?

4.2. Numerical Solution of the ELV Recycling Model. We choose the weights of the three objectives  $\theta_1 = 0.4, \theta_2 = 0.3,$  and  $\theta_3 = 0.3,$  respectively. Then, we solve model (57) by Algorithm 1. The optimal values of the decision variables are reported in Table 12 in the case that  $\delta = 0.1$  and  $\gamma = 0.5.$

Table 12 indicates the follow-ing:

(1) All the ELVs from the five sources are transported to collection centers 1, 2, 4, and 5, while the quantities of transportation between the ELV sources and collection centers 3 and 6 are zeros.

(2) The most admisible quantity of ELVs from the ELV sources is 1935, which is actualized by the first collection center. The minimal quantity of transportation occurred to the second collection center (229 units of ELV).

(3) From the optimal results, all the 4242 ELVs are transported to the collection centers from the ELV sources. That is to say, there is no direct transportation between the ELV sources and the dismantlers.

(4) All the ELVs at collection centers 1, 2, 4, and 5 are transported to Dismantlers 1, 5, and 6.

(5) All the hulks are transported from the dismantlers to Shredders 2, 3, and 5. Then, from the shredders, the total 2630 tons of ferrous material and 169.68 tons of nonferrous material are sold to the steel mills and the nonferrous smelteries, respectively.

(6) The total disposal quantity is 637.3 tons and all the ASR are disposed in the two landfills.

TABLE 4: Continued Table 3.

	Shredders					Secondary markets	
	1	2	3	4	5	1	2
<b>Dismantlers</b>							
1	123.2	25.1	39.1	194.6	42.0	10.2	2.3
2	75.3	50.7	33.4	182.8	21.9	39.5	46.7
3	85.8	42.6	27.2	183.0	17.8	28.8	36.1
4	40.4	129.2	87.1	145.2	80.1	111.6	119.8
5	160.2	185.1	128.4	31.7	134.1	155.2	160.8
6	100.7	165.6	110.9	87.1	110.3	140.1	147.7
<b>Landfills</b>							
1	148.7	39.4	59.5	204.0	65.3	-	-
2	88.1	50.9	14.4	170.2	4.7	-	-
<b>Steel mills</b>							
1	117.8	33.0	29.7	185.1	33.8	-	-
2	114.7	33.3	27.2	184.0	30.8	-	-
<b>Non-ferrous smelteries</b>							
1	149.4	33.0	65.3	214.1	69.3	-	-
2	75.7	156.5	106.2	113.9	102.7	-	-

TABLE 5: Continued Table 4.

	Oil		Battery		Rubber		Glass		Plastics	
	1	2	1	2	1	2	1	2	1	2
<b>Dismantlers</b>										
1	12.8	8.7	6.6	177.7	36.1	200.6	7.6	5.5	43.3	164.6
2	53.4	53.8	42.7	143.54	17.3	191.5	40.6	51.3	29.0	144.2
3	42.8	43.6	32.1	149.3	9.8	191.2	30.1	41.0	24.5	146.3
4	121.9	129.2	115.2	72.9	85.6	157.5	115.3	126.2	81.0	98.2
5	153.6	171.4	157.5	86.9	142.9	39.9	161.9	168.5	128.1	31.5
6	145.1	158.6	143.4	30.7	118.5	100.2	145.7	155.4	107.1	39.8

TABLE 6: Probability distribution of capacity/demand (ton).

$\bar{c}a_j$	$\bar{c}a_k$	$\bar{c}a_l$	$\bar{c}a_u$	$\bar{d}e_{1s}$	$\bar{d}e_{2s}$
$N(2000, 50^2)$	$N(2000, 50^2)$	$N(1500, 45^2)$	$N(500, 20^2)$	$N(200, 10^2)$	$N(100, 5^2)$

TABLE 7: The fixed opening cost ( $\times 10^6$  yuan).

$\bar{f}_k$	$\bar{f}_l$
(1.235,1.245,1.255,1.26)	(4.975,4.99,5.01,5.02)

TABLE 8: Unit processing cost (yuan/ton).

$\bar{p}c_k$	$\bar{p}c_l$	$\bar{p}c_u$
(1820,1920,2000,2080)	(240,260,280,300)	(440,480,520,550)

4.3. *Impacts of Violation Degree.* It is easy to see that the feasible region of model (57) is closely related to the violation degree  $\delta$ . For a smaller violation degree, the feasible region becomes smaller because a decreasing  $\delta$  implies higher restriction on capacities or demands. Thus, it is useful to analyze the impacts of violation degree on the optimal solution for the presented unified compromising optimization

approach in this paper. For this, we change the violation degree  $\delta$  with a step length of 0.01.

In Table 13, numerical results are given for different violation degrees with a fixed  $\gamma = 0.5$ .

Table 13 demonstrates the following:

(1) With an increasing violation degree, satisfactory degree of the solution by the proposed unified compromising optimization method becomes greater. On the other hand, from  $\lambda_0 > 0.5$ , it follows that the satisfaction degree of the compromising method can give a more satisfactory solution than the expectation method in the literature.

(2) Both of the center value and the lower deviation of the fuzzy objective function,  $Z_1$  and  $Z_3$ , are increasing as the violation degree takes a smaller value, while the upper deviation  $Z_2$  is decreasing.

(3) Since variations of the center value, the lower and upper deviations, seem to be greater within the same change of violation degree, it suggests that decision-makers should choose a violation degree as small as possible in practice,

TABLE 9: Unit transportation cost of each item (yuan/ton-km).

$\tilde{t}_{c_{ij}}, \tilde{t}_{c_{ik}}$	$\tilde{t}_{c_{jk}}$	$\tilde{t}_{c_{kl}}$	$\tilde{t}_{c_{lu}}$
(1.8,1.92,2.02,2.1)	(0.7,0.76,0.82,0.86)	(0.3,0.37,0.44,0.5)	(0.9,0.97,1.04,1.1)
$\tilde{t}_{c_{ks}}$	$\tilde{t}_{c_{kp}}, \tilde{t}_{c_{lm}}, \tilde{t}_{c_{ln}}$	$\tilde{t}_{c_{kq}}$	$\tilde{t}_{c_{kr}}, \tilde{t}_{c_{kv}}, \tilde{t}_{c_{kw}}$
(1.4,1.47,1.54,1.6)	(0.6,0.67,0.74,0.8)	(0.5,0.57,0.64,0.7)	(0.4,0.47,0.54,0.6)

TABLE 10: Unit selling price of each item type ( $\times 10^3$  yuan/ton).

$\tilde{s}_{1t}$	$\tilde{s}_{2t}$	$\tilde{s}_{3t}$	$\tilde{s}_{4t}$
(2.1,2.3,2.5,2.65)	(10.5,11.3,12.3,13)	(3.5,3.7,3.9,4.0)	(0.56,0.6,0.64,0.67)
$\tilde{s}_{5t}$	$\tilde{s}_{6t}$	$\tilde{s}_{7t}$	$\tilde{z}_{1t}$
(0.14,0.145,0.15,0.155)	(0.3,0.4,0.5,0.58)	(5,5.6,6.2,6.6)	(0.44,0.48,0.52,0.55)
$\tilde{z}_{2t}$			
(1.4,1.47,1.54,1.6)			

TABLE 11: Weight percentage of each item.

$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\eta$	$\eta_1$	$\eta_2$
0.81	0.06	0.04	0.017	0.013	0.03	0.015	0.015	15/81	62/81	4/81

TABLE 12: Numerical solution in the case that  $\delta = 0.1$  and  $\lambda = 0.5$ .

Variable	Value	Variable	Value	Variable	Value	Variable	Value
$B_{1,4}$	1073	$E_{1,3}$	431.2	$Q_{2,6,1}$	77.4	$Q_{6,6,1}$	29.025
$B_{2,5}$	587	$E_{5,3}$	301.3	$Q_{3,1,2}$	32.895	$Q_{7,3,2}$	29.025
$B_{3,5}$	418	$E_{6,3}$	125	$Q_{3,5,1}$	6.324	$Q_{7,5,2}$	5.58
$B_{4,1}$	1087	$E_{6,5}$	144.23	$Q_{3,6,1}$	32.895	$Q_{7,2,1}$	29.025
$B_{5,1}$	848	$F_{2,1}$	210.3975	$Q_{4,1,1}$	25.155	$Q_{8,2,2}$	869.6
$B_{5,2}$	229	$F_{3,2}$	158.8043	$Q_{4,5,1}$	4.836	$Q_{8,3,2}$	656.4
$D_{1,6}$	1935	$F_{5,2}$	267.0982	$Q_{4,6,2}$	25.155	$Q_{8,5,2}$	1104
$D_{2,5}$	229	$Q_{1,1,2}$	116.1	$Q_{5,1,1}$	58.0	$Q_{9,2,1}$	56.1065
$D_{4,1}$	930	$Q_{1,5,1}$	22.32	$Q_{5,2,2}$	11.16	$Q_{9,3,1}$	42.3478
$D_{4,5}$	143	$Q_{1,6,1}$	116.1	$Q_{5,6,2}$	58.05	$Q_{9,5,1}$	71.2262
$D_{5,1}$	1005	$Q_{2,1,2}$	77.4	$Q_{6,1,2}$	29.025		
$E_{1,2}$	1136.1	$Q_{2,5,1}$	14.88	$Q_{6,5,1}$	5.58		

TABLE 13: Optimal solutions of different violation degree ( $\gamma = 0.5$ ).

$\delta$	$\lambda$	$\lambda_0$	$Z_1$	$Z_2$	$Z_3$	$\nabla Z_1$	$\nabla Z_2$	$\nabla Z_3$
0.15	0.58486327	0.5061525	506224571	342850.6	2697172	-	-	-
0.14	0.58485733	0.5061454	50622502.3	342849.1	269718.2	45.2	-1.5	1.0
0.13	0.58485261	0.5061407	50622650.9	342848.2	269718.9	14.9	-0.1	0.7
0.12	0.58484835	0.5061360	50622728.4	342847.2	269719.6	58.4	-1.0	0.7
0.11	0.58484195	0.5061289	50622844.6	342845.7	269720.6	106.2	-1.5	1.0
0.10	0.58483555	0.5061218	50622960.9	342844.2	269721.7	116.3	-1.5	1.1
0.09	0.58482881	0.5061139	50623026.2	342842.6	269722.8	65.3	-1.6	1.1
0.08	0.58482275	0.5061077	50623193.4	342841.3	269723.7	167.1	-1.3	0.9
0.07	0.58481635	0.5061006	50623309.9	342839.8	269724.8	116.5	-1.5	1.0
0.06	0.58480776	0.5060912	50623475.2	342837.8	269726.1	165.3	-2.0	1.4
0.05	0.58479702	0.5060794	50623681.1	342835.4	269727.8	205.9	-2.5	1.7
0.04	0.58478627	0.5060676	50623888.4	342832.9	269729.6	207.3	-2.5	1.7
0.03	0.58477089	0.5060503	50624126.1	342829.3	269732.1	237.7	-3.6	2.5
0.02	0.58475481	0.5060323	50624386.1	342825.5	269734.7	260.0	-3.8	2.6
0.01	0.58473918	0.5060109	50624154.1	342821.0	269737.8	-231.9	-4.5	3.1

TABLE 14: Impacts of different compensation coefficient ( $\delta = 0.02$ ).

$\gamma$	$\lambda$	$\lambda_0$	$Z_1$	$Z_2$	$Z_3$
0.1	0.6477901	0.5051697	50571344.3	342645.3	269860.4
0.2	0.6319046	0.5056313	50605879.4	342741.7	269793.2
0.3	0.6162438	0.5060323	50624386.1	342825.5	269734.7
0.4	0.6004978	0.5059748	50618897.1	342813.5	269743.1
0.5	0.5847548	0.5060323	50624386.1	342825.5	269734.7
0.6	0.5690160	0.5060412	50624749.2	342827.4	269733.4
0.7	0.5532663	0.5060345	50624774.6	342826.0	269734.4
0.71	0.5516919	0.5060346	50624774.7	342826.1	269734.3
0.72	0.5502479	0.5156469	52976829.1	344834.1	268333.4
0.73	0.5490104	0.5156425	52976206.2	344833.2	268334.1
0.74	0.5479271	0.5330710	57636118.5	348474.4	265793.9
0.8	0.5444988	0.5330710	57636113.5	348474.3	265793.9
0.9	0.5387847	0.5330708	57636081.8	348474.3	265794.0

provided that the feasible region of model (57) is nonempty. In other words, one can obtain a better return with the same violation by this way.

4.4. *Impacts of Compensation Coefficient.* Since the compensation coefficient ( $\gamma$ ) reflects importance of the least satisfaction degree in model (57), a suitable choice of compensation coefficient is necessary to any decision-maker. Actually, a higher value for  $\gamma$  implies that more attention is paid to a greater lower bound of satisfaction degree, corresponding to more balanced compromising solutions [25].

We are in a position to study how the compensation coefficient affects the satisfaction degree. For this, we solve model (57) by changing the value of compensation coefficient with a step length of 0.01. Additionally, from the results in Section 4.3, we fix  $\delta = 0.02$ . Numerical results are listed in Table 14.

From Table 14, it follows that

(1) With an increasing compensation coefficient, the value of  $\lambda_0$  becomes greater, while the value of  $\lambda$  becomes smaller. When  $\gamma > 0.71$ , the least satisfaction degree  $\lambda_0$ , the center value  $Z_1$ , the upper deviations  $Z_2$ , and the lower deviations  $Z_3$  all generate great changes.

(2) Both of the center value and the upper deviation of the fuzzy objective function,  $Z_1$  and  $Z_2$ , are increasing as the compensation coefficient takes a greater value, while the lower deviation  $Z_3$  is decreasing. This phenomenon implies that the membership degrees of the three objective functions not always become larger or smaller simultaneously.

(3) If the decision-maker prefers to a higher satisfaction degree  $\lambda_0$ , he/she could choose a relatively higher value of  $\gamma$  for an optimal solution. Table 14 shows that there exists a threshold value ( $\gamma = 0.74$  or so for the given scenario).

### 5. Sensitivity Analysis of Model Parameters

In this section, by sensitivity analysis of model parameters, we attempt to reveal some valuable managerial implications from our model and algorithm. Specifically, we will address the following issues:

(1) How do the standard deviations (ST) of random coefficients and the variances of fuzzy coefficients affect the optimal strategy of the ELV recovery system?

(2) What are the impacts of the above deviations and variances on the satisfaction degree  $\sum_{h=1}^3 \theta_h \mu_{Z_h}(x)$ , the center value  $Z_1$ , and the deviations  $Z_2$  and  $Z_3$ ?

5.1. *Impacts of Fuzzy Cost Coefficients.* Since the cost coefficients in model (2) are fuzzy, we conduct sensitivity analysis by changing the variance of fuzzy sets so as to reveal what are their impacts on the optimal solution.

A change of cost coefficients often influences enterprise's decision. So, an interesting question is to answer whether there exist some differences among different types of cost coefficients or not.

We change the dispersion levels of different fuzzy cost coefficients in model (2) by a step length of 5% increment. An addition of 30 scenarios is conducted in order to obtain a generalization of the proposed model. The 30 scenarios are the combination of the 10 different levels of the fixed cost, 10 different levels of the transportation cost, and 10 different levels of the processing cost. Then, we implement Algorithm 1 to solve the corresponding models. Numerical results are presented in Tables 15, 16, and 17 and in Figure 7.

From Tables 15, 16, and 17 and Figure 7, it is clear that

(i) From Tables 15, 16, and 17, the increment in dispersion level of fixed and processing cost coefficients causes a change in choice of opening sites of dismantlers and shredders. But the increment in dispersion level of transportation cost coefficients only influences the choice of opening sites for the dismantlers. In any scenario, the number of opening points is not affected by the dispersion levels of fuzzy cost coefficients.

(ii) A larger dispersion level leads to a lower satisfaction degree  $\sum_{h=1}^3 \theta_h \mu_{Z_h}(x)$ . In Figure 7(a), it is seen that the processing cost has obvious influence on the satisfaction degree, compared with the fixed cost and the transportation cost in the range of 10%-25%.

TABLE 15: Impact of fixed cost coefficients' variance on choice of opening sites.

Open or not	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
K1	1	1	1	1	1	1	1	1	1	1
K2	0	0	0	0	0	0	0	0	0	0
K3	1	0	0	0	0	0	0	0	0	0
K4	0	1	1	0	0	0	1	0	0	1
K5	0	1	1	1	1	1	1	1	1	1
K6	1	0	0	1	1	1	0	1	1	0
L1	0	0	0	0	0	0	0	0	1	1
L2	1	1	1	1	1	1	1	1	1	1
L3	1	1	1	1	1	1	1	1	0	1
L4	0	0	0	0	0	0	0	0	0	0
L5	1	1	1	1	1	1	1	1	1	0

TABLE 16: Impact of processing cost coefficients' variance on choice of opening sites.

Open or not	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
k1	NA	1	1	1	1	0	0	1	1	1
k2	NA	1	0	0	0	1	1	1	1	0
k3	NA	1	0	1	0	1	0	0	0	0
k4	NA	0	1	0	1	1	1	1	0	0
k5	NA	0	1	0	0	0	0	0	0	1
k6	NA	0	0	1	1	0	1	0	1	1
L1	NA	0	0	1	0	1	1	0	0	0
L2	NA	1	0	0	0	0	0	1	1	1
L3	NA	1	1	1	1	1	1	1	1	1
L4	NA	0	1	0	1	0	0	0	0	0
L5	NA	1	1	1	1	1	1	1	1	1

TABLE 17: Impact of transportation cost coefficients' variance on choice of opening sites.

Open or not	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
K1	1	1	1	1	1	1	1	1	1	1
K2	0	1	0	0	1	0	0	1	1	0
K3	1	0	1	1	0	1	1	0	0	1
K4	1	1	1	1	1	1	1	0	0	0
K5	0	0	0	0	0	0	0	0	0	0
K6	0	0	0	0	0	0	0	1	1	1
L1	0	0	0	0	0	0	0	0	0	0
L2	1	1	1	1	1	1	1	1	1	1
L3	1	1	1	1	1	1	1	1	1	1
L4	0	0	0	0	0	0	0	0	0	0
L5	1	1	1	1	1	1	1	1	1	1

Beyond a threshold value (25% or so for the given scenario), their impacts on the satisfaction degree  $\sum_{h=1}^3 \theta_h \mu_{Z_h}^-(x)$  have no sharp distinction.

- (iii) From Figures 7(b), 7(c), and 7(d), the increment in cost dispersion level brings about the increase of the center value  $Z_1$  and the deviations  $Z_2$  and  $Z_3$ , no matter it is the fixed cost, the processing cost, or the transportation cost. The contribution from the processing costs is greater than those from the fixed cost

and the transportation cost. Therefore, we suggest that it is the most important measure to improve the ELV recovery efficiency by adopting advanced processing technology and machinery equipment, which can raise the entire satisfaction degree and reduce the system cost.

5.2. *Impacts of Fuzzy Selling Prices.* Since the selling prices in model (2) are also fuzzy, we similarly conduct the sensitivity analysis of selling price by changing the variance of fuzzy



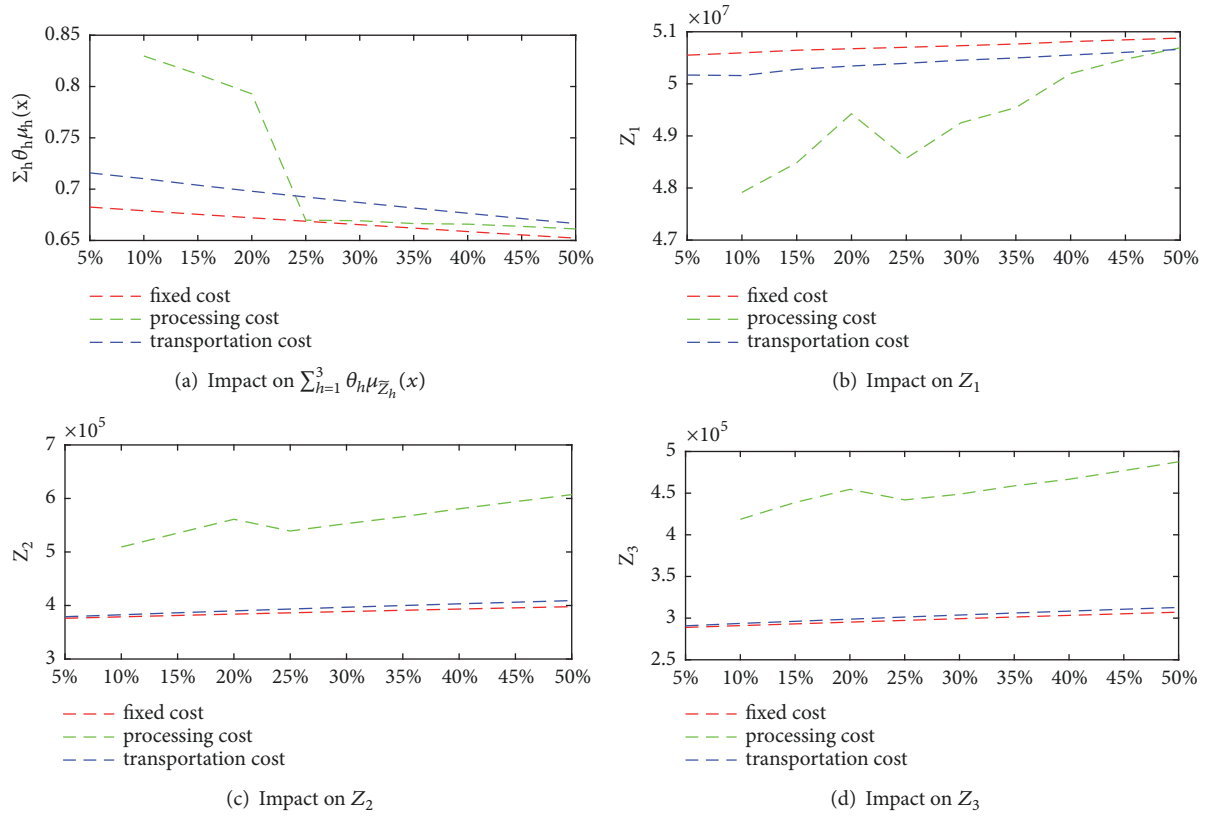


FIGURE 7: Sensitivity of fuzzy cost coefficients' variance.

TABLE 18: Impact of nonferrous components' selling price on choice of opening sites.

Open or not	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
K1	1	1	1	1	1	1	1	1	1	1
K2	0	0	0	0	0	0	0	0	0	0
K3	0	0	0	0	0	0	0	0	0	0
K4	1	0	0	0	0	0	0	0	1	0
K5	1	1	1	1	1	1	1	1	1	1
K6	0	1	1	1	1	1	1	1	0	1
L1	0	0	0	0	1	1	1	1	0	0
L2	1	1	1	1	1	1	1	0	1	1
L3	0	0	0	0	0	0	0	1	1	1
L4	1	1	1	1	0	0	0	0	0	0
L5	1	1	1	1	1	1	1	1	1	1

sets so as to reveal what are its impacts on the optimal solution.

By changing the dispersion level of the fuzzy selling prices in model (2) with a step length of 5% increment, 40 scenarios are generated, which consist of 10 different levels of the selling price of ferrous components, 10 different levels of the selling price of nonferrous components, 10 different levels of the selling price of ferrous material, and 10 different levels of the selling price of nonferrous material.

Implement Algorithm 1 to solve the corresponding models. From the results of numerical experiments, it is found that only the selling price of ferrous and nonferrous components

generates serious impact on choice of opening sites. For this reason, we only present the numerical results on fuzzy selling prices of nonferrous components and ferrous materials in Tables 18 and 19, while those on the fuzzy selling prices of ferrous components and nonferrous materials are omitted. In Figure 8, we further present the impacts of their variances on  $\sum_{h=1}^3 \theta_h \mu_{Z_h}(x)$ ,  $Z_1$ ,  $Z_2$ , and  $Z_3$ , respectively.

From Tables 18 and 19, and Figure 8, it is easy to see that

- (i) Tables 18 and 19 show that the increment in dispersion level of selling price of nonferrous components and ferrous material generates a change in choice of

TABLE 19: Impact of ferrous material' selling price on choice of opening sites.

Open or not	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
K1	1	1	1	1	1	1	1	1	1	1
K2	0	0	0	0	0	0	0	0	0	0
K3	0	0	0	0	0	0	0	0	0	0
K4	1	0	1	1	1	0	0	0	1	0
K5	1	1	1	1	1	1	1	1	1	1
K6	0	1	0	0	0	1	1	1	0	1
L1	0	1	0	1	0	1	1	0	1	0
L2	1	1	1	1	0	0	0	1	0	1
L3	0	1	0	1	1	1	1	1	1	1
L4	1	0	1	0	1	0	0	0	0	0
L5	1	0	1	0	1	1	1	1	1	1

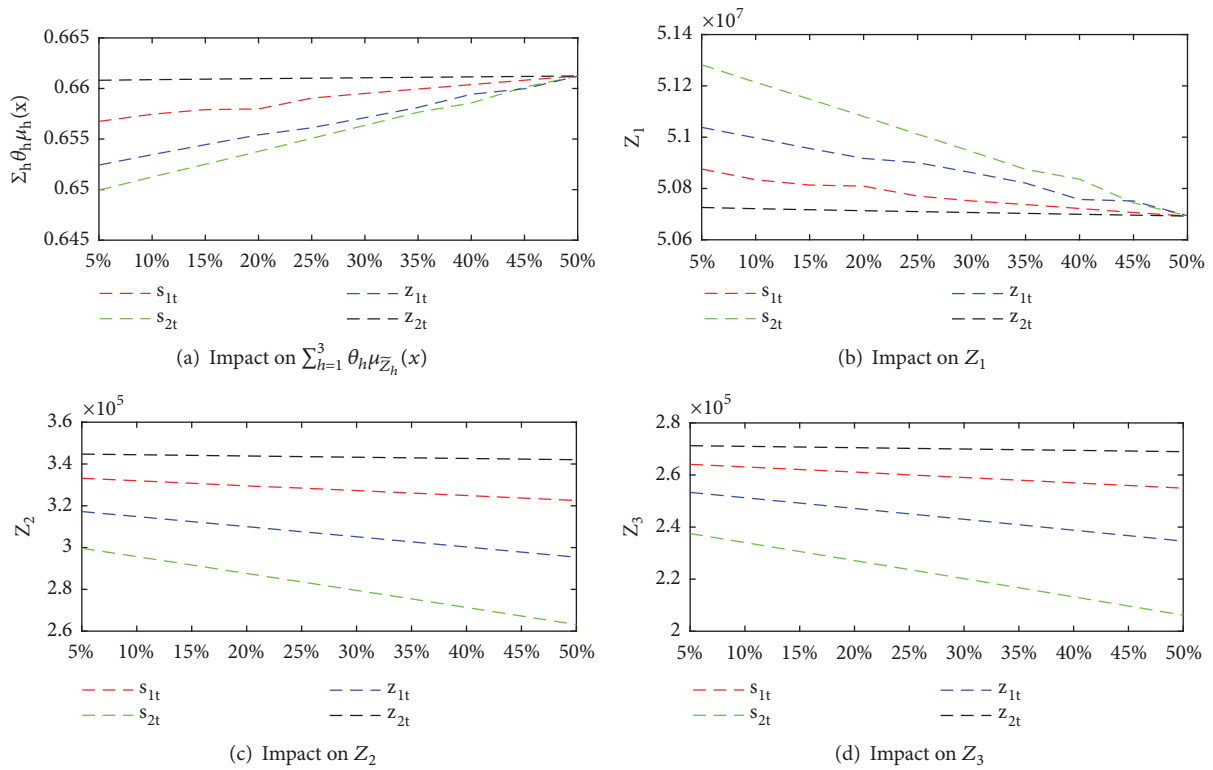


FIGURE 8: Sensitivity of selling price's variance.

opening sites for the dismantlers and shredders. In any scenario, the number of opening sites is not affected by the dispersion levels of fuzzy selling price.

(ii) Larger dispersion level leads to greater satisfaction degree  $\sum_{h=1}^3 \theta_h \mu_{Z_h}(x)$ . From Figure 8(a), it is clear that, with increasing variances, the satisfaction degree  $\sum_{h=1}^3 \theta_h \mu_{Z_h}(x)$  becomes greater. The selling price of nonferrous components causes the most obvious change, followed by ferrous material, ferrous components, and nonferrous material. When the increment reaches 50%, their satisfaction degrees are the same.

(iii) From Figures 8(b), 8(c), and 8(d), an increasing selling price dispersion level causes a drop of the center value,  $Z_1$ , and the variances,  $Z_2$  and  $Z_3$ . The contribution from the selling price of nonferrous components is greater than those from the others. When the increment reaches a threshold value (50% or so for the given scenario), the center values  $Z_1$  are the same. Therefore, more satisfactory solutions may be obtained as the selling price variances of the nonferrous components or ferrous materials properly increase.

A comparison between Figures 7 and 8 indicates that reducing the variances of fuzzy cost coefficients is more

TABLE 20: Impact of capacity's ST in dismantlers.

Open or not	5%-10%	15%	20%-35%	40%	45%-90%	95%	100%
K1	1	1	1	1	1	1	1
K2	0	0	0	0	0	0	0
K3	0	0	0	0	0	0	0
K4	0	1	0	1	0	1	0
K5	1	1	1	1	1	1	1
K6	1	0	1	0	1	0	1

TABLE 21: Impact of capacity's ST in shredders.

Open or not	5%	10%	15%-60%	65%	70%-95%	100%
K1	1	1	1	1	1	1
K2	0	0	0	0	0	0
K3	0	0	0	0	0	0
K4	0	1	0	1	0	1
K5	1	1	1	1	1	1
K6	1	0	1	0	1	0

TABLE 22: Impact of capacity's ST in landfill.

Open or not	5%-40%	45%-55%	60%	65%	70%-75%	80%	85%-90%	95%-100%
K1	1	1	1	1	1	1	1	1
K2	0	0	0	0	0	0	0	0
K3	0	0	0	0	0	0	0	0
K4	0	1	0	1	0	1	0	1
K5	1	1	1	1	1	1	1	1
K6	1	0	1	0	1	0	1	0

TABLE 23: Impact of demand's ST in secondary markets for ferrous components.

Open or not	5%-25%	30%-70%	75%	80%-85%	90%-100%
K1	1	1	1	1	1
K2	0	0	0	0	0
K3	0	0	0	0	0
K4	0	1	0	1	0
K5	1	1	1	1	1
K6	1	0	1	0	1

effective on increment of satisfaction degree and reduction of expected total cost than that by increasing the variances of fuzzy selling prices. Thus, instead of increasing the selling prices, decision-makers should pay more attention to reduction of the processing costs, the transportation costs, or the fixed opening costs.

*5.3. Impacts of Randomness.* In construction of model, we have assumed that the capacities and demands in model (2) are random. Thus, one of our concerns is to test the impact of their ST on the optimal solution.

By changing the value of ST with a step length of 5% increment, 120 scenarios are generated. Then, we implement Algorithm 1 to solve the corresponding models. From the results of numerical experiments, it is found that the capacity's ST in dismantlers, shredders, and landfills and the

demand's ST in the secondary markets for ferrous components have impacts on the choice of opening sites for the dismantlers, while none of the ST generates serious impact on the choice of opening sites for the shredders. For this reason, we only present the numerical results that cause impacts on the choice of opening sites in Tables 20, 21, 22, and 23, while the other scenarios are omitted. In Figure 9, we present the impacts of ST on  $\sum_{h=1}^3 \theta_h \mu_{\bar{z}_h}(x)$ ,  $Z_1$ ,  $Z_2$ , and  $Z_3$ , respectively.

From Tables 20, 21, 22, and 23 and Figure 9, it is concluded that

- (i) The increment of ST causes two kinds of schemes for choosing the opening sites of dismantlers. In our scenario analysis, we should open Dismantlers 1, 5, and 6 or open Dismantlers 1, 4, and 5 (see Tables 20,

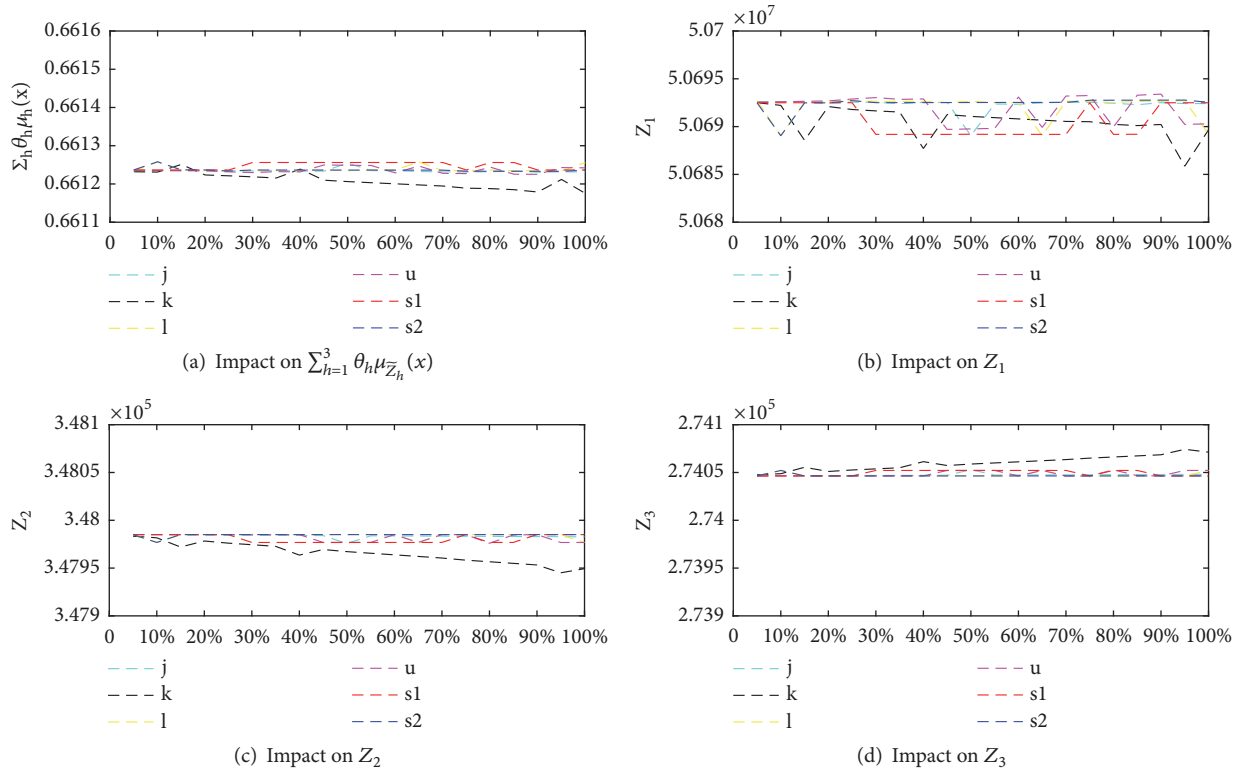


FIGURE 9: Sensitivity of standard deviation.

21, 22, and 23). In any scenario, the number of opening sites is not affected by ST.

- (ii) An increasing ST has almost the same impact on the satisfaction degree, i.e.,  $\sum_{h=1}^3 \theta_h \mu_{\bar{z}_h}(x)$ , but the impact is extremely small (see Figure 9(a)). Comparatively speaking, the impact of the capacity's ST in dismantlers is obvious than those in the other nodes.
- (iii) The center value  $Z_1$  and the deviations  $Z_2$  and  $Z_3$  are affected by the increasing ST (see Figures 9(b), 9(c), and 9(d)). However, compared with Figures 7 and 8, it is seen that the impacts caused by ST are quite smaller than that caused by that of cost coefficients or selling prices on the center value  $Z_1$  and the deviations  $Z_2$  and  $Z_3$ . That is to say, uncertainty of costs may be more critical to the decision-making than that of demand and capacity if the demands or the processing capacities are large enough, as shown in the conducted case study.

## 6. Conclusions and Directions of Future Research

In this paper, we have built an optimization model with fuzzy and stochastic parameters for the production planning problems of recycling ELVs under polymorphic uncertain environment by taking into account a number of uncertain parameters.

For this complicated PUOM, a so-called chance-constrained and multiobjective programming method has been proposed to find a compromising solution for an optimal plan of recycling ELVs. Scenario analysis and sensitivity analysis have indicated that the developed algorithm is efficient and can provide a number of valuable managerial insights from the PUOM model.

Specifically, main results in this paper include the following:

- (1) The proposed model and the developed algorithm in this paper have provided an efficient quantitative method to find a compromising (optimal) policy for the practical ELV recovery management problem in an uncertain environment. In particular, the proposed method can help decision-maker to choose an optimal number of opening recovery sites, optimal transportation quantities in the ELV recovery network.
- (2) Faced with uncertainty in ELV recovery management, decision-makers could choose a relatively greater compensation coefficient to make a greater least satisfaction degree as they apply the presented model and algorithm in this paper into the practical decision-making. Moreover, by our model and algorithm, some critical threshold values being associated with an optimal recovery policy can be found out in the case that the practical recovery environment changes.

- (3) Variance of fuzzy costs or selling prices may lead to a change of opening sites for the dismantlers or shredders. Reducing the variance of processing costs or increasing the variance of selling prices are two effective methods to make a larger satisfaction degree for the ELV recovery system. Reduction of the costs such as the processing costs and the transportation costs generate more significant impacts on the optimal recycling policy than the change of the component selling prices in the secondary markets.
- (4) Small changes of demand and capacity would not generate serious impact on the optimal recovery policies in the case that the demands or the processing capacities are large enough. It is suggested that adopting advanced processing technology and machinery equipment is the most important measure to improve the ELV recovery efficiency in this case. That is to say, fuzziness of costs may be more critical to the decision-making than randomness of demand and capacity.

For future research, the problem can be extended to a closed-loop green supply chain system. Especially, if the deterministic equivalent formulation of the original model is nonsmooth or is involved with hundreds of integer variables (the number of dismantlers and shredders), then development of heuristic algorithms is necessary since it is an NP-hard problem.

Additionally, if one considers the games between governments and recycling enterprises, it is worth further studying new models from the perspective of game theory, rather than from the centralized decision-making mode in this paper. Particularly, in the case that the collection centers and the processing centers of ELVs seek to maximize their profits as different agents, it is necessary to construct a new model in decentralized decision-making mode.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

We declare that all the authors have no any conflict of interest about submission and publication of this paper.

## Authors' Contributions

Zhong Wan and Jing Zhang conceived and designed the research plan; Jingjing Liu and Zhang Wan performed the mathematical modelling, numerical analysis and wrote the paper.

## Acknowledgments

This research is supported by the National Natural Science Foundation of China (Grant no. 71671190) and the Natural Science Foundation of Guangdong, China (Grant no. 2016A030310105).

## References

- [1] W. C. Li, H. T. Bai, J. Yin, and H. Xu, "Life cycle assessment of end-of-life vehicle recycling processes in China—take Corolla taxi for example," *Journal of Cleaner Production*, vol. 117, pp. 176–187, 2016.
- [2] V. Simic and B. Dimitrijevic, "Risk explicit interval linear programming model for long-term planning of vehicle recycling in the EU legislative context under uncertainty," *Resources, Conservation & Recycling*, vol. 73, pp. 197–210, 2013.
- [3] L. Wang and M. Chen, "Policies and perspective on end-of-life vehicles in China," *Journal of Cleaner Production*, vol. 44, pp. 168–176, 2013.
- [4] T. Z. Zhang, J. W. Chu, X. P. Wang, X. H. Liu, and P. F. Cui, "Development pattern and enhancing system of automotive components remanufacturing industry in China," *Resources, Conservation and Recycling*, vol. 55, pp. 613–622, 2011.
- [5] R. Cossu and T. Lai, "Automotive shredder residue (ASR) management: An overview," *Waste Management*, vol. 45, pp. 143–151, 2015.
- [6] E. Demirel, N. Demirel, and H. Gökçen, "A mixed integer linear programming model to optimize reverse logistics activities of end-of-life vehicles in Turkey," *Journal of Cleaner Production*, vol. 112, pp. 2101–2113, 2016.
- [7] J. Tian and M. Chen, "Assessing the economics of processing end-of-life vehicles through manual dismantling," *Waste Management*, vol. 56, pp. 384–395, 2016.
- [8] M. Chen and F. Zhang, "End-of-Life vehicle recovery in china: Consideration and innovation following the EU ELV directive," *Journal of Metals*, vol. 61, no. 3, pp. 45–52, 2009.
- [9] G. Tian, H. Zhang, Y. Feng et al., "Operation patterns analysis of automotive components remanufacturing industry development in China," *Journal of Cleaner Production*, vol. 164, pp. 1363–1375, 2017.
- [10] E. Roghanian and P. Pazhoheshfar, "An optimization model for reverse logistics network under stochastic environment by using genetic algorithm," *Journal of Manufacturing Systems*, vol. 33, no. 3, pp. 348–356, 2014.
- [11] Q. Qiao, F. Zhao, Z. Liu, S. Jiang, and H. Hao, "Cradle-to-gate greenhouse gas emissions of battery electric and internal combustion engine vehicles in China," *Applied Energy*, vol. 204, pp. 1399–1411, 2017.
- [12] X. Wang and M. Chen, "Implementing extended producer responsibility: vehicle remanufacturing in China," *Journal of Cleaner Production*, vol. 19, pp. 680–686, 2011.
- [13] J. Tian and M. Chen, "Sustainable design for automotive products: Dismantling and recycling of end-of-life vehicles," *Waste Management*, vol. 34, no. 2, pp. 458–467, 2014.
- [14] K. C. Chen, S. H. Huang, and I. W. Lian, "The development and prospects of the end-of-life vehicle recycling system in Taiwan," *Waste Management*, vol. 30, no. 8–9, pp. 1661–1669, 2010.
- [15] H. Ohno, K. Matsubae, K. Nakajima, Y. Kondo, S. Nakamura, and T. Nagasaka, "Toward the efficient recycling of alloying elements from end of life vehicle steel scrap," *Resources, Conservation & Recycling*, vol. 100, pp. 11–20, 2015.
- [16] R. Cruz-Rivera and J. Ertel, "Reverse logistics network design for the collection of end-of-life vehicles in Mexico," *European Journal of Operational Research*, vol. 196, no. 3, pp. 930–939, 2009.

- [17] E. Özceylan, N. Demirel, C. Çetinkaya, and E. Demirel, "A closed-loop supply chain network design for automotive industry in Turkey," *Computers & Industrial Engineering*, vol. 113, pp. 727–745, 2016, <http://dx.doi.org/10.1016/j.cie.2016.12.022>.
- [18] S. Ene and N. Öztürk, "Network modeling for reverse flows of end-of-life vehicles," *Waste Management*, vol. 38, pp. 284–296, 2015.
- [19] P. N. Phuc, V. F. Yu, and Y. Tsao, "Optimizing fuzzy reverse supply chain for end-of-life vehicles," *Computers & Industrial Engineering*, vol. 113, pp. 757–765, 2017.
- [20] V. Simic, "Fuzzy risk explicit interval linear programming model for end-of-life vehicle recycling planning in the EU," *Waste Management*, vol. 35, pp. 265–282, 2015.
- [21] V. Özkar and H. Baslgil, "Multi-objective optimization of closed-loop supply chains in uncertain environment," *Journal of Cleaner Production*, vol. 41, pp. 114–125, 2013.
- [22] V. Simic, "Interval-parameter chance-constraint programming model for end-of-life vehicles management under rigorous environmental regulations," *Waste Management*, vol. 52, pp. 180–192, 2016.
- [23] X. Zhang, S. Huang, and Z. Wan, "Optimal pricing and ordering in global supply chain management with constraints under random demand," *Applied Mathematical Modelling*, vol. 40, pp. 10105–10130, 2016.
- [24] Z. Wan, H. Wu, and L. Dai, "A polymorphic uncertain equilibrium model and its deterministic equivalent formulation for decentralized supply chain management," *Applied Mathematical Modelling*, vol. 58, pp. 281–299, 2018.
- [25] S. A. Torabi, M. Ebadian, and R. Tanha, "Fuzzy hierarchical production planning (with a case study)," *Fuzzy Sets and Systems*, vol. 161, no. 11, pp. 1511–1529, 2010.
- [26] Y. X. Pan and H. T. Li, "Sustainability evaluation of end-of-life vehicle recycling based on emergy analysis: a case study of an end-of-life vehicle recycling enterprise in China," *Journal of Cleaner Production*, vol. 131, pp. 219–227, 2016.
- [27] Q. H. Zhao and M. Chen, "A comparison of ELV recycling system in China and Japan and China's strategies," *Resources, Conservation and Recycling*, vol. 57, pp. 15–21, 2011.
- [28] Y. X. Li, Z. Wan, and J. J. Liu, "Bi-level programming approach to optimal strategy for VMI problems under random demand," *The ANZIAM Journal*, vol. 59, no. 2, pp. 247–270, 2017.
- [29] Z. Wan, S. Zhang, and K. L. Teo, "Polymorphic uncertain nonlinear programming approach for maximizing the capacity of V-belt driving," *Optimization and Engineering*, vol. 15, no. 1, pp. 267–292, 2014.
- [30] Y. Lai and C. Hwang, "A new approach to some possibilistic linear programming problems," *Fuzzy Sets and Systems*, vol. 49, no. 2, pp. 121–133, 1992.
- [31] M.-F. Yang and Y. Lin, "Applying fuzzy multi-objective linear programming to project management decisions with the interactive two-phase method," *Computers & Industrial Engineering*, vol. 66, no. 4, pp. 1061–1069, 2013.
- [32] K. Govindan, P. Paam, and A. Abtahi, "A fuzzy multi-objective optimization model for sustainable reverse logistics network design," *Ecological Indicators*, vol. 67, pp. 753–768, 2016.
- [33] B. Vahdani, R. Tavakkoli-Moghaddam, and F. Jolai, "Reliable design of a logistics network under uncertainty: a fuzzy possibilistic-queueing model," *Applied Mathematical Modelling*, vol. 37, no. 5, pp. 3254–3268, 2013.
- [34] S. A. Torabi and E. Hassini, "An interactive possibilistic programming approach for multiple objective supply chain master planning," *Fuzzy Sets and Systems*, vol. 159, no. 2, pp. 193–214, 2008.



## Review Article

# A Survey on the Electric Vehicle Routing Problem: Variants and Solution Approaches

Tomislav Erdelić  and Tonči Carić 

Faculty of Transport and Traffic Sciences, University of Zagreb, Vukelićeva Street 4, Zagreb, Croatia

Correspondence should be addressed to Tomislav Erdelić; [terdelic@fpz.hr](mailto:terdelic@fpz.hr)

Received 14 December 2018; Revised 5 March 2019; Accepted 2 April 2019; Published 9 May 2019

Guest Editor: Eduardo Lalla-Ruiz

Copyright © 2019 Tomislav Erdelić and Tonči Carić. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to ensure high-quality and on-time delivery in logistic distribution processes, it is necessary to efficiently manage the delivery fleet. Nowadays, due to the new policies and regulations related to greenhouse gas emission in the transport sector, logistic companies are paying higher penalties for each emission gram of CO<sub>2</sub>/km. With electric vehicle market penetration, many companies are evaluating the integration of electric vehicles in their fleet, as they do not have local greenhouse gas emissions, produce minimal noise, and are independent of the fluctuating oil price. The well-researched vehicle routing problem (VRP) is extended to the electric vehicle routing problem (E-VRP), which takes into account specific characteristics of electric vehicles. In this paper, a literature review on recent developments regarding the E-VRP is presented. The challenges that emerged with the integration of electric vehicles in the delivery processes are described, together with electric vehicle characteristics and recent energy consumption models. Several variants of the E-VRP and related problems are observed. To cope with the new routing challenges in E-VRP, efficient VRP heuristics and metaheuristics had to be adapted. An overview of the state-of-the-art procedures for solving the E-VRP and related problems is presented.

## 1. Introduction

The vehicle routing problem (VRP) is an NP-hard optimization problem that aims to determine a set of least-cost delivery routes from a depot to a set of geographically scattered customers, subject to side constraints [1]. The problem was first defined by Dantzig and Ramser [2] as the Truck Dispatching Problem. VRP is a generalization of the well-known traveling salesman problem (TSP), which aims to design one least-cost route to visit all the customers. The problem has applications in several real-life optimization problems, which has led to the definition of many problem variants over the years: limited vehicle load capacity (capacitated VRP, CVRP), customer time windows (VRP with time windows, VRPTW), multiple depots (multidepot VRP, MDVRP), pickup and delivery (VRP with pickup and delivery, VRPPD), time-dependent travel time (time-dependent VRP, TD-VRP), heterogeneous fleet (mixed fleet VRP, MFVRP), etc. [3, 4]. Due to the complexity of the problem, exact procedures are only capable of optimally solving small-sized problems: up

to 360 customers for CVRP [5] and 50-100 customers for VRPTW [6]. Over the years, a vast number of heuristics, metaheuristics, and hybrid procedures were proposed for solving different VRP problems.

In the past decade, the European Union (EU) has announced many new actions and regulations related to greenhouse gas (GHG) emissions in the transport sector [7]. External factors and the rise of social and ecological awareness have prompted green initiatives in many companies. Conventional internal combustion engine vehicles (ICEVs), which are dependent on limited fossil fuels, severely pollute the environment, especially in congested urban areas. According to WEEA [8], the EU intends to decrease GHG emissions by 20% and 40% by 2020 and 2030, respectively. Sbihi and Eglese [9] introduced the research field of green logistics, which deals with the sustainability of delivery processes by taking into account environmental and social factors. With the electric vehicle (EV) market penetration, many logistic companies evaluated the use of the EVs in their vehicle fleet in order to decrease GHG emissions and,

therefore, reduce the charges for every emission gram of  $\text{CO}_2/\text{km}$ . EVs have several advantages compared to ICEVs: (i) they do not have local GHG emissions; (ii) they produce minimal noise; (iii) they can be powered from renewable energy sources; and (iv) they are independent of the fluctuating oil price [10, 11]. There are two basic configurations of EVs: the battery electric vehicle (BEV), which is exclusively powered from batteries mounted inside the vehicle, and the hybrid electric vehicle (HEV), which can be powered from batteries inside the vehicle or by other energy sources, most commonly internal combustion engine. The plug-in HEV (PHEV) can be recharged by connecting the plug to the electric power source. In this paper mostly BEVs for logistic purposes are considered, where two main problems come to the fore: limited driving range and the need for additional recharging infrastructure.

Due to the limited battery capacity, the range that delivery BEVs can achieve with a fully charged battery is 160-240 km [12], which is much lower than the 480-650 km range of ICEVs [13]. To achieve a similar driving range as ICEVs, BEVs have to visit charging stations (CSs) more frequently. Today, there is still a lack of CSs in the road network infrastructure, and their locations and energy demand should be planned in future infrastructural plans. For an empty BEV to become operable again, battery energy has to be renewed at a CS. This can be performed in two ways: (i) by swapping empty batteries with fully charged ones at Battery Swapping Station (BSS) or (ii) by charging at CS [14, 15]. The former process can be performed in a time comparable to the refueling time of ICEVs. In the latter process BEVs recharge their batteries at CSs by plugging into the electric power source. The recharge time depends on the state of charge (SoC) when entering the CS, the desired SoC level when leaving CS, and the charging function.

*1.1. Recent Literature Reviews and Scientific Contribution.* Here we present, to our knowledge, the most recent literature reviews on the E-VRP and related problems.

Juan et al. [16] presented a review regarding the environmental, strategic, and operational challenges of EV integration in logistics and transportation activities. The authors performed a comprehensive analysis of environmental challenges including the transportation impact on the pollution and EVs' possible contribution to the reduction of carbon emissions. Regarding the strategic challenges, the authors presented issues related to CSs: battery swap technology, different charging technologies, CS location problem, limited number of charges, and charging network; issues related to the EVs: mixed fleet of ICEVs and EVs, economic challenges when integrating EVs in a fleet, and routing constraints. The solution approaches for the E-VRP were presented in general, as a class of VRP solving procedures. Similarly, Margaritis et al. [17] presented practical and research challenges of EVs focusing on battery development, lack of charger compatibility, systematic energy management, the lack of optimization procedures that could minimize the EV routing and scheduling decisions, cooling/heating usage, financial sources, novel policies, measures for EV deployment in transport services, etc.

Montoya [18] researched several variants of the E-VRP: green VRP (GVRP), E-VRP with partial recharging and nonlinear charging functions, and the technician routing problem with a mixed fleet of ICEVs and EVs. For each problem, effective solving procedures were proposed: multi-space sampling heuristic, iterated local search enhanced with heuristic concentration, and two-phase parallel metaheuristic based on solving a set of subproblems and extended set-covering formulation. The authors also formulated a fixed route vehicle charging problem (FRVCP) with and without time windows to optimize the charging decisions for a route with a fixed customer sequence. The authors did not focus as much on reviewing the E-VRP literature, especially not on the procedures for solving the problem. Compared to Montoya [18], this paper did not focus as much on the FRVCP and technician routing problem or on detailed procedures used to solve those problems.

The most recent survey on the E-VRP is presented by Pelletier et al. [19] and it includes technical background on EV types and batteries, EV market penetration, EV competitiveness and incentives, and an overview of the existing research regarding EVs in transportation science. The authors provided a comprehensive review on, at the time, the latest solving procedures for the E-VRP with mixed fleet and optimal paths and covered several key papers regarding partial recharges, hybrid vehicles, and different charging technologies.

In this paper, a survey on the E-VRP is presented, which includes approaches for solving the E-VRP and related problems that emerged with BEVs integration in the logistic processes. The focus is not on the economic and environmental challenges related to BEVs. Most of the latest literature reviews were published in 2016; hence, to the best of our knowledge, there is no published research that summarizes the state-of-the-art research in the E-VRP field. In this paper we outline the following contributions:

- (i) a review of the recent energy consumption models that could be used in BEV routing models;
- (ii) an updated literature review and a concise table summary of already reviewed E-VRP variants such as GVRP, mixed fleet, BSSs, partial recharges, and different charging technologies;
- (iii) a review of the additional emerged E-VRP variants, which include hybrid vehicles, CS siting, nonlinear charging function, dynamic traffic conditions and charging schedule optimization;
- (iv) a comprehensive analysis of operation research procedures in the E-VRP, which includes an overview of the procedures employed for solving various E-VRP variants, highlighting state-of-the-art procedures, and a concise table summary of the applied procedures.

*1.2. Organization of This Paper.* The remainder of this paper is organized as follows. In Section 2, the E-VRP is described and the literature review of the basic problem formulation is presented. In Section 3, basic characteristics and recent energy consumption models of BEVs are described together

with the BEV's application and evaluation in the delivery processes. In Section 4, variants of the E-VRP are presented with some related problems from the literature. In Section 5, approaches for solving E-VRP are presented, which include state-of-the-art exact, heuristic, metaheuristic, and hybrid procedures. The conclusion and future research directions are given in Section 6.

## 2. Electric Vehicle Routing Problem

With BEV penetration in logistic distribution processes, a problem of routing a fleet of BEVs has emerged: the E-VRP. The E-VRP aims to design least-cost BEV routes in order to serve a set of customers by taking into account often used constraints: vehicle load capacity, customer time windows, working hours, etc. [3, 20]. Additionally, BEVs have the limited driving range which directly corresponds to more frequent recharging events at CSs. CSs can be built at separate locations as public CSs or mounted at customers' locations as private CSs. The time needed to travel to a CS and the recharging time are important aspects of fleet routing, especially if customer time windows are taken into account.

To the best of our knowledge, the first research regarding the routing of an electric fleet was published by Gonçalves et al. [21], with the authors observing VRPPD using a mixed fleet of EVs and ICEVs. Refueling of the vehicle is performed at the location where the need for the refueling occurred and the total refuel time is computed, based on the total distance traveled by the EV. Conrad and Figliozzi [22] formulated recharging VRP in which vehicles with limited driving range are allowed to refuel at customers' locations during the route to up to 80% of the vehicle's battery capacity. The authors described an application of the model and analyzed the impact of different driving ranges, fixed charging times, and time windows on EV routes and solution quality. The results indicated that customer time windows greatly limit route distance when recharging time is long and vehicle range is constrained. Erdoğan and Miller-Hooks [23] formulated GVRP in which a fleet of vehicles is powered by alternative fuels (alternative fuel vehicle, AFV): biodiesel, ethanol, hydrogen, methanol, natural gas, electricity, etc. AFVs can refuel at separately located stations, with fixed refueling time. The authors did not consider customer time windows and vehicle load capacity constraints. New problem-specific instances were developed and two heuristics for solving the problem were applied. The results showed that the limitation of driving range severely increased the number of refueling stations and the total traveled distance in the solution. A similar problem was researched by Omidvar and Tavakkoli-Moghaddam [24], in which the authors added vehicle load constraint, customer time windows, and congestion management, as the vehicle can stay at the customer's location during the congestion hours. The authors minimized emission and pollution costs by applying commercial software for solving small instances and two metaheuristics for solving larger instances. Schneider et al. [25] published the first research on routing a BEV fleet by taking into account possible visits to CSs and charging time dependent on the SoC level when entering a CS. The problem was formulated as E-VRP

with time windows (E-VRPTW) as they took into account load, battery, and time window constraints. The authors formulated E-VRPTW as the mixed integer linear program (MILP) on the complete directed graph  $G$ , where customers are modeled as graph vertices and paths between customers are modeled as graph arcs. Here, the basic model formulation is given. Let  $V = \{1, \dots, N\}$  be a set of geographically scattered customers who need to be served, and let  $F$  be a set of CSs for BEVs. In order to allow multiple visits to the same CS, a virtual set of CSs  $F'$  is defined. Vertices 0 and  $N + 1$  denote the depot, and every route begins with vertex 0 and ends with vertex  $N + 1$  ( $V_{0,N+1} = V \cup \{0\} \cup \{N + 1\}$ ). Graph  $G$  is defined as  $G = (V_{0,N+1} \cup F', A)$ , where  $A$  is set of arcs,  $A = \{(i, j) \mid i, j \in V_{0,N+1} \cup F', i \neq j\}$ . Depending on the real-life constraints, different  $(i, j)$  arc values can be interpreted as distance  $d_{ij}$ , travel time  $t_{ij}$ , energy consumption  $e_{ij}$ , speed  $v_{ij}$ , cost  $c_{ij}$ , etc. The binary variable  $x_{ij} = \{0, 1\}$  is equal to 1 if arc  $(i, j)$  is traversed in the solution, and 0 otherwise. The whole MILP program for the E-VRPTW with equations for load, battery, time windows, flow, and subtour constraints is presented by Schneider et al. [25].

In the VRP, it is customary for the primary objective to minimize the total number of vehicles used (1) and then to minimize the total distance traveled (2) or some other objective functions [26]. Total vehicle number is a primary objective as generally greater savings can be achieved with fewer vehicles (vehicle fixed costs, labor cost, etc.). Such an objective is contradictory as with fewer vehicles total traveled distance increases and vice versa. By taking into account the high purchase cost of BEVs, such a hierarchical objective seems justifiable in BEV routing applications [25, 27].

$$\min \sum_{j \in V \cup F'} x_{0j} \quad (1)$$

$$\min \sum_{i \in V_0 \cup F', j \in V_{N+1} \cup F', i \neq j} d_{ij} x_{ij} \quad (2)$$

Objective functions can be complex with simultaneous minimization of vehicle number, total traveled distance [28], total travel times [29], total routing cost and planning horizon [11, 27, 30, 31], GHG emission [32, 33], energy consumption [34–36], etc. Total routing costs of BEVs usually consist of BEV acquisition cost, circulation tax, maintenance, costs related to the energy consumption (electric energy price), cost of battery pack renewal after its lifetime, labor costs, etc. Instead of a single-objective function, some authors use multiobjective function, i.e., fuel consumption and total driving time [37], fuel consumption and route cost [38], battery swapping and charge scheduling [39], etc. An overview of different objectives in E-VRP is presented in Table 1 in the column *Objective*.

## 3. Battery Electric Vehicles in Delivery Processes

The major problem that BEVs in delivery processes are facing is the limited driving range. Grunditz and Thiringer [101] analyzed over 40 globally available BEVs, which can

TABLE 1: Overview of the E-VRP variants and related problems.

Reference	Problem name	C	TW	MIX	LRP	F	Charging			Other	Objective
							L	NL	PR		
Bektaş and Laporte [40]	PRP	X	X						Emission model, speed limitation	Total costs: labor, fuel and emission as a function of load and speed	
Conrad and Figliozzi [22]	Recharging VRP	X	X			X				Vehicle number and total traveling costs: distance, service time, recharging time	
Gonçalves et al. [21]	VRPPD with mixed fleet	X		X		X			No CS location, time constraint	Total traveling costs: fixed and variable	
Demir et al. [32]	PRP	X	X						Emission model	Total traveling costs: labor, fuel and emission as a function of load and speed; speed optimization	
Erdogan and Miller-Hooks [23]	GVRP					X			AFVs, limited route duration	Vehicle number and total traveled distance	
Omidvar and Tavakkoli-Moghaddam [24]	GVRP	X	X			X			AFVs, limited fuel capacity and route duration, congestion management	Total costs: vehicle fixed costs, distance, time and emission	
Abdallah [41]	PHEVRPTW	X	X			X	X		Electric charge cost is neglected	Routing costs: time run on the fossil oil	
Barco et al. [34, 42]	E-VRP and charge scheduling	X	X			X	X		Energy consumption and battery degradation model, private and public CS, time-dependent energy rates	Energy consumption	
Davis and Figliozzi [10]		X						X	Energy consumption model with speed profiles, limit route duration and energy consumption (battery capacity), no recharging during the route	Total costs: vehicle purchase, energy, maintenance, tax incentive, battery replacement, routing	
Van Duin et al. [43]	FSMVRPTW, EVFSMVRPTW	X	X	X					No recharge, range constrained by battery, relaxed time window constraints, emission	Total costs: vehicle fixed costs, time and distance	

TABLE I: Continued.

Reference	Problem name	Charging											Objective			
		C	TW	MIX	LRP	F	L	NL	PR	DFC	BS	TD		H	Other	
Adler and Mirchandani [44]	Online routing of BEVs									X					Battery reservations, waiting for fully charged battery	The average vehicle delay time
Alesiani and Maslekar [45]	BEV routing					X									Waiting time cost at CS is proportional to the number of EVs in CS, limiting the number of CSs in route and number of vehicles in CS, energy consumption model	Traveling, charging and energy consumption costs
Demir et al. [37]	PRP	X	X												Fuel consumption and emission model	Bi-objective minimization of (1) fuel consumption and (2) total driving time
Felipe et al. [46]	GYRP-MTPR	X				X		X	X	X						Total recharging costs: fixed and variable
Preis et al. [35]	Energy-optimized routing of BEVs	X	X			X									Energy consumption model	Energy consumption
Sassi et al. [47]	HEVRP-TDMF														Time-dependent charging costs, operating windows and power limitation of CS, compatibility of BEVs with chargers, electricity grid capacity [47, 49]	Vehicle number and total costs: fixed, routing, charging and waiting costs
Sassi et al. [48]	VRP-HFCC	X		X		X		X	X	X						
Sassi et al. [49]	VRP-MFHEV															
Schneider et al. [25]	E-VRPTW	X	X			X										Vehicle number and total traveled distance
Zündorf [50]	EVRC							X	X	X	X				Battery constrained SPP, different CS types: regular, superchargers and BSS, energy consumption model	Travel time
Bruglieri et al. [29, 51]	E-VRPTW	X	X			X		X								Vehicle number and total travel, recharging and waiting time
Goeke and Schneider [30]	E-VRPTW MF	X	X	X		X									Energy consumption model: varying BEV load, road slope	Vehicle number and different objectives: (i) distance, (ii) costs: labor, driver wage, vehicle propulsion - electric energy and diesel costs, (iii) (ii) + battery replacement cost



TABLE I: Continued.

Reference	Problem name	Charging											Objective				
		C	TW	MIX	LRP	F	L	NL	PR	DFC	BS	TD		H	Other		
Lebeau et al. [52]	FSMVRPTW-EV	X	X	X			X									Energy consumption model based on the collected data, recharge only at the depot	Total costs: vehicles fixed and operating costs, labor costs
Moghaddam [53]	E-VRPTWPR	X	X				X		X							Capacitated CSs	Vehicle number and number of CSs
Pourazarm et al. [54]	Single/Multi BEV routing						X		X			X				Homogeneous and in-homogeneous CSs	Total time
Schneider et al. [55]	VRPIS (EVRPRF)	X					X										Total travel and fixed vehicle costs
Yang and Sun [56]	BSS-EV-LRP	X			X							X					Total routing and construction costs
Desaulniers et al. [57]	E-VRPTW-SF/MF/SP/MP	X	X				X		X								Vehicle number and total routing costs
Doppstadt et al. [58]	HEV-TSP						X									Four modes of travel: combustion, electric, charging and boost mode, no CSs visits	Total costs as long as maximal route duration is not overrun
Hiermann et al. [31]	E-FSMFTW	X	X	X			X										Vehicle number and total costs: vehicle fixed and routing costs
Keskin and Çatay [28]	E-VRPTWPR	X	X				X		X								Vehicle number and total traveled distance
Koç and Karaoglan [59]	GVRP							X									Vehicle number and total traveled distance
Lin et al. [60]	E-VRP	X					X									Energy consumption model and load effect	Total costs: battery charging, travel time and waiting costs
Mashiakova [61]	Routing and charging of electric buses	X					X		X							Energy consumption model, two types of buses depending on the charging event: en-route or at depot, homogeneous and in-homogeneous CSs	Investment and operations costs, a travel time of passengers
Mirmohammadi et al. [62]	Periodic green VRP	X	X	X								X				Periodic routing, primary and secondary time windows, static traffic conditions within a period	Total emissions, total service time and penalties



TABLE I: Continued.

Reference	Problem name	Charging											Other	Objective		
		C	TW	MIX	LRP	F	L	NL	PR	DFC	BS	TD			H	
Montoya et al. [63]	GVRP					X										Vehicle number and total traveled distance
Roberti and Wen [64]	E-TSPTW	X					X			X						Total traveled distance
Schiffer et al. [11, 27]	E-LRPTWPR	X	X		X		X		X							Total costs during the planning period: CS and BEV investment costs, fixed costs (tax, maintenance), distance dependent costs (energy)
Wen et al. [65]	E-VSP						X			X						Total cost: vehicles and traveling costs
Andelmin and Bartolini [66]	GVRP				X											Vehicle number and traveled distance
Çatay and Keskin [67]	E-VRPTWPR	X	X				X		X	X						Vehicle number and total recharging costs
Froger et al. [68]	E-VRP-NL-C								X	X	X					Total travel, service, charging and waiting time
Hof et al. [69]	BSS-EV-LRP	X			X						X					Total routing and construction costs
Leggieri and Haouari [70]	GVRP					X										Vehicle number and total traveled distance
Mancini [71]	HVRP										X					Total traveled distance with penalties for using internal combustion engine
Montoya et al. [72]	E-VRP-NL								X	X	X					Total travel and recharging time
Schiffer and Walther [73]	E-LRPTWPR	X	X		X		X									Total distance, number of vehicles and CSs used, total costs: investment costs of BEVs and CSs, and operational costs
Shao et al. [74]	EVRP-CTVTT	X	X		X		X									Total costs: travel, charging, penalty, and fixed vehicle costs

TABLE I: Continued.

Reference	Problem name	Charging											Objective				
		C	TW	MIX	LRP	F	L	NL	PR	DFC	BS	TD		H	Other		
Sweda et al. [75]	Adaptive routing and recharging policies of EVs						X		X							Heterogeneous CSs - the probability of being available and expected waiting time, origin-destination pairs	Traveling, waiting and recharging costs
Vincent et al. [76]	HVRP					X						X				Vertex demand in CVRP is associated with travel time for each arc in HVRP	Total costs
Amiri et al. [39]	BSS location & scheduling						X			X						Multi-objective - minimization of (1) battery charging and power loss costs; (2) deviation from nominal voltage; and (3) network capacity releasing	
Bruglieri et al. [77]	E-VReP		X													One way car sharing service, workers with bicycles go to the EVs locations and relocate them, battery level demand request	Multi-objective: (1) minimization of the workers employed; (2) minimization of the duration of the longest route; and (3) maximization of the number of served relocation requests
Joo and Lim [78]	EV routing															Energy SPP, no recharging, energy consumption model	Minimize energy consumption and average speed on the path
Keskin and Çatay [79]	E-VRPTW-FC	X	X				X	X	X	X						M/M/1 queuing system at capacitated CSs, battery capacity restriction, four planning intervals in a day, partial recharge not evident in paper	Vehicle number and total recharging costs
Keskin et al. [80]	E-VRPTW-FC	X	X					X	X							Public and private CSs, capacitated CS, single charging technology per CS	Total cost: energy cost, routing, labor and penalties for late arrivals
Kullman et al. [81]	E-VRP-PP							X	X	X						Public and private CSs, capacitated CS, single charging technology per CS	Expected time to visit all the customers
Li et al. [82]	MBFM & recharging problem			X			X									Electric, diesel, compressed natural gas and hybrid-diesel buses	Total network benefit of replacing old vehicles with new ones within the planning horizon and budget constraints

TABLE I: Continued.

Reference	Problem name	Charging										Objective			
		C	TW	MIX	LRP	F	L	NL	PR	DFC	BS		TD	H	Other
Lu et al. [83]	MTFSP	X		X	X	X					X			Travel request are known a priori in time-varying origin-destination tables, service and deadheaded trips with different consumption rate, BEVs have higher priority than ICEVs	Total operating cost of a taxi company
Masmoudi et al. [84]	DARP-EV	X	X	X						X			Energy consumption model of Genikomsakis and Mitrentsis [85] with constant speed, acceleration and road slope; different vehicle resources; accompanying person seat, handicapped person seat, stretcher and/or a wheelchair, limited user ride time	Total routing costs (distance)	
Paz et al. [86]	MDEVLRTW - BS/PR/BSPR	X	X	X	X	X	X	X	X	X	X		Three MIP models depending on the partial recharge and BSS	Total traveled distance	
Pelletier et al. [87]	EFV-CSP					X	X	X	X	X			Preemptive charging with a limited number of chargers and charging events at the depot, time-dependent energy costs, FRD charge, grid restriction, cyclic and calendar battery degradation	Total charging costs	
Poonthair and Nadarajan [38]	F-GVRP					X							Fuel consumption model, varying speed, AFVs, limited route duration	Bi-objective: (1) routing costs; and (2) fuel consumption	
Schiffer and Walther [88]	LRPIF	X	X	X	X	X	X	X					Loading and refueling facilities	Total costs: investment costs of vehicles and facilities, routing costs	

TABLE I: Continued.

Reference	Problem name	C	TW	MIX	LRP	F	Charging				Other	Objective
							L	NL	PR	DFC		
Schiffer and Walther [89]	RELRTWPR	X	X		X		X		X		Uncertain customer pattern scenarios over working days regarding the spatial customer distribution, demand and service time windows	Total costs: investment costs of vehicles and facilities, routing costs
Shao et al. [90]	E-VRP	X				X					Energy consumption model: cargo load, uncertain travel speed	Total costs: travel, charging and fixed vehicle costs
Wang et al. [91]	BEV routing						X				Parking fee, capacitated CSs - queuing time	Multi-objective minimization of (1) travel time: driving, queuing and charging time; (2) charging costs: electricity, service and parking fee; (3) energy consumption
Zhang et al. [36]	E-VRP	X									Energy consumption model, emissions, static speed, charging time unknown	Energy consumption
Basso et al. [92]	2sEVRP	X	X			X		X			Each CSs can have different charging rate, energy consumption model for road segments: speed profile, road slope, acceleration	Total energy consumption
Breunig et al. [93]	E2EVRP	X									Two echelons - first ICEVs and second BEV's, full recharge when visiting CSs, charging time unknown	Total routing costs
Bruglieri et al. [94]	GVRP					X					AFV's, limited route duration	Vehicle number and total traveled distance
Froger et al. [95]	E-VRP-NL						X	X	X			Total travel and charging time
Hiermann et al. [96]	H <sup>2</sup> E-FTW	X	X	X			X				Propulsion mode decision	Total costs: fixed and variable
Jie et al. [97]	2E-EVRP-BSS	X									Routing in two echelons, sensitivity analysis of battery driving range and vehicle emissions	Total routing costs, the battery swapping costs and the handling costs at the satellites

TABLE I: Continued.

Reference	Problem name	Charging													Objective	
		C	TW	MIX	LRP	F	L	NL	PR	DFC	BS	TD	H	Other		
Koyuncu and Yavuz [98]	MGVRP	X	X	X	X	X	X	X	X	X					ICEVs (fixed recharge time) and AFVs, node- and -arc MILP formulation, single recharge technology per CS, additional modeling: customer demands, customer vehicle restrictions, subscription or pay-as-you-go refueling costs, completely heterogeneous fleet, last-mile delivery and closed time windows	Total traveling cost
Macrina et al. [99]	GMFVRP-PRTW	X	X	X	X	X	X	X	X	X	X			Single recharge technology per CS, but different charging technologies between CSs, energy consumption model for road segments with time-dependent speeds	Cost of energy recharged during the route and at the depot, fuel costs, and cost related to traveled distance	
Macrina et al. [33]	GMFVRP-PRTW	X	X	X	X	X	X	X	X	X				Single recharge technology per CS, but different charging technologies between CSs	Recharging, routing and activation costs, limit emissions	
Normasari et al. [100]	CGVRP	X				X								AFVs, limited route duration	Vehicle number and total traveled distance	

Reference: referenced paper; Problem name: the name of the analyzed problem; 13 columns representing characteristics of the problem in the following order: C: vehicle load (cargo) capacity, TW: customer time windows, MIX: heterogeneous (mixed) fleet, LRP: location routing problem, F: fixed (constant) refuel (recharge) time, L: linear charging process, NL: nonlinear charging process, PR: partial recharge strategy, DFC: different charging technologies, BS: battery swap strategy, TD: time-dependent travel times, H: hybrid vehicles, Other: some special characteristic of the problem; Objective: the objective function for the optimization.

be categorized into small, medium-large, high-performing, and sports cars. All of the BEV models utilize lithium-based batteries, especially lithium-ion [102] with battery capacity and distance varying within the ranges 12-90 kWh and 85-528 km, respectively. An average medium-sized personal BEV has a battery capacity of 30 kWh, which is enough to travel 250 km. In the delivery process, mostly light vans and freight BEVs are used, which have a shorter driving range (160-240 km) compared to the driving range of ICEVs (480-650 km) [13, 43, 103]. The reason is that the battery has lower specific energy (130 Wh/kg) than the fossil oil (1233 Wh/kg), and the amount of energy that can be stored in the battery is much lower than in the fossil oil. Batteries mounted in BEVs are mostly the main cause of high acquisition costs and technical limitations as the battery degrades over time, resulting in decreased maximal capacity. Pelletier et al. [104] concluded that the battery should be replaced after five to ten years or after 1,000 to 2,000 cycles with large SoC variations. The authors also described factors that influence such battery degradation: overcharging, overdischarging, high and low temperatures, high SoC during storage, large depth of discharge, etc., and they presented battery degradation models that can be used in goods distribution with BEVs.

**3.1. BEV Application.** BEVs are more likely to be used on short distances and/or in urban areas where they are more effective than ICEVs due to the low driving speed, low noise production, frequent stops, and financial incentives. In cases when the average route length is short, such as the average FedEx route length in the USA, which is 68 km [12], BEVs can be applied directly and recharging can be performed on return to the depot. BEVs are already being applied in such occasions: DHL, UPS, FedEx, and Coca-Cola [105, 106] use BEVs mostly for last-mile deliveries as distances are shorter and vehicle loads are lower. Many companies are performing case studies of integrating BEVs in their delivery fleet. Lin et al. [60] were debating the use of BEVs in time-precise deliveries as the long recharging time at CS causes hard completion of an on-time delivery, which then significantly increases the overall routing costs. Davis and Figliozzi [10] and van Duin et al. [43] evaluated a wide range of scenarios to compare the routing costs of ICEVs and BEVs. Van Duin et al. [43] concluded that BEVs have the ability to efficiently perform urban freight transport and meanwhile to reduce the GHG emissions and noise nuisance. Davis and Figliozzi [10] reported that BEVs are not competitive if the solution to the same problem results in a higher number of BEVs than the number of ICEVs. For BEVs to be competitive, the authors pointed out a combination of several key elements: high daily distance (as much as maximum BEV driving range), low speeds and congestions, frequent customer stops, the reduction of a BEV's purchase cost by tax incentives or technology development, and long planning horizon. On the other hand, Schiffer et al. [27] conducted a case study using freight BEVs for deliveries with a planning horizon of five years and compared the results to the delivery done by conventional freight trucks. Several characteristics of the performed case study are important:

(i) delivery radius up to 190 km from the depot; (ii) CSs located at customers' locations, which allows simultaneous charging while unloading goods; (iii) already existing strong current at CSs; and (iv) vehicles returning to the depot at least once a day. Overall, the authors concluded that there is no operational limitation when using BEVs compared to ICEVs, as the used number of vehicles and total traveled distance are competitive, and at the same time, the overall costs are lower with almost 25% less CO<sub>2</sub> emission. A year later, Schiffer et al. [11] repeated the case study with more realistic costs when a strong current at the CS is not available and concluded that, with the higher CS investment costs, BEVs are no longer competitive. To fully assess the integration of BEVs in logistic processes, the authors pointed out three key elements: different network structures, future CO<sub>2</sub> emission policies, and future technology development (battery capacity and charging infrastructure).

**3.2. Energy Consumption.** Due to the low specific energy, energy consumption should be precisely estimated in order to achieve a BEV's maximal driving range and to reduce the overall routing costs. The energy consumption can be estimated by simulation models but due to the complexity of the E-VRP and unknown driving cycles in advance, mostly macroscopic models with several real-world approximations are applied in the BEV routing models. In the available literature, energy consumption is often estimated using longitudinal dynamics model (LDM). Here, the LDM of Asamer et al. [107] is presented. Force  $F$  needed to accelerate and to overcome resistances (grade, rolling, and air) is given by (3), where  $m$  is vehicle mass (mostly empty vehicle),  $a$  acceleration,  $v$  vehicle speed,  $g$  gravitational constant,  $f$  the inertia force of vehicle rotating parts (up to 5% of the total vehicle mass),  $\alpha$  road slope,  $c_r$  rolling friction coefficient,  $c_d$  air drag coefficient,  $\rho$  air density, and  $A$  vehicle frontal air surface. If  $F \geq 0$ , the vehicle is accelerating and power is needed for the movement of BEV (motor mode); otherwise, if  $F < 0$ , deceleration (braking) or driving downhill is occurring and energy is returned into the BEV's battery as the electric engine has the ability to return the energy (recuperating mode). By process of recuperation, up to 15% of totally consumed energy can be returned [51, 108]. Electric power that comes from the battery is divided into the auxiliary power  $P_0$  and mechanical power  $P_m = Fv$ . Auxiliary power is spent on the electronic devices in the vehicle: heating, ventilation, light, etc., which can shorten the BEV's range up to 30% [109]. Battery power  $P_b$  can be computed by (4), where  $\mu_m$  is the transmission coefficient between the electric motor and drivetrain,  $\mu_e$  is the conversion ratio from chemical energy in the battery to electric energy, and  $\mu_g$  is the conversion ratio from mechanical energy on wheels to chemical energy stored in the battery. Energy is returned into the battery only if the force  $F$  is lower than zero and speed is higher than the experimentally determined value  $v_{min}$  [107]. Energy consumption can be computed by the time integration of (4).

$$F = \underbrace{mg \sin \alpha}_{\text{Grade}} + \underbrace{c_r mg \cos \alpha}_{\text{Rolling}} + \underbrace{0.5c_d \rho A v^2}_{\text{Air}} + \underbrace{fma}_{\text{Acc.}} \quad (3)$$



$$P_b = \begin{cases} \mu_e (\mu_m Fv + P_0), & \text{if } F \geq 0 \\ 0, & \text{if } v \leq v_{min} \\ \mu_g Fv + P_0, & \text{else, if } F < 0 \end{cases} \quad (4)$$

Goeke and Schneider [30] extended the energy consumption model by taking into account variable vehicle load mass when delivering goods without acceleration and braking processes. In the CVRP variants, as customers are being served, vehicle load is decreasing. The authors concluded that actual load strongly improves the solution quality, as a large number of solutions generated without taking into account load distribution tend to be infeasible due to the violation of battery capacity or time windows.

Genikomsakis and Mitrentsis [85] presented a more realistic electric engine model: the load-efficiency curve is approximated by piecewise function and the normalization factor is added to take into account the motor size. The authors used the recuperation energy factor dependent on the vehicle speed as follows: below  $v_{min}$  there is no energy recuperation, beyond  $v_{max}$  maximum energy is recuperated, and in between linear interpolation of energy recuperation is assumed. Therefore, regarding the energy recuperation the authors observed two cases: (i) when recuperated energy exceeds the consumption of the auxiliary devices and the excess of energy is stored into the battery; (ii) when recuperated energy is not sufficient to cover the consumption of the auxiliary devices and thus the energy is drawn from the battery. The authors compared their model on nine characteristic driving cycles (short/long, congested/uncongested, highway, etc.) to the FASTSim simulation tool [110], and this resulted in a relative energy consumption error of up to 4%. The developed energy consumption model was used to create a database of coefficients for several key characteristics: motor type, motor power, battery type, road type, road slope, road speed limit, etc.

Macrina et al. [99] developed an energy model for mixed GVRP with partial recharging and time windows. The authors modeled vehicle speed on arc  $(i, j)$  with three phases  $h$ : acceleration ( $h = 1$ ), constant speed ( $h = 2$ ), and deceleration ( $h = 3$ ), resulting with either triangular function ( $1 \rightarrow 3$ ) or trapezoidal function ( $1 \rightarrow 2 \rightarrow 3$ ). The fuel consumption of ICEV is based on the similar LDM model, presented by (5) and (6), where  $u_i$  is the load (cargo) weight,  $\zeta$  fuel-to-air mass ratio,  $\kappa$  heating value of typical diesel fuel,  $\psi$  conversion factor,  $k$  engine friction coefficient,  $N_e$  radial engine speed,  $D_e$  engine displacement,  $\mu_d$  diesel engine efficiency,  $\mu_{dt}$  drivetrain efficiency, and  $\bar{t}_h$  travel spent in phase  $h$  [40, 111]. For the consumption of BEV, the authors proposed a similar model to that presented by (7), where  $\eta_h^+$  and  $\eta_h^-$  are the efficiency of the electric engine in motor and recuperating mode. The authors compared the proposed energy consumption model to the basic model, where energy consumed is proportional to the distance traveled, and the energy consumption model of Goeke and Schneider [30], where acceleration and braking processes are not taken into account. The results showed that, compared to the proposed energy consumption model, basic energy consumption model and the model of Goeke and Schneider

[30] produce an average relative error of 70% and 4%, respectively.

$$F_{ij}(u_i) = (g \sin \alpha + c_r g \cos \alpha + a(t_{ij})) (m + u_i) + 0.5c_d \rho A v (t_{ij})^2 \quad (5)$$

$$F_{ij}^M(u_i) = \sum_{h=1,2,3} \left( \frac{\zeta}{\kappa \psi} \right) \left( k N_e D_e + \frac{F_{ij}(u_i) v(t_{ij})}{P_{ij}^M(u_i)} / \mu_d \mu_{dt} \right) \bar{t}_h \quad (6)$$

$$E_{ij}^E(u_i) = \sum_{h=1,2,3} \frac{(F_{ij}(u_i) v(t_{ij}) / \eta_h) \bar{t}_h}{E_{ij}^E(u_i)} \eta_h \quad (7)$$

$$= \begin{cases} \eta_h^+ \leq 1, & E_{ij}^E(u_i) > 0 \text{ \& } 0 \leq P_{ij}^M(u_i) \leq 100 \text{ kW } (h = 1, 2) \\ \eta_h^- \geq 1, & E_{ij}^E(u_i) < 0 \text{ \& } -100 \leq P_{ij}^M(u_i) \leq 0 \text{ kW } (h = 3) \end{cases}$$

Basso et al. [92] developed an energy consumption model for two-stage E-VRP (2sEVRP) that includes detailed topography and speed profiles. The similar LDM model of Asamer et al. [107] is applied for computing the energy consumption on a road segment (link), but with time-dependent speed, varying mass, constant slope, and link distance. For each link, similar to Macrina et al. [99], the three or two characteristic phases of speed curve can be distinguished. In such a way, acceleration and braking processes before and after the intersection are taken into account. As mass changes during the delivery, the energy consumption is expressed as a linear function of mass. Then, the shortest energy paths between all vertices pairs (customers, depot, CSs) without charging constraint are determined by applying the Bellman-Ford algorithm [112]. A nominal mass value between the mass of full and empty vehicle is used in the computation. The experiments showed an average energy estimation error of 2.28% and improved energy feasibility compared to some of the previous consumption models. The authors also reported that the use of average speed consumed at least 24% less energy.

Asamer et al. [107] analyzed the data collected from the BEV trips in order to determine the dependency between BEV energy share and average trip speed. Results showed that on lower speeds  $v \leq 30$  km/h, most of the energy is spent on the acceleration and auxiliary devices, while on the higher speeds  $v \geq 80$  km/h, approx. 70% of total energy is spent on overcoming the air drag force. The slope of the terrain in total energy consumption has the lowest share, and its value is even decreasing with the increase of average trip speed. The rolling resistance energy consumption share does not depend on the average trip speed.

Preis et al. [35] analyzed the energy optimal routing of BEV, where the routes were designed in different terrain slopes and battery capacity scenarios. The authors concluded that energy savings grow linearly with the maximum altitude difference and that decreasing the battery capacity of BEVs does not affect the recharging schedule but increases the total energy consumption. Fiori et al. [113] compared the power-based consumption model of BEV and ICEV [114]. The authors concluded that BEVs and ICEVs have different fuel/energy-optimized assignment. The faster routes increase

the BEV's energy consumption while the congested and low-speed arterial routes consume less energy. Masmoudi et al. [84] compared the realistic energy consumption model of Genikomsakis and Mitrentsis [85] to the constant consumption model (237.5 W/km) on instances for the dial-a-ride problem with EVs. The authors concluded that using the realistic model is more efficient with 0.14% difference between realistic and constant energy consumption from the best-known solutions (BKS). To emphasize the importance of the energy consumption model and energy minimization, Zhang et al. [36] compared the distance and energy minimization and concluded that the distance-minimizing objective consumes 16.44% more energy than the energy-minimizing objective.

In real-life conditions, speeds on the roads are time-dependent and can be described as the speed profile over the observed time period. Speed profile depends on the road type, driver behavior, traffic (accidents, recurrent congestions), weather conditions, etc. [115, 116]. A large number of parameters make it hard to predict a BEV's energy consumption in different traffic scenarios. Therefore, some researchers apply data-driven approaches to predict the energy consumption of a BEV. De Cauwer et al. [117] developed a model for BEV energy consumption by applying multiple linear regression to real-world measured BEV data. The proposed multiple linear regression model (MLR) has seven features: distance, speed, energy consumed by auxiliary devices, positive elevation, negative elevation, temperature, and kinetic energy change per unit distance. Results showed that the prediction error of trip energy consumption is within 25%. De Cauwer et al. [115] applied a similar MLR model for predicting energy consumption on road segments, suited for BEV routing. The authors used a neural network based on the road-traffic and weather-related features to predict the speed profile of a road segment. The error prediction of the trip's energy consumption is 12-14%. Lebeau et al. [52] used real-life data to model the energy consumed and recuperated on the trip through least square analysis. The used function depends on the trip duration, measured energy consumption, temperature, and correction parameter. The  $R^2$  for the energy consumption model is 0.93 and 0.77 for the recuperated energy model. Wu et al. [118] presented an empirical and analytical method that can estimate a BEV's instantaneous power in real time and overall trip energy consumption, with the average prediction error up to 15.6%. Fiori and Marzano [119] proposed a backward microscopic power-based LDM energy consumption model based on the known driving cycle. The acceleration, speed, and regenerative braking efficiency were modeled as functions of time. Based on the real driving BEV data, for each vehicle, six parameters were optimized: three parameters regarding the drivetrain and battery efficiency, the parameter for determining when the regenerative braking is occurring, and two traction power thresholds. The authors concluded that the proposed model is flexible enough to be applied to any kind of driving cycle and BEV type with given kinematic profile and characteristics.

Pelletier et al. [104] presented battery degradation models that could be used in BEV routing applications, as batteries are a crucial part of the economic BEV routing. Several

lithium-ion battery degradation mechanisms with storage and operating conditions that affect battery lifespan were described. Barco et al. [42] applied a battery degradation model based on the temperature, SoC, and depth of discharge in their airport shuttle service with BEVs. It was observed that the consideration of the battery degradation model affects the charging patterns.

## 4. Variants of the Electric Vehicle Routing Problem

Many different VRP variants were researched over the years. With BEV appearance, researchers started to adapt them to the E-VRP context. Due to the specific characteristics of BEV routing, some new problem-specific variants emerged. Here, we present some of the most relevant E-VRP variants and related problems.

*4.1. Energy Shortest Path Problem and the Electric Traveling Salesman Problem.* The energy shortest path problem (E-SPP) and electric TSP (E-TSP) can be considered as two of the simplest forms of the E-VRP. In most of the VRP variants, graph arcs have positive weight values that can represent distance, travel time, cost, etc. To compute the shortest path between customers, the most commonly applied algorithms are Dijkstra, Bellman-Ford,  $A^*$ , contraction hierarchies, etc. [108, 120]. By taking into account the recuperated energy of BEV, some arc weights could have a negative value, which makes most of the shortest path algorithms inapplicable. To overcome this problem, Artmeier et al. [108] formalized energy-efficient routing with rechargeable batteries as a special case of the constrained SPP (CSPP) in which the battery charge is limited with its capacity, and there is no recharging at CS. The authors adapted the Bellman-Ford shortest path algorithm with time complexity  $\mathcal{O}(n^3)$  in order to solve the energy CSPP. To overcome negative graph edges, Eisner et al. [121] used Johnson's shifting technique [122] to transform negative edge cost functions into nonnegative ones and applied Dijkstra's algorithm with time complexity of  $\mathcal{O}(n \log(n) + m)$  [123]. Storandt [124] introduced CSPP for EVs with battery swapping, while Sweda and Klabjan [125] added recharging events at CSs with a differentiable charging function. Zündorf [50] solved the CSPP for BEV routing with battery constraints, different types of CSs, and nonlinear charging process. The author developed a charging function propagating algorithm in order to minimize travel time and applied CH and  $A^*$  algorithms to solve the problem. Liao et al. [103] considered the BEV's shortest travel path problem and presented a dynamic programming algorithm for solving the problem that runs in  $\mathcal{O}(kn^2)$ , where  $k$  is the upper bound on the number of BSSs. More recently, Strehler et al. [126] observed recharging events in the graph vertices and arcs for the energy-efficient shortest routes of BEVs and HEVs, while Zhang et al. [127] dealt with the electric vehicle route planning with recharging problem (EVRP), which minimizes the overall travel and charging time and takes into account CSs' locations, partial recharging, nonlinear charging functions, service time duration, and service frequency at CS.

As VRP is the generalization of the TSP, the E-VRP is closely related to the E-TSP, in which a set of customers has to be served by only one BEV. Roberti and Wen [64] formulated the E-TSP with time windows (E-TSPTW) as a compact MILP program with binary variables for recharging paths with intermediate stops. The authors observed full and partial recharge policies and applied three-phase heuristic for solving the problem. The authors solved multiple instances, among them the 13 E-VRPTW instances of Schneider et al. [25], which were solved optimally with only one vehicle. Doppstadt et al. [58] formulated the HEV-TSP with four working modes of HEV and provided new test instances. Liao et al. [103] introduced the EV touring problem, as the generalization of TSP, with the minimization of the total required time. Two scenarios with battery swap scheme were observed: the on-site station model, in which each city has a BSS; the off-site station model, in which BSS is located at an acceptable distance from the city. The authors proposed efficient polynomial time algorithms for the problem.

*4.2. Heterogeneous or Mixed Vehicle Fleet.* In today's vehicle fleets, mostly ICEVs are present. Transition to an almost wholly electric fleet is a very challenging economic task. Therefore, most companies are gradually integrating BEVs into their existing ICEV fleet. Routing algorithms have to be upgraded and adapted for electric fleet characteristics, as on-time priority is harder to achieve.

Fleet size and mix VRP (FSM-VRP) was first introduced by Golden et al. [128], where the authors considered routing a fleet of vehicles with different acquisition costs and the routing cost is dependent on the vehicle type. Goeke and Schneider [30] formulated E-VRPTW and mixed fleet (E-VRPTWMF) with equal vehicle load capacities of BEVs and ICEVs, and equal BEV battery capacities. The authors compared the percentage share of BEVs to the total traveled distance obtained with three objective functions: total traveled distance, total costs with battery costs, and total costs without battery costs. The traveled distance on one battery pack was assumed to be 241,350 km (150,000 miles) with the payment of an additional \$600 per kWh in case of replacement. The results showed that the BEV's share in total traveled distance increased significantly when battery costs are not considered, while in the other two scenarios ICEVs performed most of the deliveries. Hiermann et al. [31] analyzed a similar problem, the electric fleet size and mix VRPTW and recharging stations (E-FSMFTW), with different vehicle load and battery capacities. The authors showed the overall positive influence of the heterogeneous vehicle fleet on the generalized total cost function. In most of the instances, three to four vehicle types are used in the final solution. A similar analysis on small-sized instances with different BEVs, ICEVs, and HEVs was undertaken by Lebeau et al. [52]. The authors defined seven groups of vehicle types that could be used for the delivery, from small vans and quadricycles through diesel-only or electric-only groups to a group of all vehicle types. The results showed the following aspects of routing: (i) the fleet with different vehicle types reduced the total routing costs; (ii) in the large van group, ICEVs outperformed BEVs; and (iii) HEVs showed a great

application in deliveries solely made by trucks. Sassi et al. [47–49] also observed a heterogeneous fleet of ICEVs and BEVs with different load and battery capacities, different operating costs, time-dependent charges, and the compatibility of BEVs and chargers at CSs. The authors focused on the procedures for solving the problem, so they did not offer any comparison between the ICEV and BEV solutions. Hiermann et al. [96] introduced the Hybrid heterogeneous electric fleet routing problem with time windows and recharging stations ( $H^2E$ -FTW), in which the fleet consists of ICEVs, BEVs, and PHEVs. The authors compared the overall costs of the solutions obtained with a homogeneous fleet of ICEVs, BEVs, and PHEVs to the optimized solution with a mixed fleet. The gap for the ICEV fleet is the largest, with an average value of 60% in the extreme case when the fuel cost is the highest. In contrast, when the electric cost is the highest, the BEV fleet on average produces a 40% gap, while the PHEV fleet shows the lowest gap value, up to 25%. The authors pointed out that, in mixed solutions, BEVs are preferred for clustered instances, ICEVs for randomly distributed instances, and PHEVs for randomly clustered and distributed instances. Overall, the results showed that operational costs can be 7% lower in a mixed case compared to the homogeneous case.

*4.3. Hybrid Vehicles.* As compensation for the limited driving range of BEVs, HEVs, which have both an internal combustion engine and an electric engine, have been developed. Two main types of HEVs are present on the market: the *series hybrid*, in which only the electric motor drives the train and the internal combustion engine is used to recharge the battery pack, and the *parallel hybrid*, which uses both internal combustion engine and electric engine to drive the train, where the electric engine is more efficient in stop-and-go activities and the internal combustion engine is more efficient at high speeds. We focus on the PHEV as a version of a *parallel hybrid* in which batteries can be recharged by connecting a plug to the electric power source. The PHEVs have an option to decide during the route to run on either electric energy or fossil oil. This enables the visiting of customers far from the depot with almost no refuel/recharge during the route. As PHEVs have two engines, their load is heavier compared to the BEVs and ICEVs, and therefore they have a higher energy consumption rate. The time spent on the deliveries is shorter, which makes it easier to achieve time-precise deliveries and to reduce the costs of recharging, at the expense of higher costs due to fossil oil consumption.

One of the first papers that dealt with the PHEVs routing problem was published by Abdallah [41], where the author defined the problem as the plug-in hybrid electric VRPTW (PHEVRPTW). The objective of the proposed problem is to minimize the routing costs on the internal combustion engine while satisfying the demand and time window constraints. At each customer, the driver can either recharge the vehicle battery or go to the next open time window using an internal combustion engine when the electric energy has been depleted. Doppstadt et al. [58] defined the HEV-TSP in which the authors observed HEVs that are not plug-in and can only be charged while driving. Four working modes were observed, combustion-only mode, electric-only



mode, charging mode, and boost mode, when the combined internal combustion engine and electric engine are used. As the authors assumed that there was not a charge left from the day before, the initial charging of the battery was set to zero. The authors presented the positive effects of using HEVs: (i) compared to the ICEV routes, the overall costs for the HEV routes in the tested instances reduced up to 13% and driving time increased up to 11%; (ii) savings depend highly on the depot location, and not on the limited battery capacity, which was an a priori assumption. Mancini [71] introduced hybrid VRP (HVRP), in which PHEVs can change propulsion mode at any time and the electric engine, rather than the internal combustion engine, is promoted. Vincent et al. [76] introduced a similar HVRP problem with PHEVs and mild-HEVs, which do not have an electric-only mode of propulsion. The results showed that PHEVs have a lower average cost per mile in all scenarios compared to the mild-HEVs and ICEVs. Hiermann et al. [96], in their  $H^2E$ -FTW, used the electric motor of the PHEV as priority mode choice. When a time window of a PHEV route is violated, the mode could be changed at any time during the route by exchanging recharging time for the corresponding amount of fuel at no additional costs. The authors showed the positive impact of a PHEV fleet compared to ICEV and BEV fleets, especially for randomly clustered instances. PHEVs are usually represented with only 20% of the overall number of vehicles used in the solution due to their higher consumption and utility costs but, still, they constitute an important part of the fleet configuration due to their flexibility.

**4.4. Partial Recharging.** In the beginning, in most of the E-VRP problems, full recharge was considered when a BEV visited a CS [25]. This can be time-consuming because, depending on the SoC level, available charging technology, and battery capacity, the vehicle can charge from five minutes to eight hours [15]. Therefore, in real-life applications, partial recharging should be taken into account. The battery should be charged enough to complete the whole route or to surpass a fear that vehicle range will not be enough to perform designated tasks, the so-called range anxiety [75]. This particularly has an effect on customers with narrow time windows, where efficient charge scheduling can enable the feasibility of the route. On the economic side, significant savings can be achieved by applying partial recharging as a minimal amount of energy could be recharged during the day when electricity cost and energy network load are higher, and the rest of the energy could be replenished during the night [46, 47]. In some cases, it is natural to maintain the energy reserve. This can be done by SoC range limitation, i.e., [20, 95] [42, 47, 80]. Having an energy reserve seems even more important if energy consumption and range anxiety are taken into account because up to 30% of the consumed energy can be spent on BEV's auxiliary devices [109]. Limiting SoC value also helps to preserve the battery as battery capacity decreases by overcharging and overdischarging [104].

Several papers have analyzed strategies of partial recharging and formulated the problem as E-VRPTW with partial recharging (E-VRPTWPR). Bruglieri et al. [29, 51] and Moghaddam [53] modeled the concept of partial charging in

E-VRP. Felipe et al. [46], in GRVP with multiple technologies and partial recharges (GVRP-MTPR), and Keskin and Çatay [28], in E-VRPTWPR, ensured that after every change in the route configuration, at the previous CS, the BEV is charged only with the amount of energy sufficient to finish the segment of the route until the next CS or the depot. This results in the removal of certain CSs in the route, compared to full recharge strategy, and the arrival of a vehicle at the depot with an empty battery. The authors presented the positive impact of partial recharges on the total costs and energy savings. A similar procedure was applied by Sassi et al. [47–49] but with multiple charging technologies, different charging periods, and BEV chargers compatibility checks. Schiffer and Walther [73] compared full and partial recharge by solving small E-VRPTWPR test instances using commercial software. The authors concluded that, in some cases, a partial recharging strategy reduces the total traveled distance and number of visits to CSs. Montoya [18] and Montoya et al. [72] formulated the FRVCP for BEVs in which, for a fixed sequence of customers in the route, CS position and charging amount are optimized. The results on the newly derived test instances showed that good solutions tend to exploit partial recharges. A similar procedure was applied by Keskin and Çatay [79], Hiermann et al. [96], Froger et al. [68], and Schiffer and Walther [88, 89] to enhance the incumbent best solution by optimizing the charging decisions along the BEV route with a fixed customer sequence. Desaulniers et al. [57] presented the effects of partial recharging by optimally solving E-VRPTW instances containing up to 100 customers. In a case with a single recharge per route, the partial recharge reduced the routing costs by 0.97% and the number of vehicles by 2.25%, while in a case with multiple recharges per route these values are 1.91% and 3.80%, respectively, with a significant increase in the average number of recharges per route.

**4.5. Different Charging Technologies.** Today, multiple charging technologies are present: (i) slow, 3 kW (6–8 h); (ii) fast, 7–43 kW (1–2 h); and (iii) rapid, 50–250 kW (5–30 min) [15, 129]. To better control charging time in the E-VRP context, the selection of possible charging technology could also be optimized. This could make some customers who have narrow time windows more accessible by fast charging at previous CSs, or if the time windows are long, an economically better approach could be slow charging. Such a problem could be extended by taking into account CS working hours, time-dependent charging costs, the number of available chargers and their compatibility with BEVs, the power grid load, the charger power, etc. [46–49, 68, 80, 81, 87]. Felipe et al. [46] analyzed the effect of different charging technologies on the recharge cost. The authors concluded that none of the technologies dominated the others. The rapid and fast charging options provided slightly better results than the slow charging, but the best results were obtained when joint technologies were used as the most appropriate one could be chosen in each case. Çatay and Keskin [67] solved small-sized instances to present the insights of the quick charge option. The results showed that quick charging might reduce the fleet size and decrease the cost of energy needed to

operate BEVs. Keskin and Çatay [79] formulated E-VRPTW and fast charging (E-VRPTW-FC) with partial recharges and three different charging technologies. The authors provided MILP formulation and minimized total recharging costs while operating a minimum number of vehicles. The authors compared two mixed integer programming (MIP) models: (i) model 1, in which binary variables are used as decision variables to choose which charging technology to use; (ii) model 2, based on the model of Schneider et al. [25] and Keskin and Çatay [28], in which the authors copied each station vertex three times depending on the charging technology. The results showed that in all cases model 1 produced better results with shorter computation time. Testing on larger instances showed that the fast charging option is more beneficial when customers have narrow time windows and that multiple charging technologies improved the overall results (in 28 of 29 instances), while in the instances with larger time windows the average improvement was 0.17%.

*4.6. Nonlinear Charging Function.* In most of the E-VRP related literature, either linear or constant charging time is considered. Most of the BEVs have lithium-ion batteries installed, which are often charged in constant-current constant-voltage (CC-CV) phases: first by constant current until approx. 80% of the SoC value and then by a constant voltage. In the CC phase, SoC increases linearly, and in the CV phase, the current drops exponentially and SoC increases nonlinearly in time, which prolongs charging time [102, 104]. In only a few reviewed papers the author considered nonlinear charging process and solved the E-VRP either by linearization per segments or by estimating charging time by a data-driven approach [50, 68, 72, 87, 95]. Zündorf [50] developed a charging function propagation algorithm for determining an EV battery-constrained route by taking into account piecewise linear and concave functions of the recharging process. Montoya et al. [72] formulated the E-VRP with nonlinear charging functions (E-VRP-NL) and presented the MILP formulation. The authors fitted piecewise linear functions for 11, 22, and 44 kW CSs to the real measured data. The results showed that neglecting the nonlinear charge can lead to infeasible or overly expensive solutions: 12% of the routes in good solutions recharged the battery in the nonlinear part, after 80% of the SoC value. Froger et al. [95] proposed two new formulations for the E-VRP-NL: (i) the improved MILP version of Montoya et al. [72] by arc-based tracking of the SoC value; (ii) a path-based model without CS replication, where a sequence of vertices (a path) between a pair of customers or depots is determined. The second one yielded much better results with shorter computing time. Froger et al. [68] explored similar formulations for the E-VRP-NL with capacitated CSs (E-VRP-NL-C): (i) the CS replication-based formulation with flow and event-based formulations for CS capacity constraints; (ii) recharging path-based formulation.

*4.7. BEV Routing and CS Location.* Due to the currently low BEV market share, the number of CSs installed in the road infrastructure is also relatively low. Therefore, great potential lies in the simultaneous decision-making of CS locations

and BEV routes. The classic location routing problem (LRP) consists of determining the locations of the depots and vehicle routes supplying customers from these depots [130]. A modification of the LRP that deals with CS facilities is formulated as electric LRP (E-LRP) [11, 27, 73, 88].

Schiffer et al. [11, 27] presented a case study for solving ELRPTWPR. It was assumed that CSs can be located at customers' locations and that service time can be used for recharging. Overall, results indicated the viability of combining CSs siting and BEV routing for specific cases when a delivery range is not far from the depot. Schiffer and Walther [73] compared ELRPTWPR and E-VRPTWPR solutions on down-scaled test instances and concluded that the ELRPTWPR gives a better solution in all instances. The reason comes from the fact that BEV can be charged while serving, which reduces the overall route time, as there is no travel time wasted on traveling to and from CS. Hof et al. [69] and Yang and Sun [56] addressed the problem of E-LRP but with BSSs. Results indicated that decreasing the construction cost of BSS led to the expected increase in their number in the solution. Schiffer and Walther [88] formulated the LRP with intraroute facilities (LRPIF) as a more general problem, in which intraroute facilities can be used for refueling, loading, or unloading goods. Sun et al. [131] proposed a location model for CSs without routing based on the travel demands of urban residents. For short-distance travelers, slow CSs are utilized, while for long-distance travelers, the rapid CSs are considered. The CSs' locations are determined to maximize coverage and flow according to the concept of vehicle refueling. In the case study, the authors pointed out two critical aspects: the BEV driving range and the budget constraint.

*4.8. Battery Swap.* Instead of charging at CS, at specially designed BSS, empty or nearly empty batteries can be replaced with fully charged ones [14]. The main advantages of such procedure are the time in which it can be performed and the ability to recharge when energy network load and electricity costs are lower, i.e., during the night. A whole replacement procedure could last less than ten minutes, which is competitive to the refueling time of ICEVs and much faster than one of the fastest charging BEV technologies. The drawbacks of such procedure are the nonstandardized batteries and their installation in BEVs, which makes it hard to swap empty batteries with fully charged ones. Adler and Mirchandani [44] observed the E-VRP with swappable batteries, already determined BSS's locations, a fixed number of batteries per BSS, full recharge of four hours, and a fixed swapping time of two minutes. Full battery recharge is considered so it is possible that when the vehicle arrives at the station, there is no fully charged battery available and the vehicle has to wait. First, the total routing cost is minimized and then the battery reservations are made, so that the vehicle could avoid BSSs without available batteries. Yang and Sun [56] and Hof et al. [69] simultaneously determined the locations of BSSs and the vehicle routing plan and provided different metaheuristics for solving the problem. Yang and Sun [56] formulated the problem as BSS-EV-LRP and compared the solutions of the BSS-EV-LRP instances

to the corresponding best-known CVRP solutions. The total routing costs for 150 km and 300 km driving ranges increased to 7% and 3.5%, respectively, and in some cases, there was no gap between BSS-EV-LRP and CVRP solutions. Hof et al. [69] on the same problem reported that if construction costs of BSSs are equal to zero, the BSSs would be constructed in 83% of total available candidate sites.

**4.9. Two-Echelon Routing Problem.** Breunig et al. [93] proposed the electric two-echelon VRP (E2EVRP), in which goods are transported in two echelons: (i) in the first echelon, goods are transported by conventional freight vehicles from the depot to the satellite facilities; (ii) in the second echelon, goods are transported from the satellite facilities to the customers by light BEVs. Two vehicle types are observed in the problem: ICEVs with higher load capacity located at the depot and BEVs with lower load capacity located at satellite facilities. BEVs are used for the last-mile deliveries due to their lower pollution, noise, and size. The authors formulated the problem on the multigraph and applied exact procedures on smaller instances to solve the problem optimally and a metaheuristic procedure on larger newly developed instances. The results showed that the increase in CS density  $\rho$  decreased the detour costs following the expression  $1/\rho^{1.24}$ . Also, the vehicle range has a great impact on the solution quality: a range below 70 km produced infeasible solutions while a range higher than 150 km decreased recharging visits to almost zero. Jie et al. [97] analyzed a similar problem, the two-echelon capacitated E-VRP with BSS (2E-EVRP-BSS), in which, in both echelons, BEVs are used. The BEVs in the first echelon have higher load and battery capacities than the BEVs in the second echelon. The authors reported that the battery driving range is the most important aspect of routing and that it slightly depends on the number of BSSs used.

**4.10. Charging Schedule.** Many companies that use BEVs prefer charging the vehicles at their own facilities in order to charge the vehicles between the delivery routes and during specific periods of the day. In such occasions, there are usually a limited number of chargers at the depot, typically fewer than the fleet size; therefore, the efficient charging schedule at the depot has to be determined.

Pelletier et al. [87] considered the electric freight vehicles charge scheduling problem (EFV-CSP) at the depot for the BEVs that deliver goods to a set of customers over multiple days. The authors included multiple charging technologies at the depot, realistic charging process (piecewise linearization of nonlinear charging function), time-dependent charging costs, grid power restriction, battery degradation costs (cyclic and calendar aging), and facility-related demand (FRD) charges representing the maximal demand registered over the billing period. The fixed vehicle routes are known in advance and the number of charging events in the depot is limited to avoid impractical solutions of constantly moving a vehicle from one charger to another. The authors presented the following aspects of charge scheduling tests conducted in summer (higher electricity costs) and winter (lower electricity costs) periods: (i) model tries to keep the SoC lower when battery degradation costs are included; (ii) in

summer, vehicles are rarely charged in peak hours, which results in more vehicle charging simultaneously or the use of fast chargers that retrieve more power from the grid in nonpeak hours but incur higher FRD charges; (iii) to avoid cycling the battery in high SoC, it is preferable to split the long routes into smaller ones; (iv) fast chargers are heavily used in a high BEV utilization context; (v) grid power restriction increases overall energy costs, especially in summer months, and leads to infeasible solutions, limiting the number of vehicles simultaneously charging; and (vi) total costs are always lower with larger batteries, as smaller batteries require larger discharge cycles.

Barco et al. [34, 42] also analyzed charge scheduling for assigned routes in an airport shuttle scenario. The authors defined the set of charging actions (charge profile) over the determined programming horizon and minimized overall operating costs. Sassi et al. [47, 49] also dealt with determining the charging schedule of BEVs at the depot and included time-dependent charging costs and chargers compatibility checks with BEVs. Adler and Mirchandani [44] provided an online routing model of BEVs by taking into account battery reservations to minimize the average delay of all vehicles by occasionally detouring them. Wen et al. [65] observed the service time of CSs, meaning that a CS can be visited only in some specific time period, usually working hours. The authors proposed a battery reservation model in order to charge the reserved battery at the defined time period. Sweda et al. [75] dealt with the possibility that a CS might not be available at some point in time and rerouting should be performed. The problem is formulated as the adaptive routing and recharging policies for EVs. First, optimal a priori routing and recharging policies were determined and then heuristic procedures were applied for the adaptive routing.

**4.11. Dynamic Traffic Conditions.** Most of the E-VRP research considers static conditions on the road network. The traffic states change recurrently, depending on the time of the day, day of the week, and season, or nonrecurrently when a traffic incident occurs, such as an accident [132–134]. TD-VRP routes a fleet of vehicles by taking into account variable travel time on the road network [135, 136]. Shao et al. [74] observed BEVs in such time-dependent context in their E-VRP with charging time and variable travel time (EVRP-CTVTT). The recharging time is fixed to 30 minutes and batteries are always charged to full capacity. The authors discretized one day into two-minute intervals and applied the dynamic Dijkstra algorithm to find the shortest travel time path when weights in the graph are not constant. The authors presented a real-life problem and solved it by applying a genetic algorithm with a running time of three hours. Omidvar and Tavakkoli-Moghaddam [24] presented a model for routing AFVs that aims to avoid routing during congestion hours, when the pollution costs are high, by waiting at a customer's locations. These costs were computed based on the road speed profile. Mirmohammadi et al. [62] presented MILP formulation for the periodic green heterogeneous VRP with time-dependent urban traffic and time windows, which is green in terms of the emission minimization and not the utilization of AFVs. The planning horizon is divided into



several periods, during which static traffic conditions are considered.

#### 4.12. Related Problems

**4.12.1. Route Scheduling and Other Electric Variants.** Many researchers are dealing with the scheduling of bus/taxi routes that are fixed or could be slightly altered. Li et al. [82] addressed the mixed bus fleet management problem (MBFM) composed of electric, diesel, compressed natural gas and hybrid-diesel buses. The authors maximized the total benefit of replacement of old vehicles with new ones under budget constraints while optimizing the route assignment for each bus during the planning period. Two routing procedures were developed to solve the recharging problem: single-period routing and routing across multiple periods of a day. Analysis of the results on the case study of Hong-Kong showed a cost-effective scheme of fleet configuration in the following order: mixed fleet, electric, natural gas, diesel, and then hybrid-diesel buses. Wen et al. [65] also dealt with the scheduling of electric buses (electric vehicle scheduling problem, E-VSP) in order to efficiently serve a set of timetabled bus trips. The linear partial and full recharge were considered, with the minimization of bus numbers and the total traveled distance. Lu et al. [83] addressed the problem of optimal scheduling of a taxi fleet with mixed BEVs and ICEVs to service advance reservations. The authors presented a multilayer taxi-flow time-space network in order to minimize the total operating fleet cost.

Bruglieri et al. [77] formulated the MILP for a one-way electric car-sharing problem as the electric vehicle relocation problem (E-VReP). Relocation of BEVs is performed by workers who come with bicycles to the pickup points, put the bicycle into the BEV's trunk, and drive the BEV to one of the delivery points. The authors balance a trade-off among the customers' satisfaction, the workers' workload balance, and the car-sharing provider's objective.

Masmoudi et al. [84] formulated the dial-a-ride problem with EVs and BSSs (DARP-EV) for customers with special needs and disabilities. The authors observed a fleet of heterogeneous BEVs with equal battery capacity but different available resources for the customers: handicapped person's seat, stretcher, wheelchair, or accompanying person's seat. The authors adapted the energy consumption model of Genikomsakis and Mitrentsis [85] with constant speed, acceleration, road slope, and varying mass while loading/unloading passengers.

Paz et al. [86] addressed the multidepot electric vehicle location routing problem with time windows (MDEVLRPTW) with three variants: (i) with BSS (MDEVLRPTW-B); (ii) with partial recharging (MDEVLRPTW-PR) at CSs that could also be located at customers' locations; and (iii) with BSS and partial recharging (MDEVLRPTW-BSPR) where conventional charging is considered at customers' locations and battery swap is performed at a separately located BSSs. Three MIP models were presented for each case. The authors determined the number and location of BSSs and/or CSs (without investment costs), the number and location of the depots, the number of vehicles, and the

customers' sequence in route. The proposed model allows minimizing the number of vehicles given a fixed number of CSs or BSSs, and vice versa. For the CS, linear recharge is considered, while the battery swap time is set to 10% of the corresponding linear recharge time. Results were presented by solving small instances with a commercial solver. The results showed that BSPR and PR variants provide better results than the BS variant and that BSPR tends to act as PR in most of the instances.

Schiffer and Walther [89] extended the ELRPTWPR model of Schiffer and Walther [73] with uncertain customer patterns related to customers spatial distribution, demands, and service time windows. The authors formulated robust ELRPTWPR (RELRPTWPR) as a MILP program. Five scenarios were observed, which represent customer profile for five working days. The authors used set-based uncertainty representation, which is covered within scenarios that consider customer patterns in daily deliveries. To achieve a computationally tractable model of robust counterpart formulation, the adversarial approach was used, which uses a finite set of scenarios. The robust model was compared to the deterministic model (selected customer pattern), median model (average customer pattern), and worst-case model (worst-case customer pattern), where first the E-LRP was solved, and then the E-VRP. The deterministic models produced infeasible solutions in 35-63% of instances and incurred, on average, 5% higher costs than the robust approach. Also, the robust approach showed the most homogeneous configuration of CSs.

Keskin et al. [80] extended E-VRPTW by considering waiting times at the CSs and proposed a solution method for solving small-sized instances. The authors assume single charger at CS, Poisson arrivals, exponential distribution of service times, first-in-first-out strategy, and penalties for late arrivals. The planning horizon is split into morning, afternoon, evening, and night period. The routing decisions are determined based on time-dependent waiting times at CSs.

Kullman et al. [81] introduced E-VRP with public-private recharge strategy (E-VRP-PP), where demand at public CSs is unknown. The queuing at CS is modeled as  $M/M/c$  with first-in-first-out strategy, and  $c$  represents the number of identical chargers at CS. The problem is modeled as a Markov decision process, and an approximate dynamic programming solution is proposed. The authors extended the FRVCP formulation of Montoya et al. [72] with time-dependent waiting times and discrete charging decisions. Dynamic and static routing policies are proposed, where the static policies are incorporated as base policies for dynamic routing. Using dual bound on the optimal policy, the authors concluded that proposed routing policies are within 4.7%. The authors pointed out that the use of a public-private strategy outperformed the use of just private CSs, as overall savings were 20% higher.

**4.12.2. Green and Pollution Routing Problem.** The green vehicle routing problem focuses on the reduction of routing pollution on the environment. The key idea is to promote the use of sustainable energy sources and minimize overall emissions.

Regarding the emission of BEV, Álvarez Fernández [137] indicated that emission savings of BEVs can be significant, up to 80% in some scenarios compared to ICEVs. The authors presented a model for prediction of a BEV's GHG emissions linked to a route for each country, according to empirical measurements and prediction of routes' energy consumption.

Erdoğan and Miller-Hooks [23] first defined the GVRP with AFVs and full refuel, where also BEVs were included. Therefore, the E-VRP can be observed as a special case of the GVRP. The authors provided MILP formulation of the problem with replication of CSs vertices. Koç and Karaoglan [59] improved the MILP formulation of Erdoğan and Miller-Hooks [23] by removing dummy vertices that represent multiple visits to refueling stations. Instead, a binary variable related to traveling between two vertices by using the station in between is introduced (arc-duplicating). This formulation improved the average gap to the optimal integer solution on all scenarios compared to the model of Erdoğan and Miller-Hooks [23] by 25%. The authors proposed a combination of exact branch-and-cut procedure and simulated annealing with several moving strategies to solve the problem. Leggieri and Haouari [70] formulated the problem as a nonlinear MIP, which was linearized to derive MILP formulation by applying a reformulation-linearization technique. The authors used similar arc-duplicating to that of Koç and Karaoglan [59] but defined the binary variable as the sum of binary variables for the direct arc between the customers and binary variables for the same direct arc but with stations inserted. The proposed formulation without additional constraints was able to solve to optimality 97.5% of test instances, which substantially outperformed the formulations of Erdoğan and Miller-Hooks [23] and Koç and Karaoglan [59]. In addition, the same formulation with preprocessing consistently outperformed the exact branch-and-cut procedure of Koç and Karaoglan [59]. Andelmin and Bartolini [66] took a different approach for the exact solving of GVRP based on a set partitioning formulation in which columns correspond to feasible routes. The problem is modeled using a multigraph in which vertices correspond to customers and each arc represents a possible sequence of stations visited when traveling between customers. The authors added weak subset row inequalities, subset row inequalities, and  $k$ -path cuts, with the latter reported as the crucial part of the algorithm efficiency. The results showed tight lower bounds and high efficiency of the algorithm as instances with up to 110 customers were solved to optimality. Bruglieri et al. [94] presented a path-based approach for solving the GVRP, based on two phases. In the first phase, a set of feasible paths is generated, removing from it all those paths that are dominated. In the second phase, through MILP formulation, paths are combined to create routes for the final solution. The authors compared their results to previously observed research papers of Erdoğan and Miller-Hooks [23], Koç and Karaoglan [59], Leggieri and Haouari [70], and Andelmin and Bartolini [66]. On small-sized instances, the proposed algorithm produced optimal solutions with the lowest running time while on the larger instances it produced 2.3% worse solutions compared to Andelmin and Bartolini [66], but in much less time.

Poonthalir and Nadarajan [38] formulated biobjective fuel efficient GVRP (F-GVRP) with the minimization of route costs and fuel consumption using goal programming. The authors analyzed the impact of varying speed on fuel consumption where speed is modeled using triangular distribution. To efficiently solve the problem, the authors applied a population-based metaheuristic. The tests were performed by constraining the route cost of the solution to the BKS of GVRP and then the fuel minimization was performed. The authors pointed out that the use of varying speed instead of constant speed reduced the fuel consumption on average by 20.48%, and in half of the instances it reduced the number of vehicles. The lower number of vehicles is a consequence of the lesser amount of time spent at the refueling stations as less fuel is consumed, so the vehicle can visit more customers. The authors pointed out that if the route cost constraint was relaxed, most of the instances might have better route costs than the BKSs of GVRP. The authors also analyzed the impact of speed interval size on the fuel consumption and concluded that fuel consumption and average speed increased with larger interval size.

Normasari et al. [100] formulated capacitated GVRP (CGVRP) as a MILP program and solved the problem by applying a simulated annealing heuristic. Compared to the GVRP, the results indicated that the number of tours and total costs for the CGVRP increased by 38.8% and 32.5%, respectively.

Macrina et al. [33] investigated the green mixed fleet VRP with partial recharges and time windows (GMFVRP-PRTW), in which polluting emissions are modeled through the functions of both traveled distance and vehicle load. The problem was formulated as a MILP and solved by a commercial solver for small instances and by an iterative local search metaheuristic for large instances. The overall pollution emissions are maintained below the designated threshold value, to simulate emission limits of freight vehicles. Macrina et al. [99] presented MILP formulation for a similar GMFVRP-PRTW problem but added more realistic consumption model presented in Section 3.2, where the authors considered different charging technologies (single technology per CS), cost of energy recharged at the depot, and the fuel cost. A large neighborhood search metaheuristic is applied to solve the problem.

Koyuncu and Yavuz [98] compared the node- and arc-duplicating MILP formulation of mixed fleet GVRP (MGVRP) by taking into account the following: (i) a mixed fleet of AFVs and ICEVs; (ii) load and time window constraints; (iii) internal station visits (at customer) and external station visit (at separately located stations); (iv) fixed full, variable full, or variable partial refuel at the station; (v) site-dependent refueling rate; and (vi) the visit of two or more consecutive stations being prohibited. In the node-duplicating formulations [23, 25], each refueling vertex is replicated at most  $N$  times, where  $N$  is the number of customers. In the arc-duplicating formulation, two subsets of arcs are contained: arcs between customers (direct arcs) and refueling arcs which contain inserted station between customers [70]. In the preprocessing step for both formulations, unfeasible arcs are removed and tighter lower and upper

bounds for auxiliary variables and number of vehicles are presented. On the GVRP test instances of Erdoğan and Miller-Hooks [23], the authors reported that the node- and arc-duplicating formulations yielded an average optimality gap of 20.84% and 9.07%, respectively. On the GVRP instances of Yavuz [138], significant conclusions could not be drawn.

As BEVs have no local emissions, the E-VRP is closely related to the minimization of GHG emissions, where a problem-specific GVRP variant called the pollution routing problem (PRP) was introduced [32, 40]. Bektaş and Laporte [40] formulated the PRP as a nonlinear MIP where the main objective is the optimization of vehicle speed and GHG emissions. Demir et al. [32] concluded that speed optimization improves the PRP solution and minimizes fuel consumption and driver costs.

It can be noted that several papers are dealing with fuel consumption and pollution emissions and not AFVs, but the term *green* VRP is used to indicate a more ecologically aware routing problem [62, 139].

**4.12.3. More General Formulation of the E-VRP.** The E-VRP can be formulated more generally as the VRP with intermediate stops (VRPIS), in which the vehicle visits intermediate or intraroute facilities to replenish/unload the goods or to refuel [55]. The LRPIF extends the problem in terms of the facilities location optimization [88]. The service time depends on the load and fuel level at the arrival at the facility. Schneider et al. [55] presented the E-VRP with recharging facilities (EVRPRF) as a special case of VRPIS, where BEVs have a fixed recharging time and limited vehicle load capacity without time windows constraints.

**4.12.4. Multiobjective Variants.** Most of the researchers are using a single-objective function that represents travel distance, total costs, energy consumption, total time, etc. The multiobjective variants are still relatively scarcely applied. Generally, the solution of the multiobjective problem is not the optimal solution for all of the objectives, but rather it is satisfactory in those terms.

Demir et al. [37] introduced the biobjective PRP with the minimization of fuel consumption and total driving time. The fuel consumption is based on a model similar to the one presented by (6) but with constant speed. An adaptive large neighborhood search was used to generate nondominated/Pareto optimal solutions while an additional hybrid procedure was used to obtain the solution to the biobjective model. The experiments showed that there is no need to prolong driving time to achieve a significant reduction in fuel consumption.

The already mentioned F-GVRP of Poonthalir and Nadarajan [38] (Section 4.12.2) used biobjective minimization of route costs and fuel consumption. Wang et al. [91] considered multiobjective function to determine optimal BEV routes between origin-destination pairs. Three objectives are minimized: travel time, energy consumption, and charging cost. Travel time includes driving time, queuing time, and charging time, while charging cost includes electricity cost, service cost, and parking fee. The multiobjective model is transformed into a single objective by applying

a fuzzy programming approach for objectives' membership functions and fuzzy preference relations to obtain weighting coefficients of each objective. A genetic algorithm is designed to obtain an optimal solution. The influences of various driver trade-offs (weighting conditions) among different objectives are explored using Pareto curves. It was observed that driver trade-offs have an effect on the travel times, and almost no effect on the energy consumption and charging costs.

Amiri et al. [39] optimized the location of the battery swapping in the BSSs network and the charging schedule of depleted batteries. The multiobjective model minimizes three objectives: battery charging and power loss costs, deviation from the nominal voltages, and network capacity releasing. The nondominated sorting genetic algorithm is applied to solve the problem. The results on the test case showed that an optimal charging schedule in terms of cost and network constraints can be achieved. Besides, the electricity load profile is more leveled.

Bruglieri et al. [77] focused on the E-VReP with three objectives: minimization of workers used to relocate the BEVs (service provider objective), maximization of relocation requests served (users' objective), and minimization of the longest routes' duration (workload balance). The problem is solved by approximating the Pareto optimal front based on a two-phase approach: first, feasible solutions are generated by randomized search heuristics, and in the second phase, to find a set of nondominated solutions, the MILP formulation is solved by epsilon constraint programming.

**4.13. Summary of the E-VRP Variants.** Table 1 presents a summary of the E-VRP variants and related problems in the available literature. In total, 79 papers were presented in the table. The problem characteristics are observed in the following order: vehicle load (cargo) capacity (48), linear charging/refueling function at station (34), customer time windows (34), partial recharging strategy (30), fixed refueling/recharging time (19), different charging technologies at CSs (17), heterogeneous (mixed) fleet (16), energy consumption model (15), BSS (9), nonlinear charging function (8), dynamic traffic conditions (8), location routing (7), capacitated stations (6), and HEVs (5). We can point out that only a few papers are dealing with (i) simultaneous CS siting and BEV routing (E-LRP) where great savings can be achieved in companies whose business plans include CS investments, as E-LRP generally gives better routing configuration than the E-VRP [73, 88]; (ii) nonlinear charging function, negligence of which can lead to infeasible solutions and additional penalty costs [68, 72]; (iii) HEVs (PHEVs) that have an advantage in several routing scenarios but are being researched only recently due to the higher complexity of the problem as HEVs can change their propulsion mode at any time during the route [96]; (iv) dynamic traffic conditions, which significantly influence BEVs' energy consumption [74]; (v) different charging technologies at CSs, which could reduce the overall costs by selecting the best possible charger option in each occasion [46, 79]; (vi) CS related challenges: capacitated CS [68, 75, 80], CS reservations [44], and public-private recharge strategy [81]; and (vii) robust E-VRP variant



resistant to uncertainties in travel time, service time, and demand [89].

## 5. Problem Solving Methods

Since VRP is a well-researched problem, a large number of procedures for solving the problem have been proposed. Due to the NP-hardness of the problem and a large number of customers in real-life problems, most of the procedures used in real-life applications are heuristics, metaheuristics, and hybrid combinations. For small-sized problems, a great number of exact procedures have been proposed. Many of the VRP procedures in the available literature are with an adaptation applicable for solving the E-VRP. An overview of the applied procedures for solving the E-VRP and related problems is presented in Table 2.

**5.1. Feasibility.** When creating or modifying a VRP solution, two types of search procedures regarding the feasibility of the solution can be observed: allowing only feasible solutions or acceptance of infeasible solutions. The infeasible solution means that some customers are served without satisfying all of the problem constraints. Related, feasible procedures are searching in the space of feasible solutions, while infeasible procedures allow searching the space of infeasible solutions, which broadens the search. In the E-VRP, the feasibility mostly refers to vehicle load, customer time window, and battery (energy) constraints. In infeasible procedures, objective function is often defined with penalization coefficients, which are updated during the search process. At the beginning of the search, infeasible solutions are allowed in order to search a larger solution space. As the search process comes to an end, penalties for infeasible solutions increase. An example of such an objective function is given by (8) where  $f_{dist}(S)$  is the total traveled distance in the solution  $S$ ; penalties coefficients are  $\alpha$ ,  $\beta$ , and  $\gamma$ ;  $P_{div}(S)$  is a diversification penalty; and  $P_{cap}(S)$ ,  $P_{tw}(S)$ , and  $P_{batt}(S)$  are values of constraint violation, respectively, load capacity, time windows, and battery capacity [25]. An important part of such a function is the efficient computation of constraint violations [25, 31, 96, 140].

$$f(S) = f_{dist}(S) + \alpha P_{cap}(S) + \beta P_{tw}(S) + \gamma P_{batt}(S) + P_{div}(S) \quad (8)$$

Several researchers in the E-VRP field are allowing infeasible solutions during the search [11, 25, 30, 31, 69, 84, 88, 96] while some others are only allowing feasible solutions [23, 28, 46–49, 52, 68, 72, 79]. For example, in the E-VRPTWPR of Keskin and Çatay [28], the customer is inserted in the solution if it satisfies the load and partial time window constraints. If the solution is energy infeasible the CS is inserted to make the solution completely feasible. If the former one was not able to produce the energy feasible solution, the procedure returns to the latest feasible solution. Similarly, Montoya et al. [63] and Froger et al. [68] address the fuel infeasible route by solving a CSPP on the auxiliary graph. A list of papers allowing infeasible or feasible-only solutions is presented in Table 2 in column *NF*.

**5.2. Exact Procedures.** Exact procedures are able to find an optimal solution for a smaller number of customers: up to 360 customers for CVRP and up to 100 customers for VRPTW [5, 6]. Some of the most used exact algorithms are branch-bound-cut-and-price, dynamic programming (DP), MIP, set partitioning (SP), etc. [1]. In the E-VRP, many researchers formulate the problem as a MIP and solve the small instances with commercially available software. Koç and Karaoglan [59] applied the branch-and-cut algorithm coupled with simulated annealing and four improvement strategies to solve the GVRP. The authors reported that 22 out of 40 instances were solved to optimality. Hiermann et al. [31] solved the heterogeneous E-VRP by applying a branch-and-price algorithm on small instances up to 15 customers where the pricing problem was solved using a bidirectional labeling algorithm. The algorithm was able to find an optimal solution for all of the instances in a couple of minutes, with only three instances having computation time longer than five minutes. Desaulniers et al. [57] presented branch-price-and-cut algorithms for four E-VRPTW variants depending on the full or partial recharge and on the single or multiple recharges per route. On the 696 instances with 25, 50, and 100 customers, 98%, 90%, and 27%, respectively, were solved optimally. Schiffer et al. [11, 27], Schiffer and Walther [88], and Hiermann et al. [96] applied DP on resource CSPP (RCSPP) to find the optimal facility route configuration in order to enhance the solutions produced by metaheuristic, while Pourazarm et al. [54] applied DP to find the fastest BEV routes in the single and multivehicle environment. Similarly, Montoya et al. [63] used pulse algorithm to determine the optimal placement of refueling facilities in a route with fixed customers. Similar to the FRVCP formulation of Montoya [18], Keskin and Çatay [79] in their E-VRPTW-FC formulated the MILP program to optimize charging related decisions in BEV route with the fixed customer positions. For every  $\Omega$  iteration, the authors tried to solve this problem exactly by applying CPLEX solver. Liao et al. [103] applied a branch-and-bound algorithm to solve the EV touring problem. Montoya et al. [63, 72] and Hiermann et al. [96] used set partitioning on the pool of routes to select the best subset of routes guaranteeing no overlapping in customers assignments. Froger et al. [68] used a branch-and-bound algorithm to solve the set partitioning model and to discard the selection of routes that are infeasible or for which the total time is underestimated.

**5.3. Heuristic Procedures.** Heuristic procedures seek to solve the problem based on the specific knowledge of the problem, usually suboptimal or close enough to a satisfactory solution. In the research field of VRP, heuristic procedures can be split into constructive and improvement heuristics.

**5.3.1. Constructive Heuristics.** Constructive heuristics are often used to generate an initial solution by either serial or parallel route construction. Solutions are constructed in a greedy way, which often produces solutions of the VRP that are 10–15% far from an optimal solution [4]. In the E-VRP, constructive heuristics are modified and adapted to the BEV characteristics and feasibility checks. Some well-known general constructive heuristics used in E-VRP are presented.

TABLE 2: Overview of the applied procedures.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Bektaş and Laporte [40]	PRP					CPLEX - small	Generated	
Conrad and Figliozzi [22]	Recharging VRP		Route construction	Iterative route construction and improvement			Solomon - CVRP-TW	(1) Sequential construction of routes (2) Route improvement - improves a group or set of routes
Gonçalves et al. [21]	VRPPD with mixed fleet					CPLEX - small		ALNS destroy: Random, WorstDistance, WorstTime, Route, Shaw, ProximityBased, TimeBased, DemandBased, HistoricalKnowledge, Neighborhood, Zone, NodeNeighborhood ALNS repair: Greedy, Regret, GreedyWithNoise, RegretWithNoise, Zone Speed opt.: first optimal speeds are computed to find violations if any, and then in the second part the current speeds are revised to optimize the fuel consumption
Demir et al. [32]	PRP	CWS		ALNS + speed opt.			PRP	
Erdogan and Miller-Hooks [23]	GVRP	(1) MCWS (2) DBCA		LS		CPLEX - small	GVRP	LS operator: Exchange
Omidvar and Tavakkoli-Moghaddam [24]	GVRP			(1) SA (2) GA	SA	CPLEX - small	Generated - Solomon procedure	No description of the implementation
Abdallah [41]	PHEVRPTW	Push forward insertion heuristic + LP optimization of charges		Heuristics for feasible upper bound, TS	SA	MATLAB, CPLEX - small, DP	Adapted Solomon	Compare three procedures for small instances: (1) Master LP: three different relaxations by applying cutting plane algorithm + CPLEX for MIP subproblem; (2) Best Lagrangian bound + heuristic (3) CPLEX TS neighborhood operators: $\lambda$ - interchange ( $\lambda = 1, 2$ ), DP for determining optimal route cost (1) Determine the set of optimal routes (2) Assignment of routes over the planning horizon (3) Charge scheduling + rerouting to CSs Mutation: OR, XOR between 3 individuals Crossover: random 2-point
Barco et al. [34, 42]	E-VRP and charge scheduling	Random		Differential evolution, GA			Real	
Van Duin et al. [43]	FSMVRPTW, EVFSMVRPTW	Sequential insertion heuristics		LS				Shortrec software LS operators: Bräysy and Gendreau [26]

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Adler and Mirchandani [44]	Online routing of BEVs					Approximate DP	Real	Markov chance-decision process
Alesiani and Maslekar [45]	BEV routing	X	Random	GA			Generated	Solution: a two-dimensional binary array of routes and CSs After assigning CS to routes the TSP route is determined Initial: no charging, battery constrained SPP Crossover: 2-point, the second parent is chosen to reduce the concurrency at CSs Selection: random or based on the fitness Mutation: random - remove/insert CS, Exchange LS: removing/inserting vertices (Relocate) ALNS of Demir et al. [32] + speed optimization - to generate non-dominated Pareto optimal solutions Heuristic: adaptive weighting with $\epsilon$ -constraint method for bi-objective solutions
Demir et al. [37]	PRP			ALNS + speed opt. + heuristic			Generated	
Felipe et al. [46]	GVRP-MTPR		Constructive- $k$ -PseudoGreedy	LS(48A)	SA		GVRP-MTPR, GVRP, E-VRPTW	LS (48A): Recharge/Relocation, 2-Opt, reinsertion with four options: first or best and yes/no update of savings
Preis et al. [35]	Energy-optimized routing of BEVs	X	CWS	TS		MIP solver - small	Generated	Neighborhood operator: Relocate
Sassi et al. [47]	HEVRP-TDMF		GRH	ILS+LNS			Real	GRH - (1) Feasible charging scheme at the depot, (2) Insertion heuristics LNS destroy: Random LNS repair: Random, Regret, Score-based
Schneider et al. [25]	E-VRPTW	X	Sweep	VNS+TS	SA	CPLEX - small	E-VRPTW, MDVVRPI, GVRP	VNS - 15 neighborhood structures based on the cyclic exchange move TS - operators: 2-opt*, Relocate, Exchange, StationInRe Diversification mechanism based on the attributes frequency
Zündorf [50]	EVRC		Heuristics to speed up exact algorithms			Charging function propagating algorithm, CH,A*	Real	Battery CSPP Multi-objective search using Pareto-sets



TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Bruglieri et al. [29, 51]	E-VRPTW			VNS with branching		CPLEX	E-VRPTW - small	VNS with Hamming distance of three: removing CS, inserting CS, merging Initial - long BEV routes are converted to ICEV routes
Goetze and Schneider [30]	E-VRPTW MF	X	Modified insertion heuristics	ALNS+LS	SA		E-VRPTW MF, VRPTW, E-VRPTW	ALNS destroy: Random, Worst, Shaw, Cluster, StationVicinity ALNS repair: Greedy, Regret, GRASP LS operators: Relocate, Exchange, 2-opt*, StationInRe
Lebeau et al. [52]	FSMVRPTW- EV			CWS + tree branching			FSMVRPTW - EV	The tree is used for different combinations of vehicles - in each branch, one vehicle is selected and vertices are inserted
Moghaddam [53]	E-VRPTW PR					Gurobi	Generated	
Pourazarm et al. [54]	Single/Multi BEV routing					DP	Generated	Single - electric vehicle CSPP Grouping subsets of vehicles into sub-flows
Sassi et al. [48]	VRP-MFHEV		CRH	ITS+LNS	RR		Generalized E-VRPTW	CRH - without charge scheduling at the depot LNS destroy: Random LNS repair: Random First and best improvement
Sassi et al. [49]	VRP-HFCC		CRH	ILS+LNS	RR		Real, generalized E-VRPTW	LNS destroy: Random LNS repair: Random, Regret Perturbation: LNS random destroy and repair AVNS shaking phase - 24 neighborhood structures: 6 sequence relocation, 18 cyclic exchange
Schneider et al. [55]	VRPIS (EVRPRF)	X	MCWS	AVNS+LS	SA	CPLEX - small	EVRPRE, GVRP, VRPIRF	(1) First route selection operators: Random, Length, LengthPerDemandUnit, FacilityDensity, FacilityDetour; (2) Next route is selected from a set of spatially close routes, (3) Vertex selection: Random, DistanceToNextRoute, DistanceToNeighboringVertices LS; FacilityInsertion, 2-opt, Or-Opt, Relocate, FacilityReplacement, Exchange

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Yang and Sun [56]	BSS-EV-LRP	X	Radius covering + MCWS  Modified sweep + greedy	TS-MCWS+LS  SIGALNS	SA	CPLEX - small	BSS-EV-LRP	LS: Relocate, Exchange Neighborhood search: replace located and unlocated BSS Remove all BSSs and then apply the iterated greedy procedure to insert best BSSs ALNS destroy: Random, Worst, Related, RequestGraph, SinglePoint, TwoPoint, Binary, StationBased ALNS repair: Greedy, Regret-k, AdvanceGreedy, AdvanceRegret-k Improvement: try to remove BSSs
Desaulniers et al. [57]	E-VRPTW -SF/MF/SP/MP					Branch-price- and-cut GENCOL + CPLEX	E-VRPTW with modif. for SF/SP	
Doppstadt et al. [58]	HEV-TSP		(1) Random (2) NNH (3) Concorde - TSP	ITS		CPLEX	HEV-TSP	Initial route is driven in combustion mode + LS - 2-Opt hill-climbing 4:1 ratio - four arcs have to be driven in charging mode in order to drive the last one in electric mode LS: $\lambda$ -mode change ( $\lambda = 2, 5$ ) - changes only the mode, $\lambda$ -Opt with mode change ( $\lambda = 2, 3$ ) - changes the route configuration and the mode ITS: (1) Search: 2-opt with mode change + arcs in tabu list, (2) Hill-climbing $\lambda$ - mode change (first or best), (3) Diversification: 3-Opt with the mode change
Hiermann et al. [31]	E-FSMFTW	X	Iterative route construction	ALNS+LS	SA	Branch-and- price+DP	E-FSMFTW, E-VRPTW	Labeling algorithm for CSs and new extension functions ALNS destroy: Random, RandomAndRelated, WorstAndRelated, InefficientRouteAndNeighbour, Target ALNS repair: SequentialNode, ParallelRegret, SemiParallelConstruction, SemiParallelInsertion LS (best of first 50): 2-opt, 2-opt*, Relocate, Exchange, InsertRemoveIF (CS), Resize (type), RelocateAndResize (type)

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Keskin and Çatay [28]	E-VRPTWPR		Insertion heuristics	ALNS	SA	CPLEX - small	E-VRPTW	ALNS destroy options: remove customer only or with preceding/succeeding CS Customer: Random, WorstDistance, WorstTime, Shaw, ProximityBased, DemandBased, TimeBased, ZoneRemoval Route: Random, Greedy Station: Random, WorstDistance, WorstChargeUsage, FullCharge ALNS repair: Customer: Greedy, Regret-2/3, TimeBased, Zone; Station: Greedy, GreedyWithComparison, Best Improved MIP formulation of GVRP
Koç and Karaoglan [59]	GVRP		MCWS	SA	SA	Branch-and-cut	GVRP	SA improvement operators: Merge routes, Swap, station Add/Drop Branch-and-cut is applied to the initial solution On optimal fractional solution, the heuristic is applied to obtain a feasible solution which is further improved by SA
Lin et al. [60]	E-VRP					MATLAB	Real	The linearized model formulation including energy costs and battery charging time
Masliakova [61]	Routing and charging of electric buses		Random CSs positions + route construction	GA+AC			Real	GA: random 2-point crossover, mutation: move the CS to the nearest vertex AC - the ranking of individuals in GA
Mirmohammadi et al. [62]	Periodic green VRP					CPLEX	Generated	
Montoya et al. [63]	GVRP		((1) random NNH, (2) random nearest insertion, (3) random best insertion) + split	Modified multi-space sampling heuristic + set partitioning		Gurobi - set partitioning	GVRP	(1) Build the TSP route constructed by applying one of the construction procedures (2) Split TSP route and repair (CSPP) - pulse algorithm (3) Set partitioning over a set of routes
Roberti and Wen [64]	E-TSPTW	X	Random	Three-phase heuristic: VND + LS + DP		CPLEX, DP	E-TSPTW, TSPTW, E-VRPTW - single vehicle	(1) Initial - TSP (2) Perturbation: random customer relocation - TSP (3) TW feasible - TSPTW + (VND+LS: Relocate backward and forward, 2-Opt, Exchange) (4) Energy feasible - DP - labeling algorithm

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Schiffer et al. [11, 27]	E-LRPTWPR	X	CWS	ALNS+DP+LS		DP	Real	ALNS destroy: large (Add, Drop, SwapPerfect, SwapPerfectOut) - change CS configuration; small (Worst, Related, Route, Shaw, StationVicinity) - change route configuration ALNS repair: SequentialNode, SequentialPerturbed LS (best of 100): 2-Opt*, Or-opt, Exchange, Relocate, StationInsertion DP - optimal determination of recharging visits and charging amount
Wen et al. [65]	E-VSP		Greedy constructive	ALNS	SA	CPLEX - small and post opt.	E-VSP	ALNS destroy: Random, TimeRelated, NeighboringSchedule (problem specific) ALNS repair: regret-2,3,4 (deterministic and stochastic version) Post-optimization: a set of feasible solutions are used in set partitioning model
Andelmin and Bartolini [66]	GVRP					Exact procedure	GVRP, generated	Multigraph - arc represents possible paths between customers with inserted CS Set partitioning formulation + subset row inequalities and $k$ -path cuts (Chvatal-Gomory (CG) cuts of rank one) - MILP formulation of separation + cut-and-column algorithm
Çatay and Keskin [67]	E-VRPTWPR					CPLEX - small	E-VRPTW	First stage without CS capacity constraint: ILS - giant TSP route + perturbation (random double bridge) + split procedure on acyclic graph with heuristic for CSs + LS; 2-Opt, Relocate and MILP model for charging decisions; local optimum solutions are stored in a pool of solutions Second stage: solutions are assembled using Benders' like decomposition into route selection master problem, and CS capacity management sub-problem Three strategies for CS capacity management: (1) no revision (only feasibility check), (2) delay the charging operations to satisfy CS capacity constraints, and (3) revise the charging amount The set partitioning model is solved using the branch-and-bound algorithm and cuts for CS capacity management
Froger et al. [68]	E-VRP-NL-C		NNH-TSP	Two-stage heuristic		Gurobi - MILP models	Modified E-VRP-NL	

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Hof et al. [69]	BSS-EV-LRP	X	MCWS	AVNS+LS	SA	CPLEX - small	BSS-EV-LRP - Yang and Sun [56] and new	AVNS - Schneider et al. [55] - 32 neighborhood structures - (24) + (8) new one based on the facility-removal (4) and facility-replacement (4) LS: 2-Opt, Or-Opt, Relocate, Exchange, FacilityInsertion, FacilityReplacement, FacilityExchange
Leggieri and Haouari [70]	GVRP					CPLEX	GVRP	
Mancini [71]	HVRP		Run MILP for 5 – 60 s	LNS		Xpress	HVRP, GVRP	Two routes are destroyed and the MILP mathematical model is run again for 10 s only on the customers that were previously involved in the selected routes
Montoya et al. [72]	E-VRP-NL		NNH-TSP + split	ILS (VND)+ heuristic concentration		FRVCP-Gurobi	E-VRP-NL	Sequencing first, charging (FRVCP) second VND: Relocate, 2-Opt and GlobalChargingImprovement Perturbation: build TSP route, perturb and split Heuristic concentration: set partitioning over a set of routes FRVCP- Gurobi and a greedy heuristic
Schiffer and Walthert [73]	E-LRPTWPR					Gurobi	E-VRPTW - small	
Shao et al. [74]	EVRP-CTVTT		Encoded mode	GA			Real	Dynamic Dijkstra algorithm for time-dependent SPP, single string encoding, each individual after the crossover or mutation needs to satisfy all the constraints - if not run the procedure again, elitism
Sweda et al. [75]	Adaptive routing and recharging policies of EVs			Two specially designed heuristics			Generated	Origin-destination, (1) Optimal a priori policy - CSPP + searching for policies with a lower costs (2.1) Two-stage heuristic: select the EV path to follow and select adaptive recharging policy for the fixed path (2.2) Adjusting vehicle trajectory - find optimal policy between a priori stop locations and permit adaptive decision making for both routing and recharging
Vincent et al. [76]	HVRP	X	Modified NNH	SA	SA		Various CVRP	Neighborhood operators: Insert, Swap, Reverse Restart strategy

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics	A	Exact & Software	Instances	Description
Amiri et al. [39]	BSS location & scheduling			GA			Generated	Non-dominated sorting GA
Bruglieri et al. [77]	E-VReP		Construction heuristics	Two-phase heuristic		CPLEX	Real	(1) First phase: generate a set of feasible solutions: three random constructive heuristics (2) Second phase: $\epsilon$ -constraint programming
Joo and Lim [78]	EV routing			AC			Generated	Energy SPP AC - each ant selects next vertex by a stochastic greedy search based on the energy consumption and speed
Keskin and Çatay [79]	E-VRPTW-FC		(1) BKS (2) E-VRPTW - PR sol. FC (3) Greedy	ALNS + opt.	SA	CPLEX - small and opt.	GVRP-MTPR, E-VRPTW	ALNS of Keskin and Çatay [28] with modf.: Station destroy procedures: Random, WorstDistance, LeastUsedStation, ExpensiveStation; Customer insertion procedures - only fastest recharging option; modified station insertions algorithms which start from the cheapest and go to the more expensive charging options Post optimization of charging decisions (location, amount and type)
Keskin et al. [80]	E-VRPTW with waiting times			ALNS + opt.	SA	CPLEX - small and opt.	E-VRPTW	ALNS of Keskin and Çatay [28] Post optimization of charging decisions (location, amount and type) Markov decision process - approximated DP solution
Kullman et al. [81]	E-VRP-PP			Routing policies		Exact, FRVCP	Generated	The Benders-based branch-and-cut algorithm to solve the decomposition Integration of static policies into dynamic lookahead policies: pre-decision, post-decision, and one-step rollouts Exact labeling for FRVCP with time-dependent waiting times and discrete charging decisions Nonlinear information penalties that tighten the perfect information bound of the optimal policy



TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics	A	Exact & Software	Instances	Description
Li et al. [82]	MBFM & recharging problem					CPLEX	Real	Case study, new life additional benefit costs method for the replacement of vehicles (CPLEX); two routing methods: single/multi-period approach for solving the recharging problem Multi-layer taxi-flow time-space network
Lu et al. [83]	MTFSP			Network decomposition by vehicles and time		CPLEX - sub-models	Real	Network decomposition by vehicles: first MTFSP sub-model for $k$ EVs is solved and then the MTFSP for the rest of ICEV's Network composition by time: in each slice of the time-sliced network to each request the nearest vehicles are assigned and then the recharging of EV is performed
Masmoudi et al. [84]	DARP-EV	X	Modified insertion heuristics	EVO-VNS	CA, RR	CPLEX - small	HDARP, DARP-EV	(1) EVO - chromosome encoded as a sequence of vehicles and pickup/drop off vertices: (1a) population is divided in $m$ groups (1b) $s$ solutions are selected from each group (roulette wheel), and if the best solution was not improved for the last three iterations, merge crossover (MX1) is applied between one solution from $s$ and one from the rest of the solutions in the groups that were not selected (2) VNS - neighborhood operators: Swap, RemoveSequence, CrossExchange; three versions: VNS <sub>1</sub> : in each iteration roulette wheel selection of routes in neighborhood structures and order of LS operators with the first improvement strategy; CA acceptance function VNS <sub>2</sub> : after new solution additional $n_{loc}$ iterations of LS with only one operator in each iteration, RR acceptance function VNS <sub>3</sub> : $n_{neig}$ for applying each neighborhood structure, LS, acceptance of only better solutions LS: 2-Opt*, RemoveTwoInsertOne, Relocate, 4-Opt, Remove/Insert BatteryStation (3) Insert $s$ new solutions in the population, the rest of the solutions are generated based on the destroy and repair heuristic (2-regret)
Paz et al. [86]	MDEVLRPTW - BS/PR/BSPR					CPLEX	Modified E-LRPTW - small	

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Metaheuristics	A	Exact & Software	Instances	Description
Pelletier et al. [87]	EFV-CSP			Hybrid Metaheuristics		CPLEX	Generated	The focus is on the deriving meaningful insights and not the computation prowess
Poonthahir and Nadarajan [38]	F-GVRP		NNH + random	TVa-PSOGMO			GVRP	Particle encoding: customer sequence only which is converted to solution by inserting depots and refueling stations Personal and global best particles + velocity update Time-varying inertia and acceleration coefficients Decoding: rank of value to convert the route to the particle position GMO and archive of Pareto optimal solutions Penalty terms evaluation - corridor-based approach - a range of possible refueling times + concatenation operators ALNS and LS: Schiffer et al. [11] DP - optimal place of recharging visits - solve resource constrained SPP for each route by extending the given routes only for facility vertices + resource extension functions for feasibility and dominance checks
Schiffer and Walther [88]	LRPIF	X	MCWS	ALNS+LS+DP		DP	E-LRPTWPR (old and new), BSS-EV-LRP, E-VRPTW, E-VRPTWPR	Parallelized ALNS of Schiffer and Walther [88]: LS, DP and customer configuration performed in parallel for fixed CS configuration, while CSs configuration, adaptive evaluations and stopping criterion are executed within a critical section of the code Adversarial approach for robust formulation
Schiffer and Walther [89]	RELPTWPR	X	MCWS	ALNS+LS+DP		DP, Gurobi - small	Real	General GA + elitism LS: random Exchange
Shao et al. [90]	E-VRP			GA+LS			Real	Fuzzy programming and fuzzy preference relations are applied to transform the objective function into a single objective function GA: chromosome genes - CSs
Wang et al. [91]	BEV routing				GA + fuzzy prog.		Generated	Initial: (1) TSP - construction based on the savings and intensity trail value, (2) VRP - insertion of depots, (3) E-VRP - insertion of CSs ILS perturbation: random removal of customers and CSs, and then the best insertion procedure and CSs, and then the best insertion procedure LS: 2-Opt*, Exchange, Relocate, StationInRe Elitism
Zhang et al. [36]	E-VRP	X	Construction + insertion heuristics	AC+ILS, ALNS			Generated	ALNS of Goeke and Schneider [30] with modf.: no cluster removal operator
Basso et al. [92]	2sEVRP					CPLEX	Generated	Bellman-Ford algorithm for SPP

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Breunig et al. [93]	E2EVRP		Greedy insertion	LNS+LS+DP		Exact - route and candidate solution enumerations	E2EVRP	LNS destroy: RelatedRemoval, Random, CloseSatellite, OpenSatellite, RemoveSingleCustomerRoutes LNS repair: (i) the reinsertion of customers in the second echelon with Greedy (random or demand quantity) operators; (ii) greedy insertion on the first echelon routes (iii) DP for CSs insertions LS on second echelon: 2-Opt, 2-Opt*, Relocate, Swap, Swap2-1 Path-based approach - the route is a composition of paths for a subset of customers without station visits: (1) All feasible paths are generated, removing from set of the dominated ones (2) MILP formulation is used to combine paths in routes
Bruglieri et al. [94]	GVRP					Exact + Xpress	GVRP	(1) All feasible paths are generated, removing from set of the dominated ones (2) MILP formulation is used to combine paths in routes
Froger et al. [95]	E-VRP-NL			Heuristic for FRVCP		Exact labeling - FRVCP, GUROBI - small	E-VRP-NL	Heuristic for FRVCP: at each CS, EV is charged minimum energy required to reach next CS or depot The sequence of customers and vehicle types without CS (1) GA - assignment and sequencing of customer visits - 2 populations: feasible and infeasible (2) DP - sequencing of CS: resource CSPP - labeling algorithm for CSs insertions (3) Greedy extension policy - known sequence, partial recharging, mode selection Route evaluation: (2) & (3), extension functions GA crossover: (1) binary tournament selection, (2) each parent forms giant TSP tour, (3) OX crossover, (4) split procedure GA mutation: LNS - random selection of destroy operators: Random, RandomRoutes, Similar, Target; uniform selection of repair operators: Greedy, Regret-2 LS (only promising edges): 2-opt, 2-opt*, Swap, Relocate (single and pairs), VehicleSwap Set partitioning over a pool of routes during the search The cache memory of move evaluations for unchanged route sequences
Hiermann et al. [96]	H <sup>2</sup> E-FTW	X		GA+LS+LNS+DP		CPLEX - set partitioning	H <sup>2</sup> E-FTW, E-FSMFTW, E-VRPTW, E-VRPTWPR	(1) GA - assignment and sequencing of customer visits - 2 populations: feasible and infeasible (2) DP - sequencing of CS: resource CSPP - labeling algorithm for CSs insertions (3) Greedy extension policy - known sequence, partial recharging, mode selection Route evaluation: (2) & (3), extension functions GA crossover: (1) binary tournament selection, (2) each parent forms giant TSP tour, (3) OX crossover, (4) split procedure GA mutation: LNS - random selection of destroy operators: Random, RandomRoutes, Similar, Target; uniform selection of repair operators: Greedy, Regret-2 LS (only promising edges): 2-opt, 2-opt*, Swap, Relocate (single and pairs), VehicleSwap Set partitioning over a pool of routes during the search The cache memory of move evaluations for unchanged route sequences

TABLE 2: Continued.

Reference	Problem name	NF	Initial	Hybrid Metaheuristics Heuristics	A	Exact & Software	Instances	Description
Jie et al. [97]	2E-EVRP-BSS	X	Extended sweep	CG-ALNS	SA	CPLEX - small	2E-EVRP-BSS, 2E-VRP	(1) Second echelon - ALNS: destroy: Random, Related, SatelliteClosing, SatelliteOpening, SatelliteSwapping, RandomRoute, BSS-CustomerRoute, BSS-ProportionRoute, BSS-CostBased, BSS-BestUse; repair: basic and advance Greedy, basic and advance Regret-(2,3) (2) First echelon: column generation (CG) for split delivery EV routing problem, DP and branching scheme for the SPP with resource constraint sub-problem (3) Label-based BSS optimization
Koyuncu and Yavuz [98]	MGVRP					CPLEX	GVRP	Upper and lower bound for energy state (labeling algorithms) and arrival times Lower bound on the number of vehicles
Macrina et al. [99]	GMFVRP-PRTW		CPLEX	Hybrid LNS			GMFVRP-PRTW	Initial: CPLEX to find an initial feasible solution Random order of destroy and repair operators and tabu list to forbid the removal of recently removed customers and reinsertion into routes from which they were removed Destroy: Random, WorstDistance, WorstTime, Route Insertion: Greedy, GreedyAndCS, GreedyNewRoute, GRASP
Macrina et al. [33]	GMFVRP-PRTW		Sequential insertion heuristics	ILS		CPLEX - small	E-VRPTW	Initial: two clusters of customers for EV and ICEV routes - (1) ICEV routes, (2) EV routes - all customers from EV cluster and all un-routed customers from ICEV cluster LS with/without penalty function and perturbation operators: relocate-only EV, ICEV or both
Normasari et al. [100]	CGVRP		NNH	SA+LS	SA	CPLEX	CGVRP, GVRP	Random neighborhood structures: Swap, Insertion, station Insert/Delete

*Reference*: referenced paper; *Problem name*: the name of the analyzed problem; *NF*: allowing infeasible solutions during the search process; *Initial*: procedure used for the creation of the initial solution if any; *Hybrid heuristics*: heuristics, metaheuristics, and hybrid combinations applied for solving the problem; *A*: the acceptance criteria; *Exact & Software*: the exact procedure or software applied on the analyzed problem or subproblem; *Instances*: instances used for testing of the applied procedures where *Real* denotes the real case-study instances, *Generated* denotes the artificially created instances which are not benchmark instances, and other denotes the benchmark instances of the problem; *Description*: short description of procedure used to solve the analyzed problem.

The *modified savings method of Clarke and Wright* (MCWS) [141] with the insertion of CSs is used to generate an initial solution of several E-VRP and related problems [23, 32, 52, 69]. The algorithm starts with the creation of back-and-forth routes for each customer. Next, CSs are inserted in a least-cost manner, in either feasible or infeasible routes. Then, the routes that increase the savings the most are merged together and the procedure is repeated until all customers are served.

The *sweep algorithm* [142] inserts customers in the active route in a circular manner, resulting in efficient space division. The customers are sorted based on the value of a polar angle between the depot and randomly chosen point. Then, customers are inserted in the active route at the position causing minimal increase until a violation of route constraint, when usually a new route is opened [25, 56, 97].

*Nearest neighbor heuristic* (NNH) starts from the depot, and in each iteration the customer with the least-cost increase from the previously selected customers is added to the route. The route is terminated when some/any constraints are violated and a new route is opened [76].

*Route-first cluster-second* [143] approach constructs a giant infeasible TSP tour, usually with NNH, which is then split into several routes. The split procedure can be solved as the SPP on the acyclic graph. The procedure generates energy infeasible solutions, so the CS placement and charging schedule need to be determined to make the solution energy feasible [63, 68, 72].

The *cluster-first route-second* [144] approach divides the customers into clusters and then each cluster is solved as a subrouting problem. One example used in GVRP is the *density based clustering algorithm* (DBCA) proposed by Erdoğan and Miller-Hooks [23], which determines customer clusters based on their density, by ensuring that each customer has at least some predefined number of customers in the  $\epsilon$  radial neighborhood. In the second step to obtain the routes, the MCWS is performed.

Many researchers are adopting some form of the insertion heuristics, which have some characteristics of already reviewed heuristics. The general idea of the insertion heuristics is to iteratively insert or add customers in available routes. In each step of the algorithm, an unserved customer together with a route and its position in the route are determined to least increase the objective function. This selection can be deterministic but is often stochastic to select at random one from  $k$ -best inserts to diversify the solution, especially in population metaheuristics. In most feasibility ensured cases when the energy constraint is violated, a CS is added to the route to make it energy feasible, and when the other constraints are violated such as vehicle load capacity or time windows, the new route is opened, and the procedure is repeated. Examples of adaptations to the E-VRP are *push forward insertion heuristics* (PIH) [41, 145], *constructive- $k$ -PseudoGreedy* [46], [28, 30, 31, 33, 36] modified insertion heuristics, and *charging routing heuristic* (CRH) [47–49], in which first the charging scheme at the depot is determined, and then the least-cost feasible customer insertions are performed.

**5.3.2. Improvement Heuristics.** Improvement heuristics or local search (LS) explores the neighborhood of the incumbent

solution, searching for a better solution. The neighborhood is explored by applying perturbation moves based on the composite neighborhood operators. The local search stops when no improving solution can be found in the neighborhood of the incumbent solution, which is then called local optima. A great variety of researchers have performed LS procedures to intensify the search, coupled together with perturbation moves to escape the local optima. Often, the perturbation moves are similar to the neighborhood operators used in LS phase. Most of the classical VRP neighborhood operators [3, 4] are used in the E-VRP, but some additional neighborhood operators have been developed. Depending on whether the operators perform on only one route or between the routes, they can be divided into *intraroute* and *interoute* operators. Here, we present some of the most used LS neighborhood operators in the E-VRP literature:

- (i) *2-Opt* [146] replaces two arcs with the two new ones, with a possibility of route direction reversal.
- (ii) *2-Opt\** [147], unlike the *2-Opt* operator, avoids the reversal of a route direction.
- (iii) *4-Opt* [84, 148] replaces four consecutive arcs with four new ones.
- (iv) *Or-Opt* [149] replaces three arcs with three new ones such that a sequence of three vertices is relocated.
- (v) *Exchange (Swap)* [150] exchanges the two vertices in the solution.
- (vi) *FacilityInsertion* [55] and *StationInsertion* [11] operators insert refueling or loading facility at best position in between two consecutive facility visits where fuel or load violation occurred. The similar operator *InsertRemoveIf* [31] inserts CS into the part of the route where the violation occurred to make the route energy feasible and repair violations, but it also removes redundant CS visits. Schneider et al. [55] also proposed the facility-related removal and exchange operators *FacilityReplacement* and *Exchange*.
- (vii) *RechargeRelocation* [46] removes all visits to the CSs and then tries to insert CSs at positions leading to a better solution. Similarly, the *GlobalChargingImprovement* [72] operator removes all CS visits in the route, and if the route is energy feasible, it stops; otherwise, it solves FRVCP to optimize CSs positions and charging amount.
- (viii) *Relocate* [150] removes one vertex from the solution and inserts it into another position in the solution. Felipe et al. [46] proposed a modification of the relocation operator called *Reinsertion*,  $R_b^u$ , for solving GVRP-MTPR. Binary parameters  $b$  and  $u$  denote whether the first-improve or best-improve strategy is used and whether the values of savings are updated or not. The *RemoveTwoInsertOne* [84] operator removes two vertices at random and inserts them one by one into another vehicle at best possible positions.
- (ix) *CrossExchange* [151] removes  $n_1$  consecutive vertices from route  $r_1$  to route  $r_2$ , and  $n_2$  consecutive vertices



from route  $r_2$  to route  $r_1$  [84]. The *RemoveSequence* [84] operator is similar as it removes consecutive vertices from only one route and inserts it into another one.

- (x) *Resize* [31], *RelocateAndResize* [31],  *$\chi$ -OptWithModeChange* [58], and *VehicleSwap* [96] change the vehicle type assignment or the propulsion mode in the heterogeneous E-VRP and HEV routing problems.
- (xi) *StationInRe* [25] inserts and removes CSs from the solution in so that if an arc to the CS is already in the solution, it is removed; otherwise it is inserted.

The order of the operators affects both the solution quality and the execution time. Often, there is a question whether to make a *first better* or *the best* improvement move in the LS phase [152]. Some E-VRP researchers perform one or the other, and some combine the two approaches, i.e., best move of first 100 [11] or first 50 [31].

Neighborhood operators explore only the space of the immediate vicinity of the current solution, which often leads to local optima. In order to search a larger solution space, Shaw [153] proposed the *large neighborhood search* (LNS) heuristic, which is based on destroying and repairing the solution. Heuristic efficiency depends on the implemented destroy and, especially, repair operators. The main drawback of the heuristic is the repeated use of the same destroy and repair operators, which can lead to local optima. A modification of LNS by Christiaens and Vanden Berghe [154] is giving state-of-the-art results on a wide range of VRP variants. In the E-VRP, LNS is successfully implemented by Sassi et al. [47–49] and Zhang et al. [36], but most often it still needs some other guiding heuristic to escape the local optima. Macrina et al. [99] applied an LNS scheme to solve GMFVRP-PRTW. A list of used destroy and repair operators in the E-VRP is presented as a part of the adaptive LNS version in the appendix.

**5.3.3. Route or Move Evaluation.** Operator evaluation is an important part that highly affects execution time and has to be efficiently implemented. For example, in VRPTW, for most of the basic neighborhood operators, the load capacity check and time window check can be computed in constant time  $\mathcal{O}(1)$ , if for each customer endogenous variables, i.e., remaining load capacity, latest arrival time, etc., are tracked. In the E-VRP, there is also a constraint regarding the vehicle battery capacity (battery feasibility), which also influences the time window feasibility. Therefore, time window evaluation is more complex. For example, when the customer is trying to be inserted into the route, the load and energy check can be computed in constant time  $\mathcal{O}(1)$  for most of the basic neighborhood operators. The endogenous variables for vertices between the insertion position and next CS have to be recomputed in order to know the charging time at next CS. The complexity of such approach is  $\mathcal{O}(B)$  where  $B$  is the number of customers between the position of the insertion and next CS. If there is no CS in the second part of the route, no recalculation is needed. By storing additional information about the recharging position and its amount and by the use of general concatenation operators presented

by Vidal et al. [155], Hiermann et al. [31] reported that all of the basic neighborhood operators can be evaluated in  $\mathcal{O}(1)$ . Goeke and Schneider [30], instead of storing additional information about recharging, used a surrogate cost function whose evaluation is less demanding. In the search process, the authors store  $\epsilon$  best moves in the composite neighborhood and evaluated either exact or surrogate cost, and at the end if there is any move with the surrogate cost, its exact cost is evaluated and the best move is chosen. Hiermann et al. [96] presented a different approach for  $H^2E$ -FTW in which customer sequence without CSs is determined, and then in the route evaluation part time-consuming CSs insertions and charging schedule are performed. Whenever the route needs to be evaluated, the heuristic finds the routing cost by applying a labeling algorithm to determine CSs placement and a greedy algorithm to determine charging amount and running mode of PHEV. As such an evaluation is time-consuming, the authors performed the following: (i) use neighborhood restrictions to consider only promising arcs for each customer vertex; (ii) keep track of move evaluations in cache memory; and (iii) lower bounds to first evaluate the move as ICEV in  $\mathcal{O}(1)$ , and then if it improves the objective function, perform the route evaluation.

**5.4. Hybrid Metaheuristic Procedures.** Many researchers employ metaheuristics to continue the exploration after the first local optima occurrence. Metaheuristics can be defined as heuristics guiding other heuristics and can be divided into neighborhood-oriented metaheuristics and population metaheuristics. Most often, multiple metaheuristics and heuristics are combined together and adapted to the problem; therefore, the term hybrid is used in such occasions.

**5.4.1. Neighborhood-Oriented Metaheuristics.** Neighborhood-oriented heuristics iteratively explore the neighborhood of the incumbent solution. Here, we present the most used neighborhood heuristics in the E-VRP literature.

*Simulated annealing* (SA) [156] controls the search by the temperature parameter, which mimics the physical cooling process of the material. Most often, it is used as a control heuristic for accepting solutions produced by other heuristics. The probability of accepting solution  $S_2$ , which is worse than the current best solution  $S_1$ , is computed as  $e^{(f(S_1)-f(S_2))/T}$  (Boltzmann function [157]), where  $T$  is the temperature parameter. The higher the temperature value, the higher is the probability of accepting a deteriorating move to escape the local optima. The procedure starts with high temperatures at the beginning and then dynamically lowers the temperature as the search process comes to an end. In the E-VRP, most often initial temperature is selected such that a  $\mu\%$  worse solution than the initial solution is accepted with a probability of 0.5 [25, 28, 30, 46].

*Tabu search* (TS) [158, 159] is a well-known metaheuristic, which uses memory to forbid the exploration of the space that has already been explored. It escapes the local optima by accepting a worse solution if it is the best in the explored space. Arcs that are deleted from the solution are stored in the *tabu-list*, which prohibits the reinsertion of the deleted arcs into a specific part of the solution for a designated number of



iterations, *tabu-tenure*. In some implementation cases, when the new best solution is found, some of the moves and arcs are lifted and removed from the *tabu-list*, *aspire criterion* [25]. In the E-VRP, TS with neighborhood operators is used to intensify the search [25, 35, 48].

*Variable neighborhood search* (VNS) [160] is a metaheuristic that systematically changes the neighborhood each time there is no improvement in LS phase. The changes are based on predefined neighborhood structures. In each step, the new neighborhood operator is used, which increases the vicinity from the incumbent solution. The selection of neighborhood operators and customers can be deterministic (*variable neighborhood decent* (VND)) or probabilistic. Schneider et al. [25] successfully adapted it in the E-VRPTW for diversification of the solution based on the cyclic exchange moves, resulting in 15 different neighborhood structures. The VNS in the E-VRP was also combined with exact algorithms like branch-and-price [51], evolutionary algorithms [84], deterministic LS operators [64, 72], and adaptive mechanisms (AVNS) [11, 27, 55, 69].

*Iterated local search* (ILS) [161] is based on successively repeating LS on the incumbent solution. When LS ends up in local optima, perturbation move is applied to escape from local optima. The perturbation move can be scaled dynamically to overcome local optima. The effectiveness of the procedure highly depends on the local search procedures used. Montoya et al. [72] applied ILS together with VND for solving E-VRP-NL with perturbation performed by splitting perturbed giant TSP tour. Macrina et al. [33] used typical ILS with some of the LS operators used in the perturbation phase to solve GMFVRP-PRTW. Zhang et al. [36] used ILS and LNS in an ant colony framework for solving GVRP. The LNS is used as a perturbation mechanism while the LS is used to further improve the solution. Sassi et al. [47, 49] also used a multistart combination of ILS and LNS but without the LS phase. Sassi et al. [48] applied similar ILS paradigm, but they called it *iterated tabu search* (ITS), as they used TS for the intensification and LNS for the perturbation. A similar ITS procedure is applied by Doppstadt et al. [58] for the HEV-TSP.

*Adaptive large neighborhood search* (ALNS) [162, 163] is the extension of LNS heuristic in which, throughout the search process, different procedures for destroying and repairing the solution are selected in an adaptive manner based on their past performance. The idea is to have versatile destroy and repair operators in order to escape the local optima. For each destroy and repair operator, the adaptive weight and score are increased in the following order (scores  $\sigma_1 \geq \sigma_3 \geq \sigma_2$ ): (i)  $\sigma_1$  if a new best solution is found; (ii)  $\sigma_2$  if the solution is better than the previous one; (iii)  $\sigma_3$  if the solution is worse than the previous one but it is accepted due to the acceptance criteria; and (iv) no increase if the solution is not accepted. The score of worse solution  $\sigma_3$  is greater than the score of better solution  $\sigma_2$ , to reward unimproved solutions that diversify the search [28]. For each  $N$  iteration, the weights are updated based on the previous weight values and new score values. As a great number of the researchers used ALNS for solving the E-VRP and related problems (Table 2), a vast number of destroy and repair operators were

applied. Some of them originated from Pisinger and Ropke [162] but also some new problem-specific ones regarding the CS have been developed [28, 30–32, 79, 88]. The description of the applied destroy and repair operators in E-VRP is presented in the appendix. Schiffer and Walther [89] applied a parallel ALNS version of Schiffer and Walther [88] to solve the RELRPTWPR. The OpenMP was used to solve multiple scenarios in parallel, where the customer destroy and repair operators, as well as the LS and DP, are performed in parallel. The change of facility configuration, adaptive evaluations, and stopping criterion are not parallelized.

SA is one of the most used metaheuristics for acceptance criteria, in order to escape the local optima, but some others can also be applied. Tiwari et al. [164] proposed the Cauchy function (CA)  $T/((f(S_1) - f(S_2))^2 + T^2)$  for the acceptance of the solution as it gives more opportunities to escape from local optima. The record-to-record (RR) principle [165], which adds negative value to the objective function of the best solution so far, can also be used to escape the local optima. Several papers in the E-VRP field applied such heuristics for solution acceptance [48, 49, 84].

**5.4.2. Population Metaheuristics.** Population metaheuristics are based on the natural selection to evolve a population and survival of the fittest. They have been widely applied in the VRP field: genetic algorithm [166], path relinking [167], scatter search [168], ant colony [169], bee colony [170], particle swarm optimization [171, 172], etc.

*Genetic algorithm* (GA) consists of a pool of individuals (VRP solutions) called a population, which goes through the process of evolution. Such evolution process includes evaluation of individuals, selection of parents, crossover, mutation, and a replacement of the old population by a new one. The important aspects of the algorithm are the representation of the VRP solution, efficient management of population diversity, and fitness function to compare the individuals in the population. For further algorithm description and application, we refer to Potvin [173]. The *ant colony* (AC) algorithm is based on ant behavior when they search for food. As they search, they leave pheromones, which accumulate over the most used path, leading to the optimal path between the food and the nest. AC algorithm consists of three procedures: pheromone initialization, probability adjustment, and pheromone update. *Particle swarm optimization* (PSO) mimics the behavior of organisms such as fish schooling or bird flocking. Physical movement of the individual (particle) in the swarm is tracked and then well-balanced moves are applied to adapt the individual to the global and local exploration. Each particle has a position in the swarm and corresponding velocity in order to optimize the problem following personal best solution and global best solution [172].

Barco et al. [34, 42] applied *differential evolution* to solve the E-VRP and charge scheduling. The initial individual represents the valid route assignment and charging profile. Next, mutation based on OR and XOR operators is performed on three randomly selected individuals. The crossover is a random 2-point while the selection chooses between the target individual and new individual, which is generated

by the crossover. Alesiani and Maslekar [45] applied GA for solving a single BEV routing problem. The individual is represented as a two-dimensional binary array, where the rows represent the routes and the columns represent CSs. The initial population contains feasible individuals, individuals without CSs, and individuals containing battery CSPP routes. The authors select parents at random or based on the probability and apply 2-point crossover. Mutation is applied at random to remove, insert, or exchange CS, while the LS is performed to improve the solution. Shao et al. [74] applied GA without detailed description for solving real-life EVRP-CTVTT instances. The authors encoded the individual with all visited vertices including customers, depots, and CSs. The crossover operator tries to create a new individual that satisfies all the constraints. Also, elitism is applied to preserve the elite individuals. The authors reported the average computational time of three hours. A similar algorithm with the LS after the crossover and mutation is applied for solving E-VRP by Shao et al. [90] to obtain an optimal operation scheme including routes, charging plan, and driving paths. Masliakova [61] applied the GA to find the optimal placement of CSs for electric bus routes. The individual is represented as the sequence of CSs and bus stops visited, and a crossover is implemented as the swap between the subroutes. Mutation moves the CS to the nearest bus stop or creates a new bus stop with CS. For the fitness function, the author used an AC approach, which ranks the individuals in the population. Joo and Lim [78] used the AC approach to determine the energy consumption and speed in energy-efficient BEV routing. Each ant selects the next vertex with a greedy stochastic search, which combines heuristic information about energy consumption, speed, and pheromones. The authors concluded that AC strategy is suitable for routing BEVs in terms of energy efficiency. An efficient version of evolutionary algorithm and VNS (EVO-VNS) is applied by Masmoudi et al. [84] for solving the DARP-EV. Each population is divided into  $m$  groups in such a way that each group contains good and bad solutions. Further on,  $s$  solutions from each group are selected in the VNS phase. If the best solution in the current group is not improved for three iterations, the crossover between the selected solutions and the rest of solutions in the group is performed. The rest of the individuals in the population are generated by applying destroy-repair procedure with regret heuristic. Poonthalir and Nadarajan [38] applied particle swarm optimization with greedy mutation operator and time-varying acceleration coefficient (TVa-PSOGMO) to solve the F-GVRP. Each particle is encoded as the sequence of customers who are converted to feasible routes by inserting depots and refueling stations. Time-varying inertia and acceleration coefficients are employed to better manage diversification and intensification exploration in the multiobjective environment. Additionally, to prevent premature convergence, a greedy mutation operator (GMO) is performed. Hiermann et al. [96] used hybrid GA (HGA) with LS to obtain good assignment and routing decisions for  $H^2E$ -FTW. The authors represented the solution as a set of vehicle routes and used a biased fitness function. During the search, the authors maintained two populations of feasible and infeasible individuals. Parents

are selected using binary tournament scheme from those two populations. The recombination is performed on the parent routes combined in a giant single tour, with an OX crossover and a split algorithm to obtain a complete solution. The mutation phase is performed as LNS with destroy and repair operators similar to ones of Hiermann et al. [31]. In the end, the solution is improved by the LS procedure.

*5.5. Summary of the Applied Procedures.* The overview of the applied procedures in the E-VRP field is presented in Table 2, where 78 papers are presented. Searching in the space of infeasible solutions is applied in 17 papers, while the LS for the intensification is applied in 25 papers. Most of the researchers do not use any special criteria for accepting the solution, but 17 papers use SA as acceptance criteria, three RR and one CA. For the initial solution, mostly some form of the universal insertion heuristic is applied (18). Then follow MCWS (11), NNH (7), random (5), split procedure (4), and sweep (3). Regarding the applied heuristics for solving the problem, the order is the following: ALNS (14), GA (10), LNS (7), TS and ITS (6), ILS (6), AC (3), VNS (3), AVNS (2), VND (2), PSO (1), and differential evolution (1). From all of the exact procedures, the DP is the most applied one (12) and then there is the branch-bound-cut-and-price (5). Forty-three papers use commercially available software like CPLEX (32), Gurobi (7), MATLAB (2), and XPRESS (2) for solving either the whole problem or some subproblem during the search, mostly charging decisions.

We can note that a vast number of researchers use hybrid heuristics procedures with only a few focusing on the exact procedures [31, 57]. It can be noted that most of the procedures that are giving state-of-the-art results are using exact procedures at some point during the search, usually for (i) optimal determination of charging visits and charging amount [79, 88, 96] and (ii) set partitioning over the set of routes at the end of the search [65, 72, 95, 96].

The ALNS is one of the most applied procedures and it is giving the BKSs on various problem variants. Among them, we can point out the following: ALNS of Hiermann et al. [31] (HPH) and ALNS of Goeke and Schneider [30] (GS) for heterogeneous fleet, ALNS of Keskin and Çatay [28] (KCI) and ANLS of Keskin and Çatay [79] (KC2) for partial recharges and different charging technologies, and ALNS of Schiffer and Walther [88] (SW) originally designed for the LRPIF (ELRPTWPR) which is also giving BKSs on the E-VRPTW and E-VRPTWPR. Schiffer and Walther [88] compared their algorithm on the E-VRPTW instances to the HPH, GS, KCI, and VNS-TS of Schneider et al. [25] (SSG). The authors found nine BKSs with an average gap of 0.27% to the previous BKSs and average running time of 3.77 minutes. On the E-VRPTWPR, compared to the KCI, the authors found 41 new BKSs. The proposed ALNS of SW is efficiently parallelized to solve the RELRPTWPR [89].

Most of the population-based metaheuristics are not efficiently implemented so they are not giving high-quality solutions. But, the HGA of Hiermann et al. [96] (HHPV) gives some of the BKSs on the E-VRPTW, E-VRPTWPR, and  $H^2E$ -FTW. The authors compared their results on the E-VRPTW and E-VRPTWPR to the SSG, HPH, GS, and

KCl. On the E-VRPTW instances, the HHPV found 11 BKSs with 1.36% increase in the vehicle number, 0.46% distance gap from the previous BKSs, and an average running time of 11.59 minutes. On the E-VRPTWPR, the authors found 28 new BKSs. By comparing the results of HHPV and SW on E-VRPTW, we can conclude that in 20 instances SW outperformed HHPV, while in the opposite case the HHPV was better in 17 instances.

## 6. Conclusion and Future Research Directions

With the increase in popularity of electric-powered vehicles, many logistics companies have started to integrate electric vehicles into their delivery fleets. Routing a fleet of electric vehicles for delivering goods was formulated as the electric vehicle routing problem. Besides the often used load (cargo) capacity and time window constraints, E-VRP routing models have to account for the limited driving range of BEVs, which directly corresponds to the more frequent recharging needs at CS.

In this paper, a literature review on recent developments in the E-VRP field is presented. We consider BEVs to be vehicles powered only from batteries mounted inside the vehicle. A general overview of the BEV's characteristics for goods delivery includes the driving range, battery capacity, application, and case studies. As energy consumption estimation is an important part of BEV routing, we summarized the recent research on theoretical and data-driven energy consumption models.

Due to the BEVs specific characteristics, new E-VRP variants have emerged: a mixed fleet of electric and conventional vehicles, partial recharging, simultaneous CS siting and BEV routing, nonlinear charging, different charging technologies, battery swap technology, hybrid vehicles, green routing, etc. The development of efficient heuristics was necessary to find optimal or near-optimal solutions to the new routing problems. We reviewed the state-of-the-art exact, heuristic, and hybrid procedures applied for solving various E-VRP variants. The adaptive large neighborhood search [88] and hybrid genetic algorithm [96] produced high-quality solutions on various E-VRP variants. Most of the state-of-the-art procedures use exact algorithms during the search to determine the optimal placement of CSs and the charging amount.

From the literature review, it can be noted that the electric vehicle routing research community has grown rapidly in the last few years and many problem variants have already been explored. Nevertheless, we can highlight possible future directions as follows.

By our observation, there is a lack of papers regarding case studies and application cases, where actual E-VRP models could be evaluated, and some meaningful insights could be drawn. Several energy consumption models were reviewed, but only few are predicting realistic energy consumption at the road segment level in the road network. Only recently, have researchers started to integrate a nonlinear charging process, CS location problem, and hybrid electric vehicles into the E-VRP models. A few papers addressed the problem of CS capacity, as a limited number of BEVs can charge simultaneously at a CS. Several variations were observed:

waiting times [68, 80, 81], battery reservations [44], and adaptive routing with uncertainties in CS availability [75, 81]. Furthermore, the following real-life characteristics have not been sufficiently studied in the E-VRP field: dynamic traffic conditions [74], uncertainties in demand, travel time, time windows, service time and charging time [75, 81, 89], compatibility of BEVs with chargers in CSs [48, 49], and recharging at public or private CSs [81].

For future research regarding the solution procedures, we can highlight the following. Although there have been a couple of exact procedures developed for GVRP, only a few exact procedures have been proposed for the E-VRP and its extensions. A great number of researchers applied a population metaheuristic to solve the problem, but only a handful of them produced high-quality solutions in reasonable computation time. To improve the computation time, parallelized procedures could be used [18, 89]. Furthermore, more general procedures could be investigated as most of the state-of-the-art procedures are using complex problem-specific heuristics.

## Appendix

### ALNS Destroy and Repair Operators

Here, we present an overview of the used destroy and repair operators in ALNS metaheuristic for solving EVRP variants and related problems. We divide the destroy operators in terms of whether they include customer removal (C), station/facility removal (S), or route removal (R).

Destroy operators:

- (i) *Add*, *Swap*, *Drop*, *SwapPerfect*, and *SwapPerfectOut* (S) [88, 174] operators change the configuration of CS in the solution. As the name says, *Add*, *Swap*, and *Drop* operators add, remove, and exchange the CSs in the solution. *SwapPerfect* and *SwapPerfectOut* [88] close  $n$  arbitrarily or no longer needed CSs and iteratively open  $n$  new ones with the best insertion strategy.
- (ii) *BatteryBestUse* (C) [97] operator detects the route with the highest charge level at the return to the depot and the customers between the depot and last-used BSS are removed from the solution.
- (iii) *BSSCostBased* (C) [97] operator removes customers that have high traveling costs of arcs connecting to the BSS. First, the random BSS is selected, and the customers with the highest cost are removed from the solution.
- (iv) *BSSCustomerRoute* and *BSSProportionRoute* (R) [97] operators aim to remove routes that visit more BSSs and fewer customers (*BSSCustomerRoute*) and routes that have a relatively large cost of visiting BSSs (*BSSProportionRoute*).
- (v) *Cluster* (C,S) [30, 163] operator removes clusters of geographically close vertices. First, a random route is selected and divided into two clusters. One of the two clusters is selected at random, and all the vertices in



the cluster are removed. If more vertices need to be removed, first the random vertex from the removed cluster is chosen. Then, the route closest to the chosen vertex is selected, and the procedure is repeated.

- (vi) *ExpensiveStation* (S) [79] operator removes stations that incur the biggest costs.
- (vii) *FullCharge* (S) [28] operator removes stations at which BEVs are fully charged.
- (viii) *HistoricalKnowledge* (C,S) [32, 163] operator remembers the best position cost for each vertex during the search and then the vertex that has the highest difference from its best remembered position is removed from the solution.
- (ix) *InefficientRouteAndNeighbour* (R) [31] operator removes one route based on its average cost per unit transferred and then the next distance closest route is selected for removal.
- (x) *LeastUsedStation* (S) [79] operator aims to reduce frequent visits to CSs. This operator ranks the CSs in the nondecreasing order of charge amount and removes the top  $n$  of them.
- (xi) *Neighborhood* (C,S) [32] operator removes the vertices that have the highest shares in routes average distance.
- (xii) *NodeNeighborhood* (C,S) [32] operator randomly selects a vertex in the solution and removes it and  $n$  vertices in rectangular area around it.
- (xiii) *Random* (C,S) [32] operator randomly removes vertices.
- (xiv) *RandomRoute* and *GreedyRoute* (R) [28, 32, 174] operators randomly or greedily remove  $\omega$  routes. *GreedyRoute* operator removes routes that have the lowest number of customers.
- (xv) *Related*, *RandomAndRelated*, *WorstAndRelated* [31, 162], and *TimeRelated* [65] (C) operators remove a set of customers that are in some sense related. The first vertex is selected at random or based on the detour cost, and the next vertices are selected deterministically or by a roulette wheel selection based on the relatedness measure. The relatedness measure of Pisinger and Ropke [162] takes into account only the distance between the vertices, while Hiermann et al. [31] took into account distance and fractions of time window violations between the vertices.
- (xvi) *Remove customer options* (C,S) [28]: (i) only customer removal; (ii) removal of customer with preceding CS; and (iii) removal of customer with succeeding CS.
- (xvii) *RequestGraph* (C) [163] operator creates the undirected graph where each customer is a vertex and arc weight corresponds to the number of times the arc is used in  $N$  currently known best solutions. Then, similar to the *Shaw* procedure, related customers are removed from the solution.
- (xviii) *Shaw* (C) [153] operator tries to remove customers that are similar to each other by taking into account

geographical distance, demand difference, earliest start time difference, and assigned route difference. The first customer is selected at random while the other customers, instead of a deterministic way, can be chosen in a probabilistic way from  $B$  most similar customers to the current one [30]. Demir et al. [32] and Keskin and Çatay [28] use *ProximityBased*, *TimeBased*, and *DemandBased* as a special case of *Shaw* removal, where either only distance, time, or demand is taken into consideration. *ProximityBased* operator is used by Hiermann et al. [96] in LNS but it is called a *similar* operator.

- (xix) *SinglePoint*, *TwoPoint*, and *Binary* (C,S) [56] operators remove vertices in the BSS service zones. First, the route is selected at random. In the *SinglePoint* version, the first position in the route is selected at random, and then all the customers between the selected position and the depot are removed. The *Binary* version is similar to *SinglePoint*, but the first position is selected as the middle point in the route. In the *TwoPoint* version, two positions are selected at random, and customers between these positions are removed.
- (xx) *StationBased* (C,S) [56] operator removes all vertices connected to a randomly selected BSS.
- (xxi) *StationVicinity* (C,S) [30] operator removes customers and CS in the radial vicinity of the selected CS.
- (xxii) *Target* (R) [31] operator selects a vertex with the highest contribution to the total distance of its assigned route. Then, the routes that are closest to the selected vertex are removed from the solution.
- (xxiii) *Worst* (C,S) [163] operator removes the solution vertices that have high costs where cost is computed as routing cost between preceding and succeeding vertex. Applied examples are *WorstDistance* and *WorstTime* [28, 32] and the operator of Goeke and Schneider [30] where a more complex objective function is used. Yang and Sun [56] add a feasibility measure to the worst selection method in order to prefer the removal of vertices that improve the feasibility of a solution.
- (xxiv) *WorstChargeUsage* (S) [28] operator removes stations at which BEV arrives with a relatively high charge level.
- (xxv) *Zone* (C,S) [28, 32] operator is based on the removal of vertices in predefined zones. The whole configuration region is divided into smaller zones. The operator at random selects zones and removes customers and stations from it.

Repair operators:

- (i) *Greedy* (C,S) [162] operator iteratively inserts removed vertices at the best possible position in the solution. Modification of *Greedy* operators are (i) *greedy randomized adaptive search (GRASP)* [175] insertion, which instead of selecting the best possible insertion selects random insertion from

a pool of  $k$  best insertions (Goeke and Schneider [30] applied GRASP only on customer vertices); (ii) *GreedyWithNoise* [32] operator adds noise to the cost function in order to diversify the search (similar to RR acceptance strategy); (iii) *AdvanceGreedy* [56] operator adds feasibility value to the cost function; (iv) *SequentialNode* [31] operator inserts vertices based on the removal order followed by restricted candidate list of five best options; and (v) *SequentialPerturbed* [88] operator extends the *SequentialNode* operator by perturbing insertion costs. For the station insertions, Keskin and Çatay [28] detect the first arc in the route that is energy infeasible and then try one of the following three operators to insert the station: *Greedy* inserts the best station between the detected arc and the previous customer; *GreedyWithCopmarison* compares the *Greedy* insertions on the detected arc and on the previous arc and chooses the one that increases the cost at least; and the *Best* operator compares the *Greedy* insertions on arcs between the detected arc and the previous CS and chooses the best one.

- (ii) *Regret* (C,S) [162] computes the  $k$ -regret values for each removed vertex as a difference between the insertion cost in the best route and  $k$ -best route. This so-called *regret* value quantifies the amount of increase in routing cost if a vertex is inserted in  $k$ -best route and not the first best route, meaning that these vertices have a lower number of cost-effective insertions. In EVRP mostly regret-2 and regret-3 heuristics are used. Similar to the greedy case, noise or feasibility values can be added to the cost function: *RegretWithNoise* [32], *AdvanceRegret-k* [56].
- (iii) *SemiParallelConstruction* and *SemiParallelInsertion* (C,S) [31] operators work as the initial construction algorithms and either construct new routes or insert customers in the existing routes.
- (iv) *TimeBased* and *ZoneBased* (C,S) [28, 32] operators insert vertices in the best route positions that minimally increase the overall route time. In *ZoneBased* case, the vertices can be inserted only within a certain zone.

## Nomenclature

2E-EVRP-BSS:	Two-echelon capacitated electric vehicle routing problem with battery swapping stations	BSS-EV-LRP:	Electric vehicles battery swap stations location routing problem
2sEVRP:	Two-stage electric vehicle routing problem	CA:	Cauchy function
AC:	Ant colony algorithm	CGVRP:	Capacitated green vehicle routing problem
AFV:	Alternative fuel vehicle	CRH:	Charge routing heuristic
ALNS:	Adaptive large neighborhood search	CS:	Charging station
BEV:	Battery electric vehicle	CSPP:	Constrained shortest path problem
BKS:	Best-known solution	CVRP:	Capacitated vehicle routing problem
BSS:	Battery swap station	DARP-EV:	Dial-a-ride problem with electric vehicles and battery swapping stations
		DBCA:	Density based clustering algorithm
		DP:	Dynamic programming
		E-FSMFTW:	Electric fleet size and mix vehicle routing problem with time windows and recharging stations
		E-LRP:	Electric location routing problem
		ELRPTWPR:	Electric location routing problem with time windows and partial recharges
		E-SPP:	Energy shortest path problem
		E-TSP:	Electric traveling salesman problem
		E-TSPTW:	Electric traveling salesman problem with time windows
		E-VReP:	Electric vehicle relocation problem
		E-VRP:	Electric vehicle routing problem
		E-VRP-NL:	Electric vehicle routing problem with nonlinear charging functions
		E-VRP-NL-C:	Electric vehicle routing problem with nonlinear charging functions and capacitated CSs
		E-VRP-PP:	Electric vehicle routing problem with public-private recharging strategy
		E-VRPTW:	Electric vehicle routing problem with time windows
		E-VRPTW-FC:	Electric vehicle routing problem with time windows and fast charging

E-VRPTW-SF/MF/SP/MP:	Electric vehicle routing problem with time windows, S (single recharge), M (multiple recharges), F (full recharge), P (partial recharge)	GMO:	Greedy mutation operator
E-VRPTWMF:	Electric vehicle routing problem with time windows and mixed fleet	GRASP:	Greedy randomized adaptive search procedure
E-VRPTWPR:	Electric vehicle routing problem with time windows with partial recharging	GVRP:	Green vehicle routing problem
E-VSP:	Electric vehicle scheduling problem	GVRP-MTPR:	Green vehicle routing problem with multiple technologies and partial recharges
E2EVRP:	Electric two-echelon vehicle routing problem	H <sup>2</sup> E-FTW:	Hybrid heterogeneous electric fleet routing problem with time windows and recharging stations
EFV-CSP:	Electric freight vehicles charge scheduling problem	HDARP:	Heterogeneous dial-a-ride problem
EV:	Electric vehicle	HEV:	Hybrid electric vehicle
EVFSMVRPTW:	Electric vehicle fleet size and mix vehicle routing problem with time windows	HEV-TSP:	Hybrid electric vehicle traveling salesman problem
EVO-VNS:	Evolutionary variable neighborhood search	HEVRP-TDMF:	Heterogeneous electric vehicle routing problem with time-dependent charging costs and a mixed fleet
EVRC:	Electric vehicle route planning with recharging	HGA:	Hybrid genetic algorithm
EVRP-CTVTT:	Electric vehicle routing problem with charging time and variable travel time	HVRP:	Hybrid vehicle routing problem
EVRPRF:	Electric vehicle routing problem with recharging facilities	ICEV:	Internal combustion engine vehicle
F-GVRP:	Fuel efficient green vehicle routing problem	ILS:	Iterated local search
FRD:	Facility-related demand	ITS:	Iterated tabu search
FRVCP:	Fixed route vehicle charging problem	LDM:	Longitudinal dynamics model
FSM-VRP/MFVRP:	Fleet size and mix vehicle routing problem	LRP:	Location routing problem
FSMVRPTW:	Fleet size and mix vehicle routing problem with time windows	LRPIF:	Location routing problem with intraroute facilities
FSMVRPTW-EV:	Fleet size and mix vehicle routing problem with time windows for electric vehicles	LS:	Local search
GA:	Genetic algorithm	MBFM:	Mixed bus fleet management problem
GHG:	Greenhouse gas	MCWS:	Modified Clark and Wright savings method
GMFVRP-PRTW:	Green mixed fleet vehicle routing problem with partial battery recharging and time windows	MDEVLRTW (BS/PR/BSPR):	Multidepot electric vehicle location routing problem with time windows (battery swapping/partial recharging)
		MDVRP:	Multidepot vehicle routing problem
		MDVRPI:	Multidepot vehicle routing problem with inter-depot routes



MGVRP:	Mixed fleet green vehicle routing problem
MILP:	Mixed integer linear program
MIP:	Mixed integer program
MLR:	Multiple linear regression
MTFSP:	Mixed taxi fleet scheduling problem
NNH:	Nearest neighbor heuristic
PHEV:	Plug-in hybrid electric vehicles
PHEVRPTW:	Plug-in hybrid electric vehicle routing problem with time windows
PIH:	Push forward heuristics
PRP:	Pollution routing problem
PSO:	Particle swarm optimization
RCSP:	Resource constrained shortest path algorithm
RELRPTWPR:	Robust electric location routing problem with time windows and partial recharging
RR:	Record-to-record procedure
SA:	Simulated annealing
SoC:	State of charge
TD-VRP:	Time-dependent vehicle routing problem
TS:	Tabu search
TSP:	Traveling salesman problem
TSPTW:	Traveling salesman problem with time windows
TVa-PSOGMO:	Particle swarm optimization with greedy mutation operator and time-varying acceleration coefficient
VND:	Variable neighborhood decent
VNS:	Variable neighborhood search
VRP:	Vehicle routing problem
VRP-HFCC:	Vehicle routing problem with a mixed fleet of conventional and heterogenous electric vehicles including new constraints
VRP-MFHEV:	Vehicle routing problem with mixed fleet of conventional and heterogenous electric vehicles

VRPIRF:	Vehicle routing problem with intermediate replenishment facilities
VRPIS:	Vehicle routing problem with intermediate stops
VRPPD:	Vehicle routing problem with pickup and delivery
VRPTW:	Vehicle routing problem with time windows.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## Acknowledgments

The research has been supported by the European Regional Development Fund under grant KK.01.2.1.01.0120 and has also been partially supported by other two sources: Croatian Science Foundation under project IP-2018-01-8323 and project KK.01.1.1.01.0009 (DATACROSS) within the activities of the Centre of Research Excellence for Data Science and Cooperative Systems supported by the Ministry of Science and Education of the Republic of Croatia.

## References

- [1] G. Laporte, "Fifty years of vehicle routing," *Transportation Science*, vol. 43, no. 4, pp. 408–416, 2009.
- [2] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959.
- [3] P. Toth and D. Vigo, Eds., *The Vehicle Routing Problem*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.
- [4] T. Vidal, T. G. Crainic, M. Gendreau, and C. Prins, "Heuristics for multi-attribute vehicle routing problems: a survey and synthesis," *European Journal of Operational Research*, vol. 231, no. 1, pp. 1–21, 2013.
- [5] D. Pecin, A. Pessoa, M. Poggi, and E. Uchoa, "Improved branch-cut-and-price for capacitated vehicle routing," *Mathematical Programming Computation*, vol. 9, no. 1, pp. 61–100, 2017.
- [6] R. Baldacci, A. Mingozzi, and R. Roberti, "Recent exact algorithms for solving the vehicle routing problem under capacity and time window constraints," *European Journal of Operational Research*, vol. 218, no. 1, pp. 1–6, 2012.
- [7] European Commission, "Effort sharing: Member states emission targets," 2009, <https://ec.europa.eu/clima/policies/effort-en>.
- [8] WEEA, "Greenhouse gas data - Emissions share by sector in EU28," 2015, <http://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer>.
- [9] A. Sbihi and R. W. Eglese, "Combinatorial optimization and green logistics," *4OR*, vol. 5, no. 2, pp. 99–116, 2007.
- [10] B. A. Davis and M. A. Figliozzi, "A methodology to evaluate the competitiveness of electric delivery trucks," *Transportation Research Part E: Logistics and Transportation Review*, vol. 49, no. 1, pp. 8–23, 2013.

- [11] M. Schiffer, S. Stütz, and G. Walther, "Are ECVs breaking even?: Competitiveness of electric commercial vehicles in retail logistics," Technical Report G-2017-47, RWTH Aachen University, Aachen, Germany, 2017.
- [12] W. Feng and M. Figliozzi, "An economic and technological analysis of the key factors affecting the competitiveness of electric commercial vehicles: A case study from the USA market," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 135–145, 2013.
- [13] K. Young, C. Wang, L. Y. Wang, and K. Strunz, *Electric Vehicle Battery Technologies*, Springer, 2013.
- [14] J.-Q. Li, "Transit bus scheduling with limited energy," *Transportation Science*, vol. 48, no. 4, pp. 521–539, 2014.
- [15] J. Martínez-Lao, F. G. Montoya, M. G. Montoya, and F. Manzano-Agugliaro, "Electric vehicles in Spain: An overview of charging systems," *Renewable & Sustainable Energy Reviews*, vol. 77, pp. 970–983, 2017.
- [16] A. Juan, C. Mendez, J. Faulin, J. de Armas, and S. Grasman, "Electric vehicles in logistics and transportation: a survey on emerging environmental, strategic, and operational challenges," *Energies*, vol. 9, no. 2, p. 86, 2016.
- [17] D. Margaritis, A. Anagnostopoulou, A. Tromaras, and M. Boile, "Electric commercial vehicles: practical perspectives and future research directions," *Research in Transportation Business & Management*, vol. 18, pp. 4–10, 2016.
- [18] J.-A. Montoya, *Electric Vehicle Routing Problems: models and solution approaches [PhD Thesis]*, 2016.
- [19] S. Pelletier, O. Jabali, and G. Laporte, "50th anniversary invited article—goods distribution with electric vehicles: review and research perspectives," *Transportation Science*, vol. 50, no. 1, pp. 3–22, 2016.
- [20] M. Gendreau and C. D. Tarantilis, "Solving large-scale vehicle routing problems with time windows: The state-of-the-art," Technical Report 04, 2010, <https://www.cirrelt.ca/Documents-Travail/CIRRELT-2010-04.pdf>.
- [21] F. Gonçalves, S. Cardoso, and S. Relvas, "Optimization of distribution network using electric vehicles: A VRP problem," Technical Report, CEG-IST, Technical university of Lisbon, 2011.
- [22] R. G. Conrad and M. Figliozzi, "The recharging vehicle routing problem," in *Proceedings of the 61st Annual Conference and Expo of the Institute of Industrial Engineers*, May 2011.
- [23] S. Erdoğan and E. Miller-Hooks, "A green vehicle routing problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 1, pp. 100–114, 2012.
- [24] A. Omidvar and R. Tavakkoli-Moghaddam, "Sustainable vehicle routing: Strategies for congestion management and refueling scheduling," in *Proceedings of the IEEE International Energy Conference (ENERGYCON '12)*, pp. 1089–1094, Florence, Italy, September 2012.
- [25] M. Schneider, A. Stenger, and D. Goeke, "The electric vehicle-routing problem with time windows and recharging stations," *Transportation Science*, vol. 48, no. 4, pp. 500–520, 2014.
- [26] O. Bräysy and M. Gendreau, "Vehicle routing problem with time windows, part I: route construction and local search algorithms," *Transportation Science*, vol. 39, no. 1, pp. 104–118, 2005.
- [27] M. Schiffer, S. Stütz, and G. Walther, "Are ECVs breaking even? - Competitiveness of electric commercial vehicles in medium-duty logistics networks," Technical Report Working Paper OM-02/2016, RWTH Aachen University, Aachen, Germany, 2016.
- [28] M. Keskin and B. Çatay, "Partial recharge strategies for the electric vehicle routing problem with time windows," *Transportation Research Part C: Emerging Technologies*, vol. 65, pp. 111–127, 2016.
- [29] M. Bruglieri, F. Pezzella, O. Pisacane, and S. Suraci, "A mathematical heuristic for the electric vehicle routing problem with time windows," <https://arxiv.org/abs/1506.00211>, 2015.
- [30] D. Goeke and M. Schneider, "Routing a mixed fleet of electric and conventional vehicles," *European Journal of Operational Research*, vol. 245, no. 1, pp. 81–99, 2015.
- [31] G. Hiermann, J. Puchinger, S. Ropke, and R. F. Hartl, "The electric fleet size and mix vehicle routing problem with time windows and recharging stations," *European Journal of Operational Research*, vol. 252, no. 3, pp. 995–1018, 2016.
- [32] E. Demir, T. Bektaş, and G. Laporte, "An adaptive large neighborhood search heuristic for the pollution-routing problem," *European Journal of Operational Research*, vol. 223, no. 2, pp. 346–359, 2012.
- [33] G. Macrina, L. D. P. Pugliese, F. Guerriero, and G. Laporte, "The green mixed fleet vehicle routing problem with partial battery recharging and time windows," *Computers & Operations Research*, vol. 101, pp. 183–199, 2019.
- [34] J. Barco, A. Guerra, L. Muñoz, and N. Quijano, "Optimal routing and scheduling of charge for electric vehicles: Case study," *CoRR*, 2013 (Basque), <https://arxiv.org/abs/1310.0145>.
- [35] H. Preis, S. Frank, and K. Nachtigall, "Energy-optimized routing of electric vehicles in urban delivery systems," in *Operations Research Proceedings 2012*, S. Helber, M. Breitner, D. Rösch et al., Eds., pp. 583–588, Springer International Publishing, 2014.
- [36] S. Zhang, Y. Gajpal, S. S. Appadoo, and M. M. S. Abdulkader, "Electric vehicle routing problem with recharging stations for minimizing energy consumption," *International Journal of Production Economics*, vol. 203, pp. 404–413, 2018.
- [37] E. Demir, T. Bektaş, and G. Laporte, "The bi-objective pollution-routing problem," *European Journal of Operational Research*, vol. 232, no. 3, pp. 464–478, 2014.
- [38] G. Poonthalir and R. Nadarajan, "A fuel efficient green vehicle routing problem with varying speed constraint (F-GVRP)," *Expert Systems with Applications*, vol. 100, pp. 131–144, 2018.
- [39] S. S. Amiri, S. Jadid, and H. Saboori, "Multi-objective optimum charging management of electric vehicles through battery swapping stations," *Energy*, vol. 165, pp. 549–562, 2018.
- [40] T. Bektaş and G. Laporte, "The pollution-routing problem," *Transportation Research Part B: Methodological*, vol. 45, no. 8, pp. 1232–1250, 2011.
- [41] T. Abdallah, *The plug-in hybrid electric vehicle routing problem with time windows [Master Thesis]*, 2013, <http://hdl.handle.net/10012/7582>.
- [42] J. Barco, A. Guerra, L. Muñoz, and N. Quijano, "Optimal routing and scheduling of charge for electric vehicles: a case study," *Mathematical Problems in Engineering*, vol. 2017, Article ID 8509783, 16 pages, 2017.
- [43] J. van Duin, L. Tavasszy, and H. Quak, "Towards e(lectric)-urban freight: First promising steps in the electronic vehicle revolution," *European Transport - Trasporti Europei*, no. 54, pp. 1–19, 2013.
- [44] J. D. Adler and P. B. Mirchandani, "Online routing and battery reservations for electric vehicles with swappable batteries," *Transportation Research Part B: Methodological*, vol. 70, pp. 285–302, 2014.

- [45] F. Alesiani and N. Maslekar, "Optimization of Charging Stops for Fleet of Electric Vehicles: A Genetic Approach," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 3, pp. 10–21, 2014.
- [46] Á. Felipe, M. T. Ortuño, G. Righini, and G. Tirado, "A heuristic approach for the green vehicle routing problem with multiple technologies and partial recharges," *Transportation Research Part E: Logistics and Transportation Review*, vol. 71, pp. 111–128, 2014.
- [47] O. Sassi, W. R. Cherif, and A. Oulamara, "Vehicle routing problem with mixed fleet of conventional and heterogenous electric vehicles and time dependent charging costs," *Working Paper Or Preprint*, October 2014, <https://hal.archives-ouvertes.fr/hal-01083966>.
- [48] O. Sassi, W. R. Cherif-Khettaf, and A. Oulamara, "Iterated tabu search for the mix fleet vehicle routing problem with heterogenous electric vehicles," in *Modelling, Computation and Optimization in Information Systems and Management Sciences*, H. A. Le Thi, T. Pham Dinh, and N. T. Nguyen, Eds., pp. 57–68, Springer International Publishing, 2015.
- [49] O. Sassi, W. Ramdane Cherif-Khettaf, and A. Oulamara, "Multi-start iterated local search for the mixed fleet vehicle routing problem with heterogenous electric vehicles," in *Evolutionary Computation in Combinatorial Optimization*, vol. 9026, pp. 138–149, Springer, 2015.
- [50] T. Zündorf, *Electric vehicle routing with realistic recharging models [Master Thesis]*, Karlsruhe Institute of Technology, 2014.
- [51] M. Bruglieri, F. Pezzella, O. Pisacane, and S. Suraci, "A variable neighborhood search branching for the electric vehicle routing problem with time windows," *Electronic Notes in Discrete Mathematics*, vol. 47, pp. 221–228, 2015.
- [52] P. Lebeau, C. De Cauwer, J. Van Mierlo, C. Macharis, W. Verbeke, and T. Coosemans, "Conventional, hybrid, or electric vehicles: which technology for an urban distribution centre?" *The Scientific World Journal*, vol. 2015, Article ID 302867, 11 pages, 2015.
- [53] N. M. Moghaddam, *The partially rechargeable electric vehicle routing problem with time windows and capacitated charging stations [Master Thesis]*, Clemson University, Clemson, SC, USA, 2015.
- [54] S. Pourazarm, C. G. Cassandras, and A. Malikopoulos, "Optimal routing of electric vehicles in networks with charging nodes: a dynamic programming approach," in *Proceedings of the IEEE International Electric Vehicle Conference (IEVC '14)*, pp. 1–7, Florence, Italy, 2015.
- [55] M. Schneider, A. Stenger, and J. Hof, "An adaptive VNS algorithm for vehicle routing problems with intermediate stops," *OR Spectrum*, vol. 37, no. 2, pp. 353–387, 2015.
- [56] J. Yang and H. Sun, "Battery swap station location-routing problem with capacitated electric vehicles," *Computers & Operations Research*, vol. 55, pp. 217–232, 2015.
- [57] G. Desaulniers, F. Errico, S. Irnich, and M. Schneider, "Exact algorithms for electric vehicle-routing problems with time windows," *Operations Research*, vol. 64, no. 6, pp. 1388–1405, 2016.
- [58] C. Doppstadt, A. Koberstein, and D. Vigo, "The hybrid electric vehicle – traveling salesman problem," *European Journal of Operational Research*, vol. 253, no. 3, pp. 825–842, 2016.
- [59] Ç. Koç and I. Karaoglan, "The green vehicle routing problem: A heuristic based exact solution approach," *Applied Soft Computing*, vol. 39, pp. 154–164, 2016.
- [60] J. Lin, W. Zhou, and O. Wolfson, "Electric vehicle routing problem," *Transportation Research Procedia*, vol. 12, pp. 508–521, 2016.
- [61] K. Masliakova, *Optimal routing and charging procedures for electric buses [Master Thesis]*, The Arctic University of Norway, Narvik, Norway, 2016.
- [62] S. H. Mirmohammadi, E. B. Tirkolaee, A. Goli, and S. Dehnavi-Arani, "The periodic green vehicle routing problem with considering of time-dependent urban traffic and time windows," *International Journal of Optimization in Civil Engineering*, vol. 7, pp. 143–156, 2016.
- [63] A. Montoya, C. Guéreta, and J. E. Mendoza, "A multi-space sampling heuristic for the green vehicle routing problem," *Transportation Research Part C: Emerging Technologies*, vol. 70, pp. 113–128, 2016.
- [64] R. Roberti and M. Wen, "The electric traveling salesman problem with time windows," *Transportation Research Part E: Logistics and Transportation Review*, vol. 89, pp. 32–52, 2016.
- [65] M. Wen, E. Linde, S. Ropke, P. Mirchandani, and A. Larsen, "An adaptive large neighborhood search heuristic for the electric vehicle scheduling problem," *Computers & Operations Research*, vol. 76, pp. 73–83, 2016.
- [66] J. Andelmin and E. Bartolini, "An exact algorithm for the green vehicle routing problem," *Transportation Science*, vol. 51, no. 4, pp. 1288–1303, 2017.
- [67] B. Catay and M. Keskin, "The impact of quick charging stations on the route planning of Electric Vehicles," in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC '17)*, pp. 152–157, Heraklion, Greece, July 2017.
- [68] A. Froger, J. E. Mendoza, O. Jabali, and G. Laporte, "A math-heuristic for the electric vehicle routing problem with capacitated charging stations," *Research Report, Centre interuniversitaire de recherche sur les reseaux d'entreprise, la logistique et le transport (CIRRELT)*, June 2017, <https://www.cirrelt.ca/DocumentsTravail/CIRRELT-2017-31.pdf>.
- [69] J. Hof, M. Schneider, and D. Goeke, "Solving the battery swap station location-routing problem with capacitated electric vehicles using an AVNS algorithm for vehicle-routing problems with intermediate stops," *Transportation Research Part B: Methodological*, vol. 97, pp. 102–112, 2017.
- [70] V. Leggieri and M. Haouari, "A practical solution approach for the green vehicle routing problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 104, pp. 97–112, 2017.
- [71] S. Mancini, "The hybrid vehicle routing problem," *Transportation Research Part C: Emerging Technologies*, vol. 78, pp. 1–12, 2017.
- [72] A. Montoya, C. Guéret, J. E. Mendoza, and J. G. Villegas, "The electric vehicle routing problem with nonlinear charging function," *Transportation Research Part B: Methodological*, vol. 103, pp. 87–110, 2017.
- [73] M. Schiffer and G. Walther, "The electric location routing problem with time windows and partial recharging," *European Journal of Operational Research*, vol. 260, no. 3, pp. 995–1013, 2017.
- [74] S. Shao, W. Guan, B. Ran, Z. He, and J. Bi, "Electric vehicle routing problem with charging time and variable travel time," *Mathematical Problems in Engineering*, vol. 2017, Article ID 5098183, 13 pages, 2017.
- [75] T. M. Sweda, I. S. Dolinskaya, and D. Klabjan, "Adaptive routing and recharging policies for electric vehicles," *Transportation Science*, vol. 51, no. 4, pp. 1326–1348, 2017.



- [76] F. Y. Vincent, A. A. N. P. Redi, Y. A. Hidayat, and O. J. Wibowo, "A simulated annealing heuristic for the hybrid vehicle routing problem," *Applied Soft Computing*, vol. 53, pp. 119–132, 2017.
- [77] M. Bruglieri, F. Pezzella, and O. Pisacane, "A two-phase optimization method for a multiobjective vehicle relocation problem in electric carsharing systems," *Journal of Combinatorial Optimization*, vol. 36, no. 1, pp. 162–193, 2018.
- [78] H. Joo and Y. Lim, "Ant colony optimized routing strategy for electric vehicles," *Journal of Advanced Transportation*, vol. 2018, Article ID 5741982, 9 pages, 2018.
- [79] M. Keskin and B. Çatay, "A matheuristic method for the electric vehicle routing problem with time windows and fast chargers," *Computers & Operations Research*, vol. 100, pp. 172–188, 2018.
- [80] M. Keskin, G. Laporte, and B. Çatay, "Electric vehicle routing problem with time-dependent waiting times at recharging stations," in *Proceedings of the 7th International Workshop on Freight Transportation and Logistics, (ODYSSSEUS '18)*, vol. 06, 2018.
- [81] N. D. Kullman, J. C. Goodson, and J. E. Mendoza, "Dynamic electric vehicle routing: heuristics and dual bounds," Technical Report, November 2018.
- [82] L. Li, H. K. Lo, F. Xiao, and X. Cen, "Mixed bus fleet management strategy for minimizing overall and emissions external costs," *Transportation Research Part D: Transport and Environment*, vol. 60, pp. 104–118, 2018.
- [83] C.-C. Lu, S. Yan, and Y.-W. Huang, "Optimal scheduling of a taxi fleet with mixed electric and gasoline vehicles to service advance reservations," *Transportation Research Part C: Emerging Technologies*, vol. 93, pp. 479–500, 2018.
- [84] M. A. Masmoudi, M. Hosny, E. Demir, K. N. Genikomsakis, and N. Cheikhrouhou, "The dial-a-ride problem with electric vehicles and battery swapping stations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 392–420, 2018.
- [85] K. N. Genikomsakis and G. Mitrentsis, "A computationally efficient simulation model for estimating energy consumption of electric vehicles in the context of route planning applications," *Transportation Research Part D: Transport and Environment*, vol. 50, pp. 98–118, 2017.
- [86] J. C. Paz, M. Granada-Echeverri, and J. W. Escobar, "The multi-depot electric vehicle location routing problem with time windows," *International Journal of Industrial Engineering Computations*, vol. 9, no. 1, pp. 123–136, 2018.
- [87] S. Pelletier, O. Jabali, and G. Laporte, "Charge scheduling for electric freight vehicles," *Transportation Research Part B: Methodological*, vol. 115, pp. 246–269, 2018.
- [88] M. Schiffer and G. Walther, "An adaptive large neighborhood search for the location-routing problem with intra-route facilities," *Transportation Science*, vol. 52, no. 2, pp. 331–352, 2018.
- [89] M. Schiffer and G. Walther, "Strategic planning of electric logistics fleet networks: A robust location-routing approach," *Omega*, vol. 80, pp. 31–42, 2018.
- [90] S. Shao, W. Guan, and J. Bi, "Electric vehicle-routing problem with charging demands and energy consumption," *IET Intelligent Transport Systems*, vol. 12, no. 3, pp. 202–212, 2018.
- [91] Y. Wang, J. Bi, W. Guan, and X. Zhao, "Optimising route choices for the travelling and charging of battery electric vehicles by considering multiple objectives," *Transportation Research Part D: Transport and Environment*, vol. 64, pp. 246–261, 2018.
- [92] R. Basso, B. Kulcsár, B. Egardt, P. Lindroth, and I. Sanchez-Diaz, "Energy consumption estimation integrated into the electric vehicle routing problem," *Transportation Research Part D: Transport and Environment*, vol. 69, pp. 141–167, 2019.
- [93] U. Breunig, R. Baldacci, R. Hartl, and T. Vidal, "The electric two-echelon vehicle routing problem," *Computers & Operations Research*, vol. 103, pp. 198–210, 2019.
- [94] M. Bruglieri, S. Mancini, F. Pezzella, and O. Pisacane, "A path-based solution approach for the green vehicle routing problem," *Computers & Operations Research*, vol. 103, pp. 109–122, 2019.
- [95] A. Froger, J. E. Mendoza, O. Jabali, and G. Laporte, "Improved formulations and algorithmic components for the electric vehicle routing problem with nonlinear charging functions," *Computers & Operations Research*, vol. 104, pp. 256–294, 2019.
- [96] G. Hiermann, R. F. Hartl, J. Puchinger, and T. Vidal, "Routing a mix of conventional, plug-in hybrid, and electric vehicles," *European Journal of Operational Research*, vol. 272, no. 1, pp. 235–248, 2019.
- [97] W. Jie, J. Yang, M. Zhang, and Y. Huang, "The two-echelon capacitated electric vehicle routing problem with battery swapping stations: formulation and efficient methodology," *European Journal of Operational Research*, vol. 272, no. 3, pp. 879–904, 2019.
- [98] I. Koyuncu and M. Yavuz, "Duplicating nodes or arcs in green vehicle routing: A computational comparison of two formulations," *Transportation Research Part E: Logistics and Transportation Review*, vol. 122, pp. 605–623, 2019.
- [99] G. Macrina, G. Laporte, F. Guerriero, and L. Di Puglia Pugliese, "An energy-efficient green-vehicle routing problem with mixed vehicle fleet, partial battery recharging and time windows," *European Journal of Operational Research*, vol. 276, no. 3, pp. 971–982, 2019.
- [100] N. M. Normasari, V. F. Yu, C. Bachtiyar, and . Sukoyo, "A simulated annealing heuristic for the capacitated green vehicle routing problem," *Mathematical Problems in Engineering*, vol. 2019, Article ID 2358258, 18 pages, 2019.
- [101] E. A. Grunditz and T. Thiringer, "Performance analysis of current BEVs based on a comprehensive review of specifications," *IEEE Transactions on Transportation Electrification*, vol. 2, no. 3, pp. 270–289, 2016.
- [102] E. den Boer, S. Aarnink, F. Kleiner, and J. Pagenkopf, "An overview of state-of-the-art technologies and their potential," Technical Report, 2013, [https://www.theicct.org/sites/default/files/publications/CE\\_Delft\\_4841\\_Zero\\_emissions\\_trucks\\_Def.pdf](https://www.theicct.org/sites/default/files/publications/CE_Delft_4841_Zero_emissions_trucks_Def.pdf).
- [103] C. Liao, S. Lu, and Z. M. Shen, "The electric vehicle touring problem," *Transportation Research Part B: Methodological*, vol. 86, pp. 163–180, 2016.
- [104] S. Pelletier, O. Jabali, G. Laporte, and M. Veneroni, "Battery degradation and behaviour for electric vehicles: Review and numerical analyses of several models," *Transportation Research Part B: Methodological*, vol. 103, pp. 158–187, 2017.
- [105] DPDHL, "Electric vehicles in inner city distribution traffic," 2014, [https://www.haw-hamburg.de/fileadmin/user\\_upload/FakLS/07Forschung/FTZ-ALS/Veranstaltungen/Fuelling\\_the\\_Climate/Lohmeyer\\_FTC2014\\_VOE.pdf](https://www.haw-hamburg.de/fileadmin/user_upload/FakLS/07Forschung/FTZ-ALS/Veranstaltungen/Fuelling_the_Climate/Lohmeyer_FTC2014_VOE.pdf).
- [106] UPS, "Ups to rollout fleet of electric vehicles in california," 2013, <https://www.pressroom.ups.com/pressroom/news-assets/new-assets-subpage-landing.page?ConceptType=PressReleases>.
- [107] J. Asamer, A. Graser, B. Heilmann, and M. Ruthmair, "Sensitivity analysis for energy demand estimation of electric vehicles," *Transportation Research Part D: Transport and Environment*, vol. 46, pp. 182–199, 2016.

- [108] A. Artmeier, J. Haselmayr, M. Leucker, and M. Sachenbacher, "The shortest path problem revisited: optimal routing for electric vehicles," in *KI 2010: Advances in Artificial Intelligence*, R. Dillmann, J. Beyerer, U. D. Hanebeck, and T. Schultz, Eds., pp. 309–316, Springer Berlin Heidelberg, 2010.
- [109] J. Restrepo, J. Rosero, and S. Tellez, "Performance testing of electric vehicles on operating conditions in Bogota: DC, Colombia," in *Proceedings of the IEEE PES Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA '14)*, pp. 1–8, Medellin, Colombia, September 2014.
- [110] A. Brooker, J. Gonder, L. Wang, E. Wood, S. Lopp, and L. F. Ramroth, "A model to estimate vehicle efficiency, cost and performance," SAE Technical Papers 04, 2015.
- [111] M. Barth and K. Boriboonsomsin, "Energy and emissions impacts of a freeway-based dynamic eco-driving system," *Transportation Research Part D: Transport and Environment*, vol. 14, no. 6, pp. 400–410, 2009.
- [112] R. Bellman, "On a routing problem," *Quarterly of Applied Mathematics*, vol. 16, no. 1, pp. 87–90, 1958.
- [113] C. Fiori, K. Ahn, and H. A. Rakha, "Optimum routing of battery electric vehicles: Insights using empirical data and microsimulation," *Transportation Research Part D: Transport and Environment*, vol. 64, pp. 262–272, 2018.
- [114] C. Fiori, K. Ahn, and H. A. Rakha, "Power-based electric vehicle energy consumption model: Model development and validation," *Applied Energy*, vol. 168, pp. 257–268, 2016.
- [115] C. De Cauwer, W. Verbeke, T. Coosemans, S. Faid, and J. Van Mierlo, "A data-driven method for energy consumption prediction and energy-efficient routing of electric vehicles in real-world conditions," *Energies*, vol. 10, no. 5, 2017.
- [116] T. Erdelić, S. Vrbancic, and L. Rozic, "A model of speed profiles for urban road networks using G-means clustering," in *Proceedings of the 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO '15)*, pp. 1081–1086, Opatija, Croatia, May 2015.
- [117] C. De Cauwer, J. Van Mierlo, and T. Coosemans, "Energy consumption prediction for electric vehicles based on real-world data," *Energies*, vol. 8, no. 8, pp. 8573–8593, 2015.
- [118] X. Wu, D. Freese, A. Cabrera, and W. A. Kitch, "Electric vehicles' energy consumption measurement and estimation," *Transportation Research Part D: Transport and Environment*, vol. 34, pp. 52–67, 2015.
- [119] C. Fiori and V. Marzano, "Modelling energy consumption of electric freight vehicles in urban pickup/delivery operations: analysis and estimation on a real-world dataset," *Transportation Research Part D: Transport and Environment*, vol. 65, pp. 658–673, 2018.
- [120] H. Ortega-Arranz, D. R. Llanos, and A. Gonzalez-Escribano, *The Shortest-Path Problem: Analysis and Comparison of Methods*, Synthesis Lectures on Theoretical Computer Science, Morgan & Claypool Publishers, 2014.
- [121] J. Eisner, S. Funke, and S. Storandt, "Optimal route planning for electric vehicles in large networks," in *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAI '11)*, pp. 1108–1113, AAAI Press, 2011.
- [122] D. B. Johnson, "Efficient algorithms for shortest paths in sparse networks," *Journal of the ACM*, vol. 24, no. 1, pp. 1–13, 1977.
- [123] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 10, no. 1, pp. 269–271, 1959.
- [124] S. Storandt, "Quick and energy-efficient routes: Computing constrained shortest paths for electric vehicles," in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Computational Transportation Science (IWCTS '12)*, pp. 20–25, ACM, New York, NY, USA, November 2012.
- [125] T. M. Sweda and D. Klabjan, "Finding minimum-cost paths for electric vehicles," in *Proceedings of the IEEE International Electric Vehicle Conference (IEVC '12)*, pp. 1–4, March 2012.
- [126] M. Strehler, S. Merting, and C. Schwan, "Energy-efficient shortest routes for electric and hybrid vehicles," *Transportation Research Part B: Methodological*, vol. 103, pp. 111–135, 2017.
- [127] Y. Zhang, B. Aliya, Y. Zhou et al., "Shortest feasible paths with partial charging for battery-powered electric vehicles in smart cities," *Pervasive and Mobile Computing*, vol. 50, pp. 82–93, 2018.
- [128] B. Golden, A. Assad, L. Levy, and F. Gheysens, "The fleet size and mix vehicle routing problem," *Computers & Operations Research*, vol. 110, no. 1, pp. 49–66, 1984.
- [129] M. Yilmaz and P. T. Krein, "Review of battery charger topologies, charging power levels, and infrastructure for plug-in electric and hybrid vehicles," *IEEE Transactions on Power Electronics*, vol. 28, no. 5, pp. 2151–2169, 2013.
- [130] C. Prodhon and C. Prins, "A survey of recent research on location-routing problems," *European Journal of Operational Research*, vol. 238, no. 1, pp. 1–17, 2014.
- [131] Z. Sun, W. Gao, B. Li, and L. Wang, "Locating charging stations for electric vehicles," *Transport Policy*, 2018.
- [132] T. Erdelić, M. Ravlic, and T. Caric, "Travel time prediction using speed profiles for road network of Croatia," in *Proceedings of the International Symposium ELMAR*, pp. 97–100, Zadar, Croatia, September 2016.
- [133] J. F. Ehmke, S. Meisel, and D. C. Mattfeld, "Floating car based travel times for city logistics," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 338–352, 2012.
- [134] A. R. Güner, A. Murat, and R. B. Chinnam, "Dynamic routing under recurrent and non-recurrent congestion using real-time ITS information," *Computers & Operations Research*, vol. 39, no. 2, pp. 353–373, 2012.
- [135] J. F. Ehmke, A. Steinert, and D. C. Mattfeld, "Advanced routing for city logistics service providers based on time-dependent travel times," *Journal of Computational Science*, vol. 3, no. 4, pp. 193–205, 2012.
- [136] L. Rožić, J. Fosin, and T. Carić, "Solving the time dependent vehicle routing problem using real-world speed profiles," in *Proceedings of the Central European Conference on Information and Intelligent Systems*, 2015.
- [137] R. Álvarez Fernández, "A more realistic approach to electric vehicle contribution to greenhouse gas emissions in the city," *Journal of Cleaner Production*, vol. 172, pp. 949–959, 2018.
- [138] M. Yavuz, "An iterated beam search algorithm for the green vehicle routing problem," *Networks*, vol. 69, no. 3, pp. 317–328, 2017.
- [139] K. Karagül, Y. Sahin, E. Aydemir, and A. Oral, "A simulated annealing algorithm based solution method for a green vehicle routing problem with fuel consumption," in *Lean and Green Supply Chain Management*, vol. 273 of *International Series in Operations Research & Management Science*, pp. 161–187, Springer International Publishing, 2019.
- [140] M. Schneider, B. Sand, and A. Stenger, "A note on the time travel approach for handling time windows in vehicle routing problems," *Computers & Operations Research*, vol. 400, no. 10, pp. 2564–2568, 2013.



- [141] G. Clarke and J. W. Wright, "Scheduling of vehicles from a central depot to a number of delivery points," *Operations Research*, vol. 12, no. 4, pp. 568–581, 1964.
- [142] B. Gillett and L. Miller, "A heuristic algorithm for the vehicle dispatch problem," *Operations Research*, vol. 22, no. 2, pp. 340–350, 1974.
- [143] J. Beasley, "Route first-cluster second methods for vehicle routing," *Omega*, vol. 11, no. 4, pp. 403–408, 1983.
- [144] M. L. Fisher and R. Jaikumar, "A generalized assignment heuristic for vehicle routing," *Networks*, vol. 11, no. 2, pp. 109–124, 1981.
- [145] M. M. Solomon, "Algorithms for the vehicle routing and scheduling problems with time window constraints," *Operations Research*, vol. 35, no. 2, pp. 254–265, 1987.
- [146] S. Lin, "Computer solutions of the traveling salesman problem," *Bell Labs Technical Journal*, vol. 44, pp. 2245–2269, 1965.
- [147] J.-Y. Potvin and J.-M. Rousseau, "Exchange heuristic for routing problems with time windows," *Journal of the Operational Research Society*, vol. 46, no. 12, pp. 1433–1446, 1995.
- [148] K. Braekers, A. Caris, and G. K. Janssens, "Exact and meta-heuristic approach for a general heterogeneous dial-a-ride problem with multiple depots," *Transportation Research Part B: Methodological*, vol. 67, pp. 166–186, 2014.
- [149] I. Or, "Traveling salesman-type combinatorial problems and their relation to the logistics of regional blood banking," 1976. UMI order no. 77-10076.
- [150] M. W. P. Savelsbergh, "The vehicle routing problem with time windows: minimizing route duration," *ORSA Journal on Computing*, vol. 4, no. 2, pp. 146–154, 1992.
- [151] I. H. Osman, "Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem," *Annals of Operations Research*, vol. 41, no. 4, pp. 421–451, 1993.
- [152] P. Hansen and N. Mladenović, "First vs. best improvement: An empirical study," *Discrete Applied Mathematics*, vol. 154, no. 5, pp. 802–817, 2006.
- [153] P. Shaw, "Using constraint programming and local search methods to solve vehicle routing problems," in *Principles and Practice of Constraint Programming — CP98*, M. Maher and J.-F. Puget, Eds., vol. 1520, pp. 417–431, Springer Berlin Heidelberg, 1998.
- [154] J. Christiaens and G. Vanden Berghe, "Slack induction by string removals for vehicle routing problems," Technical Report, KU Leuven, Department of Computer Science, CODeS & imec, 2018, [https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1988666&context=L&vid=Lirias&search\\_scope=Lirias&tab=default\\_tab&lang=en\\_US](https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1988666&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US).
- [155] T. Vidal, T. G. Crainic, M. Gendreau, and C. Prins, "Timing problems and algorithms: Time decisions for sequences of activities," *Networks*, vol. 65, no. 2, pp. 102–128, 2015.
- [156] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [157] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [158] F. Glover, "Tabu search—part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.
- [159] F. Glover, "Tabu search—part II," *ORSA Journal on Computing*, vol. 20, no. 1, pp. 4–32, 1990.
- [160] N. Mladenović and P. Hansen, "Variable neighborhood search," *Computers & Operations Research*, vol. 24, no. 11, pp. 1097–1100, 1997.
- [161] H. R. Lourenço, O. C. Martin, and T. Stützle, *Iterated Local Search: Framework and Applications*, Springer, Boston, MA, USA, 2010.
- [162] D. Pisinger and S. Ropke, "A general heuristic for vehicle routing problems," *Computers & Operations Research*, vol. 34, no. 8, pp. 2403–2435, 2007.
- [163] S. Ropke and D. Pisinger, "An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows," *Transportation Science*, vol. 40, no. 4, pp. 455–472, 2006.
- [164] M. Tiwari, S. Kumar, S. Kumar, Prakash, and R. Shankar, "Solving part-type selection and operation allocation problems in an FMS: an approach using constraints-based fast simulated annealing algorithm," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 36, no. 6, pp. 1170–1184, 2006.
- [165] G. Dueck, "New optimization heuristics: the great deluge algorithm and the record-to-record travel," *Journal of Computational Physics*, vol. 104, no. 1, pp. 86–92, 1993.
- [166] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*, MIT Press, Cambridge, MA, USA, 1992.
- [167] F. Glover, "Heuristics for integer programming using surrogate constraints," *Decision Sciences*, vol. 8, no. 1, pp. 156–166, 1977.
- [168] M. G. Resende, C. C. Ribeiro, F. Glover, and R. Martí, *Scatter Search and Path-Relinking: Fundamentals, Advances, and Applications*, Springer, Boston, MA, USA, 2010.
- [169] M. Dorigo and T. Stützle, *Ant Colony Optimization*, Bradford Book, 2004.
- [170] Y. Marinakis and M. Marinaki, *Bumble Bees Mating Optimization Algorithm for the Vehicle Routing Problem*, Springer Berlin Heidelberg, 2011.
- [171] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks (ICNN '95)*, vol. 4, pp. 1942–1948, Perth, Western Australia, November–December 1995.
- [172] Y. Marinakis and M. Marinaki, "A hybrid multi-swarm particle swarm optimization algorithm for the probabilistic traveling salesman problem," *Computers & Operations Research*, vol. 37, no. 2, pp. 1446–1455, 2010.
- [173] J.-Y. Potvin, "State-of-the art review – evolutionary algorithms for vehicle routing," *INFORMS Journal on Computing*, vol. 21, no. 4, pp. 518–548, 2009.
- [174] V. C. Hemmelmayr, J.-F. Cordeau, and T. G. Crainic, "An adaptive large neighborhood search heuristic for two-echelon vehicle routing problems arising in city logistics," *Computers & Operations Research*, vol. 39, no. 12, pp. 3215–3228, 2012.
- [175] T. A. Feo and M. G. C. Resende, "Greedy randomized adaptive search procedures," *Journal of Global Optimization*, vol. 6, no. 2, pp. 109–133, 1995.

## Research Article

# Adaptive Model Predictive Control for Cruise Control of High-Speed Trains with Time-Varying Parameters

Xiaokang Xu,<sup>1,2</sup> Jun Peng ,<sup>1,2</sup> Rui Zhang,<sup>1,2</sup> Bin Chen,<sup>1,2</sup> Feng Zhou,<sup>3</sup> Yingze Yang ,<sup>1,2</sup> Kai Gao,<sup>3</sup> and Zhiwu Huang <sup>1,2</sup>

<sup>1</sup>School of Information Science and Engineering, Central South University, Changsha 410000, China

<sup>2</sup>Hunan Engineering Laboratory of Rail Vehicles Braking Technology, Changsha 410000, China

<sup>3</sup>Changsha University of Science and Technology, Changsha 410000, China

Correspondence should be addressed to Yingze Yang; yangyingze@csu.edu.cn

Received 14 December 2018; Revised 28 March 2019; Accepted 2 April 2019; Published 2 May 2019

Guest Editor: Belen M. Batista

Copyright © 2019 Xiaokang Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The cruise control of high-speed trains is challenging due to the presence of time-varying air resistance coefficients and control constrains. Because the resistance coefficients for high-speed trains are not accurately known and will change with the actual operating environment, the precision of high speed train model is lower. In order to ensure the safe and effective operation of the train, the operating conditions of the train must meet the safety constraints. The most traditional cruise control methods are PID control, model predictive control, and so on, in which the high-speed train model is identified offline. However, the traditional methods typically suffer from performance degradations in the presence of time-varying resistance coefficients. In this paper, an adaptive model predictive control (MPC) method is proposed for cruise control of high-speed trains with time-varying resistance coefficients. The adaptive MPC is designed by combining an adaptive updating law for estimated parameters and a multiply constrained MPC for the estimated system. It is proved theoretically that, with the proposed adaptive MPC, the high-speed trains track the desired speed with ultimately bounded tracking errors, while the estimated parameters are bounded and the relative spring displacement between the two neighboring cars is stable at the equilibrium state. Simulations results validate that proposed method is better than the traditional model predictive control.

## 1. Introduction

In recent years, the high-speed railway transportation has played a more and more important role in modern society. High-speed train has many more advantages such as high speed, large volume, and safe and comfortable environment than traditional railway traffic. With the speed of the high-speed trains rising, it is extremely difficult for human drivers to guarantee the safety of the operation of high-speed trains. In order to ensure the safe and effective operation of high-speed trains, the automatic train control (ATC) system is proposed, which is used to monitor, control, and adjust the train operations to guarantee safety, punctuality, and comfort [1–3].

One of the demanding control problems associated with the automatic train control (ATC) is cruise control problem in which the speed of the train is automatically controlled to

follow a desired trajectory. The methods proposed for cruise control of high-speed trains which are developed based on a motion model obtained from Newton's second law can be classified into two categories. One is to model the whole train that consists of multiple cars as a single point mass [4, 5], while the interaction force between the two cars of the train is ignored. Considering the relative movement between the two cars, the other one is to construct the high-speed train model by a cascade of masses connected by flexible couplers, which provides much more accuracy in characterizing the dynamics of the high-speed trains [6, 7].

In the most existing literature of the high-speed train, the resistance coefficients of train were often assumed to be constant [8–10]. However, for a high-speed train, the aerodynamic resistance will change large, when the train is traveling at high speed. So, the dynamic motion model of the train is a time-varying model dependent on the operating

conditions. The robust adaptive tracking control method was derived for a multiple-mass-points high-speed train dynamics model with unknown and time varying resistance coefficients [11]. Besides, the robust output feedback cruise control is developed for speed tracking with the unknown parameters [12]. Moreover, some more complex train operating conditions are considered in paper [13], and an adaptive controller is developed to deal with the problem of the uncertainty of the air resistance coefficients of the piecewise model. However, the studies of the robust adaptive controller neglected the state constraints and control input constraints. The safe speed and the saturation characteristics of traction and braking units are very important for online operation of high-speed train. The model predictive control has an advantage that fully considers the input and state constraints of the system [14, 15]. However, the influence of time-varying air resistance parameters on the system model is neglected in [15], resulting in low system model precision.

Model predictive control cannot only deal with multi-objective constraint problem, but its dynamic response is fast [16]. The goal of cruise control of high-speed trains is to track the desired target speed quickly and accurately, so, model predictive control is very suitable for the high-speed train cruise controller. According to the principle of model predictive control, it is known that model predictive control requires a prediction model with high precision. The accuracy of the prediction model determines the performance of the controller. Therefore, the starting point of this paper is how to improve the accuracy of the dynamics model of high-speed. In this paper, a multibody model of the high-speed train with time-varying air resistance coefficients and control constraints is considered. The train dynamics model set up in this paper contains time-varying resistance coefficients, and the relative movements among the connected cars of the train are considered, so the dynamics model in this paper is more accurate than that in paper [15]. This paper designs an adaptive model predictive control for cruise control of high-speed trains. Based on Lyapunov's stability theory, an adaptive updating law is given for estimated system model parameters. The closed-loop system is capable of tracking the desired speed, and the relative spring displacements between the two neighbored cars are stable at the equilibrium state.

Compared to the existing work, the proposed adaptive MPC not only solves the cruise control problem with time-varying resistance coefficients but also ensures the train operations within the range of safety constraints. The model complexity is equivalent to the traditional model. The main contributions in this paper can be summarized as follows:

(1) By considering the time-varying air resistance coefficients, this paper firstly constructs a linear multiple-mass-points dynamical model of high-speed trains with time-varying resistance coefficients.

(2) In order to improve the accuracy of the constructed model, an estimated system is proposed based on the proposed model. Then, an adaptive updating law is designed for time-varying parameters of estimated system. Based on the estimated system model, an adaptive model predictive control framework is introduced and the estimated system model is used as the prediction model. Based on the estimated

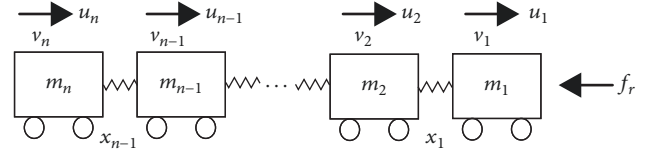


FIGURE 1: The multiple mass-point model structure of high-speed train.

system prediction model, the control problem can be formulated as an objective optimization problem with multiple constraints.

(3) According to the practical requirement, the objective optimization problem with multiple constraints is transformed into the linear quadratic programming problem, which determines the optimal cruise control for the high-speed train with time-varying resistance coefficients and control constraints to improve the safety and energy efficiency of the operation of the train.

The rest of this paper is arranged as following. In Section 2, the linear dynamical model of high-speed train with time-varying parameters is introduced. In Section 3, an adaptive model predictive controller is developed for the train with time-varying parameters to speed tracking. In Section 4, a simulation study is presented to show the performance of the proposed method. Finally, some conclusions are given in Section 5.

## 2. Dynamic Model of High-Speed Train

In this section, the nonlinear multiple mass point dynamic model of high-speed trains with time-varying resistance coefficients is established by analyzing their dynamical characteristics. It is difficult to design the system controller because of the complex characteristics of the nonlinear model. So, the linear error dynamic model of high-speed train is constructed around the equilibrium point.

*2.1. The Dynamic Model of the High-Speed Train.* Figure 1 presents the multiple mass-point model structure for high-speed train. A high-speed train contains \$n\$ cars; these cars are connected by flexible couplers. In high-speed train operation, the flexible couplers play an important role in the connected cars and transmit interaction force between two connected cars. As the couplers between two adjacent cars are not perfectly rigid, the function of couplers can be described by a spring model such that the coupler force is a function of the relative displacement \$\xi\$ between two connected cars.

$$f(\xi) = k\xi \quad (1)$$

where \$k\$ is the stiffness coefficient, which is positive. The running resistance \$f\_r\$ of high-speed train consists of the rolling mechanical resistance and aerodynamic drag, which is commonly expressed as

$$f_r = c_0 + c_1v + c_2v^2 \quad (2)$$

where \$v\$ is the car velocity and \$c\_0, c\_1, c\_2\$ are the resistance coefficients and they are bounded time-varying parameters.

The parameters change near the nominal value [17]; they can be described as  $c_0 = c_0^* + \Delta c_0$ ,  $c_1 = c_1^* + \Delta c_1$ , and  $c_2 = c_2^* + \Delta c_2$ .  $c_0^*$ ,  $c_1^*$ , and  $c_2^*$  and  $\Delta c_0$ ,  $\Delta c_1$ , and  $\Delta c_2$  are the nominal value of the resistance coefficients and the deviation value near the nominal value, respectively, where the deviation value is assumed to be time-varying and bounded, i.e.,  $|\Delta c_0| \leq \bar{c}_0$ ,  $|\Delta c_1| \leq \bar{c}_1$ , and  $|\Delta c_2| \leq \bar{c}_2$ .  $c_0$  defines the train's rolling resistance component,  $c_1$  defines the train's linear resistance coefficient, and  $c_2$  defines the train's nonlinear resistance coefficient.  $c_0 + c_1 v$  represents the rolling mechanical resistance, and  $c_2 v^2$  denotes the aerodynamic drag.

The multiple mass-point dynamic equation of a train can be described as

$$\begin{aligned} \dot{x}_i &= v_i - v_{i+1}, \quad i = 1, 2, 3, \dots, n-1 \\ m_1 \dot{v}_1 &= u_1 - kx_1 - m_1(c_0 + c_1 v) - \left( \sum_{i=1}^n m_i \right) c_2 v_1^2 \\ m_i \dot{v}_i &= u_i + kx_{i-1} - kx_i - m_i(c_0 + c_1 v), \\ & \quad i = 1, 2, 3, \dots, n-1 \\ m_n \dot{v}_n &= u_n(t) + kx_{n-1} - m_n(c_0 + c_1 v + c_2 v^2) \end{aligned} \quad (3)$$

where  $x_i(t)$  is the relative spring displacement between the neighboring cars  $i$  and  $i+1$ ,  $v_i(t)$  is the speed of  $i$ -th car, and  $u_i$  represents the traction force provided by  $i$ -th car.

It can be seen from  $c_2 v^2$  that the aerodynamic drag is proportional to the square of the speed of high-speed train. When the speed of high-speed train increases, the nonlinear characteristic between aerodynamic drag and speed becomes stronger. The nonlinearity of high-speed train model presents difficulty in solving the optimization problem. It is desirable to linearize the train dynamical equation to facilitate the controller design.

Assume that when the velocity of the high-speed train reaches the desired velocity, the current state of the train is the equilibrium state. The velocity of a high-speed train at equilibrium state is denoted as  $\bar{v}_1 = \bar{v}_2 = \dots = \bar{v}_n = v_0$ , and the relative displacements between neighboring cars are zero at the equilibrium state, which are given as  $\bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_{n-1} = 0$ . Obviously, the acceleration of each car is zero at the equilibrium state, which is given as  $\dot{\bar{v}}_1 = \dot{\bar{v}}_2 = \dots = \dot{\bar{v}}_n = 0$ , so the traction or braking force  $\bar{u}_i$  in the equilibrium state can be derived from (3) that

$$\begin{aligned} \bar{u}_1 &= c_0 m_1 + c_1 m_1 v_0 + c_2 \left( \sum_{i=1}^n m_i \right) v_0^2 \\ \bar{u}_i &= c_0 m_i + c_1 m_i v_0, \quad i = 2, 3, \dots, n \end{aligned} \quad (4)$$

So we define the error displacement variable as  $\hat{x}_i = x_i - \bar{x}_i$ , the error speed variable as  $\hat{v}_i = v_i - \bar{v}_i$ , and the error control variable as  $\hat{u}_i = u_i - \bar{u}_i$ . According to (3) and (4), the

linearized error dynamic equation around the equilibrium state is obtained as

$$\begin{aligned} \dot{\hat{x}}_i &= \hat{v}_i - \hat{v}_{i+1} \\ m_1 \dot{\hat{v}}_1 &= \hat{u}_1 - k\hat{x}_1 - c_1 v_1 m_1 - 2c_2 v_0 \left( \sum_{i=1}^n m_i \right) \hat{v}_1 \\ m_i \dot{\hat{v}}_i &= \hat{u}_i + k\hat{x}_{i-1} - k\hat{x}_i - c_1 v_i m_i, \quad i = 2, 3, \dots, n-1 \\ m_n \dot{\hat{v}}_n &= \hat{u}_n + k\hat{x}_{n-1} - c_1 v_n m_n \end{aligned} \quad (5)$$

Choosing  $x = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{n-1}, \hat{v}_1, \hat{v}_2, \dots, \hat{v}_n]^T$  as the state variable and  $u = [\hat{u}_1, \hat{u}_2, \dots, \hat{u}_n]^T$  as the control variable, the error dynamic equation (5) can be written as

$$\dot{x} = A_p x + B_p u \quad (6)$$

where  $A_p = \begin{bmatrix} 0_{n \times n} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$  and  $B_p = \begin{bmatrix} 0_{n \times n} \\ B_{21} \end{bmatrix}$ .

To be exact,  $B_{21} = \text{diag}\{1/m_1, 1/m_2, \dots, 1/m_n\}$ ,

$$\begin{aligned} A_{12} &= \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \\ A_{21} &= \begin{bmatrix} -\frac{k}{m_1} & 0 & \dots & 0 \\ \frac{k}{m_2} & -\frac{k}{m_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{k}{m_n} \end{bmatrix} \\ A_{22} &= \begin{bmatrix} -c_1 - \frac{2c_2 v_0 (\sum_{i=1}^n m_i)}{m_1} & 0 & \dots & \dots & 0 \\ 0 & -c_1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -c_1 & 0 \\ 0 & \dots & 0 & \dots & -c_1 \end{bmatrix} \end{aligned} \quad (7)$$

Then the above continuous time-domain state-space equation is discretized by the zero-order hold method with sampling period  $T_s$  to have the following form:

$$x(k+1) = Ax(k) + Bu(k) \quad (8)$$

where  $A = e^{A_p T_s}$  and  $B = \int_0^{T_s} e^{A_p \tau} d\tau B_p$ . This discrete model is then used in the following controller design in an MPC framework.

### 3. Controller Design

As shown in Figure 2, an adaptive MPC controller is proposed to achieve the closed loop stability and speed tracking







where  $\widehat{\Theta}(k) \triangleq [\widehat{A}(k), B(k)]$  and  $X(k) \triangleq [x(k)^T, u(k)^T]^T$ . Subtracting the above two equations yields

$$\widetilde{x}(k+1) = \widetilde{\Theta}x(k) \quad (13)$$

where  $\widetilde{\Theta} \triangleq \Theta - \widehat{\Theta}$  and  $\widetilde{x} \triangleq x - \widehat{x}$ .

Define a cost function for the estimated error  $\widetilde{x}$

$$\begin{aligned} J_x &= \widetilde{x}(k+1)^T \widetilde{x}(k+1) \\ &= (x(k+1) - \widehat{\Theta}X(k))^T (x(k+1) - \widehat{\Theta}X(k)) \end{aligned} \quad (14)$$

Its gradient with respect to  $\widehat{\Theta}$  can be calculated by

$$\begin{aligned} \nabla J_x(\widehat{\Theta}) &= \frac{\partial J_x}{\partial \widehat{\Theta}} = -X(k) (x(k+1) - \widehat{\Theta}X(k))^T \\ &= X(k) \widetilde{x}(k+1)^T \end{aligned} \quad (15)$$

Consequently, the updating law for  $\widehat{\Theta}(k+1)$  can be designed by

$$\begin{aligned} \widehat{\Theta}(k+1) &= \widehat{\Theta}(k) - \lambda \nabla J_x^T \\ &= \widehat{\Theta}(k) + \lambda \widetilde{x}(k+1) X(k)^T \end{aligned} \quad (16)$$

where  $\lambda > 0$  is the updating rate to be assigned. It can be proved that [18], with the proposed updating law (17), estimated parameters converge to their actual values, if  $X(k)$  is persistently exciting and satisfies the following constraint:

$$X(k)^T X(k) < \frac{2-\alpha}{\lambda} \quad (17)$$

where  $0 < \alpha < 2$ , In this brief, (19) is treated as an additional constraint. To guarantee that  $\widehat{\Theta}$  converges to  $\Theta$  exponentially fast, a strategy to determine  $\lambda$  can be suggested as [19].

**3.3. Adaptive MPC Design for the Estimated System.** In the MPC framework,  $N_p$  is the predictive horizon and  $N_c$  the control horizon. The optimization problem is to compute a trajectory of a future manipulated variable  $u$  to optimize the future behavior of the train. At the sampling time  $k$ , the current state variable  $\widehat{x}(k)$  of estimated high-speed trains system can be measured, and the predicted state variable  $\widehat{x}(k+1)$  can be calculated according to the predictive equations of the estimated system. The state-space equations of the estimated system (11) can be given by

$$\begin{aligned} \widehat{x}(k+1|k) &= \widehat{A}(k)x(k) + Bu(k) \\ \widehat{x}(k+2|k) &= \widehat{A}(k)x(k+1|k) + Bu(k+1|k) \\ &= \widehat{A}^2x(k) + \widehat{A}Bu(k) + Bu(k+1) \end{aligned}$$

$$\begin{aligned} &\vdots \\ \widehat{x}(k+N_p|k) &= \widehat{A}(k)x(k) + \widehat{A}^{N_p-1}Bu(k) \\ &\quad + \widehat{A}^{N_p-2}Bu(k+1) + \dots + \widehat{A}^{N_p-N_c}Bu(k+N_c+1) \end{aligned} \quad (18)$$

Suppose that the output of the estimated system is given by

$$\widehat{y}(k) = C\widehat{x}(k) \quad (19)$$

where  $C = [0_{(n-1) \times n} \quad I_{(n-1) \times (n-1)}]$  and the output value of the estimated system is the predicted speed of the high-speed trains. From the predicted state variables, the predicted output variables are by substitution

$$\begin{aligned} \widehat{y}(k+1|k) &= C\widehat{A}(k)x(k) + CBu(k) \\ \widehat{y}(k+2|k) &= C\widehat{A}^2(k)x(k+1|k) + C\widehat{A}Bu(k+1|k) \\ \widehat{y}(k+3|k) &= C\widehat{A}^3x(k) + C\widehat{A}^2Bu(k) \\ &\quad + C\widehat{A}Bu(k+1) + CBu(k+2) \\ &\quad \vdots \\ \widehat{y}(k+N_p|k) &= C\widehat{A}^{N_p}x(k) + C\widehat{A}^{N_p-1}Bu(k) \\ &\quad + C\widehat{A}^{N_p-2}Bu(k+1) + \dots \\ &\quad + C\widehat{A}^{N_p-N_c}Bu(k+N_c+1) \end{aligned} \quad (20)$$

Define  $\widehat{Y}(k) = [\widehat{y}^T(k+1|k), \widehat{y}^T(k+2|k), \dots, \widehat{y}^T(k+N_p|k)]$ . According to classical MPC design [18], the predictive equations for (11) can be written into a compact form

$$\widehat{Y}(k) = \widehat{F}(k)x(k) + \widehat{\Phi}(k)U \quad (21)$$

where  $\widehat{F}(k) = \begin{bmatrix} C\widehat{A} \\ C\widehat{A}^2 \\ \vdots \\ C\widehat{A}^{N_p} \end{bmatrix}$  and

$$\widehat{\Phi}(k) = \begin{bmatrix} CB & 0 & \dots & 0 \\ CAB & CB & \dots & 0 \\ \vdots & & & \vdots \\ C\widehat{A}^{N_p-1}B & C\widehat{A}^{N_p-2}B & \dots & C\widehat{A}^{N_p-N_c}B \end{bmatrix}. \quad (22)$$

where  $U = [u^T(k), u^T(k+1), \dots, u^T(k+N_c-1)]^T$  and the desired speed signals are given by  $V_r(k) = [v_r^T(k+1), v_r^T(k+2), \dots, v_r^T(k+N_p)]^T$ .

The optimization objective function can be written as

$$\begin{aligned} J &= (\widehat{Y}(k) - V_r(k))^T (\widehat{Y}(k) - V_r(k)) \\ &\quad + U^T(k)RU(k) \end{aligned} \quad (23)$$

where  $R = \text{diag}(r)_{N_c \times N_c}$  is a diagonal weight matrix with  $r > 0$ . To formulate the optimization problem, the cost function is further calculated by

$$\begin{aligned} J &= (\widehat{F}x + \widehat{\Phi}U - V_r)^T (\widehat{F}x + \widehat{\Phi}U - V_r) \\ &\quad + U(k)RU(k) \\ &= (\widehat{F}x - V_r)^T (\widehat{F}x - V_r) + 2U^T \widehat{\Phi}^T (\widehat{F}x - V_r) \\ &\quad + U^T (\widehat{\Phi}^T \widehat{\Phi} + R)U \end{aligned} \quad (24)$$

In order to minimize objective function (24) and optimal control input  $U(k)$ , we just need the formula related to  $U(k)$ . Ultimately, the train operation optimization problem in the MPC framework, which can be uniformly solved by a quadratic programming (QP) approach, is given as

$$J = U^T H U + 2U^T f \quad (25)$$

where  $H = \widehat{\Phi}^T \widehat{\Phi} + R$  and  $f = \widehat{\Phi}^T (\widehat{F}x - V_r)$ . The formula (17) can be written as follows:

$$\begin{aligned} X(k)^T X(k) &< \frac{2-\alpha}{\lambda} \\ &\Downarrow \\ x(k)^T x(k) + u(k)^T u(k) &< \frac{2-\alpha}{\lambda} \\ &\Downarrow \\ u(k)^T u(k) &< \frac{2-\alpha}{\lambda} - x(k)^T x(k) \\ &\Downarrow \\ \|u(k)\| &< \sqrt{\frac{2-\alpha}{\lambda} - x(k)^T x(k)} \end{aligned} \quad (26)$$

Each  $u(k+i|k)$  in the predictive control vector  $U(k)$  should satisfy (26). It follows that  $U$  should satisfy

$$M_1 U \leq \gamma_1 \quad (27)$$

where  $M_1 = \begin{bmatrix} I_{m \times m} & 0 & \dots & 0 \\ -I_{m \times m} & 0 & \dots & 0 \end{bmatrix}$  and

$$\begin{aligned} \gamma_1 &= \frac{\sqrt{m((2-\alpha)/\lambda - x(k)^T x(k))}}{m} \begin{bmatrix} l_m \\ l_m \end{bmatrix}, \\ l_m &= [1, \dots, 1]^T. \end{aligned} \quad (28)$$

To facilitate the MPC design, constraints (10) should be transformed into a form with respect to predictive control vector  $U$ . The constraints on the amplitude of the control signals  $u_i^l(k) \leq u_i(k) \leq u_i^u(k)$  can be formulated as

$$\begin{bmatrix} I_{(nN_c \times nN_c)} \\ -I_{(nN_c \times nN_c)} \end{bmatrix} U \leq \begin{bmatrix} U_{(nN_c \times 1)}^U \\ U_{(nN_c \times 1)}^L \end{bmatrix} \quad (29)$$

where  $U^U$  and  $U^L$  are the upper and lower limits of traction or braking force containing  $N_c$  upper and lower limit vectors ( $u^u$  and  $u^l$ ). The output value of estimated system (21) represents the speed. So, the speed signals  $0 \leq v_i(k) \leq v_{\max}$  can be written as

$$\begin{bmatrix} \widehat{\Phi} \\ \widehat{\Phi} \end{bmatrix} U \leq \begin{bmatrix} \widehat{Y}^U - \widehat{F}x(k) \\ \widehat{Y}^L - \widehat{F}x(k) \end{bmatrix} \quad (30)$$

where  $\widehat{Y}^U$  and  $\widehat{Y}^L$  are the upper and lower limits of speed containing  $N_c$  upper and lower limit vectors (0 and  $v_{\max}$ ).

The relative displacements constraints between neighboring cars are included in constraints of the state variable  $x(k)$ . Designing a matrix  $Z$  for obtaining the relative displacements, it is constructed as  $Z = [I_{(n-1) \times (n-1)} \quad 0_{(n-1) \times n}]$ . Because of the relationship  $x_i(k) = Zx(k)$ , the predicted  $x_i(k)$  within the control horizon  $N_c$  can be obtained as

$$\widehat{X}_i(k) = \widehat{F}_x(k)x(k) + \widehat{\Phi}_x(k)U \quad (31)$$

where  $\widehat{X}_i(k) = [x_i(k+1|k)^T, x_i(k+2|k)^T, \dots, x_i(k+N_c|k)^T]^T$  and

$$\begin{aligned} \widehat{F}_x(k) &= \begin{bmatrix} Z\widehat{A} \\ Z\widehat{A}^2 \\ \vdots \\ Z\widehat{A}^{N_p} \end{bmatrix}, \\ \widehat{\Phi}_x(k) &= \begin{bmatrix} ZB & 0 & \dots & 0 \\ Z\widehat{A}B & ZB & \dots & 0 \\ \vdots & & & \vdots \\ Z\widehat{A}^{N_p-1}B & Z\widehat{A}^{N_p-2}B & \dots & Z\widehat{A}^{N_p-N_c}B \end{bmatrix} \end{aligned} \quad (32)$$

Consequently, the relative displacements constraints  $x_i^l(k) \leq x_i(k) \leq x_i^u(k)$  can be defined as

$$\begin{bmatrix} \widehat{\Phi}_x \\ \widehat{\Phi}_x \end{bmatrix} U \leq \begin{bmatrix} X_i^U - \widehat{F}_x x(k) \\ X_i^L - \widehat{F}_x x(k) \end{bmatrix} \quad (33)$$

where  $X_i^U$  and  $X_i^L$  are the upper and lower limits of relative displacements between neighboring cars containing  $N_c$  upper and lower limit vectors ( $x_i^u$  and  $x_i^l$ ).

Combining constraints (27), (29), (30), and (33) yield constraints for MPC design, and the constraints in this optimization problem can be eventually constructed as

$$MU \leq \gamma \quad (34)$$

$$\text{where } M = \begin{bmatrix} M_1 \\ I_{(nN_c \times nN_c)} \\ -I_{(nN_c \times nN_c)} \\ \widehat{\Phi} \\ -\widehat{\Phi} \\ \widehat{\Phi}_x \\ -\widehat{\Phi}_x \end{bmatrix} \text{ and } \gamma = \begin{bmatrix} \gamma_1 \\ U_{(nN_c)}^U \\ U_{(nN_c)}^L \\ \widehat{Y}^U - \widehat{F}x(k) \\ \widehat{Y}^L - \widehat{F}x(k) \\ X_i^U - \widehat{F}_x x(k) \\ X_i^L - \widehat{F}_x x(k) \end{bmatrix}$$

**3.4. Adaptive MPC Algorithm.** The proposed adaptive MPC algorithm can be summarized as follows:

(1) Select a positive  $\lambda$  according to [19]; the predictive horizon and the control horizon satisfy  $N_c = N_p = N$ .

(2) Calculate optimal  $U(k)$ . With the linear state-space equations (11), the optimization problem is to minimize (25) subject to (19) and (34). This is an optimization problem with a quadratic objective function, which can be uniformly solved by a QP approach [20]. The standard quadratic programming problem has been extensively studied in the literature [21, 22]; this is a field of study in its own right; it requires a considerable effort to completely understand the relevant theory and algorithms. Optimization Toolbox in MATLAB provides functions for finding parameters that minimize or maximize objectives while satisfying constraints. The toolbox includes solvers for linear programming (LP), mixed-integer linear programming (MILP), quadratic programming (QP), nonlinear programming (NLP), constrained linear least squares, nonlinear least squares, and nonlinear equations. If the optimal control problem can be transformed into a quadratic programming problem, the quadratic programming problem can be solved by MATLAB quadprog toolbox more easily. The instructions for the toolbox can be found at <https://www.mathworks.com/help/optim/quadratic-programming.html>.

(3) Find  $u(k)$  by using receding horizon scheme:  $u(k) = [I_{m \times m}, 0, \dots, 0]U(k)$ .

(4) Update the estimated parameters  $\hat{A}(k+1)$  by using the adaptive updating law (16).

(5) Make  $k = k + 1$ , and update system states, inputs and outputs with control  $u(k)$ , and state-space equations (11). Repeat steps (1)-(4).

## 4. Simulation and Discussion

A simulation study on a high-speed train is presented to demonstrate the effectiveness of the proposed adaptive MPC algorithm. The simulation in this paper is to solve the optimization problem of model prediction with quadprog toolbox of MATLAB simulation software version 2016b under the system environment of Windows 10 operating system. The parameters of the train model are from the CRH-3 high speed train in China, which are given in Table 1. This paper investigates the advantages of using MPC to optimize the train's performance by comparing its performance under different prediction horizons. The variables  $N_p$  and  $N_c$  are set to be equal in order to investigate the prediction's impacts on the performance of the high-speed train.

**4.1. Simulation Parameter Selecting.** In order to evaluate the performance of the controller, the desired velocity curve including accelerating, decelerating, velocity step increase, velocity step decrease, and constant velocity stages, the speed command of high-speed train is given in Table 2. Here we focus on the dynamic characteristic and performance of the high-speed train in the cruise phases. The considered time horizon is  $T = 1200$ s. In this scenario, we choose the

TABLE 1: Parameters of the CRH-3 high-speed train.

Symbol	Value	Unit
$m_i$	47.5	t
$c_0^*$	$7.75 \times 10^{-3}$	$Nkg^{-1}$
$c_1^*$	$2.28 \times 10^{-4}$	$Ns(mkg)^{-1}$
$c_2^*$	$1.66 \times 10^{-5}$	$Ns^2(m^2kg)^{-1}$
$k$	$1 \times 10^4$	$kNm^{-1}$

TABLE 2: Speed Command of high-speed train.

Phase	Time(s)	Velocity(m/s)
acceleration	0 $\rightarrow$ 100	0 $\rightarrow$ 40
cruise	100 $\rightarrow$ 400	40
acceleration	400 $\rightarrow$ 450	40-70
cruise	450 $\rightarrow$ 850	70
deceleration	850 $\rightarrow$ 900	70-50
cruise	900 $\rightarrow$ 1200	50

number of trains as  $n = 4$ , and the train is comprised of all locomotives. The predictive horizon and the control horizon are given by  $N_p = 10$  and  $N_c = 10$ .  $\lambda = 0.01$  and  $R = \text{diag}(0.1)$  are assigned. Additionally, the control input  $u_i$  is subjected to the constraints  $-30kN \leq u_i \leq 30kN$ , the coupler deformation is subjected to the constraints  $-0.02 \leq x_i \leq 0.02$  and the bounded of the deviation value near the nominal value  $\bar{c}_0 = 0.01$ ,  $\bar{c}_1 = 0.00035$ , and  $\bar{c}_2 = 0.000008$ .

**4.2. Simulation Results.** Figure 3 shows the velocity curve for each car, where the abscissa is the simulation time and the ordinate is the velocity of the vehicle. From the figure, we can see that the running speed of each car almost stays the same whether in the accelerating phase or in the decelerating phase, and each car can track the reference speed well during the operation time. From  $t = 0$ s to  $t = 100$ s, each car operate at an  $0.4m/s^2$  acceleration. When  $t = 100$ s, the actual speeds of each car of the high-speed train are closed to the reference speed  $40m/s$ . From  $t = 400$ s to  $t = 450$ s, the high-speed train is running with  $0.6m/s^2$  acceleration, From the zoomed-in figure, we can see that the each car moves with the almost same velocity, and the speed error is negligible. From  $t = 440$ s to  $t = 460$ s, the speed of each car gradually achieved stability, and the velocity of each car is the same as the reference velocity basically. Based on the above simulation analysis, we can make a conclusion that the high speed train can track the target speed quickly and maintain a small steady-state tracking error, which verifies the effectiveness of proposed control method.

$y$  is output of (8), and we define estimated output errors  $\bar{y} = y - \hat{y}$ , which is shown in Figure 4. From the figure, we can see that the estimated output errors converge to zeros during the operation time, which verifies the accuracy of the estimation model is proved.

The curves for the tracking and braking forces of each car in the cruise phases are plotted in Figure 5, and the dashed line is the upper bound of traction or the lower bound of brake force. From  $t = 0$ s to  $t = 100$ s, in order to keep

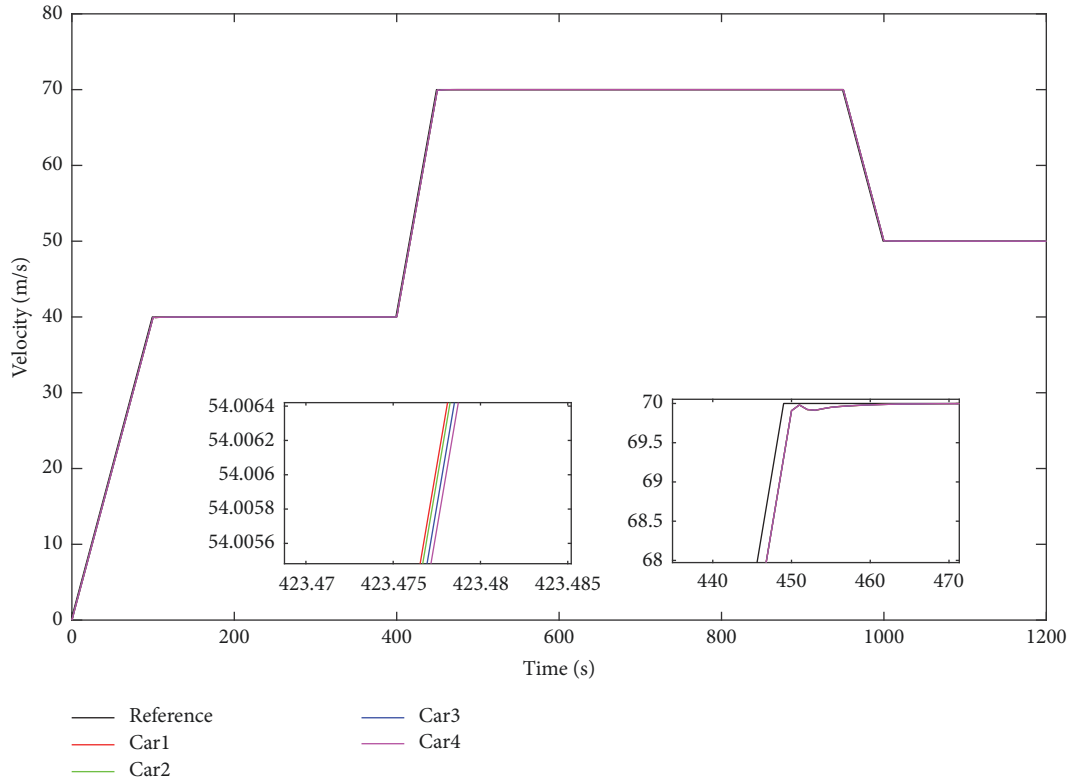


FIGURE 3: Speed curves for each car of high-speed train in the cruise phases.

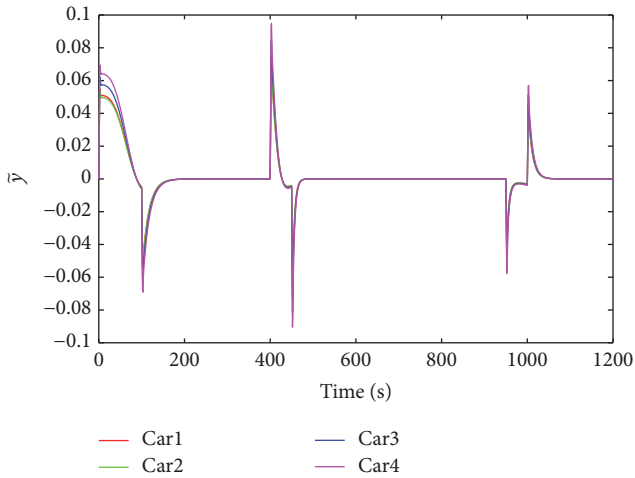


FIGURE 4: Estimated output errors.

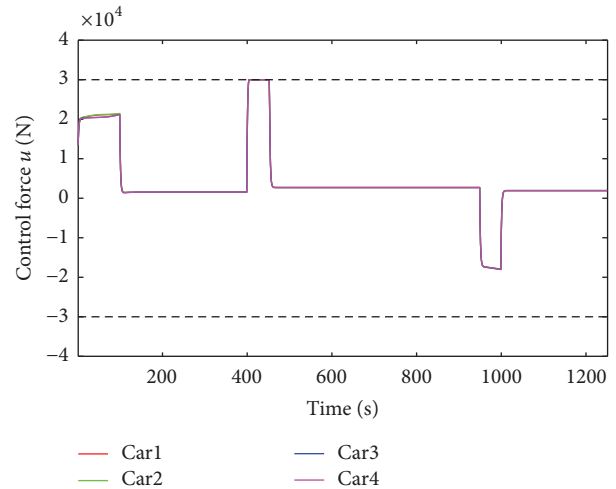


FIGURE 5: The force output of each car.

the acceleration constant, the tracking forces of high-speed train increase rapidly, because the train's resistance increases with speed. At  $t = 400s$ , the high-speed train reaches its maximum speed and the maximum tracking force of each car of the train are  $30kN$ . From  $t = 850s$  to  $t = 900s$ , the braking forces of high-speed train increase rapidly, and maintain about 20s to decrease to the desired velocity. From Figure 5, we can observe that control outputs for each car of the high speed train are almost the same. Based on the above simulation analysis, we can make a conclusion that each

car of the high-speed train can regulate the tracking force and braking force quickly based on actual speed commands and the magnitude of the tracking force and braking force is in the picture satisfies the constraints, which verifies the effectiveness of proposed control method.

The curves of relative spring displacements between the two neighboring cars are plotted in Figure 6, which shows that, under adaptive model predictive cruise control, the relative spring displacements between the two neighboring

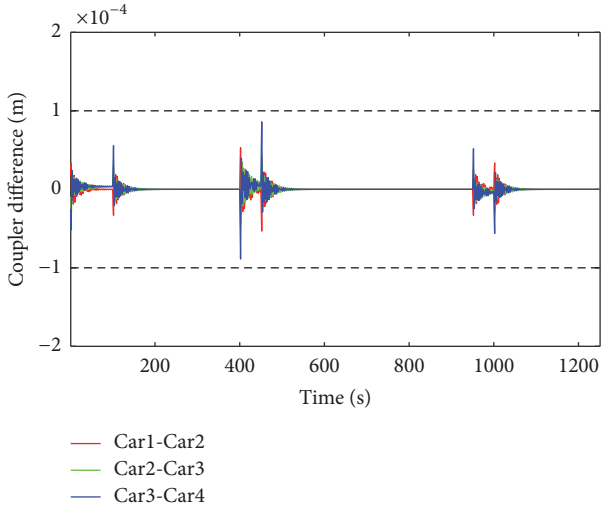


FIGURE 6: The coupler deformation between neighboring cars.

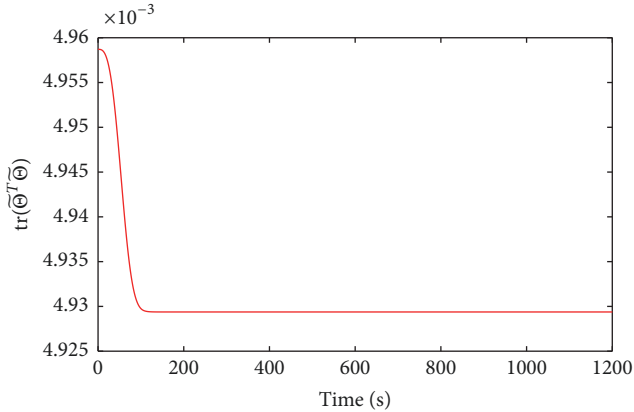


FIGURE 7: Norm of estimation errors.

cars converge to the zero point. During the train operation stage, the relative spring displacements change in a small range, which ensures the safety and comfort of the operating of high-speed train. Additionally, from the figure, we can find that the change of relative displacement is within the constraint range in the acceleration and deceleration stages, and the change of relative displacement is 0 at the equilibrium state.

The norm of estimation errors is defined by  $\text{tr}(\bar{\Theta}^T \bar{\Theta})$ , and the variation of estimated parameters are defined by  $\Delta\bar{\Theta}(k+1) = \bar{\Theta}(k+1) - \bar{\Theta}(k)$ . Because the estimated parameter  $\bar{\Theta}$  is a bounded matrix, the estimation error is a bounded matrix. Figure 7 shows the norm of estimation errors, which is bounded. The variation of estimated parameters is shown in Figure 8; from the figure, we can see that the variation of estimated parameters converges to 0 at around  $t = 100$ s.

**4.3. Further Discussions.** In this subsection, we further discuss the performance of the proposed adaptive model predictive controller in terms of superiority and computation efficiency.

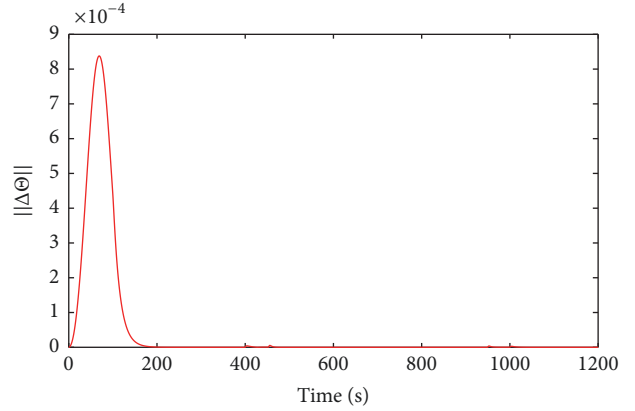


FIGURE 8: The variation of estimated parameters.

**4.3.1. Superiority.** In order to verify the superiority of the method proposed in this paper, we make a simulation comparison with the method in literature [15]; the system in [15] is a nonadaptive control system; it does not consider the impact of model parameter changes on the system model. The parameters of controller are the same of adaptive control system and nonadaptive control system. The purpose of this paper is to improve the accuracy of the prediction model, so that high-speed trains can follow the desired target speed quickly and accurately. The velocity prediction error represents the difference between the predicted model's velocity and the expected velocity; the smaller the error, the higher the accuracy of the prediction model. The velocity error of each car is plotted in the following figures. Figures 9(a) and 9(b) show the simulation results under adaptive control system and nonadaptive control system, respectively. From Figure 9(a), the car of four locomotives can track the reference velocity accurately. The maximum velocity error is no more than  $2m/s$ . From the zoomed in figure, there is little difference in the tracking velocity error of each car, because the parameters of each car are same. This paper's goal is to make sure that each car tracks the reference velocity very well with the coupling force in mind. The method of [15] presents poor control performance when the parameters of the model are uncertain; the maximum velocity error is about  $8m/s$  and the velocity of each car has severe chattering; this control performance is not conducive to the safe operation of the train.

The coupler force of each car is plotted in Figure 10. Figures 10(a) and 10(b) show the simulation results under adaptive control system and nonadaptive control system, respectively. The coupler can be damaged by too much force and excessive coupler force is not conducive to the safe operation of trains. so, when high-speed trains are running; the less coupling force between vehicles, the better. From Figure 10(a), the coupler of each car is very small; from the zoomed in Figure 10(a), the coupling force is no more than  $2KN$ . The method of [15] presents poor control performance, and the coupler of each car is very powerful. This control performance is very unfavorable to the safe operation of the train.

**4.3.2. Computational Cost.** In this subsection, by considering the different prediction horizons, this paper considers the



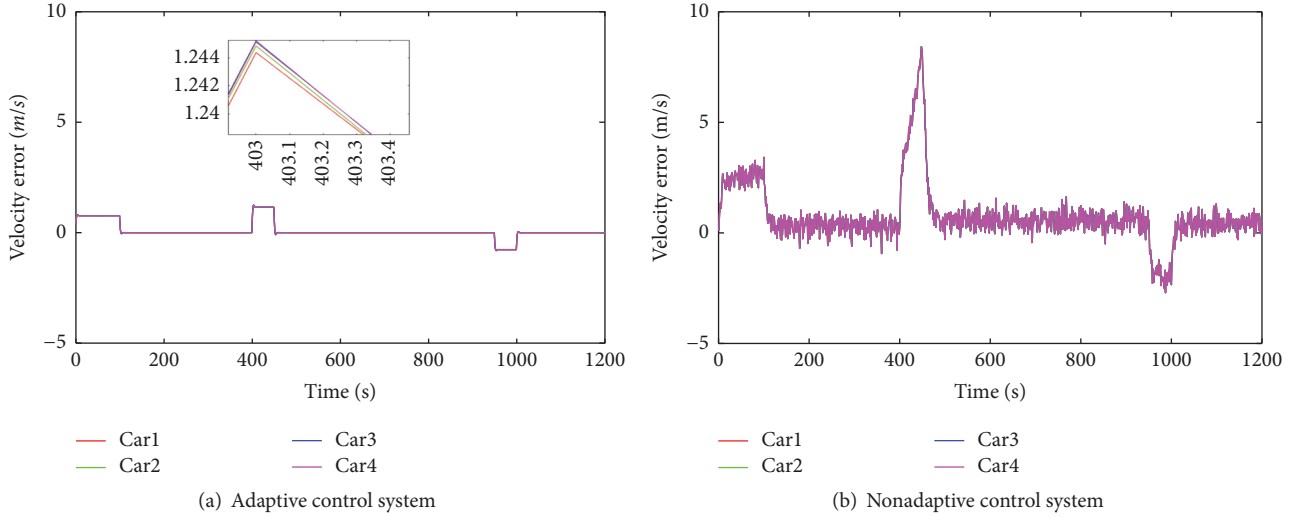


FIGURE 9: Comparison of adaptive control system with nonadaptive control system in predicted output velocity and the desired velocity error.

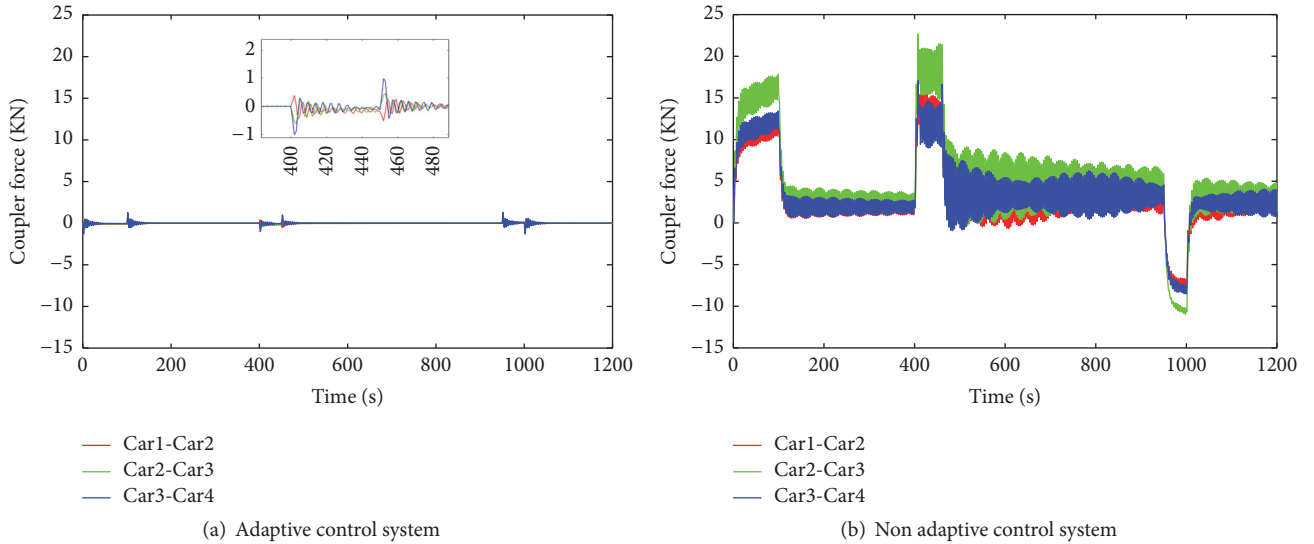


FIGURE 10: Comparison of adaptive control system with nonadaptive control system in predicted output velocity and the desired velocity error.

computation complexity of the adaptive model predictive control with different prediction horizons. Within different prediction horizons, the adaptive MPC and the traditional MPC are respectively implemented on (8). The simulations are carried out under Windows 10 operating system with Intel Core i5-4460 CPU, 8GB RAM on a notebook computer. The computation time is shown in Table 3. It can be seen that the running time of our proposed adaptive MPC is less than the traditional MPC and the computational time of the algorithm will increase with the increase of the predicted horizons from Table 3.

## 5. Conclusion

In this paper, the optimal cruise control of high-speed trains with time-varying air resistance coefficients and control

TABLE 3: Computation time of adaptive MPC and traditional MPC.

Algorithm	$N_p = 4$	$N_p = 6$	$N_p = 10$
Adaptive MPC	9.12s	10.45s	11.96s
Traditional MPC	10.08s	11.06s	13.29s

constraints is investigated. The control objective is accurate speed tracking control with minimum energy consumption and safe relative displacement between two neighbored cars. First, a multiple-mass-point model of high-speed trains is built. By considering multiple constraints and performance metrics, an adaptive MPC method is proposed to design the cruise control controller. In order to improve the accuracy of the method, a dynamic estimated system model of high-speed trains with time-varying parameters is proposed. Also,

an adaptive updating law for estimated system parameters by the Lyapunov stability theory is designed. Then the optimization objective and operation constraints are analyzed in detail. In addition, the cruising control problem is transformed into a constrained finite-time optimal control problem with quadratic objective function, which can be uniformly solved by a quadratic programming approach. Using the method in this paper, the high-speed trains track the desired speed quickly and precisely, and the relative spring displacement between the two neighbored cars is stable at the equilibrium state. Performance of the closed-loop system is substantiated by simulation results.

## Data Availability

The simulation result data used to support the findings of this study have been deposited in the <https://github.com/xuxiaokang1123/Adaptive-Model-Predictive-Control-for-Cruise-Control-of-High-Speed-Trains-with-Time-varying-Parameters.git>. The simulation result data includes the velocity curve data, the force output of each car data, the coupler deformation between neighboring car data and norm of estimation errors, and the variation of estimated parameters data. Everyone can download it through the internet.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

The research work is supported by National Nature Science Foundation of China (Grant Nos. 61772558, 61672537, 61873353, 61672539).

## References

- [1] W. Zhang, Z. Shen, and J. Zeng, "Study on dynamics of coupled systems in high-speed trains," *Vehicle System Dynamics*, vol. 51, no. 7, pp. 966–1016, 2013.
- [2] S. Li, L. Yang, K. Li, and Z. Gao, "Robust sampled-data cruise control scheduling of high-speed train," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 274–283, 2014.
- [3] P. Shakouri, A. Ordys, and M. R. Askari, "Adaptive cruise control with stopgo function using the state-dependent nonlinear model predictive control approach," *ISA Transactions*, vol. 51, no. 5, pp. 622–631, 2012.
- [4] R. Liu and I. M. Golovitcher, "Energy-efficient operation of rail vehicles," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 917–932, 2003.
- [5] Q. Song, Y. D. Song, and W. Cai, "Adaptive backstepping control of train systems with traction/braking dynamics and uncertain resistive forces," *Vehicle System Dynamics*, vol. 49, no. 9, pp. 1441–1454, 2011.
- [6] Q. Song, Y.-D. Song, T. Tang, and B. Ning, "Computationally inexpensive tracking control of high-speed trains with traction/braking saturation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1116–1125, 2011.
- [7] C. Yang and Y. Sun, "Mixed  $H_2/H_\infty$  cruise controller design for high speed train," *International Journal of Control*, vol. 74, no. 9, pp. 905–920, 2001.
- [8] H.-E. Liu, H. Yang, and B.-G. Cai, "Optimization for the following operation of a high-speed train under the moving block system," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–8, 2017.
- [9] W. Shangquan, X. H. Yan, B. G. Cai, and W. Jian, "Multiobjective optimization for train speed trajectory in CTCS high-speed railway with hybrid evolutionary algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2215–2225, 2015.
- [10] Y. Song and W. Song, "A novel dual speed-curve optimization based approach for energy-saving operation of high-speed trains," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1564–1575, 2016.
- [11] Z. Mao, G. Tao, B. Jiang, and X.-G. Yan, "Adaptive position tracking control of high-speed trains with piecewise dynamics," in *Proceedings of the 2017 American Control Conference, ACC 2017*, pp. 2453–2458, USA, May 2017.
- [12] S.-K. Li, L.-X. Yang, and K.-P. Li, "Robust output feedback cruise control for high-speed train movement with uncertain parameters," *Chinese Physics B*, vol. 24, no. 1, Article ID 010503, 2015.
- [13] H. Yang, K. P. Zhang, X. Wang, and L. S. Zhong, "Generalized multiple-model predictive control method of high-speed train," *Journal of the China Railway Society*, vol. 33, no. 8, pp. 80–87, 2011.
- [14] L. J. Zhang and X. T. Zhuan, "Optimal operation of heavy-haul trains equipped with electronically controlled pneumatic brake systems using model predictive control methodology," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 1, pp. 13–22, 2014.
- [15] Y. Yang, Z. Xu, W. Liu, H. Li, R. Zhang, and Z. Huang, "Optimal operation of high-speed trains using hybrid model predictive control," *Journal of Advanced Transportation*, vol. 2018, Article ID 7308058, 16 pages, 2018.
- [16] L. Wang, *Model Predictive Control System Design and Implementation Using MATLAB*, Springer-Verlag, London, UK, 2009.
- [17] R. S. Raghunathan, H. D. Kim, and T. Setoguchi, "Aerodynamics of high-speed railway train," *Progress in Aerospace Sciences*, vol. 38, no. 6-7, pp. 469–514, 2002.
- [18] B. Zhu and X. Xia, "Adaptive model predictive control for unconstrained discrete-time linear systems with parametric uncertainties," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 61, no. 10, pp. 3171–3176, 2016.
- [19] P. Ioannou and J. Sun, *Robust Adaptive Control*, Dover, New York, NY, USA, 2012.
- [20] C. Hildreth, "A quadratic programming procedure," *Naval Research Logistics Quarterly*, vol. 4, no. 1, pp. 79–85, 1957.
- [21] D. G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley Publishing Company, Boston, Mass, USA, 2nd edition, 1984.
- [22] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

## Research Article

# Optimal Location of Biogas Plants in Supply Chains under Carbon Effects: Insight from a Case Study on Animal Manure in North Dakota

Yong Shin Park <sup>1</sup>, Joseph Szmerekovsky <sup>2</sup>, and Alan Dybing<sup>2</sup>

<sup>1</sup>Department of Management, Bill Munday School of Business, St. Edward's University, Austin, TX 78704, USA

<sup>2</sup>Department of Transportation, Logistics, and Finance, College of Business, North Dakota State University, Fargo, ND 58108, USA

Correspondence should be addressed to Yong Shin Park; [ypark1@stedwards.edu](mailto:ypark1@stedwards.edu)

Received 12 November 2018; Revised 4 February 2019; Accepted 17 March 2019; Published 7 April 2019

Guest Editor: Belen M. Batista

Copyright © 2019 Yong Shin Park et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Faced with increasing concerns over the negative environmental impact due to human and industrial activities, biomass industry practitioners and policy makers have great interest in green supply chains to reduce carbon emissions from supply chain activities. There are many studies which model the biomass supply chain and its environmental impact. However, animal waste sourced biogas supply chain has not received much attention in the literature. Biogas from animal manure not only provides energy efficiency, but also minimizes carbon emissions compared to existing biomass products. Therefore, this study proposes a mixed integer linear program that minimizes total supply costs and carbon emissions from an animal waste sourced biogas supply chain while it also incorporates carbon price in the model to see the impact of a carbon policy on tactical and strategic supply chain decisions. To validate the model proposed, a case study of North Dakota is adopted where there is a high potential for a biogas plant to be developed. The results of our optimization experiment indicate that supply chain performance in terms of both costs and emissions is very sensitive to a carbon pricing mechanism.

## 1. Introduction

Biomethane is formed in nature by the biological degradation of biodegradable organic material such as biowaste, sludge, manure, and agroresidues under anaerobic conditions. The main components of biogas are methane and carbon dioxide which can be captured and used to generate energy in the form of heat and electricity. They can also be used as vehicle fuel in either a compressed or liquefied form and as power for fuel cell vehicles [1]. According to US Energy Information Administration (EIA), biogas could displace about 5 % and 56% of natural gas consumption in the electric power and transportation sectors, respectively. There were 242 operating anaerobic digestions (ADS) on livestock farms in the US in 2016, producing about 981 million kilowatt-hours (kWh) of energy [2]. There is growing interest in installing ADSs converting daily manure of beef cattle, cows, hogs and poultry, and other animals to biogas due to both its economic and environmental benefit. Biogas produced from ADSs is

considered methane neutral process because it has potential to capture methane that escapes into the atmosphere.

Required by the RFS2, developing a financially feasible and environmentally sustainable bioenergy supply chain across diverse feedstock harvesting, collection, storage, production, and transportation is challenging [3]. Strategic, tactical, and operational level decisions related to location, capacity, logistic issues, transportation networks, feed stock acquisition, and distribution of biomass or biofuel must be made for efficient and effective optimal network configuration [4, 5]. Traditional supply chain network design has focused on cost efficiency, but recent regulatory mandates require federal, states, and local authorities to expand their objectives beyond just economic metrics. Now, environmental performance consideration, such as carbon reduction and waste minimization, needs to be part of the project [6].

It is recognized that renewable energy is already playing a great role in reducing emissions in the energy sector in the US and many other countries. Fossil fuel-fired power plants are

the largest source of emissions accounting for 31 percent of US greenhouse gas emissions. Interest in imposing a carbon tax on carbon emissions seems to be on the rise in the US, among decision makers [7], in order to increase the cost of energy produced from fossil fuels [3]. A national carbon tax of \$40 per metric ton is expected to be raised at a rate of 5.6 percent per year and about \$2.5 trillion in revenue would be yield over a 10-year period. It would also cut US emissions by 8 percent by 2021, as well as hike gasoline and electricity prices [8].

Motivated by the evolving regulatory climate change pressures in the United States, this paper develops an optimization model and consider the strategic decisions of the number and location of biogas plants, as well as the tactical optimization of its capacity and the biogas production in order to explore how the bioenergy industry can manage its supply chain under the two carbon regulatory schemes, including carbon pricing and carbon trading mechanisms, which are two popular environmental regulatory policy schemes that have been widely implemented in different nations [9, 10]. This study provides not only practical implications associated with the modeling effort but also research implications including discussion of additional outcomes and further development. A new approach in the biogas supply chain system is also required to face ever-changing energy markets because uncertainties in climate change calculation continue to pose some of the most challenging aspects in designing sustainable bioenergy supply chains [11]. In this regard, Mixed Integer Linear Programming (MILP) is an effective optimization tool, which captures the impact of different scenarios of emission price and caps on the biogas supply chain and provides optimal strategies in designing and planning for practitioners and policy makers. The proposed biogas supply chain model contributes to the sustainable biogas plant location modeling literature through helping organizations, policymakers, and scholars evaluate the tactical and operational biogas supply chain planning.

The remainder of the paper is organized as follows: Section 2 reviews the literature on carbon regulatory schemes and biomass supply chains; Section 3 presents the problem statement and optimization model that is proposed in this research; Section 4 describes a case study; Section 5 presents the results and the discussion of research findings and potential implications for policy makers. The paper concludes by providing a summary with future research directions in Section 6.

## **2. Carbon Regulation in the US and Biomass Supply Chain**

It is recognized that renewable energy is already playing a great role in reducing emissions in the energy sector in the US and many other countries. Fossil fuel-fired power plants are the largest source of emissions accounting for 31 percent of US greenhouse gas emissions. Interest in imposing a carbon tax on carbon emissions seems to be on the rise in the US, among decision makers [7], in order to increase the cost of energy produced from fossil fuels [3]. A national carbon tax of \$40 per metric ton is expected to be raised at a rate of 5.6

percent per year and about \$2.5 trillion in revenue would be yield over a 10-year period. It would also cut US emissions by 8 percent by 2021, as well as hike gasoline and electricity prices [12].

There are significant efforts to design carbon tax and carbon cap-and-trade programs to mitigate climate change in other countries. The carbon trading scheme, also known as a cap-and-trade mechanism, is one of the significant policies for carbon emission mitigation [9]. It sets a fixed maximum level of carbon emissions, a cap, to achieve a reduction in emissions. Firms generating more emissions than the allocated allowance either pay a fine or purchase emissions allowance off the market from those firms which generated less than their allocated allowance [10]. Government regulations, community norms, and consumer expectations have all caused organizations to expand their focus beyond the economic aspect of supply chains [13].

In the last decades, researchers have been interested in deciding the location of biomass facility [6, 14–18]. Many efforts have been made to quantitatively consider biomass supply chain network design and management practices [14–18]. The objective considered takes into account economic and environmental aspects. The economic aspect identifies the cost-effective manner that minimizes the total supply chain costs regarding the number, capacity and location of biorefinery facilities, and flow of biomass [19] or maximizes the net profit [20].

On the other hand, improved life cycle performance is required to achieve sustainable biofuel supply chains that integrate environmental aspects. One of the challenges would be how to minimize the carbon footprint to maintain a low environmental impact. Recently, a number of authors have presented research on supply chain optimization of biomass that considers financial objectives as well as the environmental impact [21–23]. Different aspects such as potential GHG savings and impact of carbon tax and carbon trading on economic and environmental performance were also analyzed [24]. It was found that implementing a carbon emissions scheme was cost-effective that minimizes GHG emissions by promoting competitive advantage in biofuel technologies [25]. However, most of these studies focused on the biomass to biofuel supply chain.

Determining the optimal biogas plant location is a challenging task. Several studies related to biogas plants addressed some importance factors that influence the location decisions that includes but not limited to the current situation, potential biogas production, and biogas utilization [26–28], strategic and tactical decision level in biogas industry's supply chain management [4, 29], or the sustainable biogas plant location planning [30]. Mixed integer programming (MIP) and MILP are used extensively in the existing body of literature for strategic or tactical planning of biogas supply chains [15, 16, 31]. However, spatial distribution of supply and demand has a great influence on the design of biogas supply network [32] and optimal facility location highly affects the transportation cost. Therefore, another commonly used approach to the biofuel supply chain problem is application of geographic information system (GIS) based models, which can help to



determine the most appropriate facility location in a specific area [1].

We address the problem of facility location in biogas supply chain that use animal manure from dairy farms, simultaneously deciding the optimal capacity of the plant at each location and the amount of animal manure to be transported from the dairy farms to the biogas plant and the amount of carbon emissions from the biogas supply chain including acquisition, transportation, and production. Most of the previous studies formulated effective green supply chain design, while modeling efforts related to green supply chain design, which consider animal waste under a carbon policy strategy, are not well established in the literature. Considering these facts and research gaps, this study develops a MILP model to determine the optimal configuration of animal waste-based biogas supply chain along with the associated operational decisions that minimizes its economic and environmental performance under carbon policies.

Several factors are considered in the model: location of the manure resource, land use suitability for potential biogas plant location, practical constraints on our ability to harness it, and economic and environmental considerations with several restrictions [11]. We used this optimization model to solve a problem with real data from North Dakota in USA by integrating with GIS for spatial and network analysis.

### 3. Problem Statement and Mathematical Model

A mathematical model for biogas supply chain design under carbon policy is developed using a MILP. Biomass in the form of animal manure is considered as feedstock in the model. This biomass will then be shipped to energy conversion plants for anaerobic digestions (ADs), where the biomass is converted into biogas. Geography and distance can be important factors because biomass to energy schemes are highly geographically dependent due to the fact that manure supply and biogas demand are often widely dispersed. Thus, finding suitable locations for biogas plants, which minimize transportation distances and total supply chain costs, as well as associated carbon emissions is a key issue for sustainable biogas production. One way to serve multiple farms or ranches is to develop centralized or regional ADs, in which case it is important to decide optimal capacity of ADs and locations. The proposed model also considers the carbon pricing and trading scheme. Therefore, the biogas project either incurs costs if the carbon cap that is assigned is lower than the carbon emissions or gains revenues by selling excess carbon credits. The following supply chain inputs, decisions, and assumptions are made for the model.

#### Inputs

- (i) The annual amount of cattle manure and annual natural gas demand. Only natural gas consumption by the electric power sector in North Dakota in 2016 is loaded into the model, because natural gas consumption by vehicle fuel is unknown [33]. Upstream leg of the supply chain is considered and downstream

actors are not considered as the output from the plant is injected directly into the natural gas pipeline [29].

- (ii) The distance between each node in the supply chain is determined by GIS.
- (iii) Costs for acquiring animal manure, transporting it, and producing biogas.
- (iv) Carbon price and cap.
- (v) GHG emissions associated with acquiring manure, transporting manure and biogas, and producing biogas.

#### Decisions

- (i) Locations of biogas plants.
- (ii) Capacity levels for the biogas plants.
- (iii) Amount of biomass to be transported from the feedstock region to the biogas plant.
- (iv) Biogas production volume of each plant.
- (v) Amount of carbon emissions for the entire supply chain including acquisition, transportation, and production.

#### Assumptions

- (i) A refinery will not be shut down once it opens.
- (ii) Truck is the only mode for transporting manure and biogas.

All notation used in the model formulation is summarized in Table 1 and a complete model formulation is presented in (1)-(15). The function  $Z_1$  represents the total supply chain cost that includes the acquisition costs, investment costs including lifetime operation and maintenance costs, production costs, transportation costs of manure, penalty cost for shortage of biogas, and carbon credit generated from methane offset.

$$\begin{aligned} \min Z_1 = & \sum_{i \in I} \sum_{j \in J} c_i^{aq} X_{ij} + \sum_{k \in K} \sum_{j \in J} (c_j^k + v c_j^{om}) Z_j^k \\ & + \sum_{k \in K} \sum_{j \in J} Q_j^k c_j^{pr} \\ & + \sum_{i \in I} \sum_{j \in J} (c_{ij}^t * d_{ij} * 2) + c^{tu} X_{ij} \\ & + \sum_{k \in K} \sum_{j \in J} \lambda (Q_j^k - m^d) - \sum_{j \in J} (e_j^{moff} \alpha) \end{aligned} \quad (1)$$

where the transportation costs of manure are quantity, travel distance, and truck capacity dependent; therefore, (2) indicates the transport cost per ton mile.

$$c_{ij}^t = \frac{C^{tl}}{C^{hc}} \quad (2)$$



The objective function  $Z_2$  represents the overall supply chain carbon emissions from acquisition, production, and transportation.

$$\begin{aligned} \min Z_2 = & \sum_{i \in I} \sum_{j \in J} e_i^{aq} X_{ij} + \sum_{j \in J} \sum_{k \in K} e_j^{pr} p^k Z_j^k \\ & + \sum_{i \in I} \sum_{j \in J} e^{tr} (d_{ij} X_{ij}) \end{aligned} \quad (3)$$

Given  $Z_1$  and  $Z_2$ , the minimization of the overall supply chain cost when operating under a carbon pricing scheme or carbon trading scheme can be formulated in (4) and (5), respectively [10]:

$$\text{Carbon pricing scheme: Minimize } Z_1 + \alpha Z_2 \quad (4)$$

Carbon trading scheme:

$$\text{Minimize } Z_1 + \alpha (Z_2 - CO_2^{cap}) \quad (5)$$

Equation (4) charges a carbon price of  $\alpha$  corresponding to the amount of emissions generated in a carbon pricing situation. By adding a carbon cap in (5), in a carbon trading environment, a plant which generates more emissions than its allocated allowance ( $Z_2 > CO_2^{cap}$ ) can purchase additional allowance or permits off the market at a price of  $\alpha$ . Plants generating fewer emissions than the allowed emission allowance ( $Z_2 < CO_2^{cap}$ ) can sell their surplus to those who may be exceeding their allocated limits. In the latter case, ( $Z_2 < CO_2^{cap}$ ) would be a negative number turning carbon trading into a source of income that might help reduce the overall supply chain costs.

$$\text{s.t. } \sum_{j \in J} X_{ij} \leq a_i \beta, \quad \forall i \quad (6)$$

$$\sum_{i \in I} \theta X_{ij} = \sum_{k \in K} Q_j^k, \quad \forall j \quad (7)$$

$$\sum_{i \in I} X_{ij} \leq \sum_{k \in K} p^k Z_j^k, \quad \forall j \quad (8)$$

$$\sum_{k \in K} Z_j^k \leq 1, \quad \forall j \quad (9)$$

$$Q_j^k \leq p^k S_j^k, \quad \forall j, k \quad (10)$$

$$\sum_{j \in J} \sum_{k \in K} Q_j^k = m^d \quad (11)$$

$$\begin{aligned} \sum_{i \in I} \sum_{j \in J} e_i^{aq} X_{ij} + \sum_{j \in J} \sum_{k \in K} e_j^{pr} p^k Z_j^k \\ + \sum_{i \in I} \sum_{j \in J} e^{tr} (d_{ij} X_{ij}) = CO_2^{ce} \end{aligned} \quad (12)$$

$$X_{ij} \geq 0, \quad \forall i, j \quad (13)$$

$$Q_j^k \geq 0, \quad \forall j, k \quad (14)$$

$$Z_j^k = \{0, 1\}, \quad \forall j, k \quad (15)$$

TABLE 1: Notations used in model development.

Sets	
I	set of ranch, indexed by (i= 1,2, . . .,I)
J	set of potential biogas plant location, indexed by (j= 1,2, . . .,J)
K	set of biogas plant capacity level, indexed by (k=1,2, . . .,K)
Parameters	
$a_i$	maximum available animal manure
$c_j^{aq}$	average acquisition cost of cattle manure
$c_j^{pr}$	unit cost of biogas production at plant j (\$/m <sup>3</sup> )
$c_{ij}^t$	transport cost per ton-mile from cattle farm i to plant j
$c^{tl}$	tons per truck load
$c^{hc}$	truck hauling cost per loaded mile
$\beta$	average wet or dry content of manure (%)
$c^{lu}$	truck loading and unloading cost of (\$/tons) manure
$c_j^k$	investment cost of the plant at location j with plant capacity level k
$c_j^{om}$	annual operational and maintenance cost of the plant at location j with plant capacity level k
$v$	lifetime of biogas plant (years)
$\lambda$	penalty cost for unmet demand
$d_{ij}$	road distance (miles) between ranch i and plant j
$CO_2^{cap}$	maximum amount (tons) of carbon dioxide that can be emitted
$p^k$	annual production capacity for biogas plant size k
$e_i^{aq}$	CO <sub>2</sub> factor (CO <sub>2</sub> -eq. ton/dry ton) for animal manure acquisition
$e^{tr}$	CO <sub>2</sub> factor (CO <sub>2</sub> -eq. ton-mile/truckload) for transportation
$e_j^{pr}$	CO <sub>2</sub> factor (CO <sub>2</sub> -eq. m <sup>3</sup> /dry ton) for biogas production
$e_j^k$	amount (tons) of CO <sub>2</sub> at location j with plant capacity level k
$e_j^{moff}$	amount of offset methane at location j
$\alpha$	average expected cost of carbon price in \$/ton CO <sub>2</sub>
$\theta$	conversion efficiency to produce biogas from cattle manure (m <sup>3</sup> /dry ton)
$m^d$	annual natural gas demand
Decision Variables	
$X_{ij}$	amount of cattle manure transported to plant j from cattle farm i
$Q_j^k$	amount of biogas converted in plant j at size k
$Z_j^k$	1 if biogas plant with size k is built, 0 otherwise
$S_j^k$	size of a biogas plant, if any, to be built at site k
$CO_2^{ce}$	amount of CO <sub>2</sub> that is emitted in supply chain

The objective functions in (4) and (5) are subject to the constraints (6) to (15). Constraints (6) limit the amount of animal manure procured to the amount that is available annually in each manure producing location. Constraints (7) are flow conservation constraints at the biogas plants, which state that the amount of converted animal manure equals the

biogas produced by relating it to conversion rates at plants. Constraints (8) are logical constraints, stating that there is no flow through biogas plants unless one is open. Constraints (9) ensure that a maximum of one size can be chosen for each plant. Constraints (10) ensure that the amount of biomass that can be processed at a biogas plant is limited by the plant capacity. Constraints (11) allow biogas produced at each plant which is equal to the biogas demand. Constraints (12) calculate the carbon dioxide emissions across the whole supply chain. Constraints (13)-(15) enforce nonnegativity and binary restrictions on the decision variables.

#### 4. A Case Study: Potential Biogas Production in North Dakota

North Dakota (ND) has few anaerobic digestion facilities, although it is a significant livestock producer (ND is ranked 16<sup>th</sup> in the United States in cattle). Currently ND has only four operational biogas systems, and they involve water resource recovery and landfill. However, it is expected that there will be more than 39 new biogas plants based on ND's available resources. Upon the biogas installment, there could be enough electricity to generate 52.7 million kWh of power from biogas based natural gas enough to fuel 7,651 vehicles [34].

*4.1. Cattle Manure Resource.* A diverse set of animal waste feedstock resources are available in North Dakota for biogas production. Cattle waste is considered in this study due to its high potential for cattle manure production. Cattle are not uniformly distributed in the state; therefore cattle manure production amounts vary among regions. All cattle feedlots and inventories are collected through the ND State Feedlot Database from the Dickinson Research Extension Center [35]. Annual cattle manure is estimated by converting 1 head of cattle = 0.025 tons of manure/day [36] and multiplying by 365 and percentage of average wet or dry content of manure. This study considers the moisture content of manure and its effect on the biogas supply chain decisions. According to an expert in agricultural engineering at North Dakota State University, the moisture content of manure comprises a large portion of biomass (e.g., 30-85% on a wet basis, moisture content of cattle manure is 85%) and is a significant factor, especially for planning plant capacity and transportation. Figure 1(b) shows the geographic distribution of the cattle feedlots and quantity of solid cattle manure for each feedlot. The annual amount of cattle manure and locations has been generated using GIS. A cattle manure acquisition cost of \$10/ton is used [37].

*4.2. Potential Biogas Plants.* Twenty-two potential biogas sites were identified by performing a land use suitability analysis, considering the various factors and criteria in Figure 1(c). Table 2 presents the social, geographic, and land use criteria that were used to identify their potentially suitable sites for ADSs in ND. The default values of the criteria are based on literatures, as well as some assumptions. All criteria employ GIS analysis, such as creating buffer from lines (road, railway,

and gas grid) or point feature (urban), and clipping polygons (park and water area). Social factors include public areas that are defined as urban, geographic factor such as water (river and aquifer surface area), Bureau of Land Management (BLM), forest service, national park, and wildlife, and land use factor such as road and railway, gas grid, and well and rig [8, 38]. The criteria for wells and rigs are assumed because no studies have been found that studied suitability analysis of biogas plant within an oil producing area. This assumption can be reconsidered later by consulting considering expert opinion or actual survey.

This study considers that each plant could have one of four sizes, according to the amount of cattle manure processed and amount of natural gas produced. The four types of biogas plant are named very small, medium, large, and very large [30]. In our model, we assumed that the four types of plant have different values for the initial investment and maintenance cost. The initial investment cost and life time maintenance cost of a biogas plant are subject to economies of scale. This work considered that annual operation and maintenance costs of a biogas plant represents on average 2% of the investment cost. Operation and maintenance costs were calculated for a plant with a life time of 20 years [30]. The biogas production cost of \$4 per m<sup>3</sup> is used [39] with conversion efficiency of 23m<sup>3</sup>/ton [40].

*4.3. Transportation Data.* In this study, road transportation networks, including local, rural, urban, and highway, are used to estimate the cost of transporting cattle manure (see Figure 1(a)). The shortest path based on Dijkstra's algorithm between each node is generated using the O-D cost matrix application in ArcGIS. The hauling cost per loaded mile for cattle manure is \$4/mile, the cost of loading and unloading a truck is \$5/ton, and tons per truck load is 25 tons. Therefore, transport cost per ton mile is \$4 per mile/25 tons according to Oklahoma State University [37].

*4.4. Environmental Impact Assessment.* In terms of environmental impact analysis, the emission rate associated with biogas production, including feedstock acquisition, transportation, and production, is obtained from existing literature. The final CO<sub>2</sub>-eq value is found to be 0.008ton CO<sub>2</sub>-eq/ton of manure for acquisition [41], 0.002 ton CO<sub>2</sub>-eq/ ton manure for transportation [42], and 0.08 ton CO<sub>2</sub>-eq/ m<sup>3</sup> for biogas production [42]. The main components of biogas are carbon dioxide and methane; specifically, biogas is 60 to 70 percent methane and 30 to 40 percent carbon dioxide (CO<sub>2</sub>) with a low amount of other gases including nitrogen, hydrogen, and hydrogen sulphide. The following calculations were developed based on study from our sources and the rules of arithmetic. According to Abdeshahian et al. [40], a 1 ton of manure will produce 23 m<sup>3</sup> of biogas. Therefore, (15) converts this figure to amount of CH<sub>4</sub> produced per ton of manure using the EPA's estimate that 60% of the biogas from anaerobic digestion is methane [43]. Then, calculate the equivalent amount of CO<sub>2</sub> by assuming that 1 ton of methane is equivalent to 21 tons of carbon dioxide. Thus, multiplying the tons of methane produced per ton of manure

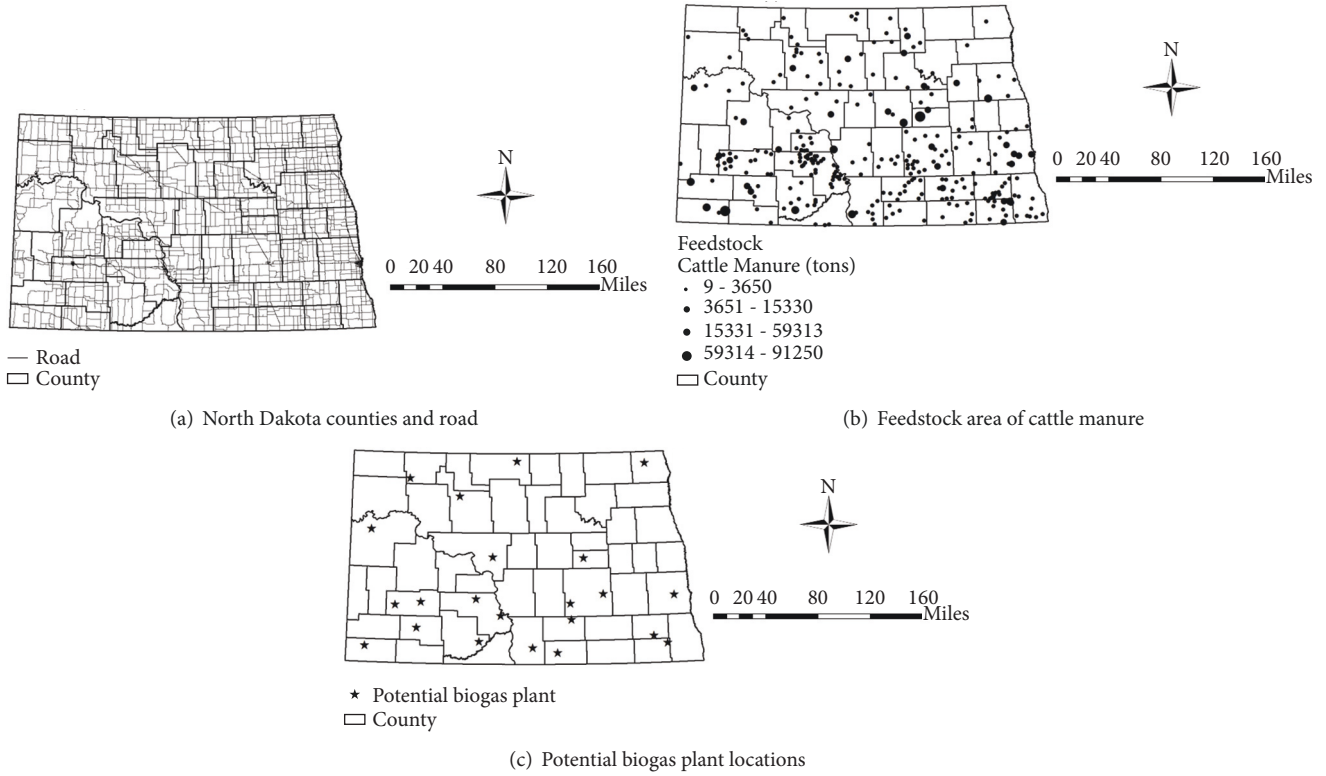


FIGURE 1: Geographic distribution of animal manure feedstock resource and potential biogas plants in North Dakota.

TABLE 2: Factor and criteria to select candidate biogas plant.

Factor	Criteria
Roads and railway	To exclude area which contain or are less than 200m away from major, county and rail network
Water (river and aquifer surface)	To exclude area which contain or are less than 150m away from water line
Bureau of Land Management (BLM)	To exclude area which contain or are less than 1km away from BLM surface land
Gas grid	To include area within 2km of gas pipeline
Forest service	To exclude area which contain or are less than 200m away from
Tribal land	To exclude area which contain or are less than 200m away from
National park	To exclude area which contain or are less than 200m away from
Wildlife	To exclude area which contain or are less than 150m away from wildlife area
Wells and rigs	To exclude area which contain or are less than 200m away from oil well and rig
Urban	To exclude area which contain or are less than 2km away

by twenty-one should provide a reasonable estimate of the amount of carbon dioxide equivalent gas (methane offset,  $e_j^{moff}$ ). Captured methane qualifies as a carbon offset, which can be a source of carbon credits ( $e_j^{moff} \alpha$ ).

$$e_j^{moff} = 0.6 \times \frac{23 \text{ m}^3}{1 \text{ ton of manure}} \times \frac{21 \text{ tons of CO}_2}{1 \text{ ton CH}_4} \quad (16)$$

$\times \text{ tons of manure converted into biogas}$

The carbon price ( $\alpha$ ) used in this case study is \$40/ton of carbon-equivalent emissions [12]. Environmental Protection Agency (EPA) indicated that 45% reduction in CO<sub>2</sub> emissions from the 2005 level by 2030 will be achieved in

North Dakota by replacing power plants with nonemitting generation resources. Using this rule, we set the initial carbon cap as 21 million metric tons of carbon emissions.

## 5. Results and Discussion

From Table 3, the cost-only and emission-only optimization scenarios without considering carbon price show what happens at the two extremes. From the analysis, it shows that the cost-only optimization ( $Z_1$ ) and emissions-only optimization ( $Z_2$ ) are two conflicting objectives. When cost-only optimization model is solved, a minimum supply chain cost of \$310,015,893 occurs which is \$60,25,757 less than when compared to the emission-only optimization. The reverse

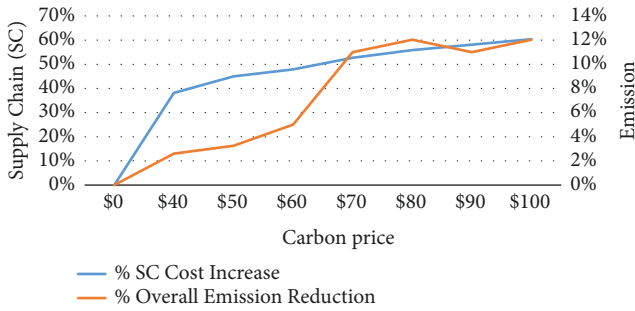


FIGURE 2: Cost increase and emission reduction performance.

situation occurs in the emission-only optimization scenario where the minimum carbon emission of 2,245,564 tons emits at the maximum cost is incurred. The results clearly indicate that, without a carbon pricing mechanism in place, the supply chain could be less costly to manage. We also observe the number of total ADs opened and their size and amount of biogas produced for each optimization scenario. Table 4 shows that the number of ADs opened increases in the emission-only optimization, which may stem from the fact that the model assigns more ADs to minimize the emissions. Also, the average size of ADs eventually decreases for emission-only optimization as the assigned demand decrease; therefore less product is allocated to the ADs.

Figure 2 illustrates the supply chain cost and emission reduction performance over the range of the carbon prices when a carbon trading scheme is in place. The y-axis values in figure 2 represent the supply chain cost percentage increase and emission percentage reduction at each carbon price when compared to the \$0 price. This perspective allows for evaluating the schemes' effectiveness over a range of carbon prices. Figure 2 shows that the supply chain cost increases steadily and relatively linearly, as the carbon price increases. However, the curve eventually flattens since, given the supply chain structure, there exist no more operational changes which impact emissions. As can be seen there is a rapid decrease in carbon emissions that occurs at the very low carbon prices of \$0-\$40 per ton. Interestingly, after this point, a slight emission reduction occurs until carbon prices reach \$60 per ton. The next significant improvement in emission reductions occurs at a carbon price of over \$60 per ton and continues to improve until carbon prices reach \$80 per ton. Increasing the carbon price provides strong motivation to reduce emission level, and, as a result, reduces system costs by the sale of offset emission credits.

Figure 3 shows the cost of carbon purchased and sold at different levels of carbon cap. At a higher cap, the firm will sell less carbon and purchase more carbon. This indicates that a change in the carbon cap will have a greater impact on the amount of carbon sold and purchased. One primary and broad-based policy question is to determine the carbon price at which the maximum environmental performance can be achieved, without substantial negative impacts on the economy and competitive position of the biogas industry. Therefore, from this analysis the price range of \$60-\$70 appears to be the most effective and efficient option in terms

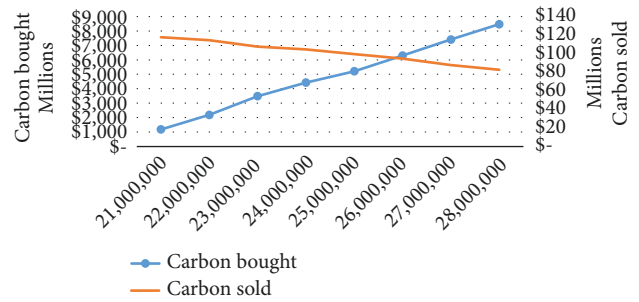


FIGURE 3: Carbon bought and sold with carbon cap variations.

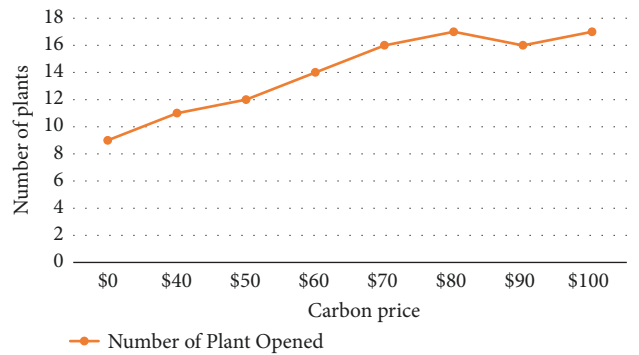


FIGURE 4: Number of plants opened with carbon price variations.

of emissions generation and cost escalation in our model. Within this range, a dollar increase in supply chain costs has the greatest positive impact on carbon pollution reduction.

Table 5 reports capacities of the plants and amounts of biogas production in each county when carbon prices vary. The results suggest that Bowman and Foster are the counties in which the largest plant is constructed at carbon price of \$0. On the other hand, when carbon price increased by \$40, Stutsman may be the county with the largest plant. Under cap and trade, the number of biogas plants is only determined by the carbon price. For a fixed carbon cap, the number of biogas plants and their relative sizes are highly dependent on carbon prices. As seen in Figures 4 and 5, the number of biogas plants opened increases in order to minimize the carbon emission due to transportation. Also, the average size of the biogas plants will eventually decrease as less cattle manure is allocated to each biogas plant.

Figure 6 shows the geographical location of biogas plants in ND for the different carbon effect. The locations of biogas plants and their different optimal capacity levels are presented. As previously mentioned, having no carbon regulatory scheme in place (i.e., a carbon price of \$0) results in 9 biogas plants being opened as the base scenario. Introducing a carbon price at the current national level of \$40 per ton results in more biogas plants being opened. When carbon price increases to \$100, the model opens 17 biogas plants in ND. An increase in the number of plants allows a reduction in transportation and emission costs, thus putting greater emphasis on more efficient and environmentally friendly transport and location decisions. It seems that the model



TABLE 3: Numerical results for the optimizations.

	Cost-only optimization ( $Z_1$ )		Emission-only optimization ( $Z_2$ )	
	Total SC costs (\$)	Total Emissions (tons)	Total SC costs (\$)	Total Emissions (tons)
Transportation	12,859,893	647,880	9,124,530	405,444
Acquisition	8,000,000	584,000	7,160,000	522,680
Production	1,656,000	1,472,000	1,482,120	1,317,440
Investment	287,500,000	-	352,500,000	-
Total	310,015,893	2,703,880	370,266,650	2,245,564

TABLE 4: Number of ADs opened and their relative size variation.

	Cost-only optimization ( $Z_1$ )	Emission-only optimization ( $Z_2$ )
Number of ADs	9	20
Total ADs Size (ton)	690,000	716,000
Average ADs Size (ton)	76,666	35,800

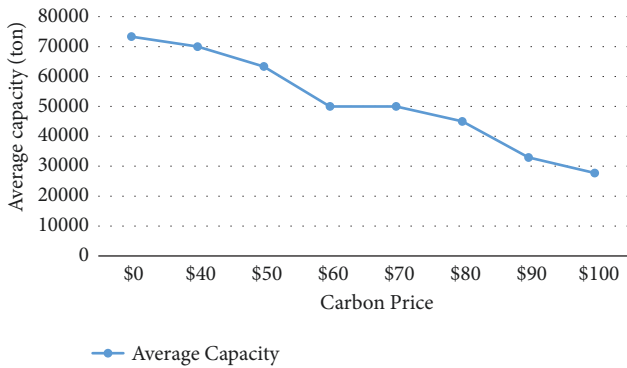


FIGURE 5: Average size of plants with carbon price variations.

locates biogas plant near the county that produces the largest amount of cattle manure. It can be concluded from the results that the location sites and plant capacities are highly dependent on per unit transportation cost for manure.

**5.1. Sensitivity Analysis.** In our model, we perform sensitivity analysis to identify the factors that are significant to the biogas supply chain, especially focusing on cost of biogas by comparing current carbon adjusted cost of natural gas. Thus, we measure cost of biogas delivered by dividing total supply chain cost by total amount of biogas produced in North Dakota as shown in Figure 7. This analysis also shows to determine the indifference point of carbon price at which unit cost of biogas and natural gas becomes equal. The cost of natural gas was calculated under carbon tax that is provided by Hafstead and Picciano [44]. The level of carbon price is varied from \$0/ton of carbon-equivalent emission to \$100/ton of carbon-equivalent emissions. Figure 7 indicates that the carbon price significantly impacts on the unit cost of biogas. Low level of carbon price results in lower cost of biogas and high level of carbon price results in higher cost of biogas. However, as carbon price increases the cost of biogas becomes higher than the cost of natural gas. The indifference

point is achieved once carbon price exceeds \$160/ton of carbon-equivalent emissions means that biogas production at current carbon price up to \$159/ton is beneficial. This may stem from the penalty generated from the plants which emitted more emissions than its allocated carbon emission allowance.

In order to understand the increase in the cost of biogas as carbon price increases, we used break-even analysis to see the relationship between carbon price and the conversion efficiency, as well as natural gas demand and cattle manure acquisition cost. Figure 8 presents the break-even point for natural gas for different values of carbon price and rate of biogas production. The current conversion rate from animal manure to biogas production is relatively low; one ton of manure produces only 23 m<sup>3</sup> of biogas. In the baseline case, the conversion efficiency of biogas was 23 m<sup>3</sup> per ton of manure. The conversion efficiency rate increases up to 188 m<sup>3</sup> per ton of manure from baseline, because it was the maximum conversion efficiency level that would have impact on number of biogas plant and capacity level. It was assumed that there is no cost with improvement of conversion efficiency. When conversion efficiency is fixed, the cost of biogas increases as carbon price increases. When carbon price remains the same, the cost of biogas decreases as the conversion efficiency increases, meaning that the cost of biogas is higher with less efficient technology is employed and a higher carbon price is applied. The increase in the cost of biogas (as the conversion rate increases) is mainly due to the increase in transportation distance and processing costs. Technological improvement of biogas conversion is necessary in order to locate fewer biogas plants that process cattle manure and serve the demand area. Increasing the number of biogas plants will decrease the cost of transportation and the cost of processing additional manure while reducing carbon emissions as carbon price increases. From the analysis results, there is a tradeoff between carbon emission and the supply chain costs.

The change in the cost of biogas at different levels of demand and carbon prices was also investigated; see Figure 9.



TABLE 5: Total manure processing and biogas production in each county at carbon price of \$0 and carbon price of \$40.

County	Carbon price	Total manure capacity (t/y)	Biogas production (m <sup>3</sup> /y)
Bowman	\$0	100,000	2,300,000
	\$40	70,000	1,610,000
Stutsman	\$0	70,000	1,610,000
	\$40	140,000	3,220,000
Sargent	\$0	-	-
	\$40	70,000	1,610,000
Foster	\$0	100,000	2,300,000
	\$40	100,000	2,300,000
Stark, Morton, McLean, Emmons, Cass, Hettinger	\$0	70,000	1,610,000
	\$40	70,000	1,610,000



FIGURE 6: Impact of carbon price on biogas plant locations and sizes.

This result presents the impact that an increase of manure supply and carbon price on cost of biogas. These experiments were inspired by the natural gas consumption trend in the United State in that natural gas consumption is expected to increase about 11% by 2040 from the 2016 level of natural

gas consumption [45]. Results show that when demand is fixed, the cost of biogas increases as carbon prices increase. It was found that the cost of biogas increases at the highest demand and at the highest carbon price. For example, the cost of biogas increases from \$1.42 to \$1.89 at carbon price of \$0

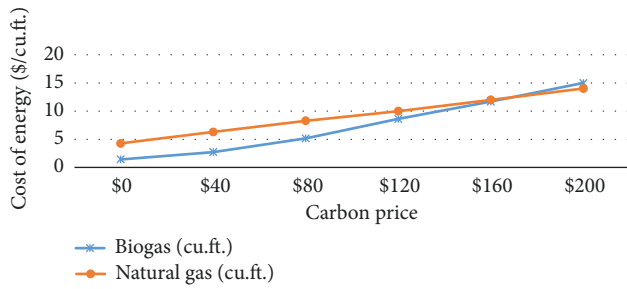


FIGURE 7: Impact of carbon price on energy cost using biogas and natural gas.

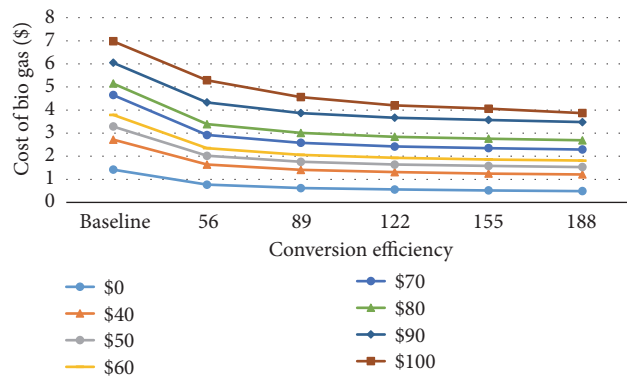


FIGURE 8: Cost of biogas by carbon price and conversion efficiency ( $m^3/ton$ ).

and \$6.98 to \$7.81 at carbon price of \$100. These results may stem from long-haul shipments that occur in high demand and locating small number of biogas plants.

In Figure 10, the impact of manure acquisition cost on unit cost of biogas was analyzed by increasing or decreasing unit cost of manure acquisition by 3% [46] as well as the relationship between carbon price and cost of manure acquisition. The results show that the cost of biogas is highly dependent on manure acquisition cost. Without carbon price being added, the unit cost of biogas is decreased by 1.8% and increased by 2.5% from base case scenario. It is also found that the unit cost of biogas increases linearly as carbon price increases. The finding indicates that overall supply chain cost of biomass related to biomass acquisition will be reduced with improved collecting and process technology. Also, short-term biomass prices are driven by the cost of the raw material, while long-term bioenergy prices are driven by fossil fuel prices. Large scale animal manure supply is also affected by an initial cost of raw material and fertilizer price. Thus, the unit cost of cattle manure acquisition is highly sensitive to fertilizer price.

This study further evaluates the impact of critical parameters on system design and cost. The effects of wet and dry content of manure are analyzed with two carbon price scenarios by assigning weight for each type of manure. The result of sensitivity analysis of wet versus dry manure content is presented in Table 6. The results indicate that wet and dry

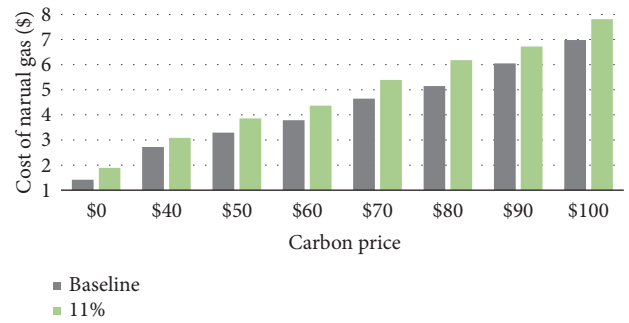


FIGURE 9: Analyzing the impact of demand variation and carbon price on cost of biogas.

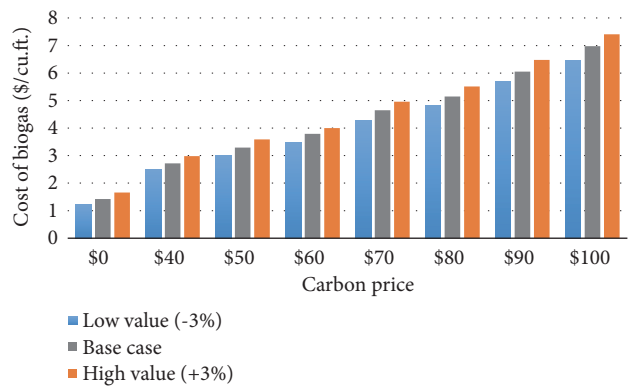


FIGURE 10: Impact of manure acquisition cost and carbon price on cost of biogas.

content of manure and carbon price have significant impacts on the total supply chain costs, carbon credit, and cost of biogas. Wet manure AD is a bit more expensive than dry ADs in terms of acquisition, transportation, and production. Cow manure is about 85% dry basis which resulted in more supply availability that reduces the overall supply chain cost. There is a dramatic change in the number of biogas plants while the average capacity of plants remains the same. The biogas plant size of 70,000 tons remains optimal when the carbon price increase \$0 to current level of carbon price.

## 6. Summary and Conclusion

Siting biogas plant that processes animal manure is a relatively unexplored field from a renewable energy supply chain point of view. In this study, we address greening the biomass supply chain for animal manure through consideration of the carbon effect along the SC and carbon strategy to provide tactical and strategic SC decisions.

This study contributes to the current literature in several ways. It proposes a mathematical model for design and management of a biomass to biogas supply chain, including anaerobic digestion as a source of renewable energy production. This study also contributes to the related body of knowledge by considering mainly waste biomass in the

TABLE 6: Impact of wet versus dry manure content.

Types of manure Scenario	Wet basis		Dry basis	
	Carbon price at \$0	Carbon price at \$40	Carbon price at \$0	Carbon price at \$40
Total cost (in \$m)	464.5	579.2	45.5	348.1
Acquisition cost (in \$m)	7.7	7.7	1.0	0.7
Investment cost (in \$m)	385.0	385.0	35.0	52.5
Production cost (in \$m)	70.8	70.8	9.2	1.8
Transport cost (in \$m)	0.9	0.8	0.3	0.3
Emission cost (in \$m)	-	114.9	-	61.0
Carbon credit (in \$m)	-	8925.8	-	231.8
Cost of biogas (\$/cu.ft.)	0.88	1.07	9.02	10.58
Number of plants	11	11	2	3
Average capacity (tons)	70,000	70,000	70,000	70,000

supply chain design model, while most of the studies focus on energy crops as a source of biomass. Therefore, waste management issues are handled by incorporating carbon policy into the bioenergy facility location problem with due consideration accorded to both monetary and environmental factors.

To validate the proposed model, computational experiments were performed on a case study using North Dakota, which is one of the significant cattle manure producers in the US. The experimental analysis shows that the biogas industry tends to reduce their carbon emissions significantly with introduction of a carbon price by locating less biogas plant to minimize the emissions from transportation and production. From sensitivity analysis, cost of biogas, size of biogas plant, and location were very responsive to different carbon prices, advanced conversion technological efficiencies, types of manure, and manure acquisition cost. This model can help supply chain practitioners devise and implement a strategy based on future expectations of a carbon policy. This model was developed mainly to determine the impact of carbon policies on biogas plant location problem. For future work, developing dedicated transportation mode, tradeoff between logistics costs of manure loss and collection of manure and the cost of transport that address vertical and horizontal relationship in supply chain management would be key area to improve the comprehensive nature of the model [47]. The proposed model can also be further improved by modeling animal waste with other biomass commodities (wood, industrial waste, crops, etc.) or using a multiple objective optimization of supply chain costs and social impact with a more comprehensive life cycle assessment.

## Data Availability

The data used to support the findings of this study are included within the article and datasets from previously reported studies have been cited.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] J. Höhn, E. Lehtonen, S. Rasi, and J. Rintala, "A Geographical Information System (GIS) based methodology for determination of potential biomasses and sites for biogas plants in southern Finland," *Applied Energy*, vol. 113, pp. 1–10, 2014, <https://doi.org/10.1016/j.apenergy.2013.07.005>.
- [2] U.S. EPA., *Renewable Fuel Standard Program*, 2016, <https://www.epa.gov/renewable-fuel-standard-program>.
- [3] J. Zhang, A. Osmani, I. Awudu, and V. Gonela, "An integrated optimization model for switchgrass-based bioethanol supply chain," *Applied Energy*, vol. 102, pp. 1205–1217, 2013, <https://doi.org/10.1016/j.apenergy.2012.06.054>.
- [4] Ş. Y. Balaman, H. Selim, S. İbrenem et al., "A network design model for biomass to energy supply chains with anaerobic digestion systems," *Applied Energy*, vol. 130, pp. 289–304, 2014, <https://doi.org/10.1016/j.apenergy.2014.05.043>.
- [5] A. Rezaee, F. Dehghanian, B. Fahimnia, and B. Beamon, "Green supply chain network design with stochastic demand and carbon price," *Annals of Operations Research*, pp. 1–23, 2015.
- [6] G. Palak, S. D. Ekşioğlu, and J. Geunes, "Analyzing the impacts of carbon regulatory mechanisms on supplier and mode

- selection decisions: an application to a biofuel supply chain," *International Journal of Production Economics*, vol. 154, pp. 198–216, 2014, <https://doi.org/10.1016/j.ijpe.2014.04.019>.
- [7] G. H. Goulder and A. Schein, "Carbon taxes vs. cap and trade: a critical review," Tech. Rep. (No. w19338), 2013, <https://doi.org/10.3386/w19338>.
- [8] E. Thompson, Q. Wang, and M. Li, "Anaerobic digester systems (ADS) for multiple dairy farms: a GIS analysis for optimal site selection," *Energy Policy*, vol. 61, pp. 114–124, 2013, <https://doi.org/10.1016/j.enpol.2013.06.035>.
- [9] T. Abdallah, A. Farhat, A. Diabat, and S. Kennedy, "Green supply chains with carbon trading and environmental sourcing: formulation and life cycle assessment," *Applied Mathematical Modelling*, vol. 36, no. 9, pp. 4271–4285, 2012.
- [10] A. Zakeri, F. Dehghanian, B. Fahimnia, and J. Sarkis, "Carbon pricing versus emissions trading: a supply chain planning perspective," *International Journal of Production Economics*, vol. 164, pp. 197–205, 2015, <https://doi.org/10.1016/j.ijpe.2014.11.012>.
- [11] Y. S. Park, *Three Essays on Sustainability of Transportation and Supply Chain (Doctoral Dissertation)*, North Dakota State University, 2018.
- [12] A. Tang, N. Chiara, and J. E. Taylor, "Financing renewable energy infrastructure: formulation, pricing and impact of a carbon revenue bond," *Energy Policy*, vol. 45, pp. 691–703, 2012, <https://doi.org/10.1016/j.enpol.2012.03.022>.
- [13] B. Fahimnia, J. Sarkis, J. Boland, M. Reisi, and M. Goh, "Policy insights from a green supply chain optimisation model," *International Journal of Production Research*, vol. 53, no. 21, pp. 6522–6533, 2015, <https://doi.org/10.1080/00207543.2014.958592>.
- [14] C.-W. Chen and Y. Fan, "Bioethanol supply chain system planning under supply and demand uncertainties," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 1, SI, pp. 150–164, 2012, <https://doi.org/10.1016/j.tre.2011.08.004>.
- [15] Y. Huang, C. W. Chen, and Y. Fan, "Multistage optimization of the supply chains of biofuels," *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 6, pp. 820–830, 2010, <https://doi.org/10.1016/j.tre.2010.03.002>.
- [16] M. S. Roni, S. D. Eksioğlu, E. Searcy, and K. Jha, "A supply chain network design model for biomass co-firing in coal-fired power plants," *Transportation Research Part E: Logistics and Transportation Review*, vol. 61, pp. 115–134, 2014.
- [17] M. Marufuzzaman, S. D. Eksioğlu, X. Li, and J. Wang, "Analyzing the impact of intermodal-related risk to the design and management of biofuel supply chain," *Transportation Research Part E: Logistics and Transportation Review*, vol. 69, pp. 122–145, 2014.
- [18] M. Marufuzzaman and S. D. Eksioğlu, "Managing congestion in supply chains via dynamic freight routing: An application in the biomass supply chain," *Transportation Research Part E: Logistics and Transportation Review*, vol. 99, pp. 54–76, 2017.
- [19] F. Zhang, D. M. Johnson, and J. Wang, "Integrating multimodal transport into forest-delivered biofuel supply chain design," *Journal of Renewable Energy*, vol. 93, pp. 58–67, 2016, <https://doi.org/10.1016/j.renene.2016.02.047>.
- [20] C. Cambero and T. Sowlati, "Assessment and optimization of forest biomass supply chains from economic, social and environmental perspectives - A review of literature," *Renewable and Sustainable Energy Reviews*, vol. 36, pp. 62–73, 2014, <https://doi.org/10.1016/j.rser.2014.04.041>.
- [21] A. Osmani and J. Zhang, "Economic and environmental optimization of a large scale sustainable dual feedstock lignocellulosic-based bioethanol supply chain in a stochastic environment," *Applied Energy*, vol. 114, pp. 572–587, 2014.
- [22] C. H. Lim and H. L. Lam, "Biomass supply chain optimisation via novel Biomass Element Life Cycle Analysis (BELCA)," *Applied Energy*, vol. 161, pp. 733–745, 2016, <https://doi.org/10.1016/j.apenergy.2015.07.030>.
- [23] J. Ren, D. An, H. Liang et al., "Life cycle energy and CO2 emission optimization for biofuel supply chain planning under uncertainties," *Energy*, vol. 103, pp. 151–166, 2016.
- [24] O. Akgul, N. Shah, and L. G. Papageorgiou, "Economic optimisation of a UK advanced biofuel supply chain," *Biomass and Bioenergy*, vol. 41, pp. 57–72, 2012, <https://doi.org/10.1016/j.biombioe.2012.01.040>.
- [25] S. Giarola, N. Shah, and F. Bezzo, "A comprehensive approach to the design of ethanol supply chains including carbon trading effects," *Bioresour. Technology*, vol. 107, pp. 175–185, 2012, <https://doi.org/10.1016/j.biortech.2011.11.090>.
- [26] A. Gómez, J. Zubizarreta, M. Rodrigues, C. Dopazo, and N. Fueyo, "Potential and cost of electricity generation from human and animal waste in Spain," *Renewable Energy*, vol. 35, no. 2, pp. 498–505, 2010, <https://doi.org/10.1016/j.renene.2009.07.027>.
- [27] C. Tricase and M. Lombardi, "State of the art and prospects of italian biogas production from animal sewage: technical-economic considerations," *Renewable Energy*, 2009, <https://doi.org/10.1016/j.renene.2008.06.013>.
- [28] Y. Ulusoy, A. H. Ulukarde, H. Ünal, and K. Aliba, "Analysis of biogas production in Turkey utilising three different materials and two scenarios," *African Journal of Agricultural Research*, vol. 4, pp. 996–1003, 2009, [https://doi.org/10.1300/J031v13n02\\_08](https://doi.org/10.1300/J031v13n02_08).
- [29] K. Kumar, C. J. C. Lopez, O. A. T. Sanchez et al., "Integrated strategic and tactical optimization of animal-waste sourced biopower supply chains," in *Proceedings of the 2015 International Conference on Industrial Engineering and Systems Management, IEEE IESM 2015*, pp. 1367–1373, Spain, 2016.
- [30] S. Silva, L. Alcada-Almeida, and L. C. Dias, "Multiobjective programming for sizing and locating biogas plants: a model and an application in a region of portugal," *Computers and Operations Research*, vol. 83, pp. 189–198, 2017, <https://doi.org/10.1016/j.cor.2017.02.016>.
- [31] Y. S. Park, J. Szmerekovsky, A. Osmani, and N. M. Aslaam, "Integrated multimodal transportation model for a switchgrass-based bioethanol supply chain," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2628, pp. 32–41, 2017.
- [32] P. Lautala, H. Pouryousef, R. Handler, and S. Chartier, "The role of railroads in multimodal woody biomass transportation in michigan," *ASME Proceedings*, vol. 473, 2012, <https://doi.org/10.1115/JRC2012-74131>.
- [33] U.S. EIA., *Natural Gas Consumption by End Use*, 2018, [https://www.eia.gov/dnav/ng/ng\\_cons.sum.dcu.SND.m.htm](https://www.eia.gov/dnav/ng/ng_cons.sum.dcu.SND.m.htm).
- [34] American Biogas Council, *Biogas State Profile: North Dakota*, 2015, [https://www.americanbiogascouncil.org/StateProfiles/ABCBiogasStateProfile\\_ND.pdf](https://www.americanbiogascouncil.org/StateProfiles/ABCBiogasStateProfile_ND.pdf).
- [35] U. Kizil, *North Dakota State Feedlot Database*, 2017, <https://www.ag.ndsu.edu/archive/dickinso/staff/radtke/index.html>.
- [36] F. Boysan, Ç. Özer, K. Bakkaloğlu, and M. T. Börekçi, "Biogas production from animal manure," *Procedia Earth and Planetary Science*, vol. 15, no. 6, pp. 908–911, 2015.
- [37] Oklahoma State University, "Value of animal waste calculator," 2009, [http://agecon.okstate.edu/faculty/publications\\_results.asp?page=1](http://agecon.okstate.edu/faculty/publications_results.asp?page=1).

- [38] S. Silva, L. Alçada-Almeida, and L. C. Dias, “Biogas plants site selection integrating multicriteria decision aid methods and GIS techniques: a case study in a portuguese region,” *Biomass and Bioenergy*, vol. 71, pp. 58–68, 2014, <https://doi.org/10.1016/j.biombioe.2014.10.025>.
- [39] USDA, *An Analysis of Energy Production Costs from Anaerobic Digestion System on U.S. Livestock Production Facilities*, 2007, [http://www.agmrc.org/media/cms/manuredigesters\\_FC5C31F0F7B78.pdf](http://www.agmrc.org/media/cms/manuredigesters_FC5C31F0F7B78.pdf).
- [40] P. Abdeshahian, J. S. Lim, W. S. Ho, H. Hashim, and C. T. Lee, “Potential of biogas production from farm animal waste in Malaysia,” *Renewable and Sustainable Energy Reviews*, 2016, <https://doi.org/10.1016/j.rser.2016.01.117>.
- [41] E. Favre, R. Bounaceur, and D. Roizard, “Biogas, membranes and carbon dioxide capture,” *Journal of Membrane Science*, vol. 328, no. 1-2, pp. 11–14, 2009, <https://doi.org/10.1016/j.memsci.2008.12.017>.
- [42] S. Esfandiari, R. Khosrokhavar, and S. Masih, “Greenhouse gas emissions reduction through a biogas plant: a case study of waste management systems at FEKA dairy farm,” in *Proceedings of the 2011 2nd International Conference on Environmental Science and Technology IPCBEE*, vol. 6, pp. 445–448, 2011.
- [43] U.S. EPA., *Basic Anaerobic Digester System Flow Diagram*, 2011, [http://www.epa.gov/agstar/documents/digester\\_flow\\_diagram.pdf](http://www.epa.gov/agstar/documents/digester_flow_diagram.pdf).
- [44] M. Hafstead and P. Picciano, *Calculating Various Fuel Prices under a Carbon Tax*, 2017, <http://www.rff.org/blog/2017/calculating-various-fuel-prices-under-carbon-tax>.
- [45] U.S. EIA., *Annual Energy Outlook*, Washington, USA, 2017, <https://www.eia.gov/outlooks/aeo/>.
- [46] USDA, *Fertilizer Use and Price*, 2018, <https://www.ers.usda.gov/data-products/fertilizer-use-and-price.aspx>.
- [47] M. Svanberg, C. Finnsgard, J. Floden, and J. Lundgren, “Analyzing animal waste-to-energy supply chains: the case of horse manure,” *Journal of Renewable Energy*, vol. 18, pp. 1–8, 2016, <https://doi.org/10.1016/j.renene.2017.04.002>.



## Research Article

# Determination of Bus Crowding Coefficient Based on Passenger Flow Forecasting

Zhongyi Zuo <sup>1</sup>, Wei Yin,<sup>1</sup> Guangchuan Yang,<sup>2</sup> Yunqi Zhang,<sup>1</sup> Jiawen Yin,<sup>1</sup> and Hongsheng Ge<sup>3</sup>

<sup>1</sup>School of Traffic & Transportation Engineering, Dalian Jiaotong University, Dalian 116028, China

<sup>2</sup>Department of Civil and Architectural Engineering, University of Wyoming, Laramie, WY 82071, USA

<sup>3</sup>College of Mechanical Engineering, Dalian Jiaotong University, Dalian 116028, China

Correspondence should be addressed to Zhongyi Zuo; [zuozy@djtu.edu.cn](mailto:zuozy@djtu.edu.cn)

Received 28 August 2018; Revised 12 January 2019; Accepted 3 February 2019; Published 1 April 2019

Guest Editor: Marcos M. Vega

Copyright © 2019 Zhongyi Zuo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To improve bus passengers' degree of comfort, it is necessary to determine the real-time crowd coefficient in the bus. With this concern, this paper employed the RBF Neural Networks approach to predict the number of passengers in the bus based on historical data. To minimize the impact of the randomness of passenger flow on the determination of bus crowd coefficient, a cloud model-based bus crowd coefficient identification method was proposed. This paper first selected the performance measurements for determining bus crowd coefficient and calculated the digital characteristics of the cloud model based on the boundary values of the selected performance measures under six Levels-of-Service (LOSs). Then the subclouds obtained under the six LOSs were synthesized into a standard cloud. According to the predicted number of passengers in the bus, the passenger density and loading frequency were calculated, which were imported into the cloud generator to set up the bus crowd coefficient identification model. By calculating the crowd degrees of identification cloud and template cloud at each site, this paper determined the crowded coefficient of each bus station. Finally, this paper took the bus line No. 10 in Dalian city as case study to verify the proposed model. It was found that the crowd coefficients of the selected route ranged from 60.265 to 109.825, and the corresponding LOSs ranged between C and F. The method of discriminating bus crowding coefficient can not only effectively determine the congestion coefficient, but also effectively avoid the fuzziness and randomness of the crowding coefficient judgment in the bus, which has strong theoretical and practical significance.

## 1. Introduction

With the development of urbanization and the rapid growth of urban population, traffic congestion problems have seriously restricted the nation's economic development and affected people's daily lives. In the process of development, we should give priority to the development of public transportation and improve the service quality and travel environment of public transportation. Developing public transportation vigorously is an important way to alleviate road traffic congestion [1]. At present, we usually use standing-passenger density and loading frequency to judge the bus congestion coefficient in the domestic and foreign countries. The standing-passenger density can not only affect the passengers' stress, but also seriously affect the personal safety of the passengers

in the bus if the standing-passenger density exceeds the normal values. Lu et al. [2] analyzed the crowded people and found that, in the case of high standing-passenger density, there is a large squeeze force between the passenger and the passenger. In addition, the accumulation and spread of the force are prone to causing traffic accidents. Ran et al. [3] explored the population density and found that the limit of the density of the Chinese population was  $9 \text{ m}^{-2}$  when the crowded accident occurred according to the individual physiological size. Liu et al. [4] analyzed the various warning indicators and optimized the population density as warning indicator of the degree of crowding in the bus. At last, she proposed the recommended value of the population density according to different situations. Chen et al. [5] used survey, simulation, and other means to select passenger flow density

as a parameter indicator and passenger flow as a weight to construct a passenger flow congestion index suitable for assessing the degree of congestion, and the congestion index is at  $[0, 1]$ . He et al. [6] proposed the fact that the congestion degree is the time characteristic of pedestrians gathering in response to passenger evacuation and can respond to the comfort during passenger evacuation, which is determined by the number of passengers per unit area. Jiang et al. [7] used the standing-passenger density to describe the congestion degree and adopted the RP\_SP method to establish a crowding degree grading system based on standing-passenger density. Xu [8] used the service level based on personal space demand to measure the congestion level of pedestrian facilities and established a crowd perception function based on neuron model. At last the paper proposed a congestion perception information representation method based on spatial environment information representation. In addition, some scholars used image processing technology for passenger flow statistics and the method had also achieved certain development. In 2008, in order to achieve passenger area detection function, Yu et al. [9] proposed a new foreground/background edge model (FBEM) detection method, which traversed all the pixels in the video image and counted statistics and learning to obtain the background area and foreground area within the image. In 2014, Tian [10] used the background difference algorithm and the closed contour fitting moving target detection algorithm in the video detection process to extract the passenger contour by using the morphological processing method, and she also used the head and shoulder classifier to count the passenger movement direction trajectory to realize the passenger detection function. In 2010, Chen [11] performed passenger edge detection in the RGB color space and then used the Hough transform to calibrate the passenger's head area. Finally, the MeanShift's target tracking was used to complete the bus passenger movement target count. In 2013, Hou [12] extracted the passenger's head area by using information such as Hough circle detection and confidence gray interval and combined with the CamShift target tracking algorithm predicted by Kalman filtering. At the same time, the images collected by the upper and lower door cameras are analyzed to calculate the passengers.

Some scholars employed video image processing technology to count the number of people in the bus, and they have achieved certain achievements. For instance, Mukherjee et al. [13] used the Hough circle to extract the passenger's head geometry and count the passengers by the number of blocks that match the set feature. In 2012, Garcia-Bunster et al. [14] corrected the viewing angle of the image and combined it with the standard linear regression model and linear discriminant parameters to find the mapping between the optimal area measurement and the count. Finally, they applied this method to queued passenger counts. In 2013, Daley et al. [15] used the infrared light of the appropriate wavelength to detect the passenger situation in the seating area and the passage area of the vehicle and realized the counting function of the passengers in the bus according to the geometric distribution of the vehicle and the passenger. Mudoj et al. [16] used the background difference method to

extract the target region in the video image, and combined with the neural network algorithm. They used the neural network algorithm to train the results to identify the object in the color and shape characteristics of the target. In 2013, Miklasz et al. [17] used facial recognition algorithm combined with passenger flow optical analysis technology to realize the statistics of passenger flow technology in the car, and the results in the experiment proved that the method has very high statistical accuracy. In 2000, Feng et al. proposed Discrete Representation Method (DRM), which is to analyze the sequence of object trigger points by analyzing discrete targets and object centerlines. This method solved the problem of overlapping objects in passenger counting research [18]. In 2008, Yahiaoui et al. simulated the stereo surveillance video sequence on the bus and the algorithm achieved 99% accuracy by passenger counting experiment [19].

The above research mainly determines the bus crowding coefficient from the two aspects, namely, standing-passenger density and the actual number of people in the bus. The discriminant index is single and has certain fuzziness and randomness. In view of this, this paper proposes a method for judging the bus crowding coefficient based on passenger flow data by using the cloud model. And this method combines the standing-passenger density and loading frequency to identify the crowding coefficient. The method of cloud model can not only avoid fuzziness and randomness of traditional method but also has a strong practical effect.

The remainder of this paper is organized as follows: Section 2 discusses the model to predict the number of passengers in the bus at each bus station. Section 3 introduces the method of bus crowding coefficient based on passenger flow forecast. In this part, we introduced the cloud model to discriminate bus crowding coefficient. Section 4 provides an experimental evaluation of the proposed enhancements. Finally, conclusions of this research are presented in Section 5.

## 2. Model Development

*2.1. Prediction of Passenger Flow Based on RBF Neural Network.* The training method of RBF neural network is simple and efficient. Besides, it has good function approximation ability, classification learning ability, and high convergence speed. The RBF neural network can deal with various intrinsic and difficult to analyze complex system regularity problems. Compared with traditional prediction methods, the use of RBF neural network for passenger flow prediction has the following advantages.

(1) *Self-Learning Ability.* The RBF neural network can adapt to the randomness and nonlinearity of passenger flow changes between stations on public transport lines through continuous training of data. And it has strong nonlinear processing ability. It also makes up for the shortcomings of traditional forecasting methods in solving nonlinear and time-varying problems.

(2) *Adaptive and Self-Organizing Ability.* The RBF neural network can automatically adjust network parameters according

to input and output samples, and it establishes a good input-output mapping relationship to achieve the prediction function.

(3) *Fault Tolerance and Self-Repairing Ability.* The RBF neural network can give correct answers to incomplete information and the system can still be in good condition when some internal faults occur. Therefore, when forecasting the number of passengers in a bus, it only provides the data of passenger flow on and off the bus to train the neural network. And the information of the distribution matrix is obtained and stored in the network. The actual situation can be predicted accurately without relying on the determined distribution matrix.

(i) *Algorithm Design.* This paper uses a three-layer neural network to predict the number of passengers in the bus. It surveyed the number of people in the bus on No. 10 in Dalian City from Monday to Friday in three weeks. Specific steps are described as follows.

*Step 1.* Collect historical passenger flow data via the Information Collection System. Then, we selected the number of passengers in the bus under normal operating conditions as the sample data. The data were divided into two subdatasets: training dataset and prediction dataset. The number of people in early rush hour in previous two weeks was trained as training dataset, and the prediction dataset was the number in early rush hour on the third Friday.

*Step 2.* To avoid the potential prediction errors that might be caused by the sample size of the collect datasets, the original data need to be normalized prior to prediction.

*Step 3.* Construct the passenger flow prediction model. The historical data of number of passengers in the bus, under various weather conditions, holidays, and weeks are selected as input variables to train the neural network and construct a predictive model.

*Step 4.* Apply the trained neural network model to predict the number of passengers in the bus at a certain time in the future.

*Step e.* Analyze prediction errors.

(ii) *Evaluation Indicators.* In order to evaluate the predict results of bus passenger traffic, this paper introduced a predictive result evaluation index. Specifically, predication errors were calculated by comparing the difference between the predicted value  $Y_{sim}$  and the actual value  $Y_{real}$ . There are four indicators for verifying the difference. Among them, the mean average error represents the deviation level between the predicted value and the actual value, and the smaller the error value is, the closer the predicted value is to the true value. The mean average relative error is a commonly used indicator for evaluating prediction results. When mean average relative error is between 20% and 50%, the prediction result is proved to be feasible.

(a) Mean Average Error

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{sim} - Y_{real}| \quad (1)$$

(b) Mean Square Error

$$MSE = \sqrt{\frac{\sum_{i=1}^n (Y_{sim} - Y_{real})^2}{n}} \quad (2)$$

(c) Mean Average Relative Error

$$MARE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_{sim} - Y_{real}|}{Y_{real}} \quad (3)$$

(d) Mean Square Relative Error

$$MSRE = \sqrt{\frac{\sum_{i=1}^n ((Y_{sim} - Y_{real}) / Y_{real})^2}{n}} \quad (4)$$

### 3. Determination of Bus Crowding Coefficient

3.1. *Measurement of Passengers' Crowding Coefficient.* The purpose of determining the crowding coefficient in the bus is to timely and reliably identify the passenger crowd in the bus, so as efficient measures could be applied to reduce the potential safety hazards to passengers and improve passengers' comfort. It has been a common practice that existing research methods use fuzziness and randomness for dividing the crowding states in buses. Therefore, it is of great significance to use a reasonable method to divide the crowded state in the bus. In this paper, the standing-passenger density and loading frequency were used to determine the crowding factor in the bus, as shown in Tables 1 and 2.

It can be seen from Tables 1 and 2 that when the service level in the bus is between A and C, the passengers can move freely in the bus and there is a large comfortable space without crowding. When the service level is worse than E, there is obvious crowd in the bus. Under this situation, it is necessary for the bus dispatchers to take corresponding countermeasures to reduce the crowdedness in the bus, such as the shuttle buses. Therefore, this paper uses the aforementioned service indicators (i.e., standing-passenger density and loading frequency) to determine the degree of crowdedness in the bus

3.2. *Method for Determining Bus Crowding Coefficient.* In order to overcome the fuzziness and randomness of discriminating bus crowding coefficient, this paper uses the cloud model to judge the crowding coefficient in the bus. The cloud model [21] combines the fuzziness in fuzzy theory with the randomness in probability statistics and it was first proposed by Academician De Yi Li and applied to the field of artificial intelligence [22]. Because the cloud model can overcome the fuzziness and randomness very well, the discriminant of the traffic flow state also has time-varying, discrete, fuzzy, and nonlinear. So it is also meaningful to apply the cloud model in the traffic field. The cloud model is used to determine the congestion coefficient in the bus, which further expands the application of the cloud model in the transportation field.

TABLE 1: Passenger standing density evaluation standards.

Service level	Comfort level	Standing-passenger density (people/m <sup>2</sup> )	Description
A	Very Comfortable	3	Satisfy the passengers' psychological comfort requirements and the comfort space is just in contact.
B	Comfortable	5	Passengers can move slightly in the vehicle to meet body comfort requirements and the passenger's body will come into contact.
C	Generally Crowded	6	Passenger's body can maintain a standing position, and the passengers can be in contact with each other without squeezing, which can satisfy the basic space.
D	Crowded	7	Passenger's body remains squeezed and have a sense of crowding, but there is no safety issue.
E	Very Crowded	7.5	Passengers squeeze each other and feel uncomfortable. They may cause safety problems.
F	Unbearable	9	Passengers need to break through the seat area and squeeze into the seat area. It is extremely crowded and unbearable. In addition, boarding and alighting bus become difficult. It is an extreme situation.

TABLE 2: Classification of crowding service quality [20].

QOS	Loading Frequency	Description
A	0~0.5	Passengers can randomly choose seats
B	0.5~0.8	Passengers can appropriately choose seats.
C	0.8~1.0	All passengers have seats but they are less selective
D	1.0~1.25	They are 20% passengers need to stand, but passengers still have personal space
E	1.25~1.5	There are only 1/3 passengers need to stand, some of them have contact, and there is pressure.
F	>1.5	Passengers are obviously crowded and have a strong sense of oppression

TABLE 3: Definition of the digital eigenvalues of the cloud.

Eigenvalues	Definition
Ex	The central value of the spatial distribution of cloud drops
En	Fuzzy measure of qualitative concept
He	Uncertainty measure of entropy

$$\mu(x) = \exp \frac{-(x - Ex)^2}{2(En')^2} \tag{5}$$

So the distribution of x on the domain U is called a normal cloud [24]. The implementation algorithm of the forward normal cloud generator is described as follows [25].

*Step a.* Generate a normal random number  $En'_i$  with En as the expectation and  $He^2$  as the variance.

*Step b.* Generate a normal random number  $x_i$  with Ex as the expectation and  $En'_i$  as the variance.

*Step c.* Calculate  $\mu_i(x) = \exp(-(x_i - Ex)^2/2(En'_i)^2)$ .

Steps (a) to step (c) are iterated until predetermined n cloud drops are produced.

*Definition 3.* There are n subclouds with the same properties in the domain.

There are n subclouds with the same properties in the domain, namely,  $C_1(Ex_1, En_1, He_1), C_2(Ex_2, En_2, He_2), \dots, C_n(Ex_n, En_n, He_n)$ . Then the process of subcloud synthesis of the parent cloud is represented by “ $\circ$ ” [15], namely,  $C = C_1 \circ C_2 \circ \dots \circ C_n$ . If each subcloud has a certain weight, a formula for the expected value, entropy, and hyperentropy of the parent cloud is

### 3.3. The Process of Establishing Cloud Model

#### (1) Definition of Cloud Model

*Definition 1.* Set U as a quantitative domain and express by an exact numerical value. Set C as a qualitative concept and  $C \in U$ . If the quantitative value  $x \in U$ , x is a random implementation of C, and the certainty of x to C is a random number  $\mu(x) : U \rightarrow [0, 1], \forall x \in U, x \rightarrow \mu(x)$  with a stable tendency. Then the distribution of x on the domain C is defined as cloud, and each x is a cloud drop [23]. The digital characteristics of the cloud are usually embodied in three aspects, namely, expected value, entropy, and hyperentropy, respectively. The specific definitions are shown in Table 3.

*Definition 2.* If x satisfies  $x \sim \text{NORM}(Ex, En')$ ,  $En' \sim \text{NORM}(En, He^2)$  and the certainty of x to C satisfies



TABLE 4: Index of each indicator.

Service level	A	B	C	D	E	F
Standing-passenger density (passenger /m <sup>2</sup> )	3	4	5	6	7	
Boarding frequency	0.5	0.8	1.0	1.25	1.5	

$$\begin{aligned}
 Ex &= \frac{\sum_{i=1}^n (En_i \times Ex_i \times \omega_i)}{\sum_{i=1}^n (En_i \times \omega_i)}, \\
 En &= \sum_{i=1}^n (En_i \times \omega_i), \\
 He &= \frac{\sum_{i=1}^n (En_i \times He_i \times \omega_i)}{\sum_{i=1}^n (En_i \times \omega_i)}
 \end{aligned} \quad (6)$$

(2) *Select Metrics.* According to Tables 1 and 2, this paper selected standing-passenger density and boarding frequency as indicators for judging the crowd coefficient in the bus. In order to reduce the singularity of the evaluation index, the same weight was given to each indicator when determining the crowd coefficient in the bus, that is,  $\omega_1 = \omega_2 = 1/2$ , as shown in Table 4.

(3) *Determine the Digital Characteristics of the Cloud.* Assume that the threshold vector of a metric is  $X = (x_1, x_2, x_3, x_4, x_5)$ , because it needs to be implemented in the same domain when synthesizing with the cloud model. When solving the digital features of the cloud, the index value needs to be standardized, and the interval is  $[0, 1]$ .

The standardized treatment formula for the larger the better indicator is

$$x_j^* = \frac{x_j - \min\{x_j\}}{\max\{x_j\} - \min\{x_j\}} \quad (j = 1, 2, 3, 4, 5) \quad (7)$$

The standardized treatment formula for the smaller the better indicator is

$$x_j^* = \frac{\max\{x_j\} - x_j}{\max\{x_j\} - \min\{x_j\}} \quad (j = 1, 2, 3, 4, 5) \quad (8)$$

where  $\max\{x_j\}$  and  $\min\{x_j\}$  are the maximum and minimum values of the  $j^{\text{th}}$  threshold, respectively, and  $x_j^*$  is the normalized value of  $x_j$ .

In addition, the service levels A and F are represented by a half-liter normal cloud and a semifalling normal cloud, respectively. Their numerical characteristics are  $Ex_{x_1}$  and  $Ex_{x_6}$ , and the entropies are  $En_{x_1}$  and  $En_{x_6}$ . The calculation formulas are described as follows:

$$\begin{aligned}
 Ex_{x_1} &= x_1^*, \\
 En_{x_1} &= En_{x_2} \\
 Ex_{x_6} &= x_5^*, \\
 En_{x_6} &= En_{x_5} \\
 He &= 0.01
 \end{aligned} \quad (9)$$

The service levels B, C, D, and E are represented by a full normal cloud, and the eigenvalue calculation formula is

$$\begin{aligned}
 Ex_{x_j} &= \frac{(x_{j-1}^* + x_j^*)}{2} \\
 En_{x_j} &= \frac{(x_{j-1}^* - x_j^*)}{6}, \quad (j = 2, 3, 4, 5) \\
 He &= 0.01
 \end{aligned} \quad (10)$$

(4) *Establish an Identified Cloud Model.* The specific process is interpreted as follows.

*Step a.* Enter the digital characteristics of the service levels of the passengers occupying space indicators in the bus and combining with the forward normal cloud generator algorithm. This paper employed MATLAB software to establish a forward normal cloud generator  $CG_{r_j}$  with metrics.

*Step b.* The actual values of collected number of passengers in the bus were normalized and recorded as  $r'$ . For the index values, if  $r$  is smaller than the  $\min\{r_j\}$ , the normalized result is 0; if  $r$  is greater than  $\max\{r_j\}$ , the normalized result is 1.

*Step c.* Substitute  $r'$  into the forward normal cloud generator  $CG_{r_1}, CG_{r_2}, CG_{r_3}, CG_{r_4}, CG_{r_5}$ , and  $CG_{r_6}$ . The output values  $\mu_{r_1}, \mu_{r_2}, \mu_{r_3}, \mu_{r_4}, \mu_{r_5}$ , and  $\mu_{r_6}$  of each cloud generator indicate the extent to which the input parameter  $r'$  belongs to  $R_j$ . Since the output value has a certain randomness, the output value  $\mu_{r_j}$  needs to be normalized to obtain the weight  $\omega_{R_j}$  of  $R_j$ .

$$\omega_{R_j} = \frac{\mu_{r_j}}{\sum_{j=1}^6 \mu_{r_j}} \quad (11)$$

(5) *Calculate Similarity of Identified Cloud and Standard Cloud.* Similarity refers to the degree of membership  $\beta_i$  of the cloud drops  $x_i$  ( $i = 1, 2, \dots, N$ ) generated by the identified cloud  $U'$  and each standard cloud  $U_j$ . The calculation process is listed as follows.

*Step a.* In identified cloud  $U'$ , generate a normal random number with  $En$  as the expected value and  $He^2$  as the variance, namely,  $En_i = NORM(En, He^2)$ .

*Step b.* In identified cloud  $U'$ , generate a normal random number with  $Ex$  as the expected value and  $En_i^2$  as the variance, namely,  $x_i = NORM(Ex, En^2)$ .

*Step c.* In standard cloud  $U_j$ , generate a normal random number with  $En_j$  as the expected value and  $He_j^2$  as the variance, namely,  $En_j = NORM(En_j, He_j^2)$ .



TABLE 5: Congestion range corresponding to the six service levels.

Service Level	Congestion $\gamma$
(A) Very Comfortable	$\gamma < 30$
(B) Comfortable	$30 \leq \gamma < 50$
(C) Generally Crowded	$50 \leq \gamma < 70$
(D) Crowded	$70 \leq \gamma < 90$
(E) Severely Crowded	$90 \leq \gamma < 110$
(F) Unbearable	$\gamma > 110$

*Step d.* Calculate the membership of  $a^{th}$  service level, namely,  $\beta_{ij} = \exp(-(x_i - Ex_j)^2/2En_j^2)$ .

*Step e.* Iterate Steps a-d until the required cloud drops  $N$  are generated.

*Step f.* The similarity between the identified cloud  $U'$  and the each standard cloud is

$$\delta_j = \frac{1}{N} \sum_{i=1}^N \beta_{ij} \quad (12)$$

*Step g.* Normalize  $j$  to get the degree to which the identified cloud belongs to the  $j^{th}$  service level, namely,

$$\lambda_j = \frac{\delta_j}{\sum_{j=1}^6 \delta_j} \quad (13)$$

The more the cloud drops  $x_i$  generated by the identified cloud  $U'$  fall within the range of a standard cloud  $U_j$ , the larger the  $\delta_j$  value as well as the  $\lambda_j$  value. This situation indicates that the closer the identified state is to the  $f^{th}$  service level.

(6) *Determination of Bus Crowding Coefficient.* This paper uses the maximum value determination method to determine the crowd level in the bus. This method determines the service level corresponding to the state to be recognized in the vehicle by using the maximum degree of possibility. It is difficult to judge the degree of crowding in the bus by this method. The in-vehicle environment with different crowding levels directly affects the behavior of passengers in the bus. Therefore, this paper introduces the crowdedness of passengers in the bus. The calculation formula is as follows:

$$\gamma = \sum_{j=1}^6 \lambda_j \times \xi_j \quad (14)$$

In the formula, the values of  $\xi_j$  represent different crowd coefficients of the service levels, and the crowd coefficients of between A and F were set to 20, 40, 60, 80, 100, and 120, respectively. The larger the congestion  $\gamma$  in the bus, the more crowded the lower service level in the bus.

The crowd range corresponding to the six service levels is shown in Table 5.

## 4. Case Study

This research selected the Dalian Bus Line No. 10 as a case study to verify the accuracy of the proposed model. The Dalian Bus Line No. 10 is a regular bus route. Typically, there are 40 seats on a bus; in addition to these fixed seats, the total effective standing area in the bus is about  $6m^2$ , making the total capacity up to about 100 passenger per vehicle. This route starts from Shahekou Railway Station to Lily Villa, with a total mileage of 16 km and the regular operation hours from 6:00 to 21:00. There are 29 stations for the inbound direction (i.e., from Shahekou Railway Station to Lily Villa) and 27 stations for the outbound direction (i.e., from Lily Villa to Shahekou Railway Station). Among them, it was found that stations 4, 7, 10, 11, and 15 were the most passenger intensive stations. A manual survey about the number of passengers getting on and off the bus at each station was conducted in the morning peak hours (i.e., 7:00 to 8:00).

*4.1. Passenger Flow Forecast in the Bus.* This paper investigated the passenger flow of bus line No. 10 during a working day morning peak hours. The data are shown in Figure 1 (inbound direction) and Figure 2 (outbound direction).

To use the RBF neural network algorithm for predicting the number of passengers in the bus, a prerequisite step is to normalize the data, described as follows:

$$y = (y_{\max} - y_{\min}) \times \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \quad (15)$$

where  $x$  and  $y$  represent the values before and after normalization, respectively; min and max represent the minimum and maximum values in the sample data, respectively. The normalized data are shown in Table 6.

The number of passengers in the bus was predicted by the RBF neural network algorithm, as shown in Figure 3.

Then, the predicted number of passengers was compared with the actual data; the following criteria, including MAE, MSE, MARE, and MSRE (see (1) to (4)), were employed for identifying the differences, as shown in Table 7.

By analyzing the evaluation results, it was found that the MAE=2.206 and MARE=0.249, indicating that the predictions were closed to field data.

*4.2. Determination of Bus Crowding Coefficient Based on Cloud Model.* According to (7) to (10), the subclouds of each evaluation indicator in the bus were calculated, as shown in Table 8.

The numerical characteristics of the standard cloud in the bus are calculated using (6), as shown in Table 9.

Based on the algorithm of the forward cloud generator that generated 5000 cloud drops and using MATLAB as simulation software, the standard cloud pattern formed is shown in Figure 4. In the figure, the green dots represent LOS A and the red dots represent LOS-F.

This section selected a section of the bus route (bus stations 8 to 15) to demonstrate how to convert the predicted number of passengers boarding at a bus station to a corresponding boarding frequency as well as the in-vehicle standing-passenger density, as shown in Table 10.

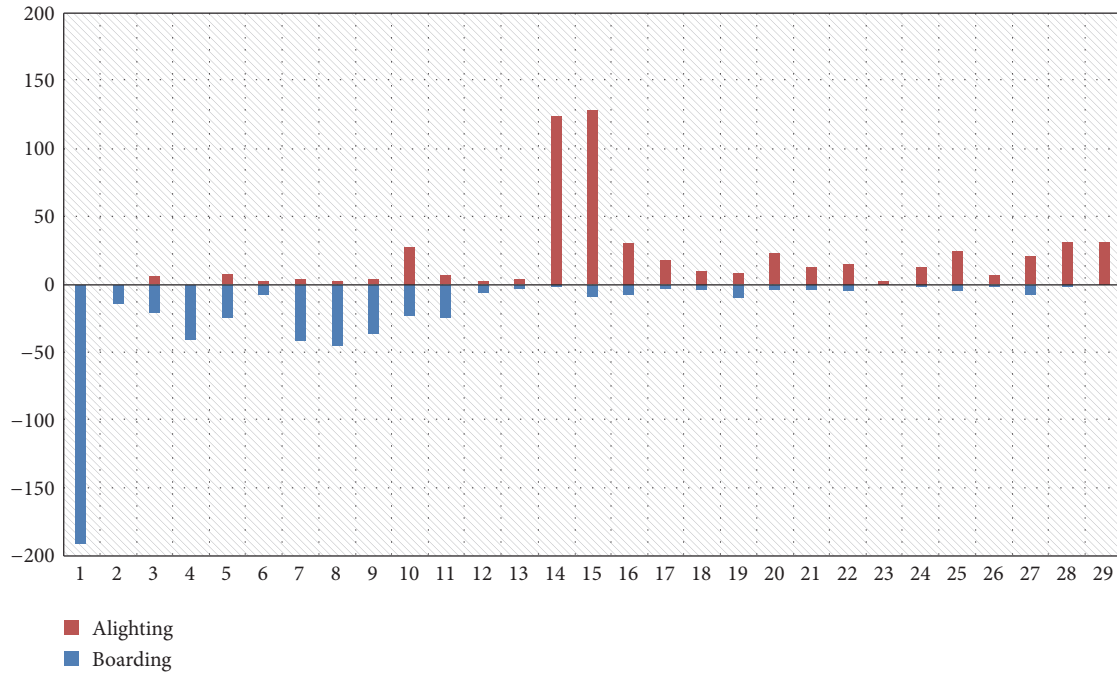


FIGURE 1: Boarding and alighting data of inbound direction.

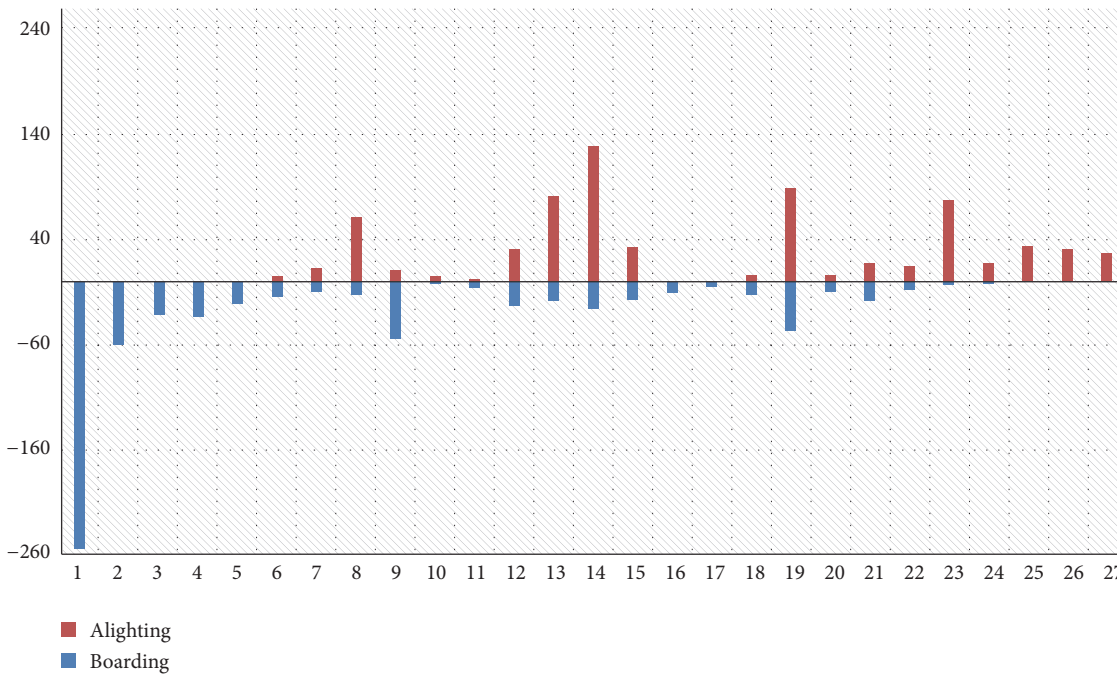


FIGURE 2: Boarding and alighting data of outbound direction.

The data presented in Table 10 were imported into (12) to (14) to calculate the similarity, possible degree, and congestion degree of the selected bus stations, as shown in Table 11.

Eventually, this paper combined with the implementation algorithm of the forward normal cloud generator based on 5000 cloud drops. The identified cloud patterns of each site are shown in Figure 5.

Based on the estimated congestion degrees, the service levels of this bus line at each station were determined, as shown in Figure 6.

Through simulation, it is found that using the cloud model algorithm to determine the crowding coefficient in the bus is a feasible method. According to the estimated congestion degree values presented in Table 8 and the determined service levels of the bus in Figure 6, it can be concluded that

TABLE 6: Normalized data in the prediction model (a part of data).

	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.3556	-0.3333	-0.2444	-0.2000	-0.1111	-0.0667	0.1556	0.1556	0.3778	0.4222	0.3111	0.3111
2	0.3111	0.3111	-0.1333	-0.3111	-0.4000	-0.4667	-0.5333	-0.5333	-0.6444	-0.6889	-0.7333	-0.7333
3	-0.7556	-0.9111	-0.8889	-0.9333	-1	-0.3556	-0.3111	-0.2	-0.0677	-0.0889	0	0.0222
4	0.1778	0.6000	0.7111	0.6889	0.6889	0.6889	0.0667	-0.4667	-0.5778	-0.6667	-0.6889	-0.7111
5	-0.7111	-0.7333	-0.7778	-0.7778	-0.8222	-0.9111	-0.9333	-0.9556	-1	-0.2222	-0.1778	0.0667
6	0.1556	0.1333	0.2444	0.6889	0.9111	0.8889	0.9333	0.9778	1	0.9333	1	-1

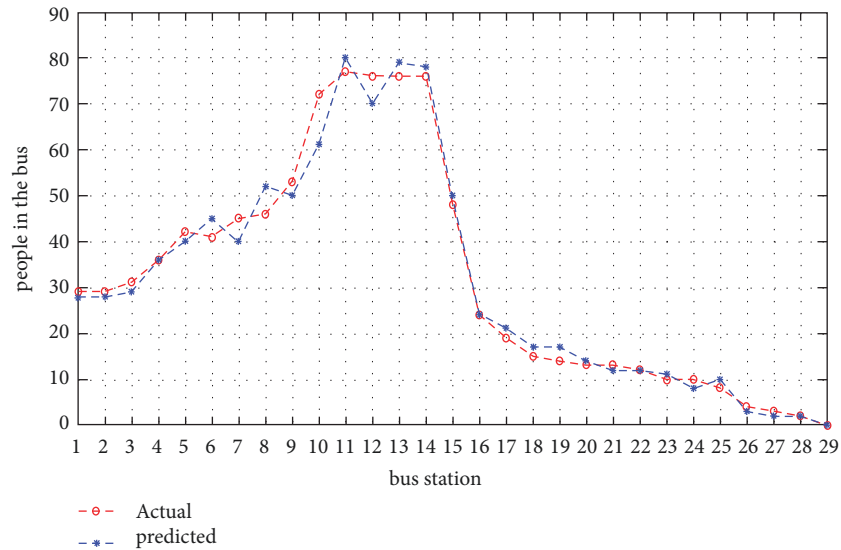


FIGURE 3: Prediction value of RBF neural network.

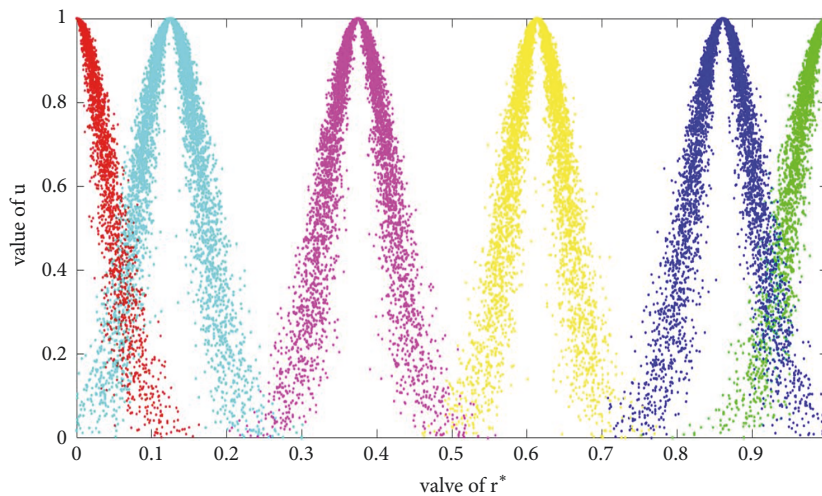


FIGURE 4: The standard cloud map of the bus.

the crowding level in the bus increases from Gaizhou Street to Zhilin Park, with the service level deteriorated from LOS-C to LOS-F. From Zhiyuan Park to Software Park, the crowding level in the bus has been decreased, and the service level changed from LOS-F to LOS-C.

## 5. Conclusion

(1) This paper employed the cloud model to combine the standing-passenger density with the loading frequency to determine the crowded coefficient. The method realized the change of the crowded coefficient from qualitative analysis to quantitative analysis.

(2) The cloud model can overcome the singularity of the index and ambiguity of the congestion state division, and it has certain adaptability in the congestion state division.

(3) Future research needs to make further exploration. It needs to cover a boarder range of bus routes that

have different route lengths, numbers of stations, passenger demand levels, and vehicle capacities. In addition, it is necessary for future research to develop algorithms for applying the estimated crowding coefficient to the optimization of bus scheduling. According to the different congestion levels of bus stations, it is a new direction for future research to reasonably control the departure interval of vehicles and improve the service levels in the bus during peak hours.

## Data Availability

The data used to support the findings of this study are included within the article. The data are shown in Figures 1 and 2. No external data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.



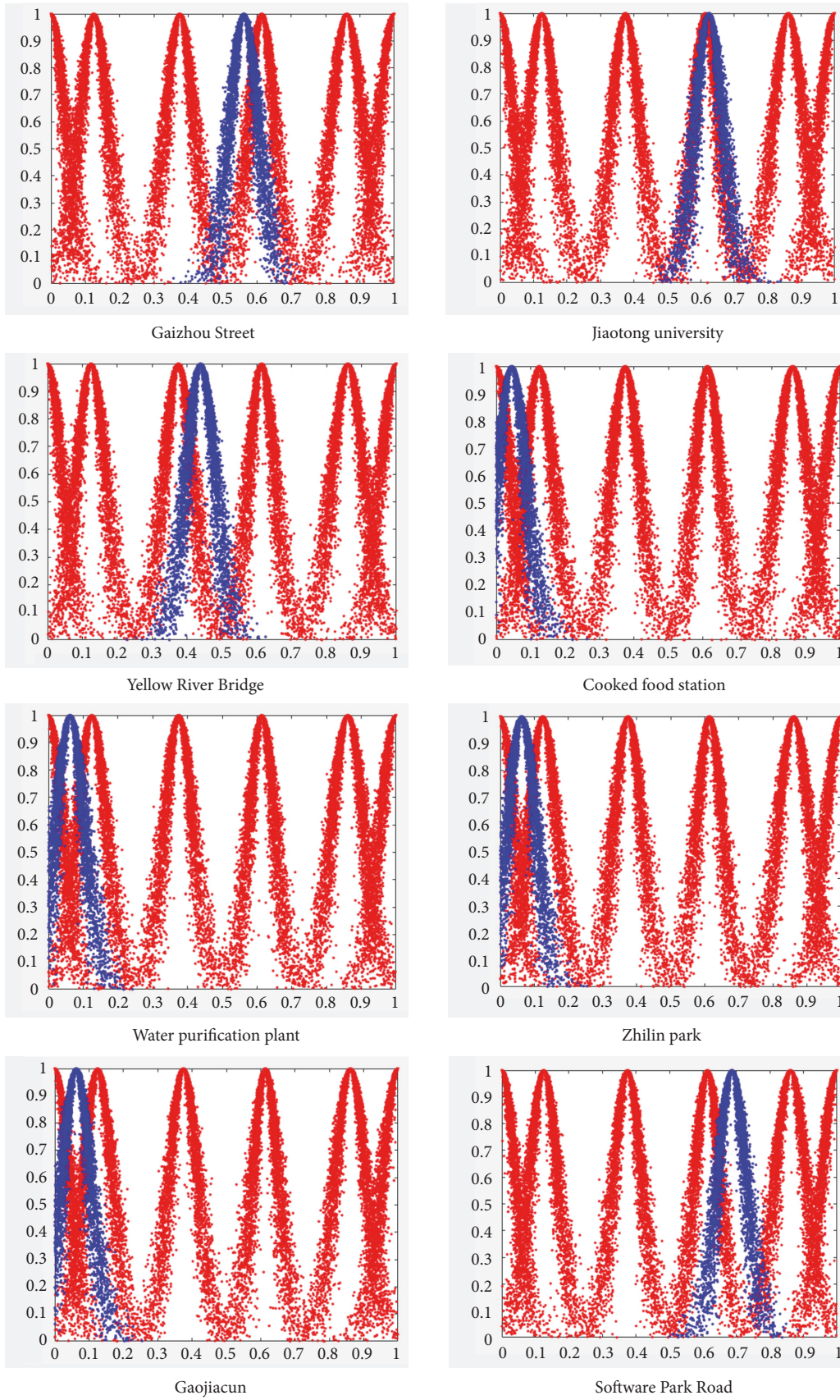


FIGURE 5: Identified cloud and template cloud within the bus.



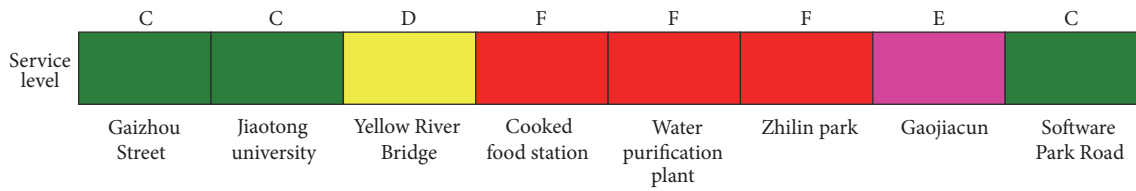


FIGURE 6: Division of service level within the bus.

TABLE 7: Performance evaluation results.

Evaluation Method	MAE	MSE	MARE	MSRE
Results	2.206	3.118	0.249	0.421

TABLE 8: In-vehicle congestion coefficient evaluation index cloud.

indicator 1	LOS	Ex	En	He	indicator 2	LOS	Ex	En	He
Standing-passenger Density	A	1	0.042	0.010	Loading Frequency	A	1	0.050	0.010
	B	0.875	0.042	0.010		B	0.85	0.050	0.010
	C	0.625	0.042	0.010		C	0.6	0.033	0.010
	D	0.375	0.042	0.010		D	0.375	0.042	0.010
	E	0.125	0.042	0.010		E	0.125	0.042	0.010
	F	0	0.042	0.010		F	0	0.042	0.010

TABLE 9: In-vehicle evaluation index standard cloud.

LOS	Ex	En	He
A	1	0.046	0.010
B	0.861	0.046	0.010
C	0.614	0.0375	0.010
D	0.375	0.042	0.010
E	0.125	0.042	0.010
F	0	0.042	0.010

TABLE 10: The load factor and standing-passenger density corresponding to the predicted number of people.

Station number	Station name	Loading Frequency	Standing-passenger Density
8	Gaizhou Street	1.3	2
9	Jiaotong University	1.25	1.7
10	Yellow River Bridge	1.525	3.5
11	Cooked food station	2.0	6.7
12	Water purification plant	1.975	6.5
13	Zhilin park	1.95	6.3
14	Gaojiacun	1.975	6.5
15	Software Park Road	1.25	1.7

TABLE 11: Similarity, possibility degree, and crowding degree of the selected bus stations.

Bus Station Number	Similarity	Possible Degree	Crowding Degree
8	[0.00, 0.00, 0.4284, 0.0065, 0.00, 0.00]	[0.00, 0.00, 0.985, 0.0148, 0.00, 0.000]	60.2645
9	[0.00, 0.0011, 0.6578, 0.00, 0.00, 0.00]	[0.00, 0.0017, 0.9979, 0.00, 0.00, 0.00]	59.9761
10	[0.00, 0.00, 0.0079, 0.3895, 0.00, 0.00]	[0.00, 0.00, 0.0200, 0.9800, 0.00, 0.00]	79.6019
11	[0.00, 0.00, 0.00, 0.00, 0.2721, 0.5259]	[0.00, 0.00, 0.00, 0.00, 0.3410, 0.6590]	113.18
12	[0.00, 0.00, 0.00, 0.00, 0.3937, 0.4005]	[0.00, 0.00, 0.00, 0.00, 0.4957, 0.5043]	110.059
13	[0.00, 0.00, 0.00, 0.00, 0.3951, 0.3978]	[0.00, 0.00, 0.00, 0.00, 0.4983, 0.5017]	110.034
14	[0.00, 0.00, 0.00, 0.00, 0.4043, 0.3904]	[0.00, 0.00, 0.00, 0.00, 0.5088, 0.4912]	109.825
15	[0.00, 0.000, 0.6631, 0.000, 0.00, 0.00]	[0.00, 0.0014, 0.9981, 0.00, 0.00, 0.00]	59.9800

## Acknowledgments

This study is supported by the Liaoning Provincial Department of Education Research Projects (JDL2017034).

## References

- [1] Morpace International and Cambridge Systematics, "A handbook for measuring customer satisfaction and service quality," TCRP Report 47, Transportation Research Board, National Research Council, Washington, DC, USA, 1999.
- [2] C. Lu, "The analysis of compressed force in crowds," *Journal of Transportation Systems Engineering and Information Technology*, vol. 7, no. 2, pp. 98–103, 2007.
- [3] X. Liu, J. Li, and W. Li, "Crowd density and advance warning management," *Safety*, vol. 30, pp. 19–22, 2009.
- [4] L. Ran and M. Liu, "Effect of crowded people density on crushing fatalities," *Journal of Safety and Environment*, vol. 7, pp. 135–138, 2007.
- [5] Y. Chen, P. Li, and G. Zhang, "Real-time evaluation method of passenger flow congestion in integrated transportation hubs," *Journal of Highway and Transportation Research and Development*, vol. 29, pp. 75–80, 2017.
- [6] J. He and X. Liu, "Optimization method of subway emergency evacuation path based on congestion degree," *China Safety Science Journal*, vol. 23, pp. 166–171, 2013.
- [7] S. Jiang, Y. Sun, and Y. Du, "The influence of congestion degree on the willingness to choose public transportation modes," *Journal of Tongji University (Natural Science)*, vol. 40, pp. 1831–1835, 2012 (Chinese).
- [8] Q. Xu, *Passenger Crowding Perception and Behavioral Dynamics Modeling of Urban Rail Transit Stations*, Beijing Jiaotong University, 2014.
- [9] S. Yu, X. Chen, and W. Sun, "A robust method for detecting and counting people," in *Proceedings of the ICALIP 2008 - 2008 International Conference on Audio, Language and Image Processing*, pp. 1545–1549, IEEE, July 2008.
- [10] S. Tian, *Research on Moving Human Body Recognition and Tracking Algorithm Based on Video Image*, Changsha University of Science and Technology, 2014.
- [11] X. Chen, *Research on Target Recognition of Bus Passengers Based on Color Image*, Beijing Jiaotong University, 2011.
- [12] L. Hou, *Research on Automatic Passenger Counting Algorithm Based on Video*, Chang'an University, 2013.
- [13] S. Mukherjee, B. Saha, I. Jamal, R. Leclerc, and N. Ray, "A novel framework for automatic passenger counting," in *Proceedings of the 2011 18th IEEE International Conference on Image Processing, ICIP 2011*, pp. 2969–2972, IEEE, September 2011.
- [14] G. Garcia-Bunster, M. Torres-Torriti, and C. Oberli, "Crowded pedestrian counting at bus stops from perspective transformations of foreground areas," *IET Computer Vision*, vol. 6, no. 4, pp. 296–305, 2012.
- [15] W. Daley, C. Usher, and O. Arif, "Detection of vehicle occupants in HOV lanes: exploration of image sensing for detection of vehicle occupants," in *Proceedings of the IS&T/SPIE Electronic Imaging*, pp. 86630S–86630S-15, International Society for Optics and Photonics, 2013.
- [16] M. Miklasz, P. Olszewski, and A. Nowosielski, "Pedestrian traffic distribution analysis using face recognition technology," in *Proceedings of the International Conference on Transport Systems Telematics*, pp. 303–312, Springer, Berlin, Germany, 2013.
- [17] D. Mudoj and P. A. Kashyap, "Vision based data extraction of vehicles in traffic," in *Proceedings of the 1st International Conference on Signal Processing and Integrated Networks, SPIN 2014*, pp. 202–208, IEEE, February 2014.
- [18] Y. Feng, K. N. Ngan, T. Sikora, A. L. Harvey, and M. Sun, "New discrete representation method for passenger counting system," in *Proceedings of the Visual Communications and Image Processing 2000*, pp. 1057–1065, International Society for Optics and Photonics, Perth, Australia, 2000.
- [19] T. Yahiaoui, C. Meurie, L. Khoudour, and F. Cabestaing, "A people counting system based on dense and close stereovision," in *Proceedings of the International Conference on Image and Signal Processing*, vol. 5099 of *Lecture Notes in Computer Science*, pp. 59–66, Springer, Berlin, Germany, 2008.
- [20] KFH Group, *Transit Capacity and Quality of Service Manual*, TRB, Washington, DC, USA, 3th edition, 2013.
- [21] D. Li, H. Meng, and X. Shi, "Membership clouds and membership cloud generators," *Journal of Computer Research and Development*, vol. 32, pp. 15–20, 1995.
- [22] D. Y. Li, C. Y. Liu, and L. Y. Liu, "Study on the universality of the normal cloud mode," *International Journal of Engineering Science*, vol. 6, no. 8, pp. 28–34, 2004.
- [23] D. Li and H. Meng, "Subordinate to cloud and subordinate cloud generator," *Journal of Computer Research and Development*, vol. 32, pp. 15–20, 1995.
- [24] J. Lu and Z. Qian, "The prediction method of normal cloud association rules," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, pp. 383–386, 2000.
- [25] H. Huang and W. Wang, "Study on subjective trust evaluation model based on membership cloud theory," *Journal of Communications*, vol. 29, pp. 13–19, 2008.

## Research Article

# A Two-Step Approach for Airborne Delay Minimization Using Pretactical Conflict Resolution in Free-Route Airspace

Ramazan Kursat Cecen <sup>1</sup> and Cem Cetek<sup>2</sup>

<sup>1</sup>Anadolu University, Eskisehir, Turkey

<sup>2</sup>Eskisehir Technical University, Eskisehir, Turkey

Correspondence should be addressed to Ramazan Kursat Cecen; ramazankursatcecen@anadolu.edu.tr

Received 27 November 2018; Accepted 6 March 2019; Published 1 April 2019

Guest Editor: Jesica de Armas

Copyright © 2019 Ramazan Kursat Cecen and Cem Cetek. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study proposes a two-step solution approach for aircraft conflict resolution and fuel consumption due to resolution maneuver occurring in free-route airspace. This model aims to provide a mathematical basis for a decision-support system that is used during the pretactical conflict resolution in air traffic management. Mathematical model of the first step presents alternative entry points on both sides of existing sector entry points to minimize delays by directing aircraft to the most convenient entry points. The second step suggests a vector deflection maneuver to minimize extra fuel consumption caused by conflict resolution. GAMS/CPLEX solver is used to solve the first step of the model but the solution is not produced in a reasonable time. To obtain feasible solutions, genetic algorithm and tabu search algorithms are implemented in the first step. Small size test problems are generated to evaluate the metaheuristic algorithms, and results are compared with GAMS/CPLEX solver solutions. According to this comparison, both metaheuristics algorithms produce near optimal solutions in a reasonably short time. The proposed approach has made significant improvements for airborne delays and extra fuel consumption caused by aircraft conflicts resolution in large-scaled airspaces.

## 1. Introduction

Air transportation has evolved to be a global industry with its large-scaled, complex, and rapidly growing nature. It has a unique impact on commercial and economic growth not only because it provides rapid world-wide transportation critical for international business, trade, and tourism, but also because it facilitates regional economic and social growth. Airlines carried around 4.1 billion passengers with revenue passenger kilometers over 7.7 trillion as well as 6 trillion-US\$-value of cargo in 2017 [1]. The total air traffic passenger demand has doubled in the last fifteen years, and 4.4% average annual growth is expected over the next twenty years [2]. This rapid growth in the demand overwhelms the current airspace capacities of air traffic management (ATM) system which is responsible for safe, efficient, and economic operations of flights within this global network. This imbalance between the capacity and demand leads to more airborne delays and congestions inducing increased operational costs, environmental impacts, customer dissatisfaction, and air traffic

controllers' workload. Therefore, the improvement of the current airspace capacities is addressed as a critical issue for the sustainable growth of the industry in ICAO's long-term Global Air Navigation Plan [3] as well as European Union's Single European Sky ATM Research (SESAR) objectives [4].

Free-routing of aircraft is one of the effective methods to enhance airspace capacities as well as efficiency of flight operations based on the use of user-preferred trajectories. Free-route airspace (FRA) allows airlines to freely plan their flights between a defined entry and exit points without reference to the current air traffic service route networks while flights still oblige to follow the instructions of air traffic controllers [5]. Despite its advantages, FRA increases the traffic complexity and conflict detection due to the increase in the number of intersection points; therefore, the entire airspace becomes a potential "hot spot" [6]. Pretactical conflict detection and resolution (CDR) using improved decision-support systems can be one of the alternatives to handle this traffic complexity efficiently within FRAs.

Aircraft operating in the same airspace should maintain predefined safe separation distances between each other. These minimum separation distances are accepted as 5 nautical miles horizontally and 1000 feet vertically for en-route airspace. If air traffic controllers detect any loss of separation, they apply suitable conflict resolution maneuvers to aircraft in order to prevent any risk of collision. There are three main resolution maneuvers used for the collision avoidance between aircraft pairs: heading change, airspeed change, and flight level change. While heading and speed change maneuvers involve the adjustment of the direction or the magnitude of the airspeed vector in the horizontal plane, flight level change deals with climbing or descending aircraft in the vertical plane. Conflict resolution (CR) problem involves the search of one (or combination) of these maneuvers to ensure safe separation between aircraft. Numerous studies have proposed different mathematical models and approaches for CR problems, and detailed reviews were provided by Kuchar and Yang [7] and Martin-Campo [8]. Rodionova and Sridhar [9] classified these approaches as tactical (performed within 30 minutes prior to the conflict), pretactical (performed up to 2 hours prior to the conflict), and strategic (performed more than 2 hours prior to the conflict) CR problems.

This study addresses pretactical CR between aircraft pairs in the horizontal plane using heading change maneuvers within the time period of 20 to 60 minutes prior to the potential conflicts within the FRA. The model focuses on deterministic aircraft motion. A two-step approach is developed to minimize total airborne delay and extra fuel consumption per aircraft due to the required heading change resolution maneuver, respectively. The first step presents a mixed-integer linear optimization model along with two metaheuristics: genetic algorithms and tabu search, while the second step uses a nonlinear programming (NLP) model. Mixed-integer programming has been extensively applied to aircraft CDR problems. Pallottino et al. [10] developed two different mixed-integer linear models using either airspeed change or heading change resolution maneuvers. Christodoulou and Costoulakis [11] proposed a mixed-integer linear model combining these two resolution maneuvers for small-scale problems. Vela et al. [12, 13] presented mixed-integer linear models to minimize fuel burn using combined airspeed and altitude change resolutions and airspeed and heading angle change resolutions, respectively. Alonso-Ayuso et al. [14] introduced a mixed 0-1 linear optimization model resolving conflicts and returning aircraft to their original routes via airspeed and altitude changes. Later on Alonso-Ayuso et al. [15] also proposed two MILP models using just altitude change and combined altitude and airspeed changes. Cafieri and Durand [16] proposed a MINLP formulation allowing aircraft to change their airspeeds to avoid conflicts. Omer [17] formulated a MILP model using both airspeed and heading change based on space discretization of aircraft trajectories. Cafieri and Rey [18] proposed a MINLP model that adapts speed adjustment method to maximize the largest conflict free aircraft set. Besides these exact solution models, metaheuristics approaches such as genetic algorithm [19], ant colony optimization [20, 21], and particle swarm optimization [22] were presented to obtain

good and feasible solutions in a short time; nonetheless they cannot guarantee globally optimal solution.

This study presents an alternative approach for pretactical CR based on flexible airspace entry point assignment which basically attempts to resolve conflict with no airborne delay by changing conflict geometries in the first step. The proposed approach ensures the safe separation using space discretization technique which allows us to focus on the critical points including entry, exit, and route intersection points instead of searching the whole airspace. The model checks all possible pairwise conflicts on trailing, crossing, merging, and diverging routes using these critical points. If any airborne delay is required to resolve aircraft conflicts, the second step searches a feasible vector deflection model corresponding to the airborne delay with minimum total fuel consumption per each aircraft before entering to the airspace. The vector deflection maneuver is a nonlinear model which includes the effects of bank angle changes as well as aerodynamic and propulsive characteristics on fuel consumption rate. The model determines bank angle, deflection angle, and vector maneuver distance within the lower and upper bounds to provide fuel optimal vector maneuver for the given airborne delay. The proposed two-step model aims to provide a mathematical basis for a decision-support system that can be used during pretactical control of flights in air traffic management and, therefore, metaheuristics are implemented to the first step of the model to obtain good and feasible solutions in a reasonable time.

## 2. Problem Statement

An airspace is any volume of the earth's atmosphere of defined dimensions which accommodates flights with or without air traffic control services. Airspaces can be classified as controlled, uncontrolled, and special airspaces according to air traffic services provided and flight requirements. Flight operations taking place in controlled airspaces receive air traffic control services according to the airspace types such as airport zones (CTR), terminal control areas (TMA), and en-route airspaces [23]. En-route airspaces cover the largest portion of the controlled airspace where flights in the climb, cruise, and descend phases are monitored and controlled by the relevant Area Control Center (ACC). The conventional en-route airspaces include a network of fixed waypoints and routes which aircraft have to follow during their flights. FRA, on the other hand, allows aircraft to choose their routes freely between the predefined airspace entry and exit points. Figure 1 presents a generic FRA with predefined boundaries, entry points (i.e.,  $EP_1, EP_2, \dots, EP_k$ ), and exit points (i.e.,  $XP_1, XP_2, \dots, XP_l$ ).

*2.1. Flexible Entry Point Approach for FRA.* This study adopts a flexible entry point approach which suggests adding two alternative entry points on both sides of the existing entry points with a certain distance (i.e., 10 NM), while preserving the existing boundaries and existing entry and exit points of FRA (Figure 2). Each existing entry point with its alternatives forms an entry area (i.e.,  $EA_1, EA_2, \dots, EA_j$ ). This approach, therefore, enables aircraft to avoid all potential pairwise

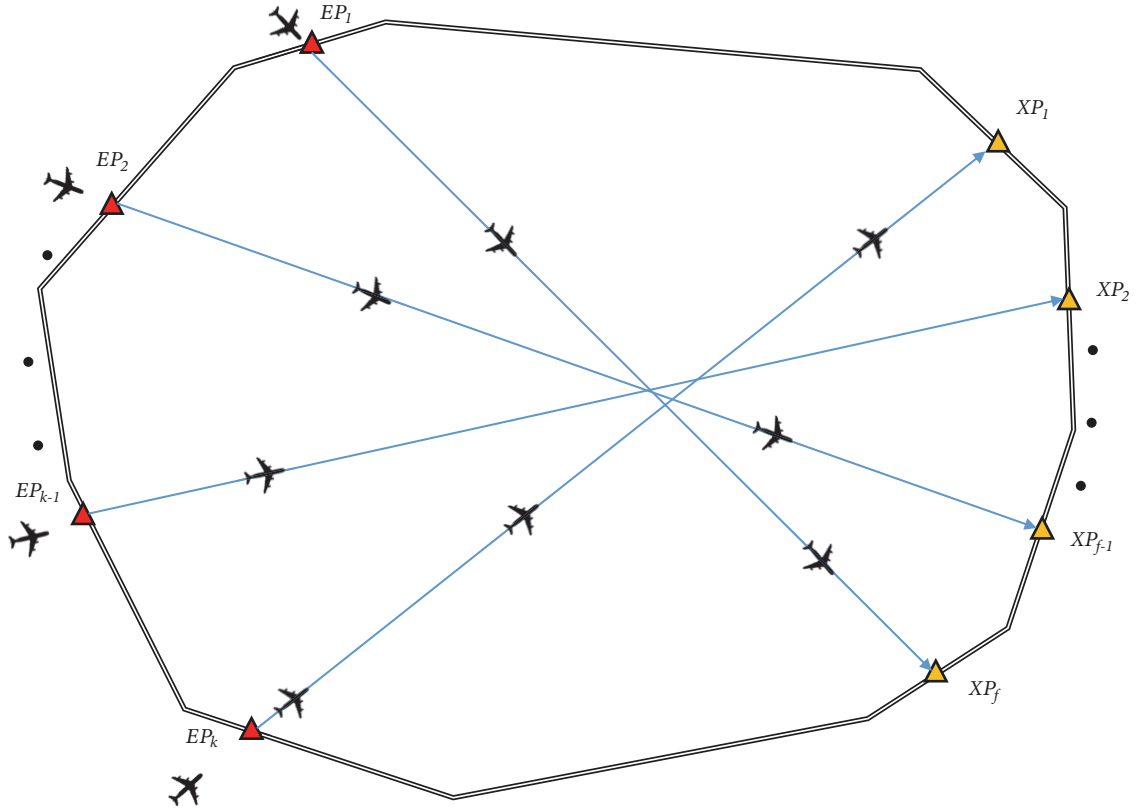


FIGURE 1: A generic free-route airspace (FRA) configuration.

conflicts prior to entering airspace by assigning them appropriate entry points within the entry areas. The flexible entry point assignment resolves pairwise aircraft conflicts through change of route intersection geometry which alters minimum separation time between aircraft pairs. A suitable entry point assignment combination ensures the reduction of airborne delays required for conflict resolution between aircraft pairs.

Two different types of conflicts can emerge between aircraft during their flight within the described airspace in the horizontal plane: crossing conflicts and trailing conflicts (Figure 3). The crossing conflicts are checked for three different intersection geometries: conflicts in intersecting, converging, and diverging routes (Figures 3(a)–3(c)) to calculate the minimum separation time between aircraft that fly over the same conflict points.

Each conflict point requires a specific time separation between aircraft pairs depending on aircraft speeds and encounter geometry. This time separation is described by [24] as follows.

$$T_{ii'} = \frac{D_{\min}}{V_i V_{i'} |\sin(\theta_{ii'})|} \sqrt{(V_i)^2 + (V_{i'})^2 - 2V_i V_{i'} \cos(\theta_{ii'})} \quad (1)$$

In (1),  $D_{\min}$  is the minimum separation distance,  $V_i$  and  $V_{i'}$  are airspeeds of leading and trailing aircraft, respectively, and  $\theta_{ii'}$  is the route crossing angle. In order to ensure safe separation between aircraft, one aircraft should be delayed by  $T_{ii'}$  prior to airspace entry point (pretactical level). Airspeeds  $V_i$  and  $V_{i'}$

are assumed to be optimal (best range) airspeeds at the given flight level and they depend on aircraft performance category (APC). In this study, aircraft are classified into three different performance categories: regional jet (RJ), narrow-body jet (NB), and wide-body jet (WB). Trailing conflicts may occur between aircraft flying on the same routes. No overtaking is allowed within the airspace; therefore the separation is maintained by delaying trailing aircraft in pretactical level. Aircraft are assumed to fly with optimal cruise airspeeds at the given flight level.

**2.2. Vector Deflection Maneuver.** A vector deflection maneuver is proposed to resolve aircraft conflicts in pretactical level (Figure 4). The maneuver consists of two phases: steady coordinated turns and steady straight flights with zero bank angle in the horizontal plane. The vector maneuver uses deflection angle ( $\psi_i$ ), bank angle ( $\phi_i$ ), and maneuver distance ( $l_i$ ) as decision variables. Other variables such as turning radius ( $r_i$ ), distance traveled along the arc ( $a_i$ ), projected arc distance on the undeflected route ( $b_i$ ), and deflected straight route ( $l_{di}$ ) are calculated based on these decision variables which are limited in the model with upper and lower values due to operational constraints. Bank angle variations for vectored and nonvectored aircraft are presented in Figure 5.

Aircraft fuel consumption for a given airspeed during the cruise operation depends on distance traveled and propulsive characteristics of aircraft. Turning maneuvers, on the other



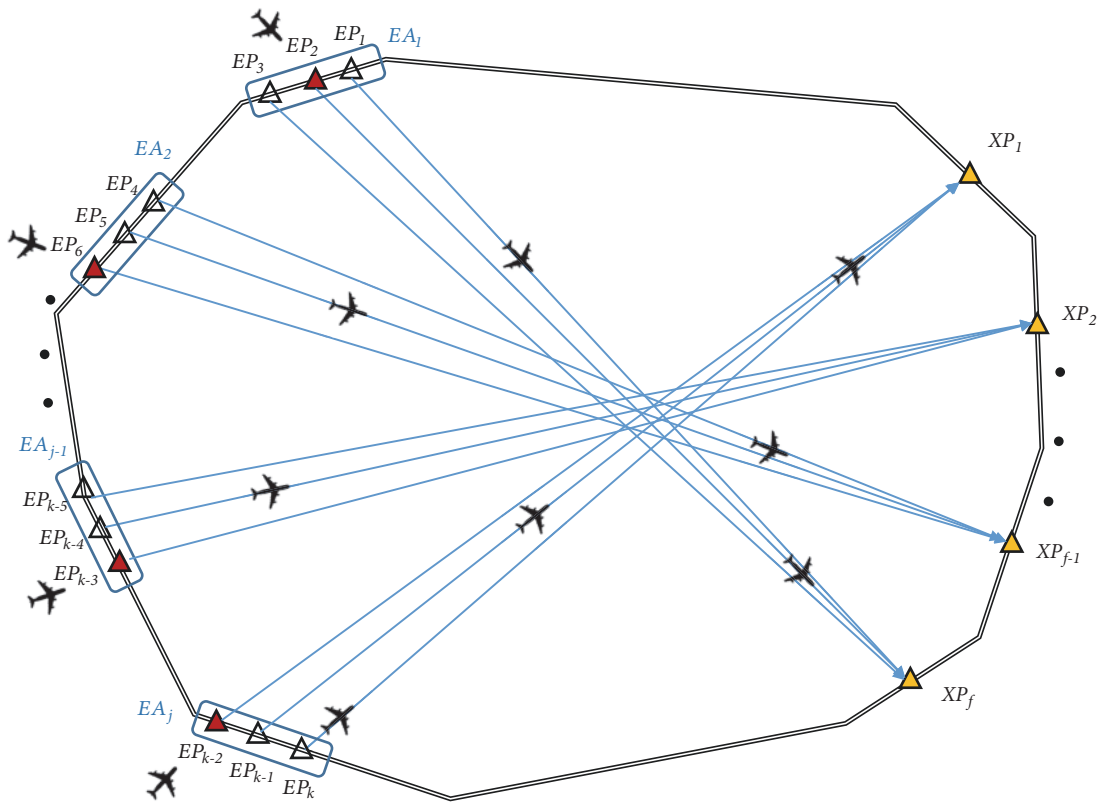


FIGURE 2: The flexible entry point airspace geometry.

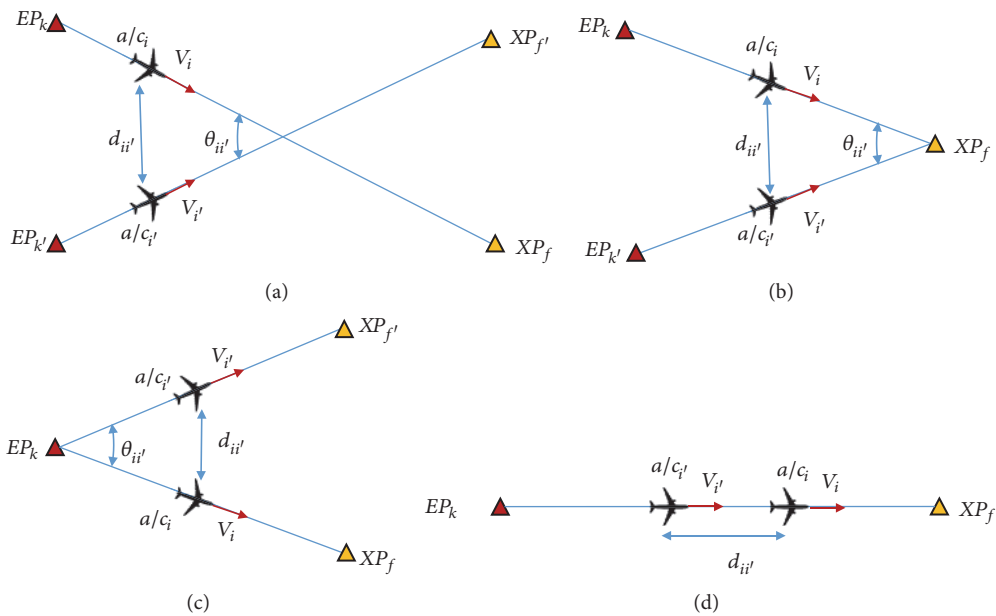


FIGURE 3: Conflict geometries: crossing conflicts (a) in intersecting routes, (b) converging routes, and (c) diverging routes and (d) trailing conflict in the same route.

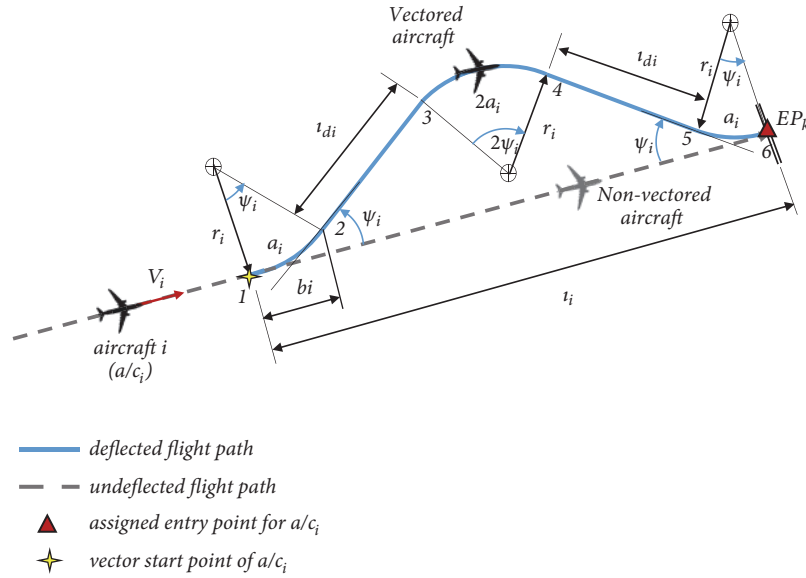


FIGURE 4: Vector deflection maneuver.

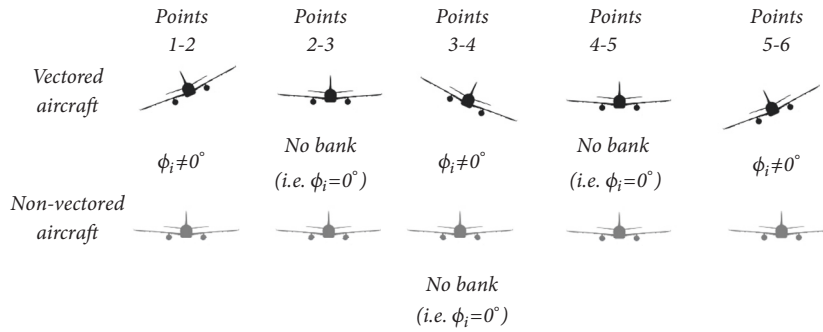


FIGURE 5: Bank angle variation of vectored and nonvectored aircraft along their flight paths.

hand, create extra fuel burnt not only due to the extended flight path but also due to the increase in the bank angle which results in higher load factors. The proposed model takes bank angle effects into consideration while calculating the fuel consumption. The extra fuel burnt per aircraft along the deflected flight path can be described as follows.

$$w_f = FC_t \cdot 4a_i + FC_{nb} \cdot 2l_{di} - FC_{ud} \cdot l_i \quad (2)$$

While the first two terms in (2) correspond to the amount of fuel burnt along the deflected flight path including total distance traveled along arcs and deflected straight routes, the latter corresponds to the amount of fuel burnt along vector maneuver distance. Therefore,  $FC_{nb}$ , and  $FC_{ud}$  are fuel consumption rates per unit distance during the straight level flight. For the given aircraft performance category, airspeed, and flight level, these values are constant such that

$$FC_{nb} = FC_{ud} = \chi_{v,o} \quad (3)$$

$FC_t$  is the fuel consumption rate per unit distance during the coordinated turn and it can be expressed in terms of a

third order polynomial of bank angle using curving fitting technique such that

$$FC_t = \chi_{v,o} + \chi_{v,1} \cdot \phi_i + \chi_{v,2} \cdot \phi_i^2 + \chi_{v,3} \cdot \phi_i^3 \quad (4)$$

In (3) and (4),  $\chi_{v,o}$ ,  $\chi_{v,1}$ ,  $\chi_{v,2}$ , and  $\chi_{v,3}$  are regression coefficients found in kg/NM for the performance category  $v$  at the given altitude.

### 3. Mathematical Model

Figure 6 presents the general methodology of the proposed two-step solution approach for minimum airborne delay and extra fuel consumption due to vector deflection maneuvers. In order to evaluate the proposed flexible entry point approach, a baseline case representing fixed entry point FRA is generated as the reference. Estimated time of arrivals (ETA), aircraft performance category (APC), and airspace exit points (XP) of each flight are generated randomly as inputs for both cases. A predefined entry point is provided for each flight in the baseline case while flexible entry point approach assigns each flight to the most suitable entry point within the predefined entry area. In the baseline case, the

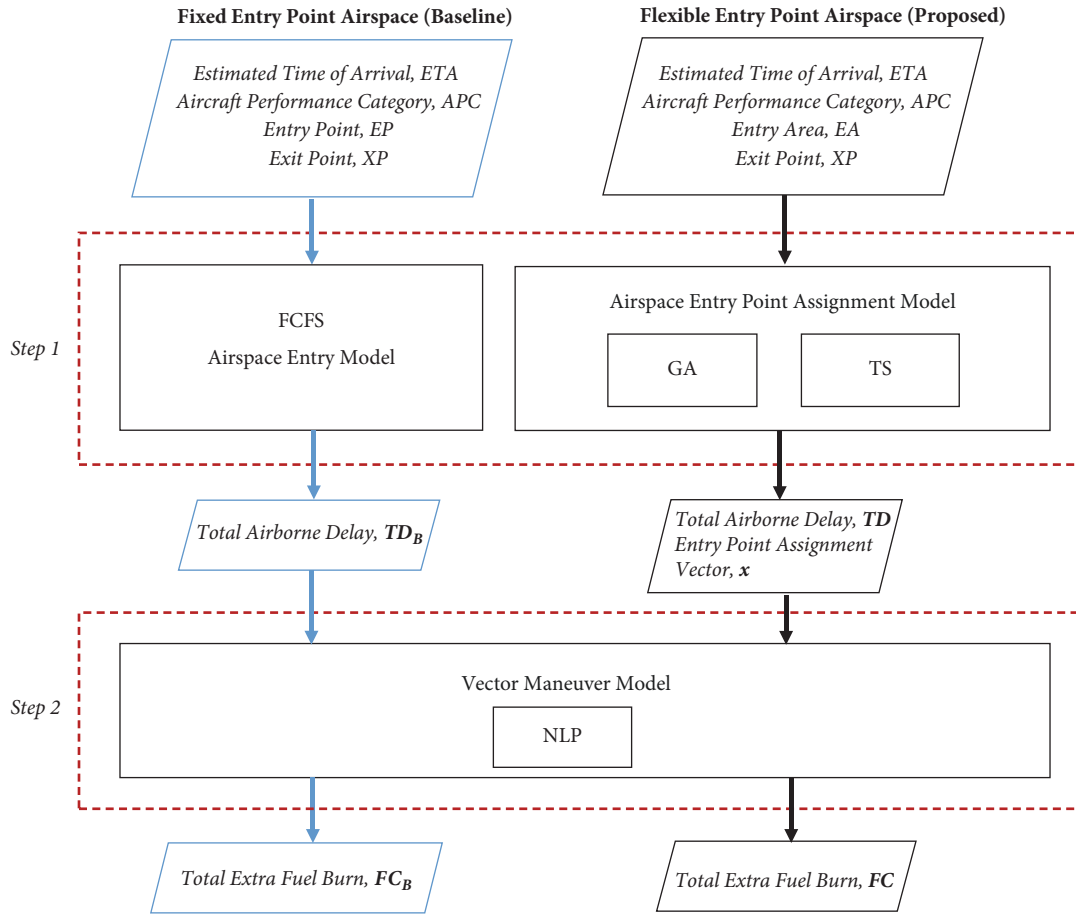


FIGURE 6: Flow diagram of the two-step model.

total and individual airborne delays are estimated based on first come first served (FCFS) discipline. Flexible entry point assignment also obtains the total and individual airborne delays as well as entry point assignment vector using genetic algorithms (GA) and tabu search (TS) metaheuristics. Upon the estimation of airborne delays, minimum fuel consumption per aircraft is estimated using vector deflection model for both cases.

**3.1. First Step: Entry Point Assignment Model.** In this section entry point assignment model is presented. The aim is to assign each aircraft to the most suitable entry point according to their entry area to avoid conflict when aircraft is entering into the sector. The main idea of avoiding conflicts is changing conflict geometries just by adjusting entry point. The model seeks for the most appropriate entry point assignment for aircraft, and if it is necessary, the model imposes airborne delay to aircraft. The following assumptions are imposed on the model in addition to the ones described in Section 2.1:

- (1) Each aircraft should be assigned to an entry point within its predefined entry area.
- (2) Adjacent entry points in the same entry area are located 10 NM apart.
- (3) Airspeed and altitude of each aircraft are constant during their flight within the FRA.

Therefore, sets, parameters, variables, objective function, and constraints are given as follows:

*Sets*

$I$ : Set of aircraft  $i, i_1, i_2 \in I$

$J$ : Set of entry areas  $j, \in J$

$K$ : Set of entry points  $k, k_1, k_2 \in K$

$L$ : Set of exit points  $l, l_1, l_2 \in L$

$V$ : Set of aircraft performance category  $v, v_1, v_2 \in V$

$N$ : Set of intersection points  $n, \in N$ .

*Parameters*

$M$ : Big number enough

$D_{min}$ : Minimum separation distance between aircraft

$g_i$ : Scheduled airspace entry time of aircraft  $i$

$t_i$ : Performance category of aircraft  $i$

$r_i$ : Entry area of aircraft  $i$

$e_i$ : Exit point of aircraft  $i$

$bg_i$ : Entry point of aircraft  $i$  for baseline scenario

$h_v$ : Airspeed of performance category  $v$

$d_{kn}$ : Distance between entry point  $k$  and intersection point  $n$

$td_{kl}$ : Distance between entry point  $k$  and exit point  $l$

$sreq_{v_1 v_2 n}$ : Separation time on intersection point  $n$  between aircraft performance category  $v_1$  and  $v_2$

$treq_{k_1 k_2 v_1 v_2 l}$ : Separation time at exit point  $l$  between aircraft flying from entry point  $k_1$  and  $k_2$  with performance category  $v_1$  and  $v_2$

$breq_{l_1 l_2 v_1 v_2 k}$ : Separation time at entry point  $k$  between aircraft flying to exit point  $l_1$  and  $l_2$  with performance category  $v_1$  and  $v_2$

$u_{jk}$ : 0-1 parameter that is one if entry point  $k$  is in entry area  $j$ ; otherwise, it is zero

$o_{kln}$ : 0-1 parameter that is one if an aircraft flies from entry point  $k$  to exit point  $l$  via intersection point  $n$ ; otherwise, it is zero.

### Variables

$q_{ik}$ : Actual entry time of aircraft  $i$  at entry point  $k$

$w_i$ : Airborne delay of aircraft  $i$

$p_{ikn}$ : Fly over time of aircraft  $i$  from entry point  $k$  at intersection point  $n$

$c_{ikl}$ : Fly over time of aircraft  $i$  from entry point  $k$  at exit point  $l$

$e_{1i_1 i_2}$ : 0-1 variable that is 1 if aircraft  $i_2$  flies over intersection point before aircraft  $i_1$ ; otherwise, it is zero

$b_{1i_1 i_2}$ : 0-1 variable that is 1 if aircraft  $i_2$  enters and leaves the airspace before aircraft  $i_1$ ; otherwise, it is zero

$b_{2i_1 i_2}$ : 0-1 variable that is 1 if aircraft  $i_2$  exits the airspace before aircraft  $i_1$ ; otherwise, it is zero

$b_{3i_1 i_2}$ : 0-1 variable that is 1 if aircraft  $i_2$  enters the airspace before aircraft  $i_1$ ; otherwise, it is zero

$x_{ik}$ : 0-1 variable that is 1 if aircraft  $i$  is assigned to entry point  $k$ ; otherwise, it is zero.

The first step of mathematical formulation is as follows.

$$\min \sum_i w_i \quad (5)$$

$$\text{Subject to } \sum_{k|u_{jk}=1} x_{ik} = 1 \quad \forall i, j \mid j = r_i \quad (6)$$

$$\sum_{k|u_{jk}=1, k=bg(i)} x_{ik} = 1 \quad \forall i, j \mid j = r_i \quad (7)$$

$$q_{ik} = g_i + w_i \quad \forall i, k \quad (8)$$

$$p_{ikn} = q_{ik} + \frac{d_{kn}}{h_v} \quad \forall i, k, n, v \mid v = t_i \quad (9)$$

$$c_{ikl} = q_{ik} + \frac{TD_{kl}}{h_v} \quad \forall i, k, l, v \mid v = t_i \quad (10)$$

$$q_{i_2 k} - q_{i_1 k} \geq \frac{D_{min}}{h_v} - (2 - x_{i_1 k} - x_{i_2 k}) \cdot M - b_1(i_1, i_2) \cdot M \quad \forall i_1, i_2, k, v \mid i_1 \neq i_2, v = t_{i_1}, e_{i_1} = e_{i_2} \quad (11)$$

$$q_{i_1 k} - q_{i_2 k} \geq \frac{D_{min}}{h_v} - (2 - x_{i_1 k} - x_{i_2 k}) \cdot M - b_1(1 - i_1, i_2) \cdot M \quad \forall i_1, i_2, k, v \mid i_1 \neq i_2, v = t_{i_2}, e_{i_1} = e_{i_2} \quad (12)$$

$$c_{i_2 kl} - c_{i_1 kl} \geq \frac{D_{min}}{h_v} - (2 - x_{i_1 k} - x_{i_2 k}) \cdot M - b_1(i_1, i_2) \cdot M \quad \forall i_1, i_2, k, v, l \mid i_1 \neq i_2, v = t_{i_2}, e_{i_1} = e_{i_2} \quad (13)$$

$$c_{i_1 kl} - c_{i_2 kl} \geq \frac{D_{min}}{h_v} - (2 - x_{i_1 k} - x_{i_2 k}) \cdot M - b_1(1 - i_1, i_2) \cdot M \quad \forall i_1, i_2, k, v, l \mid i_1 \neq i_2, v = t_{i_1}, e_{i_1} = e_{i_2} \quad (14)$$

$$q_{i_2 k} - q_{i_1 k} \geq breq_{l_1 l_2 v_1 v_2 k} - (2 - x_{i_1 k} - x_{i_2 k}) \cdot M - b_3(i_1, i_2) \cdot M \quad (15)$$

$$\forall i_1, i_2, l_1, l_2, k, v_1, v_2 \mid i_1 \neq i_2, v_1 = t_{i_1}, v_2 = t_{i_2}, e_{i_1} \neq e_{i_2}, l_1 = e_{i_1}, l_2 = e_{i_2}$$

$$q_{i_1 k} - q_{i_2 k} \geq breq_{l_1 l_2 v_1 v_2 k} - (2 - x_{i_1 k} - x_{i_2 k}) \cdot M - b_3(1 - i_1, i_2) \cdot M \quad (16)$$

$$\forall i_1, i_2, l_1, l_2, k, v_1, v_2 \mid i_1 \neq i_2, v_1 = t_{i_1}, v_2 = t_{i_2}, e_{i_1} \neq e_{i_2}, l_1 = e_{i_1}, l_2 = e_{i_2}$$

$$c_{i_2 k_2 l} - c_{i_1 k_1 l} \geq \text{treq}_{k_1 k_2 v_1 v_2 l} - (2 - x_{i_1 k_1} - x_{i_2 k_2}) \cdot M - b_2(i_1, i_2) \cdot M \quad (17)$$

$$\forall i_1, i_2, k_1, k_2, l, v \mid i_1 \neq i_2, v_1 = t_{i_1}, v_2 = t_{i_2}, e_{i_1} = e_{i_2}, k_1 \neq k_2, l = e_{i_1}$$

$$c_{i_1 k_1 l} - c_{i_2 k_2 l} \geq \text{treq}_{k_1 k_2 v_1 v_2 l} - (2 - x_{i_1 k_1} - x_{i_2 k_2}) \cdot M - b_2(1 - i_1, i_2) \cdot M \quad (18)$$

$$\forall i_1, i_2, k_1, k_2, l, v \mid i_1 \neq i_2, v_1 = t_{i_1}, v_2 = t_{i_2}, e_{i_1} = e_{i_2}, k_1 \neq k_2, l = e_{i_1}$$

$$p_{i_2 k_2 n} - p_{i_1 k_1 n} \geq \text{sreq}_{v_1 v_2 n} - (2 - x_{i_1 k_1} - x_{i_2 k_2}) \cdot M - e_1(i_1, i_2) \cdot M \quad (19)$$

$$\forall i_1, i_2, l_1, l_2, k_1, k_2, v_1, v_2 \mid i_1 \neq i_2, v_1 = t_{i_1}, v_2 = t_{i_2}, e_{i_1} \neq e_{i_2}, k_1 \neq k_2, l_1 = e_{i_1}, l_2 = e_{i_2}, o_{k_1 l_1 n} = 1, o_{k_2 l_2 n} = 1$$

$$p_{i_1 k_1 n} - p_{i_2 k_2 n} \geq \text{sreq}_{v_1 v_2 n} - (2 - x_{i_1 k_1} - x_{i_2 k_2}) \cdot M - e_1(1 - i_1, i_2) \cdot M \quad (20)$$

$$\forall i_1, i_2, l_1, l_2, k_1, k_2, v_1, v_2 \mid i_1 \neq i_2, v_1 = t_{i_1}, v_2 = t_{i_2}, e_{i_1} \neq e_{i_2}, k_1 \neq k_2, l_1 = e_{i_1}, l_2 = e_{i_2}, o_{k_1 l_1 n} = 1, o_{k_2 l_2 n} = 1$$

The objective function (5) of the model is to minimize the total airborne delay. Constraint set (6) ensures that every aircraft is assigned to one entry point which belongs to its entry area. Constraint set (7) ensures that every aircraft is assigned to its own baseline entry point. Constraint sets (8), (9), and (10) calculate the flyover time at the entry point, intersection point, and exit point of the airspace, respectively. Constraint sets (11), (12), (13), and (14) maintain the required separation time between all aircraft pairs for trailing conflicts. Constraint sets (15) and (16) maintain the required separation time between all aircraft pairs for conflicts in diverging routes. Constraint sets (17) and (18) maintain the required separation time for all conflicts in converging routes. Constraint sets (19) and (20) maintain the required separation time between all aircraft pairs for conflicts in intersecting routes. Therefore all possible conflicts between all aircraft pair combinations within the airspace are controlled by the model using the provided constraints.

**3.2. Second Step: Vector Deflection Model.** Although the first step calculates the minimum airborne delay to resolve conflicts, it does not specify which maneuver to be implemented to aircraft in order to avoid conflicts. The entry point assignment is capable of resolving many potential conflicts without requiring extra resolution maneuvers. In case of a necessity of an airborne delay, the proposed vector deflection model in Section 2.2 can be applied to aircraft prior to their entrance to the airspace. Certainly, there are infinitely many vector deflection maneuvers which can resolve the conflict for the given airborne delay. In this step, the model searches vector deflection resolutions with minimum fuel consumption per aircraft under given constraints as described in Section 2.2. The following assumptions are imposed on the model in addition to the ones described in Section 2.2:

- (1) Weight of all aircraft is constant during their vector deflection maneuver.
- (2) Airspeed and altitude of each aircraft are constant during the vector deflection maneuver.
- (3) Bank angle changes are done instantly during the coordinated turning maneuvers.

(4) Standard atmospheric conditions are valid.

(5) Wind speed and acceleration are zero.

The second step of mathematical formulation is as follows.

*Sets*

$I$ : Set of aircraft  $i, \in I$

$V$ : Set of aircraft performance category  $v, \in V$ .

*Parameters*

$g$ : gravitational acceleration

$t_i$ : Performance category of aircraft  $i$

$dx_i$ : Extra distance flown due to airborne delay  $i$

$h_v$ : Airspeed of performance category  $v$

$\chi_{v,o}$ : Fuel consumption coefficient with no bank angle for aircraft performance category  $v$

$\chi_{v,1}$ : Fuel consumption coefficient with first order bank angle for aircraft performance category  $v$

$\chi_{v,2}$ : Fuel consumption coefficient with second order bank angle for aircraft performance category  $v$

$\chi_{v,3}$ : Fuel consumption coefficient with third order bank angle for aircraft performance category  $v$ .

*Variables*

$R_i$ : Turn radius of aircraft  $i$

$\psi_i$ : Deflection angle of aircraft  $i$

$\phi_i$ : Bank angle of aircraft  $i$

$l_i$ : Maneuver distance of aircraft  $i$

$a_i$ : Arc distance traveled by aircraft  $i$  during turning maneuver

$b_i$ : Total distance traveled by aircraft  $i$  projected on the undeflected route during turning maneuver

$l_{di}$ : Distance traveled by aircraft  $i$  along the deflected route with no bank

$f_{i_i}$ : Fuel consumption of aircraft  $i$  during turning maneuver



TABLE 1: Cruise airspeed and fuel consumption rates of aircraft performance categories at 33000 feet.

Aircraft Performance Category	Airspeed, $h_v$ (knots)	Fuel Consumption Regression Coefficients (kg/NM)			
		$\chi_{v,o}$	$\chi_{v,1}$	$\chi_{v,2}$	$\chi_{v,3}$
Regional jet, RJ ( $v=1$ )	388	1.284	0.9147	0.05794	4.243
Narrow-body jet, NB ( $v=2$ )	426	1.371	0.9771	0.06189	5.282
Wide-body jet, WB ( $v=3$ )	482	6.416	4.572	0.2896	17.09

$f_{2_i}$ : Fuel consumption of aircraft  $i$  along the deflected route with no bank

$f_{3_i}$ : Fuel consumption of aircraft  $i$  along the undeflected route.

### Objective

$$\min \sum_i FC_{t_i} \cdot 4a_i + FC_{nb_i} \cdot 2l_{di} - FC_{ud_i} \cdot l_i \quad (21)$$

$$\text{Subject to } r_i = \frac{h_v^2}{g \cdot \tan(\phi_i)} \quad \forall i, v \mid v = t_i \quad (22)$$

$$a_i = r_i \psi_i \quad \forall i \quad (23)$$

$$l_{di} = \frac{l_i - b_i}{2 \cdot \cos(\psi_i)} \quad \forall i \quad (24)$$

$$b_i = r_i \cdot \sin(\psi_i) \quad \forall i \quad (25)$$

$$4a_i + 2l_{di} - l_i - dx_i = 0 \quad \forall i \quad (26)$$

$$4b_i \leq l_i \quad \forall i \quad (27)$$

$$FC_{t_i} = (\chi_{v,o} + \chi_{v,1} \cdot \phi_i + \chi_{v,2} \cdot \phi_i^2 + \chi_{v,3} \cdot \phi_i^3) \quad \forall i, v \mid v = t_i \quad (28)$$

$$FC_{nb_i} = (\chi_{v,o}) \quad \forall i, v \mid v = t_i \quad (29)$$

$$FC_{ud_i} = (\chi_{v,o}) \quad \forall i, v \mid v = t_i \quad (30)$$

$$0 \leq \psi_i \leq \frac{\pi}{2} \quad \forall i \quad (31)$$

$$0 \leq \phi_i \leq \frac{\pi}{6} \quad \forall i \quad (32)$$

$$0 \leq l_i \leq 40 \quad \forall i \quad (33)$$

The objective function of the second model (21) is to minimize total extra fuel consumption. Constraint set (22) calculates the turn radius for each aircraft. Constraint sets (23)-(25) calculate the distance traveled along the arc ( $a_i$ ), deflected straight route ( $l_{di}$ ), and projected arc distance ( $b_i$ ) on the undeflected route, respectively. Constraint set (26) guarantees that the distance traveled during vector deflection maneuver satisfies the required airborne delay. Constraint set (27) ensures that the sum of arc distances projected on undeflected route cannot be longer than vector maneuver distance.

While constraint set (28) estimates fuel consumption rates per unit distance during the coordinated turns, constraint sets (29)-(30) provide fuel consumption rates per unit distance during straight level flights along deflected and undeflected flight paths. Table 1 presents the cruise airspeed and fuel consumption rates per unit distance estimated based on the values provided in BADA [25] for each aircraft performance category at 33000 feet (FL330).

## 4. Metaheuristic Algorithms

Metaheuristic algorithms are an effective way of using trial and error methods to produce acceptable solutions to a complex problem. The complexity of the problems makes it difficult to evaluate all possible solutions. These algorithms aim to obtain good solution within an acceptable period of time, but they do not guarantee achieving the global optimum result. In this study, genetic algorithms and tabu search are proposed to solve the mathematical model presented in the first step.

**4.1. The Proposed Genetic Algorithm.** The entry point assignment model is a complex model so it is difficult to solve this problem by MIP solver within the reasonable time period. Thus, we present a genetic algorithm to deal with this problem. Genetic algorithms (GA) are search algorithms based on Darwin's theory of evolution and described by John Holland [26]. Genetic algorithms try to achieve the best solution by imitating natural selection. In genetic algorithms, each chromosome represents a solution and multiple chromosomes come together to form a population. Chromosomes pass their selection, crossing, and mutation steps to transfer their genes to the next generation. Highly compatible chromosomes are more likely to transfer genes to subsequent generations.

**4.1.1. Chromosomes Structure and Initial Population.** Chromosomes come from genes, and each gene refers to entry point for each aircraft. An example of chromosomes structure for six aircraft and twelve entry points can be seen in Table 2. Each gene belongs to only one aircraft and each aircraft is assigned to one entry point according to its entry area. During the process of obtaining initial population, each aircraft is assigned to an entry point randomly according to its entry area; therefore, infeasible gene structures can be prevented.

**4.1.2. Fitness Function.** According to entry point assignment, fitness function of all chromosomes is calculated. The algorithm checks for any conflicts between each aircraft pair.

TABLE 2: Chromosomes structure.

Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
EP11	EP 7	EP 4	EP 2	EP 9	EP 6

Parent1	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 7	EP 4	EP 2	EP 9	EP 6
Random Value	0.3	0.6	0.1	0.7	0.9	0.45
Parent2	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP12	EP 5	EP 6	EP 1	EP 10	EP 4

Offspring 1	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 5	EP 4	EP 1	EP 10	EP 6
Offspring 2	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP12	EP 7	EP 6	EP 2	EP 9	EP 4

FIGURE 7: An example of uniform crossover operator.

there is a conflict, the algorithm randomly decides which aircraft should be delayed. For this process, genetic algorithm generates a random number ranging from 0 to 1 for both aircraft, and the aircraft having the lower value is imposed to delay. This method allows different solutions in the search space to be reached. Finally, the total airborne delay of each chromosome is calculated.

**4.1.3. Selection Process and Elitism.** The selection process is a critical step to determine population diversity and selection bias that affect performance of GA significantly. If the selection pressure increases, the population diversity decreases. In the opposite case, if diversity increases in the population, good solutions are beginning to decrease. Keeping these two factors in balance is an important step for GA's success [27]. In this study, the selection is performed using the roulette wheel method that allows chromosomes with good adaptability to transmit more genes to the next generation. Elitism is used to reduce genetic drift by copying the best chromosomes into future generations in the final step.

**4.1.4. Crossover.** The crossover operator allows for the generation of new individuals as a result of mutual exchange of the genes of the chromosomes which will be crossed after the selection. With this method, some features of previous generations can be transferred to the new generation. The success of the crossover process depends on the proper coding of the chromosomes according to the probing. In this study, new chromosomes are produced through the use of reciprocal displacement of genes belonging to the same aircraft between selected chromosome pairs. This process is included in the literature as uniform crossover (Figure 7). In the crossover process, a value between 0 and 1 is produced randomly for each pair of genes, and if this value is greater than 0.5, the displacement takes place between the gene pairs.

**4.1.5. Mutation and Elitism.** The mutation operator modifies one or more gene values randomly to increase the diversity of the solution. Furthermore, the mutation operator enables genetic algorithms to converge a feasible solution more rapidly. As a result of the mutation process, a new chromosome structure is obtained by replacing the selected genes with another entry point in its entry region (Figure 8).

**4.2. The Proposed Tabu Search.** The tabu search (TS) algorithm was developed by Glover in 1986 [28]. The TS algorithm is an intuitive method that finds the best value by moving away from the local optimum. TS algorithm, which can be used to solve many different problems, has a flexible structure to produce the best or nearest solutions. By using the memory structure, some solutions are banned from being produced over a certain number of iterations. These solutions are added to tabu list and called "tabu". The primary task of tabu list is to ensure that the algorithm achieves the overall best value by avoiding local best values. In some cases, if a tabu move solution gives a better solution than the best found so far, this solution is no longer classified as tabu. By this way, the TS algorithm helps to improve the harmonization value by using aspiration criterion.

**4.2.1. Initial Solution.** The TS algorithm first needs an initial solution to start searching for a solution. Generating a good initial solution significantly improves the success of the algorithm. In this study, baseline airspace entry points are chosen as the initial solution (Table 3).

**4.2.2. Neighborhood Structure.** Generating neighborhood solutions helps to improve the existing solution by changing airspace entry points randomly within the entry area for selected aircraft. Generation process is shown in Table 4.

Individual	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 7	EP 4	EP 2	EP 9	EP 6
↑						
Offspring	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 8	EP 4	EP 2	EP 9	EP 6

FIGURE 8: An example of mutation operator.

TABLE 3: Initial solution representation.

Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
EP11	EP 5	EP 5	EP 2	EP 8	EP 11

TABLE 4: Generating neighborhood.

Existing Solution	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 8	EP 5	EP 2	EP 8	EP 5
Neighborhood Solution 1	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 7	EP 5	EP 2	EP 8	EP 5
Neighborhood Solution 2	Aircraft 1	Aircraft 2	Aircraft 3	Aircraft 4	Aircraft 5	Aircraft 6
	EP11	EP 8	EP 5	EP 3	EP 8	EP 4

TABLE 5: Tabu list demonstration.

Aircraft	Entry Points											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	0	0	0	0	0	0	0	2	0
2	0	0	0	0	0	0	3	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0

Table 4 shows how we obtain two neighborhood solutions. To achieve this process, the second and fourth aircraft are randomly selected, and two different neighbor solutions are produced in addition to the existing solution. In the first neighboring solution, the entry point of the second aircraft is changed randomly while the entry point of the fourth aircraft remains the same. Similarly, in the second neighboring solution, the entry point of the fourth aircraft is changed randomly while the entry point of the second aircraft remains the same in the second neighboring solution.

*4.2.3. Short-Term Memory.* Tabu list allows the search for the different points in the solution space by banning searched solutions for a certain number of iterations. The number of iterations specifies the length of the tabu list. Tabu list does not interfere to produce neighborhood solution production. However, after the fitness values are calculated, the solutions on the tabu list are called tabu and removed from neighboring solutions. This process is performed using short-term memory [28]. Both the aspiration criterion and the short-term memory structure are used in this study. Tabu list demonstration is shown in Table 5.

In Table 5, tabu length size is chosen as three. To explain the use of tabu list, an example of tabu list structure is given.

The entry point 7 for the second aircraft will be kept on the tabu list for three iterations. It is also forbidden to assign the first aircraft on the list to the entry point 11 and the fifth aircraft to the entry point 7. The number of entry points kept on the list varies according to the length of the tabu list.

## 5. Computational Results

The proposed GA and TS metaheuristics are developed to solve more complex airspace structures of which GAMS/CPLEX solver cannot reach a solution within reasonable time period. A set of small-sized test problems are generated in order to evaluate these metaheuristics. For all test problems, total airborne delays are calculated using GAMS/CPLEX solver, GA, and TS algorithm. The selected parameters of metaheuristics are presented in Table 6. These parameters are determined experimentally. The solution time of the all approaches is set to 900 seconds in order to provide a feasible solution in pretactical time window. A computer with 2.3 GHz Intel Core i7 processor and 16 GB RAM is used in all computations.

Thirty test problems are generated for 180 nm<sup>2</sup> en-route airspace (as presented in Figure 2) at FL330 with three

TABLE 6: The parameters of metaheuristics.

Genetic Algorithm		Tabu Search Algorithm	
Population Size	50	Neighborhood Solution Size	9
Selection Process	Roulette	Tabu List Size	7
Crossover Rate	0.8	Number of Generation	1110
Mutation Rate	0.1		

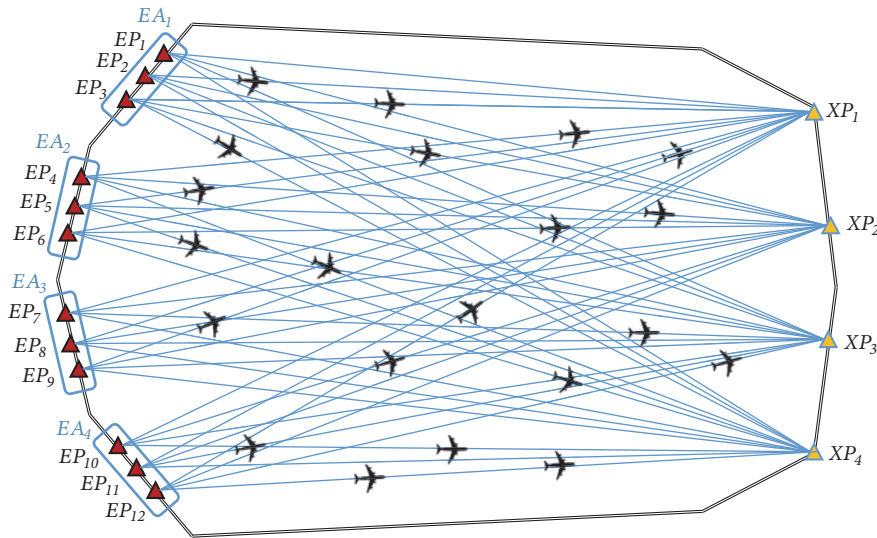


FIGURE 9: Complex airspace structure.

combinations of 12 routes between four entry areas with 3 entry points and 4 exit points. The numbers of the intersection points for these airspace configurations (i.e., I, II, and III) are 27, 36, and 45, respectively. Hourly traffic flow entering to the airspace is set to 20 aircraft in three different performance categories: regional jets, narrow-body aircraft, and wide-body aircraft. Estimated times of arrival (ETA) based on exponential distribution, aircraft performance categories (APC), entry areas (EA), and exit points (XP) are generated as input parameters using MATLAB. Entry points for the baseline scenario are chosen as the mid-points of each entry area (i.e.,  $EP_2$  for  $EA_1$ ,  $EP_5$  for  $EA_2$ ,  $EP_8$  for  $EA_3$ , and  $EP_{11}$  for  $EA_4$ ). The results of GAMS/CPLEX, GA, and TS are presented for 30 test problems in Table 7. While GAMS/CPLEX results include the values of objective function ( $z$ ) and CPU time, GA and TS results contain the average and minimum objective function values of three runs for each test problem as well as their average CPU times. The minimum airborne delays found using GA and TS are equal to those of GAMS/CPLEX solver solutions for each test problem. These results demonstrate that the proposed metaheuristics can achieve optimal solutions of GAMS/CPLEX for all test problems.

The complex route network for the generic airspace includes all possible 48 route combinations between the existing entry and exit points (Figure 9). The length of each route is shown in Table 8. In order to test this complex network, two different traffic flow rates are selected as 20 and

25 aircraft per hour. For each traffic flow rate thirty different scenarios are generated.

*5.1. Results of Flexible Entry Point Assignment Model.* Table 9 provides a comparison of the results found for baseline case of fixed entry points and flexible entry point approach for all scenarios. While baseline results include the values of objective function ( $z$ ) calculated via using FCFS model in MATLAB, flexible entry point results present the total airborne delays found by using GA and TS as the minimum of three separate runs.

No airborne delays occur in three scenarios (i.e., 6, 20, and 30) with the traffic flow of 20 aircraft per hour for the baseline case. While GA and TS find the same values in 25 scenarios for this traffic flow rate, GA achieved slightly better values than TS in two scenarios (i.e., 14 and 25). When the traffic flow rate is increased to 25 aircraft per hour, GA and TS find the same total airborne delay in 23 scenarios. GA provides better solutions in the remaining scenarios. The average results are demonstrated in Figure 10.

Figure 10 shows that the baseline average delay time for 20 aircraft decreases from 127.2 seconds to 8.4 seconds for GA and 8.6 seconds for TS. The percentages of recovery for GA and TS are 93.4% and 93.3%, respectively. For the 25 aircraft, the average delay in the baseline case is reduced from 193.1 seconds to 21.4 seconds for GA and to 23.2 seconds for TS. The recovery percentages for GA and TS are 88.9% and 88%, respectively. Both algorithms have yielded successful results.

TABLE 7: Test problem results using GAMS/CPLEX, GA, and TS.

Test Problems	GAMS/CPLEX		TS			GA		
	Z (sec.)	t(sec.)	$z_{ave}^*$ (sec.)	$z_{min}$ (sec.)	$T_{avg}$ (sec.)	$z_{ave}$ (sec.)	$z_{min}$ (sec.)	$T_{avg}$ (sec.)
1	<u>0</u>	111	0.0	<u>0</u>	2.8	0.0	<u>0</u>	0.7
2	<u>0</u>	112	0.0	<u>0</u>	0.3	0.0	<u>0</u>	0.7
3	<u>0</u>	109	0.0	<u>0</u>	6.0	0.0	<u>0</u>	2.1
4	<u>2</u>	110	2.0	<u>2</u>	67.7	2.0	<u>2</u>	108.6
5	<u>0</u>	113	0.0	<u>0</u>	18.4	1.7	<u>0</u>	41.9
6	<u>18</u>	114	18.0	<u>18</u>	67.5	18.0	<u>18</u>	112.6
7	<u>0</u>	113	0.0	<u>0</u>	4.4	0.0	<u>0</u>	1.0
8	<u>0</u>	114	0.0	<u>0</u>	0.4	0.0	<u>0</u>	0.8
9	<u>34</u>	113	55.0	<u>34</u>	73.2	34.0	<u>34</u>	124.3
10	<u>22</u>	114	22.7	<u>22</u>	75.8	24.7	<u>22</u>	122.5
11	<u>0</u>	144	0.3	<u>0</u>	40.8	0.0	<u>0</u>	3.8
12	<u>0</u>	151	0.0	<u>0</u>	2.8	0.0	<u>0</u>	1.8
13	<u>0</u>	149	0.0	<u>0</u>	4.0	0.0	<u>0</u>	3.1
14	<u>0</u>	150	0.0	<u>0</u>	2.3	0.0	<u>0</u>	1.3
15	<u>24</u>	148	24.0	<u>24</u>	94.7	26.0	<u>24</u>	154.1
16	<u>10</u>	148	10.7	<u>10</u>	88.6	10.0	<u>10</u>	148.9
17	<u>4</u>	149	4.0	<u>4</u>	82.4	6.7	<u>4</u>	131.8
18	<u>5</u>	150	7.7	<u>5</u>	95.5	7.3	<u>5</u>	153.4
19	<u>10</u>	148	13.3	<u>10</u>	114.1	23.3	<u>10</u>	180.2
20	<u>3</u>	152	3.0	<u>3</u>	97.7	6.0	<u>3</u>	148.4
21	<u>0</u>	190	0.0	<u>0</u>	20.1	0.0	<u>0</u>	8.3
22	<u>17</u>	187	17.0	<u>17</u>	109.4	17.0	<u>17</u>	192.0
23	<u>0</u>	188	1.3	<u>0</u>	66.2	0.0	<u>0</u>	25.0
24	<u>0</u>	191	0.0	<u>0</u>	4.6	0.0	<u>0</u>	1.2
25	<u>38</u>	190	38.0	<u>38</u>	92.2	38.0	<u>38</u>	165.9
26	<u>1</u>	191	1.0	<u>1</u>	108.6	1.0	<u>1</u>	187.0
27	<u>0</u>	189	0.0	<u>0</u>	1.0	0.0	<u>0</u>	1.0
28	<u>0</u>	186	0.0	<u>0</u>	1.4	0.0	<u>0</u>	1.0
29	<u>12</u>	190	12.0	<u>12</u>	103.7	12.0	<u>12</u>	183.9
30	<u>10</u>	187	10.0	<u>10</u>	102.8	10.0	<u>10</u>	174.8

TABLE 8: Routes distances in complex airspace.

Entry Points	Route Distances (nm)			
	XP1	XP2	XP3	XP4
1	180.0	189.7	216.3	247.5
2	180.2	186.8	210.9	240.8
3	181.1	184.3	205.9	234.3
4	186.8	180.2	193.1	216.3
5	189.7	180.0	189.7	210.9
6	193.1	180.2	186.8	205.9
7	205.9	184.3	181.1	193.1
8	210.9	186.8	180.2	189.7
9	216.3	189.7	180.0	186.8
10	234.3	201.2	182.4	181.1
11	240.8	205.9	184.3	180.2
12	247.5	210.9	186.8	180.0



TABLE 9: The airborne delays (sec.) and CPU times (sec.) of the all scenarios with 20 and 25 aircraft per hour.

Scenario Number	20 Aircraft					25 Aircraft				
	Baseline	GA	t <sub>ort</sub>	TS	t <sub>ort</sub>	Baseline	GA	t <sub>ort</sub>	TS	t <sub>ort</sub>
1	94	0	14.7	0	22.0	242	26	252.4	26	254.6
2	284	3	173.8	3	171.6	195	17	231.6	17	246.0
3	271	1	173.5	1	157.0	13	0	5.7	0	4.1
4	86	0	4.4	0	21.8	412	61	325.9	84	337.4
5	36	0	10.9	0	19.1	18	0	5.1	0	5.6
6	0	0	4.3	0	0.5	91	27	254.7	27	258.8
7	29	0	4.5	0	5.1	305	21	244.9	21	251.9
8	352	36	158.3	36	163.5	247	28	283.6	28	280.4
9	41	0	5.9	0	5.1	122	11	228.3	11	223.6
10	59	0	4.4	0	10.7	7	0	5.6	0	3.3
11	39	0	4.4	0	5.6	162	0	8.7	0	50.1
12	65	14	171.5	14	159.4	180	23	276.6	23	274.3
13	291	32	172.2	32	175.0	110	25	236.9	25	224.0
14	265	21	161.4	22	166.4	512	121	372.4	129	347.4
15	312	2	175.9	2	170.4	206	34	316.7	34	314.8
16	46	0	6.9	0	11.2	221	23	240.6	23	252.6
17	81	0	4.4	0	3.7	319	62	301.5	70	305.5
18	301	20	148.3	20	141.3	83	0	6.3	0	42.8
19	61	18	149.0	18	149.7	266	27	327.3	33	325.7
20	0	0	4.6	0	0.5	419	0	33.7	6	267.9
21	230	35	181.0	35	164.3	119	0	7.9	0	74.0
22	61	0	5.7	0	6.8	143	25	219.1	25	212.2
23	5	0	5.6	0	4.6	32	0	12.5	0	17.4
24	26	0	5.8	0	2.2	194	0	11.8	0	26.4
25	289	38	171.7	43	149.1	94	12	258.4	14	258.7
26	49	5	180.9	5	155.7	146	0	6.2	0	15.3
27	141	20	168.7	20	162.1	498	48	335.2	50	307.6
28	113	4	137.5	4	130.4	278	50	287.8	50	272.1
29	189	2	157.9	2	160.9	3	0	5.9	0	2.8
30	0	0	4.5	0	0.5	157	0	19.0	0	35.4

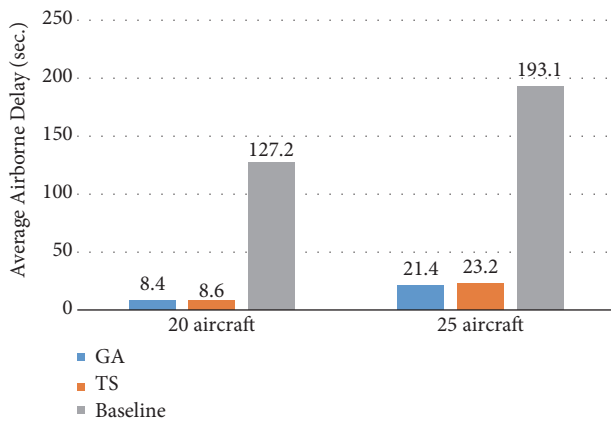


FIGURE 10: The average airborne delay for two different aircraft flow rates.

The results show that airborne delays due to conflicts can be reduced significantly using flexible entry point assignment.

5.2. *Results of Vector Deflection Model.* Table 10 presents the total fuel consumption for each scenario due to vector maneuvers based on the total airborne delay calculated in the first step of the model. Similar to Table 9, the baseline case represents the results found for fixed entry point configuration based on FCFS discipline, while GA and TS columns provide the results of flexible entry point approach for all scenarios calculated in GAMS/CONOPT solver and each scenario is estimated in less than 10 seconds.

The fuel consumption results based on GA and TS airborne delays are the same or very close to each other in all scenarios for the traffic flow rate of 20 aircraft per hour. GA-based results, on the other hand, are better than TS-based results in five scenarios and worse than TS-based results in one scenario. The fuel consumption averages are presented in Figure 11.

Aircraft vector deflection model is implemented in both baseline and metaheuristics-based airborne delays. The baseline, GA-based, and TS-based fuel consumption averages are found to be 201.4, 9.3, and 9.4 kg, respectively for 20 aircraft

TABLE 10: The delayed aircraft vector deflection model fuel consumption (kg).

Scenario Number	20 Aircraft			25 Aircraft		
	Baseline	GA-based	TS-based	Baseline	GA-based	TS-based
1	182.28	0	0	439.03	11.9	11.9
2	638.64	1.9	1.9	354.47	10.7	10.7
3	170.69	0.6	0.6	8.23	0	0
4	53.12	0	0	384.05	91	105.5
5	16.78	0	0	11.41	0	0
6	0	0	0	152.33	17	17
7	18.3	0	0	452.02	24.6	24.6
8	403.88	83.6	83.6	211.66	31.4	31.4
9	95.19	0	0	127.42	5.4	5.4
10	64.43	0	0	16.42	0	0
11	90.57	0	0	331.98	0	0
12	103.62	20.9	20.9	170.02	12.58	12.58
13	473.5	20.1	20.1	197.55	15.7	15.7
14	498.86	15.2	15.3	699.73	138	158.9
15	568.93	7.1	7.1	235.45	50.4	50.4
16	106.74	0	0	439.81	14.5	14.5
17	50.96	0	0	430.29	32.4	35.8
18	618.74	12.6	12.6	86.38	0	0
19	39.42	8.3	8.3	367.17	40.9	55.01
20	0	0	0	270.53	0	0
21	444.4	25.2	25.2	260.96	0	0
22	141.64	0	0	198.04	15.7	15.7
23	11.76	0	0	74.39	0	0
24	60.5	0	0	398.77	0	0
25	276.37	47.9	49.3	105.24	16.2	8.8
26	66.59	12.5	12.5	188.11	0	0
27	146.51	9.3	9.3	564.78	89.8	94.4
28	261.83	9.4	9.4	600.61	115.6	115.6
29	437.79	4.7	4.7	7.09	0	0
30	0	0	0	249.33	0	0

per hour, while for the 25 aircraft, the baseline, GA-based, and TS-based fuel consumption averages are found to be 267.7, 24.5, and 26.1 kg, respectively. The flexible entry point assignment obtained from GA and TS reduces the extra fuel consumption considerably. This is mainly because this approach resolves most of the possible conflicts via switching the entry points rather than implementing vector maneuvers. Therefore, both the number of aircraft conflicts and the required average airborne delays for conflict resolution are reduced significantly. The vector maneuvers are only implemented to fewer number of aircraft than those of the baseline scenarios. Reduced airborne delay leads to less extra fuel consumption caused by vector maneuvers.

## 6. Conclusion

The proposed two-step approach provided aircraft conflict resolutions with minimum airborne delay through flexible entry point assignment using GA and TS metaheuristics and offered vector deflection maneuvers with minimum extra fuel

consumption for delayed aircraft within FRA using NLP. The optimal solutions in both steps are obtained in reasonable times for a wide range of scenarios including test problems and more complex and realistic airspace routes structures. The proposed approach can be a good candidate for a decision-support system for effective conflict management in pretactical level. Through the entry point assignment and vector resolution advisories, air traffic controller's occupancy time can be reduced dramatically. It increases the airspace capacity through the reduction of airborne delays as well. The economic and environmental efficiency of the flight operations can be further enhanced through reduction of extra fuel consumption. The proposed vector deflection model also enables air traffic controller to choose limit values of maneuvering distance, bank angle, and deflection angle constraints. In future studies, the problem can be formulated as a multiobjective mathematical model that aims to find optimal airborne delay and fuel consumption simultaneously. Aircraft conflict resolution can also be extended to airspeed and altitude change maneuvers and, therefore, the model can

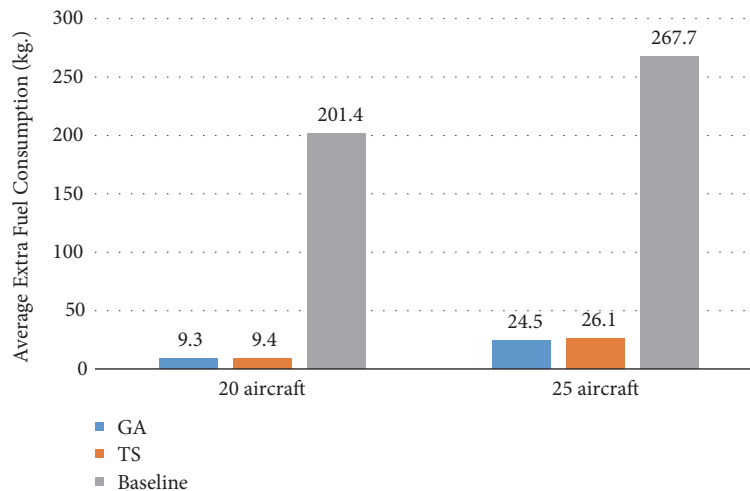


FIGURE 11: The average extra fuel consumption rates.

be applied to terminal airspaces. Inclusion of uncertainties such as wind variations in this approach can be a useful extension for tactical CD&R at low level en-route and terminal airspaces.

### Data Availability

The aircraft input parameters (ETA, APC, EA, EP, and XP) are randomly generated in MATLAB, and the data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

This study was supported by Anadolu University Scientific Research Projects Commission under grant number 1707F454.


### References

- [1] "Aviation. benefits beyond borders report," Tech. Rep., Air Transport Action Group, 2018.
- [2] Airbus Group, "Global market forecast 2017/2036, growing horizons".
- [3] ICAO, *2013–2028 Global Air Navigation Plan*, Doc 9750-AN/963, 4th edition, 2013.
- [4] SESAR, *Step 1 V3 UDPP Validation Report (D67), Optimized Airspace User Operations*, 2015.
- [5] [https://www.eurocontrol.int/eec/gallery/content/public/documents/other\\_documents/Air\\_traffic\\_conflict\\_resolution.pdf](https://www.eurocontrol.int/eec/gallery/content/public/documents/other_documents/Air_traffic_conflict_resolution.pdf)
- [6] B. Antulov-Fantulin, T. Rogošić, B. Juričić, and P. Andraši, "Air traffic controller assessment of the free route airspace implementation within zagreb area control centre," in *Proceedings of the International Scientific Conference Science and Traffic Development-ZIRP*, 2018.
- [7] J. K. Kuchar and L. C. Yang, "A Review of Conflict Detection and Resolution Modeling Methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 4, pp. 179–189, 2000.
- [8] M. Campo and F. Javier, *The collision avoidance problem. Methods and algorithms [Doctoral dissertation]*, Universidad Rey Juan Carlos, 2010.
- [9] O. Rodionova and B. Sridhar, *Efficient Planning of Wind-Optimal Routes in North Atlantic Oceanic Airspace*, 2017.
- [10] L. Pallottino, E. M. Feron, and A. Bicchi, "Conflict resolution problems for air traffic management systems solved with mixed integer programming," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 1, pp. 3–11, 2002.
- [11] M. Christodoulou and C. Costoulakis, "Nonlinear mixed integer programming for aircraft collision avoidance in free flight," in *Proceedings of the 12th IEEE Mediterranean Electrotechnical Conference, MELECON 2004*, vol. 1, pp. 327–330, IEEE, May 2004.
- [12] A. Vela, S. Solak, W. Singhose, and J. P. Clarke, "A mixed integer program for flight-level assignment and speed control for conflict resolution," in *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference*, pp. 5219–5226, IEEE, 2009.
- [13] A. E. Vela, S. Solak, J.-P. B. Clarke, W. E. Singhose, E. R. Barnes, and E. L. Johnson, "Near real-time fuel-optimal en route conflict resolution," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 826–837, 2010.
- [14] A. Alonso-Ayuso, L. F. Escudero, and F. J. Martín-Campo, "Collision avoidance in air traffic management: a mixed-integer linear optimization approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 47–57, 2011.
- [15] A. Alonso-Ayuso, L. F. Escudero, P. Olaso, and C. Pizarro, "Conflict avoidance: 0-1 linear models for conflict detection & resolution," *TOP*, vol. 21, no. 3, pp. 485–504, 2013.
- [16] S. Cafieri, L. Cellier, F. Messine, and R. Omhni, "Combination of optimal control approaches for aircraft conflict avoidance via velocity regulation," *Optimal Control Applications and Methods*, vol. 39, no. 1, pp. 181–203, 2018.

- [17] J. Omer, "A space-discretized mixed-integer linear model for air-conflict resolution with speed and heading maneuvers," *Computers & Operations Research*, vol. 58, pp. 75–86, 2015.
- [18] S. Cafieri and D. Rey, "Maximizing the number of conflict-free aircraft using mixed-integer nonlinear programming," *Computers & Operations Research*, vol. 80, pp. 147–158, 2017.
- [19] N. Durand, J.-M. Alliot, and J. Noailles, "Automatic aircraft conflict resolution using genetic algorithms," in *Proceedings of the 1996 ACM Symposium on Applied Computing, SAC 1996*, pp. 289–298, ACM, February 1996.
- [20] N. Durand and J.-M. Alliot, "Ant Colony Optimization for air traffic conflict resolution," in *Proceedings of the 8th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2009*, pp. 182–187, July 2009.
- [21] G. Meng and F. Qi, "Flight conflict resolution for civil aviation based on ant colony optimization," in *Proceedings of the 2012 Fifth International Symposium on Computational Intelligence and Design (ISCID)*, vol. 1, pp. 239–241, IEEE, 2012.
- [22] D. Alejo, J. A. Cobano, G. Heredia, and A. Ollero, "Particle swarm optimization for collision-free 4d trajectory planning in unmanned aerial vehicles," in *Proceedings of the 2013 International Conference on Unmanned Aircraft Systems, ICUAS 2013*, pp. 298–307, IEEE, May 2013.
- [23] M. Janic, *Air Transport System Analysis and Modelling*, CRC Press, 2014.
- [24] J. Carlier, D. Nace, V. Duong, and H. H. Nguyen, "Using disjunctive scheduling for a new sequencing method in multiple-conflicts solving," in *Proceedings of the 2003 IEEE Intelligent Transportation Systems*, vol. 1, pp. 708–714, IEEE, 2003.
- [25] C. Sheehan, *Coverage of 2012 European Air Traffic for the Base of Aircraft Data (BADA)-Revision 3.11*, 2007.
- [26] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–72, 1992.
- [27] M. Affenzeller, A. Beham, M. Kofler, G. Kronberger, S. A. Wagner, and S. Winkler, "Metaheuristic optimization," in *Hagenberg Research*, pp. 103–155, Springer, Berlin, Germany, 2009.
- [28] F. Glover and M. Laguna, "Tabu search," in *Handbook of Combinatorial Optimization*, pp. 2093–2229, Springer, Boston, MA, USA, 1998.

## Research Article

# Equilibrium and Optimization in a Double-Ended Queueing System with Dynamic Control

Yuejiao Wang <sup>1,2</sup> and Zaiming Liu<sup>2</sup>

<sup>1</sup>College of Mathematics and Computational Science, Hunan First Normal University, 410205 Changsha, China

<sup>2</sup>School of Mathematics and Statistics, Central South University, 410083 Changsha, China

Correspondence should be addressed to Yuejiao Wang; [yuejiao.wang@csu.edu.cn](mailto:yuejiao.wang@csu.edu.cn)

Received 13 October 2018; Revised 21 December 2018; Accepted 10 February 2019; Published 19 March 2019

Guest Editor: Marcos M. Vega

Copyright © 2019 Yuejiao Wang and Zaiming Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we consider a double-ended queueing system which is a passenger-taxi service system. In our model, we also consider the dynamic taxi control policy which means that the manager adjusts the arrival rate of taxis according to the taxi stand congestion. Under three different information levels, we study the equilibrium strategies as well as socially optimal strategies for arriving passengers by a reward-cost structure. Furthermore, we present several numerical experiments to analyze the relationship between the equilibrium and socially optimal strategies and demonstrate the effect of different information levels as well as several parameters on social benefit.

## 1. Introduction

The taxi service is an important component in the comprehensive transportation hubs. However, many travelers encounter such a situation that there are no taxis in the taxi stand during peak hour and there are many taxis in the taxi stand during the nonpeak hour. In order to make more efficient use of the taxi resource, we consider optimization problems in the passenger-taxi service system under dynamic taxi control. The passenger-taxi service system can be described as a double-ended queue: a queue for passengers, a queue for taxis. Clearly, the two queues cannot exist at the same time. In this paper, we will give some efficient strategies to ensure passengers' and taxis' utilities and reduce the taxi stand congestion.

Kendall [1] first studied the double-ended queue. The passenger-taxi service system was introduced as an example. Dobbie [2] found transient behavior under the nonhomogeneous Poisson arrival of passengers and taxis. Given [3] showed the asymptotic behavior of the double-ended queueing system under when the mean rates of arrival of passengers and taxis vary. Kashyap [4, 5] considered a double-ended queue with limited waiting space for taxis and for passengers. He studied the expected queue lengths of taxis and passengers

under the general arrival of passengers and the Poisson arrival of taxis. Wong, Wong, Bell, and Yang [6] adopted an absorbing Markov chain to model the searching process of taxi movements and proposed a useful formulation for describing the urban taxi services in a network. Conolly, Parthasarathy, and Selvaraju [7] studied double-ended queues with an impatient server or customers. Crescenzo, Giorno, Kumar, and Nobile [8] discussed a double-ended queue with catastrophes and repairs and obtained both the transient and steady-state probability distributions. Moreover, the double-ended queue can be applied to many other areas, for example, computer science, perishable inventory system, and organ allocation system. Zenios [9] illustrated a double-ended matching problem between several classes of organs and patients who would renege due to death. Wong, Szeto, and Wong [10] adopted the sequential logit approach to modeling bilevel decisions of vacant taxi drivers in customer-search. However, the above references discussed the performance measures of the double-ended queue. In this paper, we study the strategic behaviors of the passengers.

The study of queueing systems with strategic behavior of customers was first done by Naor [11], who analyzed the strategic behavior of customers under an observable queue by a linear reward-cost structure. Edelson and Hildebrand



[12] investigated the same problem following with Naor (1969). However, in this model, arriving customers do not know the queue length before his decision. Yang, Leung, Wong, and Bell [13] presented an equilibrium model for the problem of bilateral searching and meeting between taxis and passengers in a general network. Burnetas and Economou [14] discussed strategic behavior in a single server Markovian queue with setup times. Burnetas, Economou, and Vasiliadis [15] studied strategic customer behavior in a queueing system with delayed observations. Guo and Hassin [16] illustrated strategic behavior of customers and social optimization in Markovian vacation queues. Wang, Zhang, and Huang [17] considered strategic behavior of customers and social optimization in a constant retrial queue with the N-policy. The optimization problems of passenger-taxi service system under strategic behavior of passengers were first considered by Shi and Lian [18] who studied the arriving passengers' equilibrium strategies and socially optimal strategies under a limited waiting space of taxis and the same taxis' arrival rate. Shi and Lian [19] discussed a double-ended queueing system with limited waiting space for arriving taxis and arriving passengers. A passenger-taxi service system with a gated policy was considered by Wang, Wang, and Zhang [20] who studied the arriving passengers' equilibrium strategies and socially optimal strategies in fully observable, almost unobservable and fully unobservable cases, while they considered the model with same arrival rate of taxis. In order to balance the relationship between long passenger delays and high taxi's company costs, we consider a passenger-taxi service system with dynamic taxi control. The taxi control is to improve the arrival rate of taxis when the queue length of passengers is large so as to reduce delays and decrease it at times of increased queue length of taxis so as to reduce the costs of taxis' drivers. In this model, we study the (Nash) equilibrium strategies and socially optimal strategies of arriving passengers. Our model will improve the social benefit in the observable case and unobservable case if the waiting space of taxis is large enough, compared with the results in [18].

In the passenger-taxi service system with dynamic taxi control, arriving passengers decide whether to join the taxi stand or balk based on a linear reward-cost structure. We discuss the equilibrium strategies and socially optimal strategies under three different information levels: (1) fully observable case: the arriving passengers are noticed the number of passengers and taxis in the taxi stand; (2) almost unobservable case: the arriving passengers are only informed of the state of taxis; (3) fully unobservable case: the arriving passengers are not informed of the number of passengers or taxis in the taxi stand. The passenger's strategic behavior is under two different types: "selfishly optimal" and "socially optimal". "Selfishly optimal" is the strategy under (Nash) equilibrium conditions. "Socially optimal" is the strategy to maximize the social benefit. The contribution of the present paper is as follows: (1) study the passenger's selfishly optimal threshold and socially optimal threshold in fully observable case; (2) obtain the selfishly optimal joining probabilities and socially optimal joining probabilities in the almost unobservable case; (3) investigate the selfishly optimal joining probabilities

and social benefit function in the fully unobservable case; (4) present several numerical experiments to analyze the relationship between the equilibrium and socially optimal strategies and demonstrate the effect of different information levels as well as several parameters on social benefit.

The rest of the paper is organized as follows. In Section 2, we describe precisely the passenger-taxi service system. Sections 3, 4, and 5 discuss the equilibrium strategies and socially optimal strategies of passengers in three different information levels. In Section 6, we present some numerical examples to show how different information levels and parameters impact passenger's strategic behavior and social benefit. Section 7 concludes the paper with a summary.

## 2. Model Description

In this paper, we consider a passenger-taxi service system which is a double-ended queueing system. Now we give a precise description of the model. Passengers (one to four passengers traveling together and will arrive at the same destination can be seen as one passenger) arrive according to a Poisson process with rate  $\lambda_1$ . Taxis arrive according to a Poisson process. The arrive rate sets to  $\lambda_0$  whenever the number of passengers in the system equals to 0 and sets to  $\lambda_2$  otherwise, where  $\lambda_0 < \lambda_2$ . Passengers and taxis are served according to first-in-first-out discipline and leave the taxi stand at once if a taxi takes one passenger. In the taxi stand, the taxis' capacity is  $N$  which means that a taxi cannot join the taxi stand if there are  $N$  taxis waiting for passengers. The passengers can join the taxi stand without any limit. Let  $N(t)$  represent the queue length of passengers or taxis in taxi stand in time  $t$ . If  $N(t) > 0$ , it shows that passengers are waiting for taxis. If  $N(t) = 0$ , it shows that the system is empty. If  $N(t) < 0$ , it shows that taxis waiting for new passengers. Obviously, we know that  $\{N(t), t \geq 0\}$  is a one-dimensional continuous time Markov chain with state space  $\mathcal{F} = \{-N, -N + 1, \dots, -1, 0, 1, \dots\}$ . The state transition diagram is shown in Figure 1.

We assume that every joining passenger incurs a waiting cost  $C_1$  per unit time of waiting in the passenger queue, pays a taxi fare of  $p_1$ , and obtains a reward of  $R$  after arriving at his definition. Let  $C_2$  be the waiting cost of a taxi per unit time. Finally, we assume that joining passengers are not allowed to retrial and renege.

## 3. Almost Unobservable Case

In this section, we consider the almost unobservable case where an arriving passenger is only informed the state of taxis. If the number of taxis is more than zero, an arriving passenger will take a taxi immediately, to join the taxi stand without a doubt. But if the number of taxis equals 0, the passenger is not informed the number of passengers. The joining probability of an arriving passenger is  $q_{au}$  and the balking probability is  $1 - q_{au}$ . The state transition diagram is shown in Figure 2. For stability, let  $\rho_0 = \lambda_1/\lambda_0$ ,  $\rho_1 = \lambda_1/\lambda_2$ , and  $\rho_{1,2} = \lambda_1 q_{au}/\lambda_2 < 1$ .

Let  $\pi_i^{au} = \lim_{t \rightarrow \infty} \mathbb{P}(N(t) = i)$ , ( $i \in \mathcal{F}$ ) be the steady-state probability of state  $i$  in the almost unobservable case. We

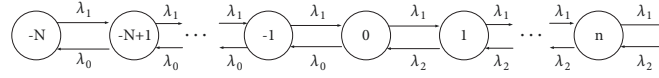


FIGURE 1: State transition diagram for the passenger-taxi service system.

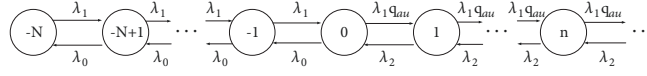


FIGURE 2: State transition diagram for the almost unobservable case.

can obtain the stationary distribution by solving the balance equations.

**Proposition 1.** *The stationary distribution in the almost unobservable case is given by*

$$\pi_{-N}^{au} = \frac{(1 - \rho_0)(1 - \rho_{1,2})}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}}; \quad (1)$$

$$\pi_i^{au} = \frac{\rho_0^{N+i}(1 - \rho_0)(1 - \rho_{1,2})}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}}, \quad -N + 1 \leq i \leq 0; \quad (2)$$

$$\pi_i^{au} = \frac{\rho_0^N \rho_{1,2}^i (1 - \rho_0)(1 - \rho_{1,2})}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}}, \quad i \geq 1. \quad (3)$$

*Proof.* The balance equations can be written as follows:

$$\lambda_1 \pi_{-N}^{au} = \lambda_0 \pi_{-N+1}^{au};$$

$$(\lambda_1 + \lambda_0) \pi_i^{au} = \lambda_1 \pi_{i-1}^{au} + \lambda_0 \pi_{i+1}^{au}, \quad -N + 1 \leq i \leq 1; \quad (4)$$

$$(\lambda_0 + \lambda_1 q_{au}) \pi_0^{au} = \lambda_1 \pi_{-1}^{au} + \lambda_2 \pi_1^{au};$$

$$(\lambda_2 + \lambda_1 q_{au}) \pi_i^{au} = \lambda_1 q_{au} \pi_{i-1}^{au} + \lambda_2 \pi_{i+1}^{au}, \quad i \geq 1.$$

Therefore, by the normalization condition, we obtain (1), (2), and (3).  $\square$

By (1), (2), and (3), we can get the expected queue length of passengers  $\mathbb{E}L_{au1}$  and expected queue length of taxis  $\mathbb{E}L_{au2}$ , respectively,

$$\begin{aligned} \mathbb{E}L_{au1} &= \sum_{i=0}^{\infty} i \pi_i^{au} \\ &= \frac{\rho_0^N \rho_{1,2} (1 - \rho_0)}{(1 - \rho_{1,2})(1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2})}; \end{aligned} \quad (5)$$

$$\begin{aligned} \mathbb{E}L_{au2} &= \sum_{i=-N}^0 (-i) \pi_i^{au} \\ &= \frac{1 - \rho_{1,2}}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}} \left[ N - \frac{\rho_0 (1 - \rho_0^N)}{1 - \rho_0} \right]. \end{aligned} \quad (6)$$

The effective arrival rate of passengers is

$$\begin{aligned} \lambda_c^* &= \lambda_1 \sum_{j=-N}^{-1} \pi_j^{au} + \lambda_1 q_{au} \sum_{j=0}^{\infty} \pi_j^{au} \\ &= \frac{\lambda_1 (1 - \rho_0^N)(1 - \rho_{1,2}) + \lambda_1 q_{au} \rho_0^N (1 - \rho_0)}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}}. \end{aligned} \quad (7)$$

The effective arrival rate of taxis is

$$\begin{aligned} \lambda_T^* &= \lambda_0 \sum_{j=-N+1}^0 \pi_j^{au} + \lambda_2 \sum_{j=1}^{\infty} \pi_j^{au} \\ &= \frac{\lambda_1 (1 - \rho_0^N)(1 - \rho_{1,2}) + \lambda_1 q_{au} \rho_0^N (1 - \rho_0)}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}}. \end{aligned} \quad (8)$$

Therefore, by Little's law, we obtain the expected waiting time of a joining passenger  $\mathbb{E}W_{au1}$  and the expected waiting time of a taxi  $\mathbb{E}W_{au2}$ , respectively

$$\mathbb{E}W_{au1} = \frac{\mathbb{E}L_{au1}}{\lambda_c^*}; \quad (9)$$

$$\mathbb{E}W_{au2} = \frac{\mathbb{E}L_{au2}}{\lambda_T^*}. \quad (10)$$

**3.1. Equilibrium Strategies of Passengers.** Now we consider the equilibrium strategy of an arriving passenger in almost unobservable case. We first consider the average waiting time of a joining passenger; see the below proposition.

**Proposition 2.** *When the queue length of taxis in the taxi stand is zero, the expected waiting time for a joining passenger is*

$$W(q_{au} | N(t) \geq 0) = \frac{1}{\lambda_2 (1 - \rho_{1,2})}. \quad (11)$$

*Proof.* By Little's law and (5), we get the expected waiting time of a joining passenger

$$\begin{aligned} W(q_{au} | N(t) \geq 0) &= \frac{(\sum_{j=0}^{\infty} j \pi_j) / (\lambda_1 q)}{\sum_{j=0}^{\infty} \pi_j} \\ &= \frac{1}{\lambda_2 (1 - \rho_{1,2})}. \end{aligned} \quad (12)$$

$\square$

Therefore, the utility of an arriving passenger is

$$U_{au}(q_{au} | N(t) \geq 0) = R - p_1 - C_1 \frac{1}{\lambda_2(1 - \rho_{1,2})}. \quad (13)$$

An equilibrium strategy for an arriving passenger who decides whether to join or balk is represented by  $q_{au}^e$  which is the joining probability for an arriving passenger and  $1 - q_{au}^e$  is the balking probability.  $q_{au}^e$  is also called selfishly optimal joining probability.

**Theorem 3.** *In the almost unobservable case, the equilibrium strategy for an arriving passenger is given by*

$$q_{au}^e = \begin{cases} 0, & \text{if } R < p_1 + C_1 \frac{1}{\lambda_2}; \\ q_{au}^{e*}, & \text{if } p_1 + C_1 \frac{1}{\lambda_2} \leq R \leq p_1 + C_1 \frac{1}{\lambda_2(1 - \rho_1)}; \\ 1, & \text{if } R > p_1 + C_1 \frac{1}{\lambda_2(1 - \rho_1)}, \end{cases} \quad (14)$$

where  $q_{au}^{e*} = (\lambda_2(R - p_1) - C_1)/(R - p_1)\lambda_1$ .

*Proof.* By Proposition 2, we have

$$W_c'(q_{au} | N(t) \geq 0) = \left( \frac{1}{\lambda_2(1 - \rho_{1,2})} \right)' = \frac{\lambda_1}{(\lambda_2(1 - \rho_{1,2}))^2}, \quad (15)$$

so that  $W_c'(q_{au} | N(t) \geq 0) > 0$ ; therefore,  $W_c(q_{au} | N(t) \geq 0)$  is increasing for  $q_{au} \in [0, 1]$ .

If  $R < p_1 + C_1(1/\lambda_2)$ , then  $U_c(q_{au} | N(t) \geq 0) < 0$  for  $q_{au} \in [0, 1]$ . Therefore, the best choice for an arriving passenger is balking, so that  $q_{au}^e = 0$ .

If  $R > p_1 + C_1(1/\lambda_2(1 - \rho_1))$ , then  $U_c(q_{au} | N(t) \geq 0) > 0$  for  $q_{au} \in [0, 1]$ , so that an arriving passenger's best choice is  $q_{au}^e = 1$ .

Since  $U_c(q_{au} | N(t) \geq 0)$  is a decreasing function for  $q_{au} \in [0, 1]$ , so that there exists a unique solution of the equation  $U_c(q_{au}^e) = 0$  within  $(0, 1)$  for  $p_1 + C_1(1/\lambda_2) \leq R \leq p_1 + C_1(1/\lambda_2(1 - \rho_1))$ .  $\square$

**3.2. Socially Optimal Strategies of Passengers.** Now, we consider the socially optimal strategy of an arriving passenger. By (5), (6), (7), and (8), we obtain the social benefit  $S_{au}$  in the almost unobservable case

$$\begin{aligned} S_{au}(q_{au}) &= \lambda_c^*(R - p_1 - C_1 \mathbb{E}W_{au1}) \\ &+ \lambda_T^*(p_1 - C_2 \mathbb{E}W_{au2}) = R \\ &\cdot \frac{\lambda_1(1 - \rho_0^N)(1 - \rho_{1,2}) + \lambda_1 q_{au} \rho_0^N(1 - \rho_0)}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}} - C_1 \end{aligned}$$

$$\begin{aligned} &\cdot \frac{\rho_0^N \rho_{1,2}(1 - \rho_0)}{(1 - \rho_{1,2})(1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2})} - C_2 \\ &\cdot \frac{1 - \rho_{1,2}}{1 - \rho_{1,2} - \rho_0^{N+1} + \rho_0^N \rho_{1,2}} \left[ N - \frac{\rho_0(1 - \rho_0^N)}{1 - \rho_0} \right]. \end{aligned} \quad (16)$$

We investigate the socially optimal strategy which is represented by  $q_{au}^*$  to maximize social benefit in almost unobservable case.

**Theorem 4.**  $S_{au}(q_{au})$  is a concave function in  $q_{au} \in [0, 1]$  and reaches maximum at  $q_{au}^* = \min((dS_{au}(q_{au})/dq_{au})|_{q_{au}=q_{au}^*} = 0; 1)$ .

*Proof.* To simplify, let  $C = C_2[N - \rho_0(1 - \rho_0^N)/(1 - \rho_0)](1/(1 - \rho_0))$ . Therefore,

$$\begin{aligned} S_{au}(q_{au}) &= \lambda_2 R \\ &- \left[ \left( \frac{\lambda_2}{1 - \rho_0} - \frac{\lambda_2 \rho_0^{N+1}}{1 - \rho_0} - \frac{\lambda_1(1 - \rho_0^N)}{1 - \rho_0} \right) R \right. \\ &\left. + C_1 \frac{\rho_0^N \rho_{1,2}(1 - \rho_0)}{(1 - \rho_0)(1 - \rho_{1,2})^2} + C \right] \pi_{-N}^{au}. \end{aligned} \quad (17)$$

The first derivative of  $S_{au}(q_{au})$  is

$$\begin{aligned} \frac{dS_{au}(q_{au})}{dq_{au}} &= \frac{\rho_1 \rho_0^N}{(1 - \rho_{1,2})^2} (\pi_{-N}^{au})^2 \left[ R \left( \frac{\lambda_2}{1 - \rho_0} \right. \right. \\ &- \frac{\lambda_2 \rho_0^{N+1}}{1 - \rho_0} - \frac{\lambda_1(1 - \rho_0^N)}{1 - \rho_0} \left. \left. \right) + C_1 \frac{\rho_0 \rho_{1,2}}{(1 - \rho_{1,2})^2} + C \right. \\ &- \frac{1 + \rho_{1,2}}{1 - \rho_{1,2}} \frac{C_1}{\pi_{-N}^{au}} \left. \right] = \frac{\rho_1 \rho_0^N}{(1 - \rho_{1,2})^4} (\pi_{-N}^{au})^2 \left[ R \left( \frac{\lambda_2}{1 - \rho_0} \right. \right. \\ &- \frac{\lambda_2 \rho_0^{N+1}}{1 - \rho_0} - \frac{\lambda_1(1 - \rho_0^N)}{1 - \rho_0} \left. \left. \right) (1 - \rho_{1,2}) + C_1 \rho^N \rho_{1,2} \right. \\ &\left. + C(1 - \rho_{1,2})^2 - (1 - \rho_{1,2}^2) \frac{C_1}{\pi_{-N}^{au}} \right] \\ &= \frac{\rho_1 \rho_0^N}{(1 - \rho_{1,2})^4} (\pi_{-N}^{au})^2 \\ &\cdot \left[ \left[ R \left( \frac{\lambda_2}{1 - \rho_0} - \frac{\lambda_2 \rho_0^{N+1}}{1 - \rho_0} - \frac{\lambda_1(1 - \rho_0^N)}{1 - \rho_0} \right) + C \right. \right. \\ &\left. \left. + C_1 \frac{1 - \rho_0^N}{1 - \rho_0} \right] \rho_{1,2}^2 \right. \end{aligned}$$

$$\begin{aligned}
& -2\rho_{1,2} \left( R \left( \frac{\lambda_2}{1-\rho_0} - \frac{\lambda_2 \rho_0^{N+1}}{1-\rho_0} - \frac{\lambda_1 (1-\rho_0^N)}{1-\rho_0} \right) \right. \\
& \left. + C \right) + \left( R \left( \frac{\lambda_2}{1-\rho_0} - \frac{\lambda_2 \rho_0^{N+1}}{1-\rho_0} - \frac{\lambda_1 (1-\rho_0^N)}{1-\rho_0} \right) \right. \\
& \left. + C - C_1 \frac{1-\rho_0^{N+1}}{1-\rho_0} \right) \Big]. \tag{18}
\end{aligned}$$

Since

we have

$$\frac{R(\lambda_2/(1-\rho_0) - \lambda_2 \rho_0^{N+1}/(1-\rho_0) - \lambda_1 (1-\rho_0^N)/(1-\rho_0)) + C + \sqrt{\Delta/4}}{R(\lambda_2/(1-\rho_0) - \lambda_2 \rho_0^{N+1}/(1-\rho_0) - \lambda_1 (1-\rho_0^N)/(1-\rho_0)) + C + C_1 ((1-\rho_0^N)/(1-\rho_0))} > 1. \tag{20}$$

Therefore, we know that  $dS_{au}(q_{au})/dq_{au}$  is a decreasing function at  $q_{au} \in [0, 1]$ . Hence, we obtain the maximum of  $S_{au}(q_{au})$  in  $q_{au}^* = \min((dS_{au}(q_{au})/dq_{au})|_{q_{au}=q_{au}^*} = 0, 1)$ .  $\square$

#### 4. Fully Observable Case

We first consider the fully observable case in which arriving passengers are informed both the number of passengers and taxis in taxi stand. In this case, we consider the equilibrium strategies and socially optimal strategies for arriving passengers.

**4.1. Equilibrium Strategies of Passengers.** In fully observable case, the equilibrium joining strategy of an arriving passenger who decides whether to join the taxi stand or balk is represented by threshold type that is an arriving passenger will join the taxi stand if the queue length of passengers is less than threshold and balking otherwise. If there exists a threshold  $n^0$  such that the passengers will join the taxi stand if  $N(t) \leq n^0$  and balk otherwise, then  $n^0$  is called selfishly optimal threshold. The value of  $n^0$  is given by the following Theorem 5.

**Theorem 5.** *In the fully observable passenger-taxi system, there exists a unique selfishly optimal threshold*

$$n^0 = \left\lfloor \frac{\lambda_2(R-p_1)}{C_1} \right\rfloor \tag{21}$$

which is the equilibrium strategy of an arriving passenger.

*Proof.* By the passenger's utility,  $n^0$  should satisfy the following conditions:

$$\begin{aligned}
R - p_1 - C_1 \frac{n^0}{\lambda_2} &\geq 0; \\
R - p_1 - C_1 \frac{n^0 + 1}{\lambda_2} &< 0.
\end{aligned} \tag{22}$$

$\square$

**4.2. Socially Optimal Strategies of Passengers.** Then we consider the socially optimal strategy for an arriving passenger. That is specified by threshold which means that there exists a unique  $n^*$  (is called socially optimal threshold) such that the social benefit reaches maximum. Clearly, we know that the system follows a one-dimensional continuous time Markov chain with state space  $\mathcal{F}^0 = \{-N, -N+1, \dots, -1, 0, \dots, n\}$ , where  $n$  is the passenger buffer size. The transition rate diagram is shown in Figure 3. For simplicity, let  $\rho_0 = \lambda_1/\lambda_0$  and  $\rho_1 = \lambda_1/\lambda_2$ . Let  $\pi_k$  be the stationary distribution of state  $k \in \mathcal{F}^0$  in the fully observable case. We can obtain the stationary distribution by solving the balance equations.

**Proposition 6.** *The stationary distribution in the fully observable case is as follows:*

$$\begin{aligned}
\pi_{-N} &= \frac{(1-\rho_0)(1-\rho_1)}{1-\rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1} - \rho_0^N \rho_1^{N+1} + \rho_0^{N+1} \rho_1^{N+1}}; \\
\pi_i &= \frac{\rho_0^{N+1} (1-\rho_0)(1-\rho_1)}{1-\rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1} - \rho_0^N \rho_1^{N+1} + \rho_0^{N+1} \rho_1^{N+1}}, \\
&\quad -N+1 \leq i \leq 0; \\
\pi_i &= \frac{\rho_0^N \rho_1^i (1-\rho_0)(1-\rho_1)}{1-\rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1} - \rho_0^N \rho_1^{N+1} + \rho_0^{N+1} \rho_1^{N+1}}, \\
&\quad 1 \leq i \leq n.
\end{aligned} \tag{23}$$

By the same method of (16), we obtain the social benefit function:

$$\begin{aligned}
S_{ob}(n) &= \lambda_c (R - p_1 - C_1 \mathbb{E}W_{obc}) \\
&\quad + \lambda_T (p_1 - C_2 \mathbb{E}W_{obt}) = \lambda_c R \\
&\quad - C_1 \mathbb{E}L_{obc} - C_2 \mathbb{E}L_{obt} \\
&= \frac{1}{1-\rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1} - \rho_0^N \rho_1^{N+1} + \rho_0^{N+1} \rho_1^{N+1}} \left[ R \lambda_1 (1 \right.
\end{aligned}$$

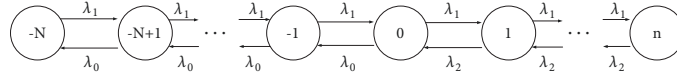


FIGURE 3: State transition diagram for the fully observable case.

$$\begin{aligned}
& -\rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1} - \rho_0^N \rho_1^n + \rho_0^{N+1} \rho_1^n) - C_1 \rho_0^N (1 \\
& - \rho_0) \left[ \frac{\rho_1 (1 - \rho_1^n)}{1 - \rho_1} - n \rho_1^{n+1} \right] - C_2 (1 - \rho_1) \left( N \right. \\
& \left. - \frac{\rho_0 (1 - \rho_0^N)}{1 - \rho_0} \right) \Bigg], \tag{24}
\end{aligned}$$

where  $\mathbb{E}W_{obc}$  and  $\mathbb{E}W_{obt}$  represent the expected waiting time of passengers and taxis respectively,  $\mathbb{E}L_{obc}$  and  $\mathbb{E}L_{obt}$  represent the mean queue length of passengers and taxis respectively, and  $\lambda_c = \lambda_T$  represents the effective arrival rate of passengers or taxis.

By definition of the socially optimal threshold, we know that  $n^*$  should follow the below two inequalities:

$$\begin{aligned}
S_{ob}(n^*) - S_{ob}(n^* + 1) &\geq 0; \\
S_{ob}(n^*) - S_{ob}(n^* - 1) &\geq 0. \tag{25}
\end{aligned}$$

By calculation, we have the following inequalities which are equal to the condition (25)

$$\begin{aligned}
& n^* (1 - \rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1}) (1 - \rho_1) \\
& - \rho_0^N \rho_1 (1 - \rho_0) (1 - \rho_1^{n^*}) \leq \frac{M}{C_1 \rho_1} \\
& \leq (n^* + 1) (1 - \rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1}) (1 - \rho_1) \\
& - \rho_0^N \rho_1 (1 - \rho_0) (1 - \rho_1^{n^*+1}), \tag{26}
\end{aligned}$$

where

$$\begin{aligned}
M &= R \lambda_1 (1 - \rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1}) (1 - \rho_1)^2 \\
&+ C_2 \rho_1 (1 - \rho_1)^2 \left[ N - \frac{\rho_0 (1 - \rho_0^N)}{1 - \rho_0} \right]. \tag{27}
\end{aligned}$$

Let

$$\begin{aligned}
f(x) &= x (1 - \rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1}) (1 - \rho_1) \\
&- \rho_0^N \rho_1 (1 - \rho_0) (1 - \rho_1^x). \tag{28}
\end{aligned}$$

In the following proposition, we study the monotonicity of the function  $f(x)$  for  $x \geq 1$ .

**Proposition 7.** If  $\rho_0 < 1$ ,  $f(x)$  is an increasing function in  $x \geq 1$ . If  $\rho_0 > 1$  and  $\rho_1 \neq 1$ ,  $f(x)$  is a decreasing function in  $x \geq 1$ .

*Proof.* The first order derivative of  $f(x)$  is

$$\begin{aligned}
g(\rho_1) &:= \frac{df(x)}{dx} \\
&= (1 - \rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1}) (1 - \rho_1) \\
&+ \rho_0^N (1 - \rho_0) \rho_1^{x+1} \log \rho_1. \tag{29}
\end{aligned}$$

The second order derivative of  $f(x)$  is

$$\frac{d^2 f(x)}{dx^2} = \rho_0^N (1 - \rho_0) \rho_1^{x+1} (\log \rho_1)^2 > 0. \tag{30}$$

The first order derivative of  $g(\rho_1)$  is

$$\begin{aligned}
\frac{dg(\rho_1)}{d\rho_1} &= -2(1 - \rho_1) + \rho^N (1 - \rho_1) + \rho_0^N (\rho_0 - \rho_1) \\
&+ (x + 1) \rho_0^N (1 - \rho_0) \rho_1^x \log \rho_1 \\
&+ \rho_0^N (1 - \rho_1) \rho_1^x. \tag{31}
\end{aligned}$$

(1) If  $0 < \rho_1 \leq \rho_0 < 1$  and  $1 < \rho_1 \leq \rho_0$ , by (31), we have

$$\begin{aligned}
\frac{dg(\rho_1)}{d\rho_1} &< -2(1 - \rho_1) + \rho_0^N (1 - \rho_1) + \rho_0^N (1 - \rho_1) \\
&+ \rho_0^{N+x} (1 - \rho_0) - \rho_0^N (1 - \rho_0) \\
&= (1 - \rho_1) (-2 + 2\rho_0^N) \\
&+ (1 - \rho_0) \rho_0^N (\rho_0^x - 1) < 0. \tag{32}
\end{aligned}$$

Therefore,  $g(\rho_1)$  is a decreasing function for  $\rho_1 \in (0, 1) \cup (1, \infty)$ . Then,

$$g(1) = \left. \frac{df(x)}{dx} \right|_{\rho_1=1} = 0. \tag{33}$$

Hence,

$$\begin{aligned}
\frac{df(x)}{dx} &> 0, \quad \forall \rho_1 \in (0, 1) \\
\text{and } \frac{df(x)}{dx} &< 0, \quad \forall \rho_1 \in (1, \infty). \tag{34}
\end{aligned}$$

(2) If  $\rho_0 > 1$  and  $0 < \rho_1 < 1$ , we have

$$\begin{aligned}
\frac{dg(\rho_1)}{d\rho_1} &> 2\rho_0^N (1 - \rho_1) - 2(1 - \rho_1) \\
&= 2(1 - \rho_1) (\rho_0^N - 1) > 0. \tag{35}
\end{aligned}$$



Therefore,  $g(\rho_1)$  is an increasing function for  $\rho_1 \in (0, 1)$  and  $\rho_0 \in (1, \infty)$ . Then,

$$g(1) = \left. \frac{df(x)}{dx} \right|_{\rho_1=1} = 0. \quad (36)$$

Thus,

$$\frac{df(x)}{dx} < 0, \quad \forall \rho_1 \in (0, 1). \quad (37)$$

By (34) and (37), for  $x \geq 1$ , we obtain that  $f(x)$  is an increasing function, when  $\rho_0 < 1$  and  $f(x)$  is a decreasing function when  $\rho_0 > 1$  and  $\rho_1 \neq 1$ .  $\square$

Let

$$\begin{aligned} n^* (1 - \rho_1 + \rho_1 \rho_0^N - \rho_0^{N+1}) (1 - \rho_1) \\ - \rho_0^N \rho_1 (1 - \rho_0) (1 - \rho_1^*) = \frac{M}{C_1 \rho_1}. \end{aligned} \quad (38)$$

In the following theorem, we consider the socially optimal strategy which is specified by the socially optimal threshold  $n^*$  such that the social benefit reaches maximum.

**Theorem 8.** *In fully observable case, the socially optimal strategy is given as follows:*

- (1) If  $\rho_1 = 1$  or  $\rho_0 = 1$ , there is no solution to (38).
- (2) For  $\rho_0 \in (0, 1)$ ,
  - (a) if  $f(1) > M/C_1 \rho_1$ , there is no solution to (38).
  - (b) if  $f(1) < M/C_1 \rho_1$ , there exists a unique solution  $n^* > 1$  of (38).
  - (c) if  $f(1) = M/C_1 \rho_1$ , there exists a unique solution  $n^* = 1$  of (38).
- (3) For  $\rho_0 \in (1, \infty)$ ,
  - (a) if  $f(1) < M/C_1 \rho_1$ , there is no solution to (38).
  - (b) if  $f(1) > M/C_1 \rho_1$ , there exists a unique solution  $n^* > 1$  of (38).
  - (c) if  $f(1) = M/C_1 \rho_1$ , there exists a unique solution  $n^* = 1$  of (38).

## 5. Fully Unobservable Case

Now we consider fully unobservable case where arriving passengers are not informed the number of passengers or taxis, but they know the arrival rates of passengers and taxis. The joining probability of an arriving passenger is  $q$  and the balking probability is  $1 - q$ . The state transition diagram is shown in Figure 4. Let  $\rho_{1,1} := \lambda_1 q / \lambda_2$  and  $\rho_{0,1} := \lambda_1 q / \lambda_0$ . Suppose the system is stable, so that  $\rho_{1,1} = \lambda_1 q / \lambda_2 < 1$ .

Let  $\pi_i^{fu}$  be the steady-state probability of state  $i$  in the fully unobservable case. In the following proposition, we obtain the stationary distribution by balance equations.

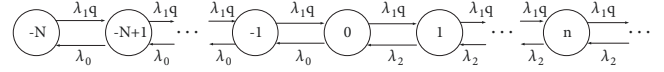


FIGURE 4: State transition diagram for the fully unobservable case.

**Proposition 9.** *The stationary distribution in fully unobservable case is given by*

$$\pi_{-N}^{fu} = \frac{(1 - \rho_{0,1})(1 - \rho_{1,1})}{1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N}; \quad (39)$$

$$\pi_i^{fu} = \frac{\rho_{0,1}^{N+i} (1 - \rho_{0,1})(1 - \rho_{1,1})}{1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N}, \quad -N \leq i \leq 0; \quad (40)$$

$$\pi_i^{fu} = \frac{\rho_{0,1} \rho_{1,1}^i (1 - \rho_{0,1})(1 - \rho_{1,1})}{1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N}, \quad i \geq 1. \quad (41)$$

By the same method, we obtain the expected queue length of passengers  $\mathbb{E}L_{fu1}$  and the expected queue length of taxis  $\mathbb{E}L_{fu2}$ , respectively:

$$\mathbb{E}L_{fu1} = \sum_{i=0}^{\infty} i \pi_i^{fu} \quad (42)$$

$$= \frac{\rho_{1,1} \rho_{0,1}^N (1 - \rho_{0,1})}{(1 - \rho_{1,1})(1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N)};$$

$$\mathbb{E}L_{fu2} = \sum_{i=-N}^0 (-i) \pi_i^{fu} \quad (43)$$

$$= \left( N - \frac{\rho_{0,1} (1 - \rho_{0,1}^N)}{1 - \rho_{0,1}} \right) \frac{1 - \rho_{1,1}}{1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N}.$$

Obviously, the effective arrival rate of passengers is  $\lambda_1 q$ . Let the effective arrival rate of taxis be  $\alpha$ . Then

$$\alpha = \sum_{i=-N+1}^0 \lambda_0 \pi_i^{fu} + \sum_{i=1}^{\infty} \lambda_2 \pi_i^{fu} = \lambda_1 q. \quad (44)$$

The expected waiting time of a passenger  $\mathbb{E}W_{fu1}(q)$  and the expected waiting time of a taxi  $\mathbb{E}W_{fu2}(q)$  are, respectively,

$$\mathbb{E}W_{fu1}(q) = \frac{\mathbb{E}L_{fu1}}{\lambda_1 q} \quad (45)$$

$$\text{and } \mathbb{E}W_{fu2}(q) = \frac{\mathbb{E}L_{fu2}}{\lambda_1 q}.$$

**5.1. Equilibrium Strategies of Passengers.** We now consider the equilibrium strategy of an arriving passenger. By (45), we get the utility of passengers

$$U_{fu}(q) = R - p_1 - C_1 \mathbb{E}W_{fu1}(q). \quad (46)$$

The equilibrium strategy of an arriving passenger is specified by the joining probability, denoted by  $q_e$  such that a passenger will choose to take a taxi with probability  $q_e$  and balk with probability  $1 - q_e$ .

**Theorem 10.** *In the unobservable queue case, the equilibrium strategy for each passenger is as follows:*

$$q_e = \begin{cases} 0, & \text{if } R < p_1; \\ q_e^*, & \text{if } p_1 \leq R \leq p_1 + C_1 \frac{\rho_1 \rho_0^N (1 - \rho_0)}{\lambda_1 (1 - \rho_1) (1 - \rho_1 - \rho_0^{N+1} + \rho_1 \rho_0^N)}; \\ 1, & \text{if } R > p_1 + C_1 \frac{\rho_1 \rho_0^N (1 - \rho_0)}{\lambda_1 (1 - \rho_1) (1 - \rho_1 - \rho_0^{N+1} + \rho_1 \rho_0^N)}, \end{cases} \quad (47)$$

where  $q_e^*$  is the unique solution of equation  $U_{fu}(q_e^*) = 0$ .

*Proof.* By (45), we know that  $W_{fu}(q)$  is an increasing function for  $q \in [0, 1]$ ; thus,  $U_{fu}(q)$  is a decreasing function for  $q \in [0, 1]$ .

If  $R < p_1$ ,  $U_{fu}(q) < 0$ , then the best response for an passenger is balking, so that  $q_e = 0$ .

If  $R > p_1 + C_1(\rho_1 \rho_0^N (1 - \rho_0)/\lambda_1(1 - \rho_1)(1 - \rho_1 - \rho_0^{N+1} + \rho_1 \rho_0^N))$ ,  $U_{fu}(q) > 0$ ; hence, his best choice is  $q_e = 1$ .

Since  $U_{fu}(q)$  is a decreasing function for  $q$ , so that there exists a unique solution to the equation  $U_{fu}(q_e^*) = 0$  within  $(0, 1)$  if  $p_1 \leq R \leq p_1 + C_1(\rho_1 \rho_0^N (1 - \rho_0)/\lambda_1(1 - \rho_1)(1 - \rho_1 - \rho_0^{N+1} + \rho_1 \rho_0^N))$ .  $\square$

**5.2. Social Benefit Function.** In this section, we study the social benefit function in fully unobservable case. By (42), (43), and (44), we obtain the social benefit function

$$\begin{aligned} S_{fu} &= \lambda_1 q R - C_1 \\ &\cdot \frac{\rho_{1,1} \rho_{0,1}^N (1 - \rho_{0,1})}{(1 - \rho_{1,1}) (1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N)} \\ &- C_2 \left( N - \frac{\rho_{0,1} (1 - \rho_{0,1}^N)}{1 - \rho_{0,1}} \right) \\ &\cdot \frac{1 - \rho_{1,1}}{1 - \rho_{1,1} - \rho_{0,1}^{N+1} + \rho_{1,1} \rho_{0,1}^N}. \end{aligned} \quad (48)$$

Since the social benefit function is complicated, the first and second order derivatives are too difficult to analyze. So, we study the socially optimal strategies which is represented by a joining probability (is called socially optimal joining probability) such that the social benefit reaches maximum by numerical experiments.

## 6. Numerical Experiments

In this section, we will show some tables to find the relationship between the equilibrium strategies and socially optimal strategies of arriving passengers under three different information levels (fully observable case, almost unobservable case, and fully unobservable case). Moreover, we will present figures to compare the socially optimal joining probabilities in the almost unobservable case with those in the fully

unobservable case. We also find the effect of three different information levels as well as several parameters: taxi buffer size  $N$ , arrival rate of passengers  $\lambda_1$ , the low arrival rate of taxis  $\lambda_0$ , and the high arrival rate of taxis  $\lambda_2$  on social benefit.

We first study the equilibrium joining probabilities in almost unobservable and fully unobservable cases, respectively. These results are shown in Figure 5. From the left of Figure 5, we can see that the equilibrium joining probabilities in the fully unobservable case is increasing as  $N$  and  $\lambda_0$  increase, respectively. It is obvious that the equilibrium joining probabilities in the fully unobservable case is always larger than that in the almost unobservable case. From the right of Figure 5, we know that the equilibrium joining probability in the almost unobservable case decreases with respect to  $\lambda_1$ . Moreover, the equilibrium joining probability in the almost unobservable case is increasing as  $\lambda_2$  increases.

We second consider the socially optimal strategies in three different information levels. These results are shown in Figures 6 and 7. In the left of Figure 6, we find that socially optimal joining probabilities in the almost unobservable case are decreasing as  $N$  increases. When  $\lambda_0 < \lambda_2$ , the socially optimal joining probabilities is larger than that in the case  $\lambda_0 = \lambda_2$ . From the right of Figure 6, we can observe that socially optimal joining probabilities in the fully unobservable case is increasing with respect to  $N$ . Furthermore, the socially optimal joining probabilities in the case  $\lambda_0 < \lambda_2$  is less than that in the case  $\lambda_0 = \lambda_2$ . In Figure 7, we study the socially optimal threshold in two cases. When  $\lambda_2 > \lambda_1$ , the socially optimal threshold is increasing as  $N$  increases. The socially optimal threshold in the case  $\lambda_0 = \lambda_2$  is larger than that in the case  $\lambda_0 < \lambda_2$ . When  $\lambda_2 < \lambda_1$ , the relationship is opposite.

We then consider the optimal social benefit in the fully unobservable and fully observable cases, respectively. These results are shown in Figure 8. In the left of Figure 8, we find that the optimal social benefit in the observable case is decreasing with respect to  $N$ . If  $N$  is less than 5, the best choice is  $\lambda_0 = \lambda_2$  which is the case in [18]. If  $N$  is larger than 5, the social benefit in the case  $\lambda_0 < \lambda_2$  is larger than that in the case  $\lambda_0 = \lambda_2$ . From this behavior, we know that our model can be used to improve the optimal social benefit. The right of Figure 8 shows the relationship between the social benefit and the arrival rates of taxis.

In the last numerical example, we investigate the effect of three different information levels as well as several parameters

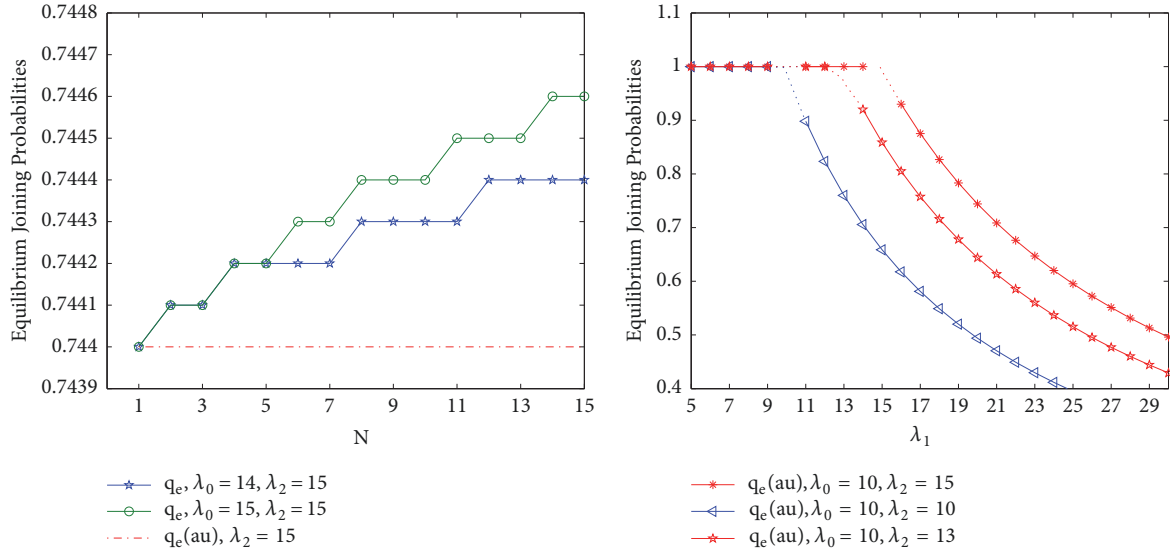


FIGURE 5: Left: equilibrium joining probabilities for the almost and fully unobservable cases vs.  $N$  for  $R = 30$ ;  $p_1 = 5$ ;  $C_1 = 3$ ;  $\lambda_1 = 20$ . Right: equilibrium joining probabilities for the almost and fully unobservable cases vs.  $\lambda_1$  for  $R = 30$ ,  $p_1 = 5$ ,  $C_1 = 3$ ,  $N = 5$ .

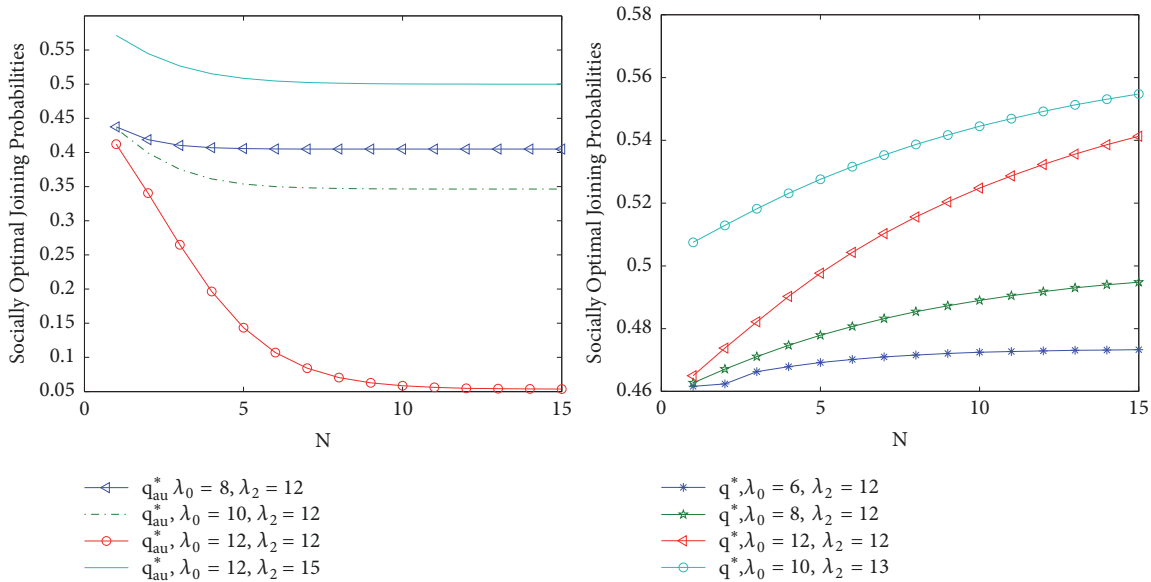


FIGURE 6: Left: socially optimal joining probabilities for the almost unobservable case vs.  $N$  for  $R = 30$ ,  $C_1 = 15$ ,  $C_2 = 10$ ,  $\lambda_1 = 20$ . Right: socially optimal joining probabilities for the unobservable case vs.  $N$  for  $R = 30$ ,  $C_1 = 20$ ,  $C_2 = 15$ ,  $\lambda_1 = 20$ .

(taxi buffer size  $N$ , the arrival rate of passengers  $\lambda_1$ , and the low arrival rate of taxis  $\lambda_0$ ) on social benefit. These results are shown in Figures 9, 10, and 11. From the left of Figure 9, when  $\lambda_1 > \lambda_2$ , we can see that the social benefit in fully observable case and almost unobservable case increase as  $\lambda_0$  increases. In the fully unobservable case, we can see that the social benefit function is unimodal; then, we get the optimal low taxi arrival rate. In the right of Figure 9, when  $\lambda_1 < \lambda_2$ , the social benefits in three cases are convex functions for  $\lambda_0$ . However, social benefits in three cases are basically the same. Moreover, from Figure 9 we know that the dynamic taxi control policy can improve the social benefit in several cases compared with the same taxi arrival rate case. In the left of Figure 10, when

$\lambda_1 < \lambda_2$ , we know that the social benefit in three cases is not much difference. When  $\lambda_1 > \lambda_2$ , the social benefits in fully observable case and almost unobservable case are larger than that in fully unobservable case. From this behavior we know that providing the taxi stand information is an efficiency policy to improve the social benefit. In the right of Figures 10 and 11, the social benefit functions are convex in three cases. In other words, we obtain an optimal taxi buffer size which maximizes the social benefit. From the left of Figure 11, it can be seen that the social benefit in the fully observable case is more than that in the fully unobservable case. Moreover, the gap between the social benefit in the almost unobservable case and fully unobservable case becomes smaller as  $N$

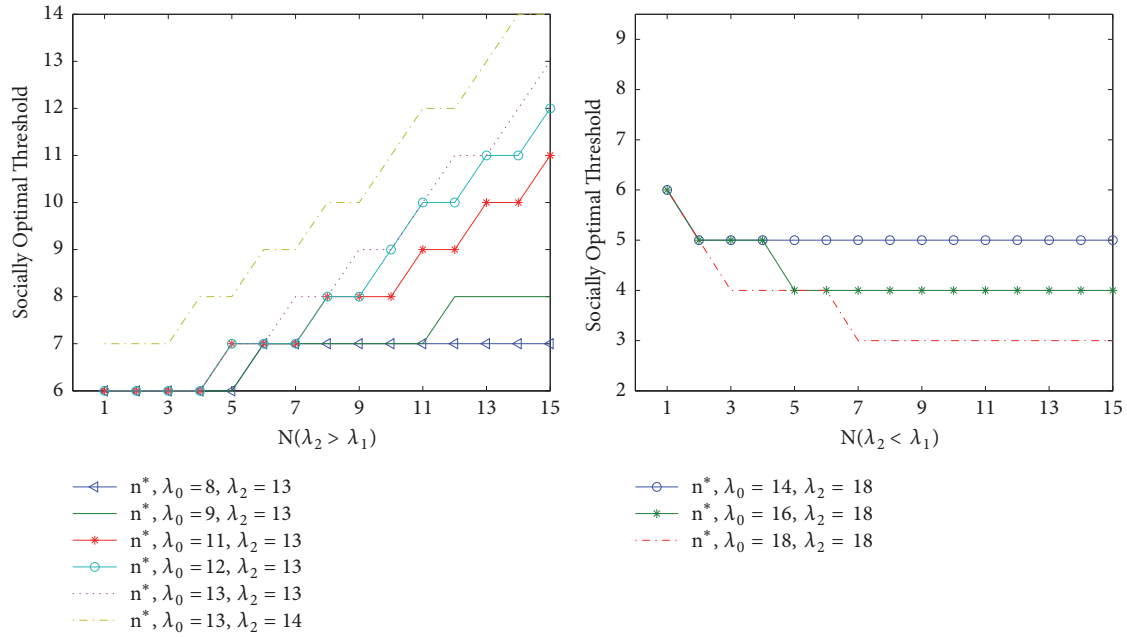


FIGURE 7: Left: socially optimal threshold in the observable case vs.  $N(\lambda_2 > \lambda_1)$  for  $R = 30, C_1 = 20, C_2 = 15, \lambda_1 = 10$ . Right: socially optimal threshold in the observable case vs.  $N(\lambda_2 < \lambda_1)$  for  $R = 30; C_1 = 15; C_2 = 10; \lambda_1 = 20$ .

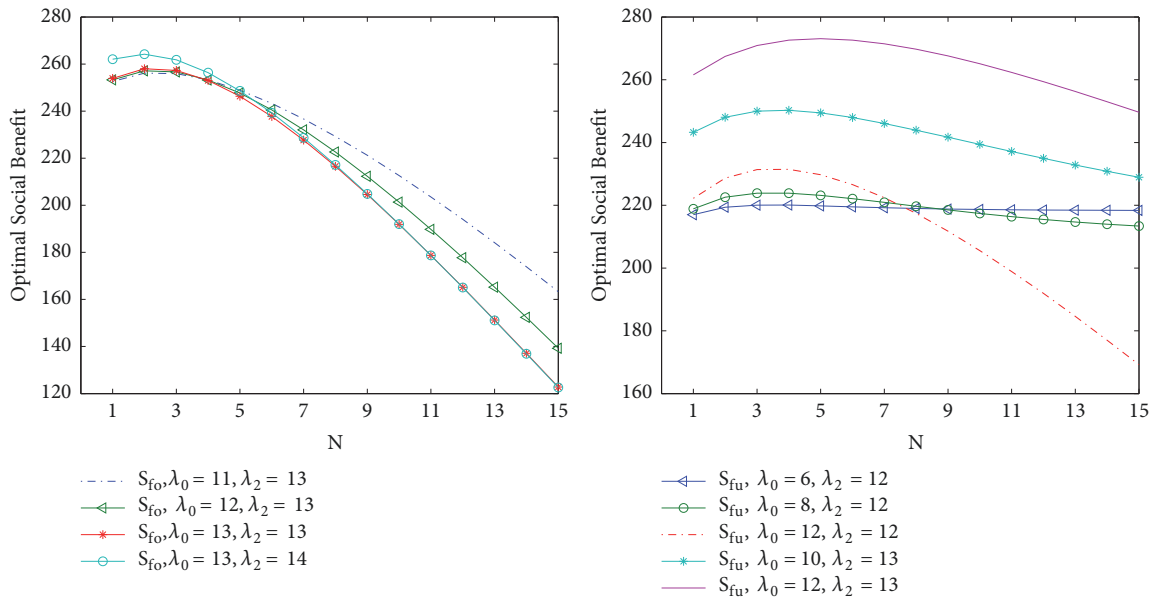


FIGURE 8: Left: optimal social benefit in observable case vs.  $N$  for  $R = 30, C_1 = 10, C_2 = 10, \lambda_1 = 15$ . Right: optimal social benefit in the unobservable case vs.  $N$  for  $R = 30, C_1 = 10, C_2 = 10, \lambda_1 = 15$ .

increases. This behavior indicates that providing the queue length of passengers and taxis can improve the social benefit. From the right of Figure 11, when  $\lambda_0 < \lambda_2 < \lambda_1$ , we know that the social benefits in the fully observable case and almost unobservable case are larger than that in the fully unobservable case. However, the gap between the social benefit in the fully observable case and almost unobservable case becomes smaller as  $N$  increases. This phenomenon indicates that when  $\lambda_0 < \lambda_2 < \lambda_1$ , if the cost of providing

the fully information of the system is large, announcing the state of taxis can also greatly improve the social benefit.

### 7. Conclusions

In this paper, we study the passenger-taxi service system with dynamic taxi control by a double-ended Markovian queueing system. The taxi control is to improve the arrival rate of taxis

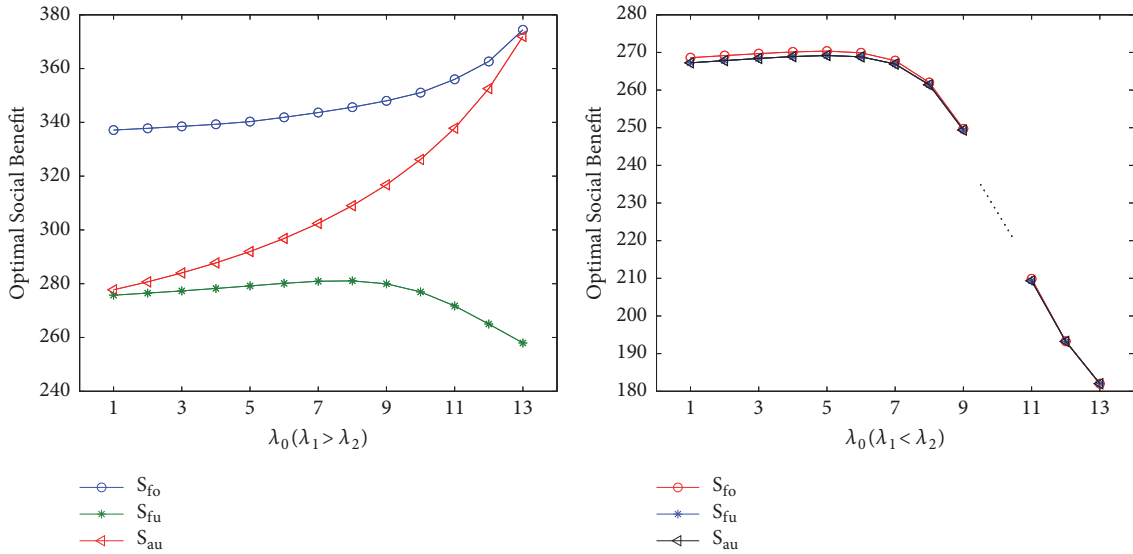


FIGURE 9: Left: Optimal social benefit under three different information levels vs.  $\lambda_0$  ( $\lambda_1 > \lambda_2$ ) for  $R = 30, C_1 = 10, C_2 = 10, \lambda_1 = 20, \lambda_2 = 13, N = 15$ ; Right: Optimal social benefit under three different information levels vs.  $\lambda_0$  ( $\lambda_1 < \lambda_2$ ) for  $R = 30, C_1 = 10, C_2 = 10, \lambda_1 = 10, \lambda_2 = 13, N = 15$ .

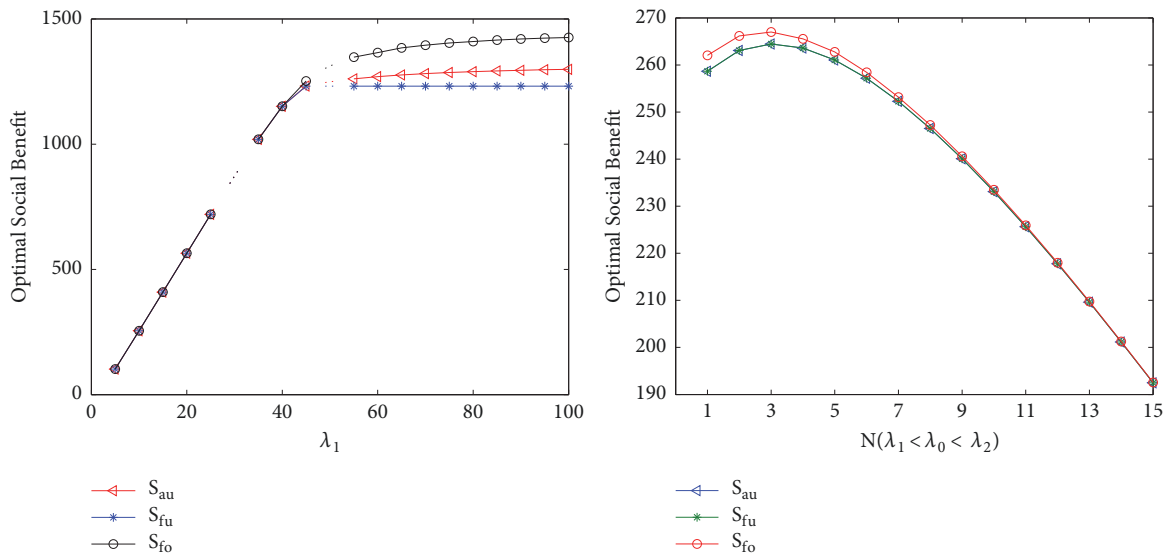


FIGURE 10: Left: optimal social benefit under three different information levels vs.  $\lambda_1$  for  $R = 30, C_1 = 15, C_2 = 10, \lambda_0 = 30, \lambda_2 = 50, N = 5$ . Right: optimal social benefit under three different information levels vs.  $N$  ( $\lambda_1 < \lambda_0 < \lambda_2$ ) for  $R = 30, C_1 = 15, C_2 = 10, \lambda_1 = 10, \lambda_0 = 12, \lambda_2 = 13$ .

when the queue length of passengers is large so as to reduce delays and decrease it at times of increased queue length of taxis so as to reduce the costs of taxis. We derive the passenger's and taxi's expected waiting times in three different information levels. By the reward-cost structure, we analyze the strategic behavior of arriving passengers from their individual utility and social benefit under three different information levels.

We study the (Nash) equilibrium strategies and socially optimal strategies in three different information levels, respectively. We obtain the selfishly and socially optimal thresholds of passengers in the fully observable

case. In the almost unobservable case and the fully unobservable case, we consider the selfishly and socially optimal joining probabilities for arriving passengers. Furthermore, the numerical results showed that dynamic taxi control policy can greatly improve the social benefit compared with the model with the same arrival rate of taxis.

In order to reduce the waiting time of passengers, a possible extension to this work can be to consider a model with priority. In another direction, the extension of the study to non-Markovian models with general interarrival times seems also important.



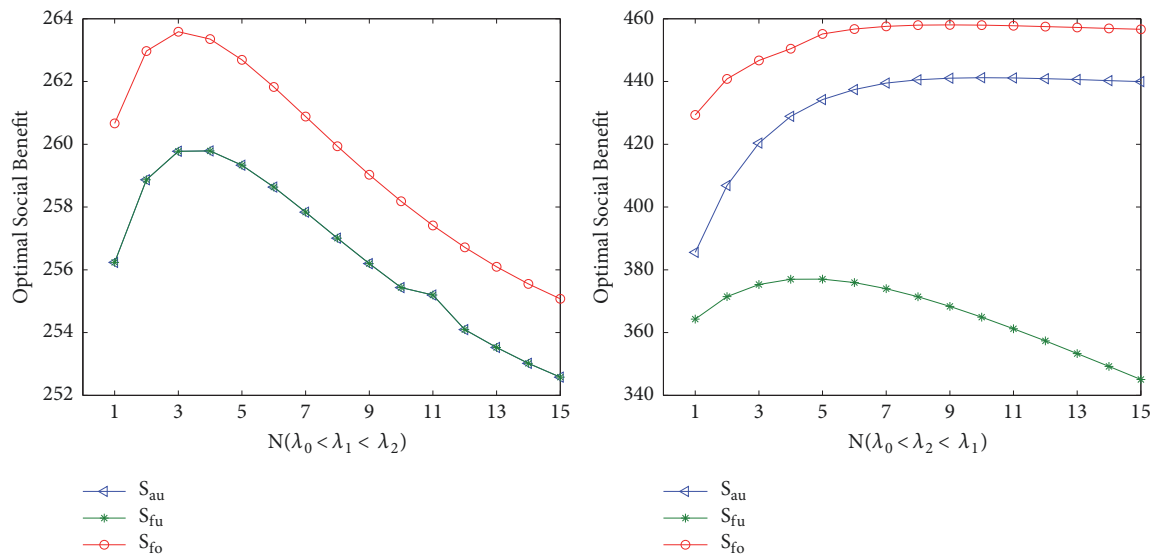


FIGURE 11: Left: optimal social benefit under three different information levels vs.  $N(\lambda_0 < \lambda_1 < \lambda_2)$  for  $R = 30, C_1 = 15, C_2 = 10, \lambda_1 = 10, \lambda_0 = 8, \lambda_2 = 13$ . Right: optimal social benefit under three different information levels vs.  $N(\lambda_0 < \lambda_2 < \lambda_1)$  for  $R = 30, C_1 = 20, C_2 = 15, \lambda_1 = 20, \lambda_0 = 16, \lambda_2 = 18$ .

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research is partially supported by the National Natural Science Foundation of China (11671404).

## References

- [1] D. G. Kendall, "Some problems in the theory of queues," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 13, no. 2, pp. 151–185, 1951.
- [2] J. Dobbie, "A double-ended queueing problem of Kendall," *Operations Research*, vol. 9, no. 5, pp. 755–757, 1961.
- [3] S. M. Giveen, "A taxicab problem with time-dependent arrival rates," *SIAM Review*, vol. 5, pp. 119–127, 1963.
- [4] B. R. Kashyap, "The double-ended queue with bulk service and limited waiting space," *Operations Research*, vol. 14, pp. 822–834, 1966.
- [5] B. R. Kashyap, "Further results for the double ended queue," *Metrika. International Journal for Theoretical and Applied Statistics*, vol. 11, pp. 168–186, 1966/1967.
- [6] K. I. Wong, S. C. Wong, M. G. H. Bell, and H. Yang, "Modeling the bilateral micro-searching behavior for Urban taxi services using the absorbing Markov chain approach," *Journal of Advanced Transportation*, vol. 39, no. 1, pp. 81–104, 2005.
- [7] B. W. Conolly, P. R. Parthasarathy, and N. Selvaraju, "Double-ended queues with impatience," *Computers & Operations Research*, vol. 29, no. 14, pp. 2053–2072, 2002.
- [8] A. Di Crescenzo, V. Giorno, B. Krishna Kumar, and A. G. Nobile, "A double-ended queue with catastrophes and repairs, and a jump-diffusion approximation," *Methodology and Computing in Applied Probability*, vol. 14, no. 4, pp. 937–954, 2012.
- [9] S. A. Zenios, "Modeling the transplant waiting list: a queueing model with renegeing," *Queueing Systems*, vol. 31, no. 3-4, pp. 239–251, 1999.
- [10] R. Wong, W. Szeto, and S. Wong, "Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: moling methodology and policy implications," *Transport Policy*, vol. 33, pp. 73–81, 2014.
- [11] P. Naor, "The refulation of queue size by levying tolls," *Econometrica*, vol. 37, no. 1, pp. 15–24, 1969.
- [12] N. M. Edelson and D. K. Hildebrand, "Congestion tolls for Poisson queueing processes," *Econometrica*, vol. 43, no. 1, pp. 81–92, 1975.
- [13] H. Yang and T. Yang, "Equilibrium properties of taxi markets with search frictions," *Transportation Research Part B: Methodological*, vol. 45, no. 4, pp. 696–713, 2011.
- [14] A. Burnetas and A. Economou, "Equilibrium customer strategies in a single server Markovian queue with setup times," *Queueing Systems*, vol. 56, no. 3-4, pp. 213–228, 2007.
- [15] A. Burnetas, A. Economou, and G. Vasiliadis, "Strategic customer behavior in a queueing system with delayed observations," *Queueing Systems*, vol. 86, no. 3-4, pp. 389–418, 2017.
- [16] P. Guo and R. Hassin, "Strategic behavior and social optimization in Markovian vacation queues," *Operations Research*, vol. 59, no. 4, pp. 986–997, 2011.
- [17] J. Wang, X. Zhang, and P. Huang, "Strategic behavior and social optimization in a constant retrial queue with the  $N$ -policy," *European Journal of Operational Research*, vol. 256, no. 3, pp. 841–849, 2017.

- [18] Y. Shi and Z. Lian, "Optimization and strategic behavior in a passenger–taxi service system," *European Journal of Operational Research*, vol. 249, no. 3, pp. 1024–1032, 2016.
- [19] Y. Shi and Z. Lian, "Equilibrium strategies and optimal control for a double-ended queue," *Asia-Pacific Journal of Operational Research*, vol. 33, no. 3, Article ID 1650022, p. 18, 2016.
- [20] F. Wang, J. Wang, and Z. G. Zhang, "Strategic behavior and social optimization in a double-ended queue with gated policy," *Computers & Industrial Engineering*, vol. 114, pp. 264–273, 2017.

## Research Article

# Evaluation of Transportation Network Reliability under Emergency Based on Reserve Capacity

Xiongfei Zhang <sup>1</sup>, Qi Zhong,<sup>2</sup> and Qin Luo <sup>1</sup>

<sup>1</sup>College of Urban Traffic and Logistics, Shenzhen Technology University, Shenzhen 518118, China

<sup>2</sup>College of Urban Rail Transit, Shenzhen University, Shenzhen 518060, China

Correspondence should be addressed to Qin Luo; [luoqin82@126.com](mailto:luoqin82@126.com)

Received 14 December 2018; Accepted 18 February 2019; Published 7 March 2019

Guest Editor: Marcos M. Vega

Copyright © 2019 Xiongfei Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There are differences between the requirements for traffic network for traffic demand in daily and emergency situations. In order to evaluate how the network designed for daily needs can meet the surging demand for emergency evacuation, the concept of emergency reliability and corresponding evaluation method is proposed. This paper constructs a bilevel programming model to describe the proposed problem. The upper level problem takes the maximum reserve capacity multiplier as the optimization objective and considers the influence of reversible lane measures taken under emergency conditions. The lower level model adopts the combined traffic distribution/assignment model with capacity limits, to describe evacuees' path and shelter choice behavior under emergency conditions and take into account the traits of crowded traffic. An iterative optimization method is proposed to solve the upper level model, and the lower level model is transformed into a UE assignment problem with capacity limits over a network of multiple origins and single destination, by adding a dummy node and several dummy links in the network. Then a dynamic penalty function algorithm is used to solve the problem. In the end, numerical studies and results are provided to demonstrate the rationality of the proposed model and feasibility of the proposed solution algorithms.

## 1. Introduction

In recent decades, the frequent occurrence of emergencies around the world has caused certain casualties and property losses, posing a great challenge to public security. Especially when these emergencies occur in densely populated urban areas, they are undoubtedly becoming a burden to the already crowded urban traffic. This premise puts forward higher requirements for traffic network under emergency.

Apparently, different from the daily demand of the traffic network, the traffic demand under emergency conditions has an evident feature of unidirection and will surge within a short time. As a result, the existing capacity of road infrastructures is usually not able to meet the demand of evacuation.

To investigate the maximum emergency evacuation demand that can be satisfied by the network designed for daily needs, this paper analyzes the traffic demand under emergency conditions and takes the emergency reliability of traffic network as the research target.

Meanwhile, so as to take full advantage of the finite road resources, ease the traffic pressure which exceeds the load of the road network, increase the evacuation capacity of the road network, and improve the evacuation efficiency, traffic managers will take certain emergency traffic management measures, such as lane reversal. Hence, when studying the traffic network emergency reliability, we should not only consider the particularity of emergency traffic demand, but also consider the impact of these measures on the network service capacity at the same time. As a result, the network under this situation can be deemed as a variable one with capacity constraints, which is more in line with the traits of traffic supply and demand in an emergency.

The remainder of this paper is organized as follows: Section 2 is the literature review. Section 3 is the problem statement. Section 4 establishes the mathematical model and Section 5 proposes the solution algorithm. Section 6 is the case study. Section 7 summarizes the conclusions.

## 2. Literature Review

Traffic reliability research began in the 1980s. Over the past decade, transportation reliability has been a new avenue for advancing transportation research in terms of both theories and practical applications, and has attracted tremendous effort over the past few years [1]. Chen first introduced capacity reliability as a new network performance index [2]. In general, it is defined as one of the important indexes to measure traffic network quality and evaluate network operation state. The higher the reliability of the road network, the more stable the overall operation of the road network, indicating that the structure of the road network is reasonable, with strong anti-interference ability. Existing reliability studies of road networks at home and abroad mainly include the following three categories: connectivity reliability, travel time reliability, and capacity reliability.

Chang [3] analyzed with different paths failure strategy quantitatively by measuring the betweenness centrality of the paths so as to study the reliability of highway and transportation network. Li [4] proposed a method to calculate connectivity reliability of shortest path in the network model of complex system. Based on the concept of network connectivity reliability, Harun [5] discussed the assessment of road maintenance and network capacity expansion planning, revealed the relationship between network capacity expansion and connectivity reliability.

Wael [6] investigated the impact of roadworks undertaken on a given road link over wider parts of the network and assessed network travel time reliability by setting up a network assignment model. Sabyasachee [7] proposed a method based on empirically observed travel time data, to measure and determine the value of travel time reliability, estimate changes in it, and incorporate reliability in the transportation planning process. Vincenza [8] proposed a methodology for estimating travel time reliability of an extended road network, through the calibration of empirical relations and the finding's representation on GIS maps with a dynamic simulation model. Taking into account the influence of travel time variability on the congestion profile, Roxana [9] addressed the valuation of travel time reliability in the presence of endogenous congestion and the role of scheduling preferences.

Researches in capacity reliability mainly concern about the max-flow of the network in early stage, especially in transportation network. Capacity reliability is defined as the largest multiplier applied to a given existing (or basic) OD demand matrix that can be allocated to a network without violating the link capacities [10]. Chen [2] extended the capacity reliability analysis by providing a comprehensive methodology, which combines reliability and uncertainty analysis, network equilibrium models, and sensitivity analysis of equilibrium network flow, to assess the performance of a degradable road network. To evaluate capacity reliability, Kuang [11] built up a bilevel programming model based on travel time reliability, concerning OD traffic demand multiplier, and investigated the interaction between two kinds of reliability. Lee [12] proposed a capacity reliability algorithm with Monte Carlo simulation in communication

network based on the shortest delay which can indicate the delay's transient variation caused by flow fluctuation. Qin [13] considered quantity of emergency resources, reserve capacity, and number of locations and built an optimization model for emergency resource layout considering service reliability, in which the objective function is to minimize the system cost and various constraints are considered.

The reliability of road network under extreme conditions is very important. In the case of not considering the road damage, how much margin a network has to accommodate emergency traffic demand is an important question that needs to be considered in traffic network design. Based on capacity reliability, this paper evaluates traffic network reliability under emergency conditions from planning perspective.

## 3. Problem Statements

This section first analyzes the characteristics of evacuation traffic flow, then summarizes emergency traffic management measures under emergency conditions, and finally leads to the establishment of emergency reliability concept.

*3.1. Traffic Flow Characteristics under Emergency Conditions.* Different from the traffic state under daily conditions, the traffic state under emergency evacuation has the characteristics of suddenness, high risk, contingency, and so on. In general, emergency evacuation traffic has the following characteristics.

(1) *Large and Centralized Demand.* From the perspective of space, the origin node of emergency evacuation traffic demand is concentrated in the influenced area. And from the perspective of time, traffic demand for emergency evacuation is basically concentrated in a short period of time before and after the disaster, and people need to be evacuated to a safe area in a short time. This leads to the uneven spatial and temporal distribution of traffic flow directly [14]. In the case of emergencies, the peak of travel demand is several times higher than the peak of daily travel demand, which brings great difficulties to the traffic management of emergency evacuation.

(2) *Characteristics of Crowded Traffic Flow.* Due to the large and centralized traffic demand, a large number of crowded points and congested sections will inevitably appear in the traffic network. The traffic flow on the road is basically saturated or oversaturated, and the evacuation traffic flow is basically in the following state, so it is difficult to find an acceptable gap for overtaking [15]. Therefore, emergency evacuation traffic network is a network with congestion influence.

(3) *High Reliability Requirement.* Sudden disasters threaten people's life and property safety, so emergency evacuees have higher requirements on the reliability of whether they can be quickly evacuated to a safe area. The reliability requirements

are as follows: network connectivity reliability, road capacity reliability, travel time reliability, etc.

(4) *Uneven Flow of Both Directions*. In the process of evacuation, the traffic flow from the affected area to other refuge areas is usually large, resulting in oversaturation and serious traffic congestion. However, the traffic flow into the evacuation zone is small, and most lanes are idle.

3.2. *Emergency Traffic Management strategies*. Depending on traffic flow characteristics under emergency conditions, there are several emergency traffic management measures that are commonly used in practice. The purpose is to improve evacuation efficiency and reduce casualties and property losses.

(1) *Staged Evacuation Strategy*. Staged evacuation strategy is widely used in emergency situations when different parts of the target network may suffer different levels of severity over different time windows. By evacuating those populations in the network via an optimized sequence, the staged evacuation strategy can best utilize available roadway capacity, optimally distribute the total demand over the evacuation time horizon, and thus minimize the network congestion level [16, 17].

(2) *Lane Reversal*. In order to alleviate the one-way traffic demand in emergency evacuation which is far beyond the capacity of the road network, the direction of one or more lanes of the road can be reversed to increase the capacity of evacuation direction [15].

(3) *Crossing Elimination Strategy*. The delay of evacuation vehicles mainly occurs at intersections, which is caused by the conflict and interweaving of evacuation flows from different directions. The interference of evacuation traffic flow can be reduced by eliminating or reducing the confluence point and conflict point of intersections or changing the conflict point into confluence point, which can improve the maximum traffic flow of evacuation network significantly.

The most widely used emergency traffic control strategy is lane reversal.

3.3. *Establishment of Emergency Reliability Concept*. Given a traffic network which is designed for daily traffic demand, in order to evaluate the extent how the network can meet the evacuation requirements under emergency conditions, the concept of emergency reliability is proposed here. In this study, the concept of network reserve capacity is used to evaluate emergency reliability of traffic network. Network reserve capacity refers to the maximum traffic demand multiplier that can be applied to the given existing (or basic) traffic OD matrix in the process of trip assignment when the road capacity constraints can be satisfied [9].

In this paper, traffic demand multiplier  $\mu$  is used to express emergency reliability of the traffic network. First, the given emergency evacuation demand is expanded  $\mu$  times to obtain a new evacuation demand. Then it is assigned to the traffic network according to evacuees' travel behavior. The

impact of lane reversal implementation on road capacity is considered during trip assignment. Since the capacity of each road section has a certain upper limit, there exists a maximum multiplier  $\mu$ .

With the constant adjustment of traffic demand multiplier  $\mu$ , the trip assignment will approach the network equilibrium state continuously, until the traffic flows allocated to certain road sections reach the upper limit of their capacity. At that time, the maximum value of traffic demand multiplier  $\mu$  is obtained. The maximum value of  $\mu$  reflects the extent of network reliability. If multiplier  $\mu \geq 1$ , it indicates that the traffic network is reliable. And the larger the multiplier is, the better the traffic network works, and the more reliable the traffic network is. At the same time, it also means that the traffic network can accommodate greater extra traffic demand.

In the process of trip assignment, it is assumed that the evacuation demand from each origin node is known, and the evacuation demand to each destination depends on the choice of evacuees. Each evacuation destination corresponds to a shelter. And the evacuation demand that each shelter can accommodate is limited. It is assumed that, in case of emergency evacuation, evacuees tend to choose the fastest routes and the nearest shelter in order to reach the safe area as soon as possible, considering the congestion on the routes and in the shelters.

## 4. Mathematical Formulation

In this section, the mathematical description for the problem of interest is formally proposed with a bilevel programming model to analyze the aforementioned multiplier  $\mu$  quantitatively.

4.1. *Upper Level Model*. Given an existing traffic network  $G(N, A)$ , where  $N$  denotes the set of nodes and  $A$  denotes the set of existing links.  $a, a' \in A$ , represent for the two opposite links of the same road segment respectively.  $n_a$  and  $n_{a'}$  denote the number of lanes on the corresponding links.  $N_{aa'}$  is the total number of lanes of the segment.

The upper level model aims to maximum the variable  $\mu$  while searching for a feasible lane distribution scheme under unbalanced distribution of two-way traffic flow. The formulation is as follows:

$$\max \quad \mu \quad (1a)$$

$$\text{s.t.} \quad n_a + n_{a'} = N_{aa'} \quad (1b)$$

$$n_a, n_{a'} \geq 0 \quad (1c)$$

In this model, the total number of lanes of a certain segment,  $N_{aa'}$ , should be a constant on the basis of the existing traffic network. And the number of lanes of all the segments in each direction in the lane distribution scheme could not exceed the corresponding upper bounds, according to the constraint conditions.

The upper level model influences the lower level model through two sets of variables. The variable  $\mu$  changes the actual evacuating demand calculated in the lower level model,



and the number of lanes on each link,  $n_a$  and  $n_{a'}$ , influences the capacity of each link and ultimately affects the capacity of the entire road network.

**4.2. Lower Level Model.** The lower level model is a traveler behavior model describing route and shelter choices of evacuees in emergency situations. The total number of trips generated at origin nodes is given, and the capacity of a single lane and each shelter is also known. Evacuees will choose the shelter that they can arrive as soon as possible and choose the route by which they can reach the safe area in the shortest time, with the consideration of congestion influence on the routes and in the shelters under emergency circumstance. The number of evacuees reaching each shelter is finally decided by the lower level model. Therefore the lower level model is formulated as a joint UE distribution/assignment model with capacity constraints:

$$\min Z(x, \mu, n) = \sum_a \int_0^{x_a} t_a(\omega, n_a) d\omega \quad (2a)$$

$$\text{s.t. } \sum_k f_k^{rs} = q_{rs} \quad (2b)$$

$$\sum_s q_{rs} = \mu D_r \quad (2c)$$

$$\sum_r q_{rs} \leq \overline{D}_s \quad (2d)$$

$$x_a \leq C_a(n_a) \quad (2e)$$

$$f_k^{rs} \geq 0 \quad (2f)$$

$$q_{rs} \geq 0 \quad (2g)$$

The definitional constraints are as follows:

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \quad (2h)$$

where

$x_a$ : traffic flow on link  $a$ ;

$t_a$ : travel time on link  $a$ , where  $t_a(\bullet)$  is the impedance function which represents the relationship between link travel time and link traffic flow;

$C_a$ : capacity of link  $a$ , where  $C_a(n_a)$  represents the relationship between link capacity and the number of lanes on the link;

$D_r$ : total number of existing trips generated at origin node  $r$ ;

$\overline{D}_s$ : capacity of destination  $s$  (in this paper, destination refers to shelter);

$q_{rs}$ : trip rate between origin  $r$  and destination  $s$ ;

$f_k^{rs}$ : flow on path  $k$  connecting origin  $r$  and destination  $s$ ;

$\delta_{a,k}^{rs}$ : binary variable, if link  $a$  is on path  $k$  between OD pair  $r$  and  $s$ ,  $\delta_{a,k}^{rs} = 1$ , otherwise,  $\delta_{a,k}^{rs} = 0$ .

In this model, (2a) is the objective function, which is the sum of the integrals of the link performance functions. This function does not have any intuitive economic or behavioral interpretation. It should be viewed strictly as a mathematical construct that is utilized to solve equilibrium problems. Equations (2b) and (2c) represent a set of flow conservation constraints. Equation (2b) states that the flow on all paths connecting each OD pair has to equal the OD trip rate, and (2c) indicates that the total number of trips from each origin node to all the destination nodes has to equal the amplified traffic demand with variable  $\mu$  at each origin node. Equation (2d) indicates that the total number of trips end with a certain shelter is constrained by the capacity of the shelter. Equation (2e) states that the capacity of a link constrains the flow on it. Equations (2f) and (2g) are nonnegative conditions to ensure that the solution of the program will be physically meaningful.

The objective function of program (2a)–(2h),  $Z(x, \mu, n)$ , is formulated in terms of link flows, traffic demand multiplier and number of lanes on each link, whereas the flow conservation constraints are formulated in terms of path flows. Equation (2h) expresses the incident relationships between link flow and path flow, which is determined by the network structure itself.

**4.3. Equivalency Conditions of Lower Level Model.** To demonstrate the equivalence conditions of lower level model established above, it has to be shown that any flow pattern that solves the mathematical program (2a)–(2h) also satisfies the equilibrium conditions. This equivalency is demonstrated in this section by proving that the first-order conditions for the minimization program are identical to the equilibrium conditions. These conditions can be derived by forming and analyzing the Lagrangian [15, 18, 19].

The Lagrangian of the equivalent minimization problem with respect to all the equality and inequality constraints can be formulated as

$$\begin{aligned} L(f, q, \mu, m, \alpha, \beta) &= Z(x(f)) + \sum_{rs} \mu_{rs} \left( q_{rs} - \sum_k f_k^{rs} \right) \\ &+ \sum_r m_r \left( D_r - \sum_s q_{rs} \right) \\ &+ \sum_s \alpha_s \left( \sum_r q_{rs} - \overline{D}_s \right) + \sum_a \beta_a \left( \sum_{rs} \sum_k f_k^{rs} \delta_{a,k}^{rs} - C_a \right) \end{aligned} \quad (3)$$

where  $\mu_{rs}$ ,  $m_r$ ,  $\alpha_s$ , and  $\beta_a$  are Lagrangian multipliers.

According to Karush-Kuhn-Tucker (KKT) conditions, Lagrangian function (3) must satisfy the following conditions at its extreme points:

$$\begin{aligned} \frac{\partial L}{\partial f_k^{rs}} f_k^{rs} &= 0 \\ \text{and } \frac{\partial L}{\partial f_k^{rs}} &\geq 0 \end{aligned} \quad (4a)$$

$$\frac{\partial L}{\partial q_{rs}} q_{rs} = 0$$

$$\text{and } \frac{\partial L}{\partial q_{rs}} \geq 0$$
(4b)

$$\frac{\partial L}{\partial \alpha_s} \alpha_s = 0$$

$$\text{and } \frac{\partial L}{\partial \alpha_s} \leq 0$$
(4c)

$$\frac{\partial L}{\partial \beta_a} \beta_a = 0$$

$$\text{and } \frac{\partial L}{\partial \beta_a} \leq 0$$
(4d)

$$\frac{\partial L}{\partial \mu_{rs}} = 0$$
(4e)

$$\frac{\partial L}{\partial m_r} = 0$$
(4f)

One can solve the partial derivatives in expression (3) and substitute them into expressions (4a) to (4f), which yields

$$(\bar{c}_k^{rs} - \mu_{rs}) f_k^{rs} = 0$$
(5a)

$$\bar{c}_k^{rs} - \mu_{rs} \geq 0$$
(5b)

$$(\mu_{rs} - m_r + \alpha_s) q_{rs} = 0$$
(5c)

$$\mu_{rs} - m_r + \alpha_s \geq 0$$
(5d)

$$\left( \sum_r q_{rs} - \bar{D}_s \right) \alpha_s = 0$$
(5e)

$$\sum_r q_{rs} - \bar{D}_s \leq 0$$
(5f)

$$(x_a - C_a) \beta_a = 0$$
(5g)

$$x_a - C_a \leq 0$$
(5h)

$$q_{rs} - \sum_k f_k^{rs} = 0$$
(5i)

$$D_r - \sum_s q_{rs} = 0$$
(5j)

$$f_k^{rs} \geq 0$$
(5k)

$$q_{rs} \geq 0$$
(5l)

where

$$\bar{t}_a = t_a(x_a) + \beta_a$$
(5m)

$$\bar{c}_k^{rs} = \sum_{rs} \sum_a \bar{t}_a(x_a) \delta_{a,k}^{rs}$$
(5n)

Here,  $\bar{t}_a$  and  $\bar{c}_k^{rs}$  both denote the generalized travel time, while  $\alpha_s$  denotes the penalty coefficient related to shelter

capacity, and  $\beta_a$  denotes queuing delay due to limited link capacity respectively.

Equations (5a) and (5b) spell out the familiar user-equilibrium condition over the network. With (5k), it can be inferred that the flow on path connecting any origin and destination satisfies

- (1) if  $f_k^{rs} > 0$ ,  $\bar{c}_k^{rs} = \mu_{rs}$ ;
- (2) if  $f_k^{rs} = 0$ ,  $\bar{c}_k^{rs} \geq \mu_{rs}$ .

That is to say, the paths connecting any O-D pair can be divided into two categories: those carrying flow, on which the generalized travel time equals the minimum O-D generalized travel time; and those not carrying flow, on which the generalized travel time is greater than (or equal to) the minimum O-D generalized travel time. This is satisfied with the first equilibrium principle of Wardrop.

Similarly, with (5c), (5d), and (5l), the following conditions can be speculated:

- (1) if  $q_{rs} > 0$ ,  $\mu_{rs} + \alpha_s = m_r$ ;
- (2) if  $q_{rs} = 0$ ,  $\mu_{rs} + \alpha_s \geq m_r$ .

That is to say, when shelter  $s$  is chosen by evacuees from origin  $r$ , the generalized impedance from origin  $r$  to shelter  $s$  is the smallest one among the generalized impedances from origin  $r$  to all the shelters. Otherwise, if none of evacuees choose shelter  $s$  as a destination, it means the generalized impedance from origin  $r$  to shelter  $s$  is greater than (or equal to) the minimum generalized travel time from origin  $r$  to all the shelters. This is consistent with the assumption that all evacuees tend to choose the nearest shelter.

From (5e) and (5f), it can be inferred that, for shelter  $s$ , its capacity penalty coefficient satisfies

- (1) if  $\sum_r q_{rs} = \bar{D}_s$ ,  $\alpha_s \geq 0$ ;
- (2) if  $\sum_r q_{rs} < \bar{D}_s$ ,  $\alpha_s = 0$ .

The penalty coefficient of shelter  $s$  depends on the amount of demand reaching this destination. It only exists when shelter  $s$  has reached its capacity. Penalty coefficient of shelter  $s$  is caused by congestion in the shelter, but the penalty coefficient will not increase indefinitely, as evacuees will change their destinations to avoid excessive delays.

Similarly, from (5g) and (5h), it can be inferred that, for link  $a$ , its queuing delay satisfies the following:

- (1) if  $x_a = C_a$ ,  $\beta_a \geq 0$ ;
- (2) if  $x_a < C_a$ ,  $\beta_a = 0$ .

That is, the queuing delay on link  $a$  depends on its traffic flow. It only exists when the link traffic volume achieves its capacity. Queue delay is caused by congestion on the link and it causes the traffic flow to be redistributed spatially so that the traffic flow of all links does not exceed their capacity.

Equations (5i) and (5j) are the flow conservation constraints.

By solving the lower programming model, the link flows and path flows (the result of path flows is not unique) that meet the above conditions can be obtained. But if the demand is too large and exceeds the capacity of the whole traffic network, the model will have no solution.

**Step 1: Initialization.**

Set  $\mu^0 = 1, k = 1$ . Initialize parameters including step size  $\sigma$  ( $\sigma > 0$ ), population size  $m$ , maximum number of iterations  $max$  and mutation probability  $p$ .

**Step 2: Generate initial lane distribution schemes.**

Generate  $m$  lane distribution schemes randomly, including the original lane distribution scheme. All the schemes are represented by  $\mathbf{u} = \{u_1, u_2, \dots, u_m\}$ , which is the change in the number of lanes in each pair of sections.

**Step 3: Find the feasible lane distribution schemes.**

For each lane distribution scheme:

(3.1) Calculate the number of lanes on each link  $a, n_a$ , and its capacity  $C_a$ ;

(3.2) Run the lower programming model according to demand  $\mu^k D_r$ ;

(3.3) If any solution exists, store the corresponding lane distribution scheme into the set of historical feasible schemes  $\mathbf{P}$ .

Turn to Step 5.

**Step 4: Renew the schemes.**

If all schemes can't find a feasible solution in the lower level model, they should be renewed according to following procedure:

(4.1) Renew the update count of lane schemes,  $n$ , and if  $n > max$ , turn to Step 6;

(4.2) If there are more than  $m$  schemes in set  $\mathbf{P}$ , randomly pick  $m$  schemes and do mutation processing by using local search algorithm, then obtain  $m$  new schemes.

(4.3) If there are less than  $m$  schemes in set  $\mathbf{P}$ , suppose it is  $w$ , do mutation processing by using local searching algorithm to all of these schemes. Besides, generate  $m - w$  schemes at random.

(4.4) Turn to Step 3.

**Step 5:** Let  $\mu^{k+1} = \mu^k + \sigma, k = k + 1$ , and turn to Step 3.

**Step 6: Termination.**

Let  $\mu = \mu^k - \sigma$ . The algorithm ends.

ALGORITHM 1: Iterative optimization algorithm.

## 5. Solution Algorithm

This section aims to design an algorithm to effectively solve the proposed bilevel model. Since bilevel programming is a NP-hard problem, which is difficult to solve by traditional optimization algorithms or analytical methods, the usual solving methods are based on heuristic algorithms. Therefore, in this paper, iterative optimization algorithm is adopted in the upper model. Besides, the addition of constraint conditions in the lower model changes the original Descartes form of UE model, which results in the fact that the lower model cannot be solved by traditional F-W method. Commonly used methods include penalty function method and Lagrange multiplier method. In this paper, the lower model is transformed and then solved by dynamic penalty function algorithm proposed by Zhang [19]. Step-by-step procedures of two algorithms are also given below, respectively.

**5.1. Solution Algorithm for Upper Level Model.** The specific steps of the iterative optimization algorithm are shown in Algorithm 1.

The process of local searching algorithm in Algorithm 1 is as follows. Lane schemes are generally generated in a monotonous way, such as random generating method, which can easily get trapped in local optimality. In order to obtain better newly generated schemes, this paper starts from historically feasible lane distribution schemes in set  $\mathbf{P}$  and searches their neighborhoods to get new schemes based on the characteristics of specific problems.

For each road section, judge the flow distribution results of a certain historical feasible scheme; if  $t_a/C_a \geq t_{a'}/C_{a'}$ ,  $u_a$

increases by 1 with probability  $p$  while ensuring that it does not exceed its upper limit; otherwise,  $u_a$  decreases by 1 with probability  $p$  while ensuring it does not exceed its lower limit. This is the principle of generating a new lane distribution scheme.

**5.2. Solution Algorithm for Lower Level Model.** As for the lower level model, the topology of the origin network is firstly transferred. Augment the original network with a dummy node, denoted  $s'$ , and connect all the destination nodes (i.e., shelters) to node  $s'$ , deriving several dummy links. The total number of dummy links is the product of the number of shelters. Assume that free flow travel time on each dummy link is 0, and capacity of each dummy link equals the capacity of the corresponding shelter.

This simple modification of the network topology can transform the combined distribution/assignment problem with constraint of traffic capacity into an equivalent UE problem with constraint of traffic capacity. Based on the modification, a dynamic penalty function algorithm is used to solve the problem.

The basic idea of the dynamic penalty function algorithm is to transform equivalent UE problem with constraint into traditional UE problem by adding penalty function. In UE problem with constraint, the generalized travel time  $\overline{C}_k^{rs}$  at equilibrium is composed of travel time calculated by impedance function and queuing delay denote by Lagrangian multipliers  $\beta_a$ . However, the Lagrangian multipliers cannot be calculated directly, thus a penalty function  $p(x_a, \lambda_a) = \lambda_a f(x_a/C_a)$  is used to replace it. Here,  $\lambda_a$  is penalty function coefficient related to link  $a$ . Let  $y$  denotes the ratio of  $x_a$  to

**Step 1: Initialization.**

(1.1) Set  $n = 1$ , initialize parameters including penalty function parameters  $\rho$ , penalty function coefficients  $\lambda_a^0$  and convergence criteria  $\varepsilon_1$  and  $\varepsilon_2$ ;

(1.2) Set valid paths of each O-D pair  $K_w^{rs} = \emptyset$ .

**Step 2: Calculate initial solution  $f$** 

(2.1) Calculate the generalized travel time  $\bar{t}_a$  according to the initial link flow  $\{x_a = 0, \forall a\}$ , and find the shortest path between each OD pair;

(2.2) Apply all-or-nothing assignment to get the initial path flow  $f$ .

**Step 3: For each OD pair  $rs$ :**

(3.1) Update the set of valid paths  $K_w^{rs}$  and calculate the generalized travel time  $\bar{t}_a$ , obtain the shortest path  $k_{rs}$  and its travel time  $u_{rs}$  between each OD pair  $rs$ . If the convergence criterion Eq. (\*) is not satisfied, then  $K_w^{rs} = K_w^{rs} \cup \{k_{rs}\}$ ;

$$\frac{\min_{k \in K_w^{rs}} \bar{c}_k^{rs}(f) - u_{rs}}{u_{rs}} \leq \varepsilon_1 \quad (*)$$

(3.2) Linearize the NCP problem composed of effective paths at  $\{f_k^{rs}, \forall k \in K_w^{rs}\}$  and solve it by improved Lemke algorithm [22] to get new  $\{f_k^{rs}, u_{rs}\}$ ;

(3.3) If the convergence criterion Eq. (\*\*) is not satisfied, return to (3.2).

$$\frac{\max_{k \in K_w^{rs}, f_k > 0} \bar{c}_k^{rs}(f) - u_{rs}}{\max_{k \in K_w^{rs}, f_k > 0} \bar{c}_k^{rs}(f)} \leq \varepsilon_1 \quad (**)$$

**Step 4:** Update the link flow  $x^n = x(f)$  and penalty function parameter  $\lambda^n$ .

**Step 5: Termination criterion.**

If the convergence criterion Eq. (\*\*\*) is satisfied,  $x^n$  is the answer. Otherwise, let  $n = n + 1$  and turn to Step 3.

$$\frac{|x^n - x^{n-1}|}{x^{n-1}} \leq \varepsilon_2 \quad (***)$$

ALGORITHM 2: Dynamic penalty function algorithm.

$C_a$  so as to simplify the instruction. The following form of function  $f(y)$  is adopted:

$$f(y) = \begin{cases} \frac{\rho}{2(1-y)} & y < 1 - \rho \\ \frac{y-1}{2\rho} + 1 & y \geq 1 - \rho \end{cases} \quad (6)$$

where  $\rho$  can be a very small constant between 0 and 1.

Function  $f(y)$  has the following characteristics:

- (1) monotone, continuous, and derivable;
- (2) when  $y = 1$ ,  $f(y) = 1$  and  $p(x_a, \lambda_a) = \lambda_a$ ; when  $y > 1$ ,  $f(y)$  is very big; when  $y < 1$ ,  $f(y)$  is very small.

During the problem solving process, both  $\lambda_a$  and  $f(x_a/C_a)$  will be constantly updated according to the new link traffic flow obtained from each iteration, so that the penalty function  $p(x_a, \lambda_a)$  will finally approach Lagrange multiplier  $\beta_a$ . Update rule for  $\lambda_a$  is as follows:

$$\lambda_a^n = \lambda_a^{n-1} f\left(\frac{x_a^n}{C_a}\right) \quad (7)$$

Due to the poor convergence performance of the traditional Frank-Wolfe method for solving UE model, the linear decomposition algorithm proposed by Aashitiani and Magnanti [20, 21] is used here to solve UE problem. Based on the path flow, the linear decomposition algorithm starts from the nonlinear complementarity problem (NCP) of UE problem and transforms it into a series of linear complementarity problems.

The specific steps of the dynamic penalty function algorithm for the lower level model are shown in Algorithm 2.

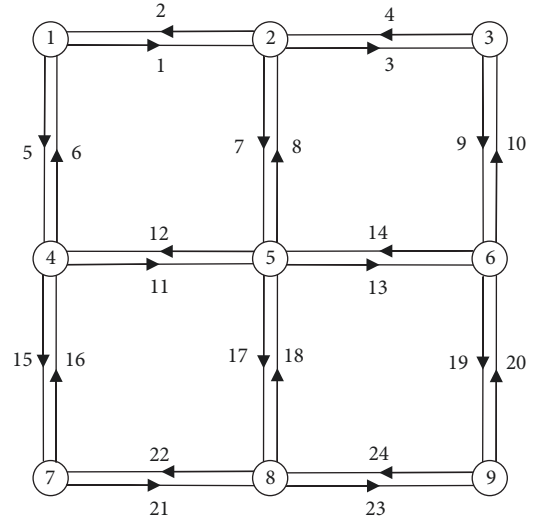


FIGURE 1: Topology of tested network.

## 6. Numerical Experiments

In this section, the above proposed model and algorithm will be tested and evaluated in an experimental network. First the topology of network and relevant information is given, then the analysis of the results.

**6.1. Topology of Network.** Take the test network in the Figure 1 as the example, which has 9 nodes and 24 directed links. Here node 1 and node 4 are the origin points of evacuation, and node 3, 6, and 9 are shelters. Information of link properties

TABLE 1: Information of link properties.

Link no.	Starting node	Ending node	Number of lanes	Free-flow travel time ( <i>min</i> )	Capacity ( <i>pcu/h</i> )
1	1	2	3	1.5	900
2	2	1	3	1	900
3	2	3	3	0.8	900
4	3	2	3	0.8	900
5	1	4	2	1	600
6	4	1	2	0.8	600
7	2	5	4	1.25	1200
8	5	2	4	1	1200
9	3	6	3	1.25	900
10	6	3	3	2	900
11	4	5	4	1.5	1200
12	5	4	4	1.25	1200
13	5	6	4	0.75	1200
14	6	5	4	0.6	1200
15	4	7	3	0.8	900
16	7	4	3	0.9	900
17	5	8	4	1.25	1200
18	8	5	4	1.2	1200
19	6	9	3	1.1	900
20	9	6	3	1	900
21	7	8	3	0.95	900
22	8	7	3	0.9	900
23	8	9	3	1.2	900
24	9	8	3	1.3	900

TABLE 2: Information of evacuation demand.

No.	Origin node	Demand ( <i>pcu/h</i> )
1	1	1500
2	4	1200

TABLE 3: Information of shelter capacity.

No.	Destination node	Capacity ( <i>pcu/h</i> )
1	3	1000
2	6	800
3	9	1200

in the network, including link free flow travel time and link capacities, is listed in Table 1. Information of evacuation demand of each origin point is listed in Table 2, and Table 3 gives the information of the capacity of shelters.

Modify the original network by adding a dummy node 10 and connect it with three shelters by three dummy links. The topology of modified network is shown in Figure 2. As mentioned above, free flow travel time of each dummy link is always 0, and capacity of each dummy link equals to the capacity of the corresponding shelter. On the basis of the modification above, the problem of interest is transformed

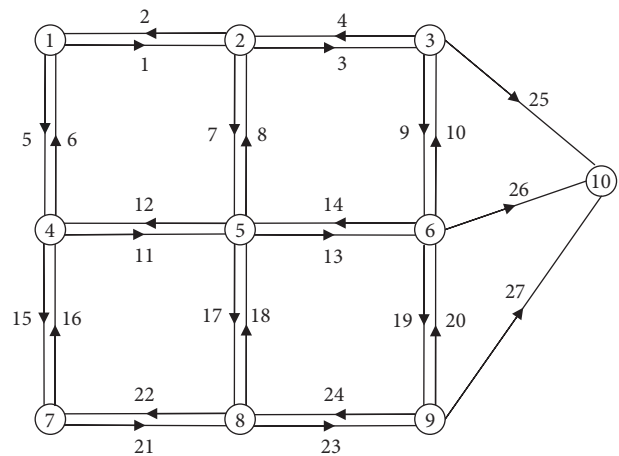


FIGURE 2: Topology of modified network.

into UE assignment problem of a multiorigin and single-destination network with link capacity constraints.

### 6.2. Result Analysis

(1) *Result Analysis of the Lower Model.* The trip assignment results are analyzed when  $\mu = 1$  without changing the



TABLE 4: Path details of OD pair 1-10 at equilibrium.

Path No.	Path flow ( <i>pcu/h</i> )	Composition of path (link no.)	Travel time ( <i>min</i> )	Queuing delay ( <i>min</i> )	Generalized travel time ( <i>min</i> )
1	900	1,3,25	2.645	2.1993	4.8443
2	557.6044	5,15,21,23,27	4.2837	0.5606	4.8443
3	42.3956	5,11,13,26	3.7375	1.1068	4.8443
4	0	5,11,7,3,25	5.0450	1.0999	6.1449
5	0	1,7,13,26	3.8375	2.2060	6.0435
6	0	5,15,21,17,13,9,25	6.4106	0.5606	6.9712

TABLE 5: Path details of OD pair 4-10 at equilibrium.

Path No.	Path flow ( <i>pcu/h</i> )	Composition of path (link no.)	Travel time ( <i>min</i> )	Queuing delay ( <i>min</i> )	Generalized travel time ( <i>min</i> )
1	757.6044	11,13,26	2.5875	1.1065	3.6940
2	42.3956	15,21,23,27	3.1337	0.5603	3.6940
3	400	11,13,19,27	3.6939	0.0001	3.6940
4	0	11,13,10,25	4.5875	0	4.5875
5	0	6,1,3,25	3.4450	2.1993	5.6443
6	0	15,21,23,19,9,25	5.4901	0.5603	6.0504

TABLE 6: Detailed information for OD pairs.

No.	Origin node	Destination node	Demand ( <i>pcu/h</i> )	Minimum generalized travel time ( <i>min</i> )
1	1	3	900	4.8443
2	1	6	42.3956	4.8443
3	1	9	557.6044	4.8443
4	4	3	0	3.8375
5	4	6	757.6044	3.6940
6	4	9	442.3956	3.6940

number of lanes in each link. In the experiment, values of each parameter are as follows:  $\rho = 0.01$ ,  $\varepsilon_1 = \varepsilon_2 = 0.001$ , the initial penalty function coefficient  $\lambda_a^0 = 0.1t_0$ , where  $t_0$  denotes the average of free flow travel time on all links. The BPR formula is employed as impedance function and two parameters in the function are set as 0.15 and 4, respectively.

(a) *From the Perspective of Route Choice.* Tables 4 and 5 show the path details of OD pair 1-10 and 4-10 at equilibrium. From the perspective of route choice, for OD pair 1-10, three paths are chosen by evacuees, and the generalized travel time for these three paths is all equal to 4.8443min, which is the minimum generalized travel time of this OD pair. The generalized travel time of unselected paths including path 4, 5, and 6 is 6.1449, 6.0435, and 6.9712min, respectively, all larger than 4.8443min. Similarly, for OD pair 4-10, three paths are chosen by evacuees, and the generalized travel time for these three paths is all equal to 3.6940min. The generalized travel time of unselected paths including path 4, 5, and 6 is 4.5875, 5.6443, and 6.0504min, respectively.

Although the path flow results of traffic assignment in crowded traffic networks with queuing delays are not

necessarily unique, the result satisfies that generalized travel time of all the selected paths is equal, which meets the conditions of UE equilibrium.

(b) *From the Perspective of Shelter Choice.* According to the path flow results, the demand from origins to all the shelter can be referred, shown in Table 6. For origin node 1, the number of evacuees choosing shelter 3, 6, and 9 is 900, 42.3956, and 557.6044, respectively. And the generalized travel time from node 1 to all the shelters is equal to 4.8443. For origin node 4, the number of evacuees choosing shelter 6 and 9 is 757.6044 and 442.3956, respectively. No one chooses shelter 3 because its generalized travel time is 3.8375, a little larger than other shelters.

That means for a certain origin, generalized travel time of all the selected shelters is equal, and generalized travel time of shelters which are not chosen is larger than the chosen shelters.

Table 7 shows the final shelter demands. The number of evacuees choosing the three shelters is 900, 800, and 1000, respectively. And among them, only shelter 6 reaches its capacity, and the queuing delay is 1.1064 min. The other two

TABLE 7: Detail information for shelter choice.

No.	Destination node	Capacity ( <i>pcu/h</i> )	Demand ( <i>pcu/h</i> )	Queuing Delay ( <i>min</i> )
1	3	1000	900	0
2	6	800	800	1.1064
3	9	1200	1000	0

TABLE 8: Information of different demand and shelter capacity scenarios.

Demand scenarios			Shelter capacity scenarios		
	Origin node	Demand ( <i>pcu/h</i> )		Destination node	Capacity ( <i>pcu/h</i> )
D1	1	1500	S1	3	1000
	4	1200		6	800
D2	1	700	S2	9	1200
	4	2000		3	3000
D3	4	700		6	800
	7	2000		9	3000

TABLE 9: Information of combination of different scenarios.

No.	Demand scenario	Shelter capacity scenario
Case 1	D1	S1
Case 2	D1	S2
Case 3	D2	S2
Case 4	D3	S2

shelters are not saturated and the queuing delays are both 0. This is consistent with the previous conclusion that the queuing delay associated with shelters only exists when the shelter has reached its capacity.

All the experimental results are consistent with the above analysis of first-order conditions of the Lagrangian function.

(2) *Results Analysis of Bilevel Model.* In order to illustrate the rationality of the model and the effectiveness of the algorithm proposed in this paper, three different demand scenarios and two different shelter capacity scenarios (shown in Table 8) are constructed, and four cases of combination of different scenarios are generated (shown in Table 9).

In the solving algorithm, set step size  $\sigma = 0.01$ , population size as  $m = 10$ , maximum iteration count  $max = 50$ , and mutation probability  $p = 0.8$ .

(a) *Results Analysis of Different Cases.* In case 1, the maximum multiplier  $\mu$  that the solution algorithm can obtain is 1.11. The corresponding feasible lane distribution scheme and link flow are shown in Table 10. Only links 5, 15, and 19 are saturated, but there is still plenty of potential room for the capacity of these links. Table 11 shows the corresponding shelter demands. Three shelters are chosen by 997, 800, and 1200 evacuees, respectively. All the shelters are almost saturated. It turns out that the emergency reliability of the entire road network in the case here is limited to the shelter capacity.

TABLE 10: Feasible lane distribution scheme and link flow in Case 1 when  $\mu = 1.11$ .

Link No.	Number of lanes	Capacity	Link flow ( <i>pcu/h</i> )
1	5	1500	1065
2	1	300	0
3	4	1200	1000
4	2	600	0
5	2	600	600
6	2	600	0
7	6	1200	65
8	2	600	0
9	5	1500	0
10	1	300	0
11	5	1500	1332
12	3	900	0
13	7	2100	1100
14	1	300	0
15	2	600	600
16	4	1200	0
17	4	1200	297
18	4	1200	0
19	1	300	300
20	5	1500	0
21	3	900	600
22	3	900	0
23	3	900	897
24	3	900	0

In case 2, by increasing the shelter capacity, the maximum multiplier  $\mu$  comes up to 2.00. The corresponding results at optimal are shown in Tables 12 and 13. Although shelter 6 is

TABLE 11: Detailed information for shelter choice in Case 1 when  $\mu = 1.11$ .

No.	Destination node	Capacity (pcu/h)	Demand (pcu/h)
1	3	1000	997
2	6	800	800
3	9	1200	1200

TABLE 12: Feasible lane distribution scheme and link flow in Case 2 when  $\mu = 2.00$ .

Link No.	Number of lanes	Capacity	Link flow (pcu/h)
1	6	1800	1800
2	0	0	0
3	6	1800	1800
4	0	0	0
5	4	1200	1200
6	0	0	0
7	4	1200	0
8	4	1200	0
9	1	300	0
10	5	1500	0
11	7	2100	2100
12	1	300	0
13	8	2400	2100
14	0	0	0
15	5	1500	1500
16	1	300	0
17	0	0	0
18	8	2400	0
19	5	1500	1300
20	1	300	0
21	6	1800	1500
22	0	0	0
23	6	1800	1500
24	0	0	0

TABLE 13: Detailed information for shelter choice in Case 2 when  $\mu = 2.00$ .

No.	Destination node	Capacity (pcu/h)	Demand (pcu/h)
1	3	3000	1800
2	6	800	800
3	9	3000	2800

saturated, but there is still plenty of room in shelter 3 and 9, to accommodate additional evacuation demands. Some typical links are saturated including links 1, 3, 5, 11, and 15. Meanwhile, the opposite links of link 1, 3, and 5 have no remaining capacity, which means the capacity of these three links cannot be enlarged by applying reversal lane strategy. In other words, the emergency reliability of the entire road network is limited.

TABLE 14: Comparison between calculated optimal solution and true optimal solution.

	Calculated optimal	Computation time	True optimal
Case 1	1.11	125	1.111111
Case 2	2.00	1125	2
Case 3	2.22	1340	2.222222
Case 4	1.80	940	1.8

In conclusion, the emergency reliability of the entire road network can be limited by both shelter capacity and road capacity. Although improving road capacity is a good way to improve network emergency reliability, sometimes, increasing the shelter capacity, there would also be a significant increase in the network emergency reliability.

(b) *Performance Evaluation of the Proposed Algorithm.* Case 3 and 4 are obtained by adjusting the pattern of evacuation demand. Both cases can be solved effectively by the proposed algorithm.

By analyzing the solution results of all the above cases and combining the enumeration method, the true optimal value of multiplier  $\mu$  in different cases can be obtained. The true optimal can also be checked by the network topology under feasible lane distribution schemes. The results are shown in Table 14. It can be seen that, in different cases, the optimal multiplier calculated by the proposed algorithm is almost equal to its true optimal value, and the slight difference in some cases is caused by the step size. Therefore, the algorithm proposed in this paper is proved to be of high accuracy.

The computation time of the proposed algorithm is mainly consumed in solving the lower level joint UE distribution/assignment problem. Therefore, in this paper, the times of solving the lower assignment problem are used to evaluate the solving efficiency of the proposed algorithm. Each case runs 10 times, and the average times to solve the lower assignment problem are listed in Table 14. It can be seen that the calculation times are acceptable in those numerical experiments. In addition, the research results of this paper are applicable to evaluate the existing road network at the level of management planning, and the solving efficiency of the model can meet the actual needs.

## 7. Conclusions

Reliability is an important feature of transportation network. In order to evaluate the reliability of road network under emergency evacuation, a bilevel programming model is proposed in this paper. The concept of reserve capacity is adopted, and the influence of reversible lane measures taken under emergency conditions, as well as the capacity limits of links and shelters, are considered. An iterative optimization method is proposed to solve the upper level model, and the lower level model is transformed and then solved by a dynamic penalty function algorithm. Finally, the numerical example demonstrates the rationality of the proposed model and feasibility of the proposed solution algorithms.

The proposed lower level model can effectively reflect the route choice and shelter choice behavior of evacuees under emergency condition. Evacuees may choose to evacuate to another shelter if the shelter with smaller travel time is full. Besides, when the shortest path has reached its maximum capacity, evacuees are forced to choose other subprime paths. The assignment result accords with UE equilibrium.

Although the reserve capacity of the network can be improved by reversible lane measures, but both link capacity and shelter capacity could be the restriction of emergency reliability of the entire road network. In addition, the efficiency of the proposed algorithms is also verified in the case study.

The model proposed here is a little complicated and many practical problems have been simplified. In further study, the proposed model can be not difficult to contain congestion and delays at intersections. Additionally, uncertainty factors should be considered, including demand uncertainty, disaster location uncertainty, link capacity uncertainty, and so on.

## Data Availability

All the data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This paper is supported by the Research Projects of Natural Science Foundation of Guangdong Province under Grant no. 2015A030310341, the Research Projects of Natural Science Foundation of Guangdong Province under Grant no. 2018A030313119, and the Research Project of Shenzhen Technology University.

## References

- [1] W. H. K. Lam, H. K. Lo, and S. C. Wong, "Advances in equilibrium models for analyzing transportation network reliability," *Transportation Research Part B: Methodological*, vol. 66, pp. 1–3, 2014.
- [2] A. Chen, H. Yang, H. K. Lo, and W. H. Tang, "Capacity reliability of a road network: an assessment methodology and numerical results," *Transportation Research Part B: Methodological*, vol. 36, no. 3, pp. 225–252, 2002.
- [3] S. C. Chang and L. X. Li, "Reliability analysis of highway and transportation network with paths failure," in *Proceedings of the IEEE Workshop on Advanced Research and Technology in Industry Applications*, Ottawa, ON, Canada, 2014.
- [4] L. Li, L. Jia, Y. Wang, and J. Li, "Reliability evaluation for complex system based on connectivity reliability of network model," in *Proceedings of the International Conference on Logistics, Informatics and Service Sciences (S)*, Barcelona, Spain, 2015.
- [5] A. L. Harun, P. Erriza, S. Idwan et al., "Connectivity reliability in road maintenance and network capacity expansion," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 8, pp. 1542–1554, 2010.
- [6] W. Jamous and C. Balijepalli, "Assessing travel time reliability implications due to roadworks on private vehicles and public transport services in urban road networks," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 5, no. 4, pp. 296–308, 2018.
- [7] S. Mishra, L. Tang, S. Ghader, S. Mahapatra, and L. Zhang, "Estimation and valuation of travel time reliability for transportation planning applications," *Case Studies on Transport Policy*, vol. 6, no. 1, pp. 51–62, 2018.
- [8] T. Vincenza, I. Matteo, and I. Giuseppe, "Estimating travel time reliability in urban areas through a dynamic simulation model," *Transportation Research Procedia*, vol. 27, pp. 857–864, 2017.
- [9] R. J. Javid and R. Jahanbakhsh Javid, "A framework for travel time variability analysis using urban traffic incident data," *IATSS Research*, vol. 42, no. 1, pp. 30–38, 2018.
- [10] J. B. Wu, *The theory and methodology for evacuation traffic network aggregation design [Master thesis]*, Southeast University, 2015 (Chinese).
- [11] A. W. Kuang, Z. Q. Tang, and L. C. Shan, "Road network capacity reliability considering travel time reliability," in *Proceedings of the 13th COTA International Conference of Transportation Professionals (CICTP)*, vol. 96, pp. 1818–1827, 2013.
- [12] D. P. Lee, N. Huang, and Z. Liu, "Capacityreliability algorithm in communication network based on the shortest delay," in *Proceedings of the 4th IEEE International Conference on Network Infrastructure and Digital Content*, Beijing, China, 2014.
- [13] J. Qin, Y. Ye, C. Y. Shen et al., "Optimization method for emergency resource layout for transportation network considering service reliability," *Journal of Railway Science and Engineering*, vol. 15, no. 2, pp. 506–514, 2018 (Chinese).
- [14] J. Yang, H. M. Zhou, K. Liu et al., "Research review of emergency traffic evacuation," *Northern Communications*, vol. 11, pp. 73–76, 2011 (Chinese).
- [15] M.-X. Gao, W.-Y. Hong, and Q.-Q. Si, "Optimization of contraflow links in traffic evacuation based on the critical edge for improving maximum flow of a network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 3, pp. 194–199, 2016.
- [16] Y. Liu, X. R. Lai, and G. L. Chang, "Cell-based network optimization model for staged evacuation planning under emergencies," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1964, no. 1, pp. 127–135, 2006.
- [17] J. Cao, J. Gong, and X. K. Yang, "Analysis of resolving reversible traffic congestion in Beijing," *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, vol. 33, no. 6, pp. 1116–1119, 2009.
- [18] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*, Prentice-Hall, Inc., Englewood Cliffs, NJ, USA, 1985.
- [19] X. F. Zhang, R. M. Li, X. M. Guo et al., "Traffic assignment problem model withturning delays for congested networks," *Journal of Tsinghua University (Science and Technology)*, vol. 52, no. 6, pp. 848–853, 2012.
- [20] H. Z. Aashtiani and T. L. Magnanti, "A linearization anddecomposition algorithm for computing urban traffic equilibria," in *Proceedings of the 1982 IEEE International Large-Scale System Symposium*, Virginia Beach, Va, USA, 1982.
- [21] H. P. Lu and Y. F. Yin, "A method of variational inequalityfor analyzing the urban traffic network equilibrium," *Journal of*

*Highway and Transportation Research and Development*, vol. 14, no. 2, pp. 24–29, 1997 (Chinese).

- [22] I. Adler, R. W. Cottle, and J.-S. Pang, “Some LCPs solvable in strongly polynomial time with Lemke’s algorithm,” *Mathematical Programming*, vol. 160, no. 1-2, pp. 477–493, 2016.



## Research Article

# ST-LSTM: A Deep Learning Approach Combined Spatio-Temporal Features for Short-Term Forecast in Rail Transit

Qicheng Tang, Mengning Yang , and Ying Yang

*School of Big Data & Software Engineering, Chongqing University, Chongqing, China*

Correspondence should be addressed to Mengning Yang; [mnyang@cqu.edu.cn](mailto:mnyang@cqu.edu.cn)

Received 7 November 2018; Revised 29 December 2018; Accepted 21 January 2019; Published 6 February 2019

Guest Editor: Helena Ramalhinho

Copyright © 2019 Qicheng Tang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The short-term forecast of rail transit is one of the most essential issues in urban intelligent transportation system (ITS). Accurate forecast result can provide support for the forewarning of flow outburst and enables passengers to make an appropriate travel plan. Therefore, it is significant to develop a more accurate forecast model. Long short-term memory (LSTM) network has been proved to be effective on data with temporal features. However, it cannot process the correlation between time and space in rail transit. As a result, a novel forecast model combining spatio-temporal features based on LSTM network (ST-LSTM) is proposed. Different from other forecast methods, ST-LSTM network uses a new method to extract spatio-temporal features from the data and combines them together as the input. Compared with other conventional models, ST-LSTM network can achieve a better performance in experiments.

## 1. Introduction

With the development of urban scale, the short-term traffic forecast has become a core issue of ITS. Accurate short-term traffic forecast can provide technical support for the surveillance and the forewarning of passenger flow. Therefore, over the past few decades, many data analysis models have been proposed to promote the forecast accuracy. Among these models, LSTM network is widely recognized as the most suitable model to deal with traffic forecast. LSTM unit has three gates, namely, input gate, forget gate, and output gate, which can adjust the state of unit dynamically, so LSTM network is able to capture the features on longer time span. Therefore, LSTM network can provide a higher accuracy in traffic forecast because traffic data is usually collected according to time series.

In recent years, researchers pay more attention to the spatial features of traffic flow. It is widely acknowledged that traffic forecast is a problem with spatio-temporal complexity, i.e., the problem of spatial transportation in temporal dimension. In [1], Zheng Zhao et al. establish a network by connecting several LSTM units, which aimed to imitate the

structure of urban traffic. However, it failed to imitate the structure of large urban scale. Xiaobo Chen et al. proposed a new method to process spatial features by using sparse hybrid genetic algorithm [2]. Liu Qingchao et al. proposed a model based on manifold similarity to capture the spatial regularity from freeway data [3]. These two approaches are sensitive to the spatial features, but compared with LSTM network, they cannot process temporal features well.

In this paper, the object of study is the short-term forecast of rail transit. In the research, we find that differing from other transportation, rail transit has stations with fixed position, vehicles with uniform speed, and regular schedule. Because of these characteristics, the spatial correlation between stations can be transformed into the time cost. Based on this analysis, this paper proposes a new method to capture spatio-temporal features from rail transit data and input the features into a new model named spatio-temporal long short-term network (ST-LSTM), which is based on LSTM network. Compared with most existing methods, the proposed model has a better performance on accuracy and meets the real-time requirement.

## 2. Related Work

There are many methods that have been proposed to improve traffic forecast, including historical average and smoothing [4, 5], dynamic linear methods [6, 7], traffic theory-based methods [8, 9], and machine learning methods [10, 11]. These forecast approaches can be divided into two categories, namely, parametric approaches and nonparametric approaches. Autoregressive integrated moving average (ARIMA) model is widely recognized as a classic method in parametric approaches. As early as the 1970s, Levin and Tsao found that ARIMA model was the most statistically significant in traffic forecast [12]. Parametric approaches have favorable properties and capture regular variations very well. However, traffic data usually shows irregular variations. To solve this problem, researchers also paid attention to nonparametric approaches, such as nonparametric regression models [13], support vector machine (SVM) [14, 15], and recurrent neural network [16, 17]. Afterwards, recurrent neural network [18] was proposed to process temporal features, such as evolutionary neural network (ENN) [19], dynamic neural network (DNN) [20], and nonlinear autoregressive models with exogenous inputs (NARX) [21]. Among them, RNN is widely recognized as a suitable method to capture the temporal features of passenger flow. However, previous studies proved that RNNs failed to capture the long-term features because of vanishing gradient and exploding gradient. To solve these problems, long short-term memory neural network (LSTM NN) [22] was applied in the traffic forecast. In recent years, some approaches have been proposed to deal with the spatio-temporal complexity of traffic data, which are mentioned in Section 1.

Different from these methods, this paper proposes a new method to capture spatio-temporal features and a new network based on LSTM to forecast the exit passenger flow of rail transit. The remainder of this paper is as follows. Section 3 introduces the architecture of ST-LSTM network. Experiments based on the data of Chongqing rail transit are shown in Section 4. Section 5 is composed of the analysis of experiment result, and future work is at the end of this paper.

## 3. Methodology

Short-term forecast for rail transit is a problem with spatio-temporal complexity. Suppose the exit passenger flow of station  $j$  is needed to be predicted. The temporal features are the correlation between historical data and current data, i.e., the previous exit passenger flow of station  $j$ . These features can be extracted directly because the rail transit data is collected according to the temporal dimension. The spatial features are the transportation of passenger flow on geographic position; i.e., the summation of estimated passenger flows from the other stations. For every two stations, the spatial features include the volume and the cost of transportation between them. The volume of transportation is reflected by the passenger flow between two stations. In the proposed model, spatial correlation matrix (SCM) is integrated to calculate the volume of transportation. The cost of transportation is reflected by several factors, such as time cost, economic cost,

and distance. Among them, time cost is the most suitable factor to reflect the spatial correlation, which is mentioned in Section 1. Therefore, time cost matrix (TCM) is introduced to calculate the cost of transportation. The proposed model is based on the technologies, including passenger information system (PIS), features extraction method, and ST-LSTM network. The detail of the technologies will be explained in this section.

**3.1. Passenger Information System.** Sufficient data is the basis of accurate forecast, and PIS can provide us with comprehensive data. PIS is a huge and complex network. Various rail transit data can be collected in real time through the gate system, ticketing system, and vehicle scheduling system. With the development of data acquisition technology, PIS is able to provide sufficient support for short-term forecast. Based on the card records, the entrance and exit passenger flow of stations are calculated with a frequency of 10 min, which can be denoted by

$$i = \{cardid, otime, ostation, dtime, dstation, date\} \quad (1)$$

$$x_{in_{j,t}} = \sum_{i \in M} \{1 \mid i.ostation = j, i.otime = t\} \quad (2)$$

$$x_{out_{j,t}} = \sum_{i \in M} \{1 \mid i.dstation = j, i.dtime = t\} \quad (3)$$

where  $i$  is a card record.  $cardid$ ,  $otime$ ,  $ostation$ ,  $dtime$ ,  $dstation$ , and  $date$  are the attributes of  $i$ , which represent card identification, origin time, origin station, destination time, destination station, and date, respectively.  $M$  is the database of card records.  $x_{in_{j,t}}$  is the entrance passenger flow of station  $j$  in time  $t$ , and  $x_{out_{j,t}}$  is the exit passenger flow of station  $j$  in time  $t$ .

**3.2. Feature Extraction Method.** The extraction of spatio-temporal features is one of the core problems of the proposed model. The proposed model extracts temporal features and spatial features, respectively, and then put, them together into the ST-LSTM network. The temporal features can be extracted directly, because rail transit data is recorded according to the temporal dimension. To extract the spatial features, TCM matrix and SCM matrix are integrated into the method.

**3.2.1. Time Cost Matrix.** Time cost is the most suitable factor to reflect the spatial correlation between stations, so the time cost between all stations constitutes the TCM matrix. Due to the changes of schedule and passenger flow, TCM matrix is dynamic with time going on. Suppose there are  $m$  stations in the rail transit system; then the size of TCM matrix is  $m \times m$ , which can be denoted by

$$TCM_t = \begin{pmatrix} \Delta T_{1,1} & \Delta T_{1,2} & \cdots & \Delta T_{1,m} \\ \Delta T_{2,1} & \Delta T_{2,2} & \cdots & \Delta T_{2,m} \\ \cdots & \cdots & \Delta T_{k,j} & \cdots \\ \Delta T_{m,1} & \Delta T_{m,2} & \cdots & \Delta T_{m,m} \end{pmatrix} \quad (4)$$

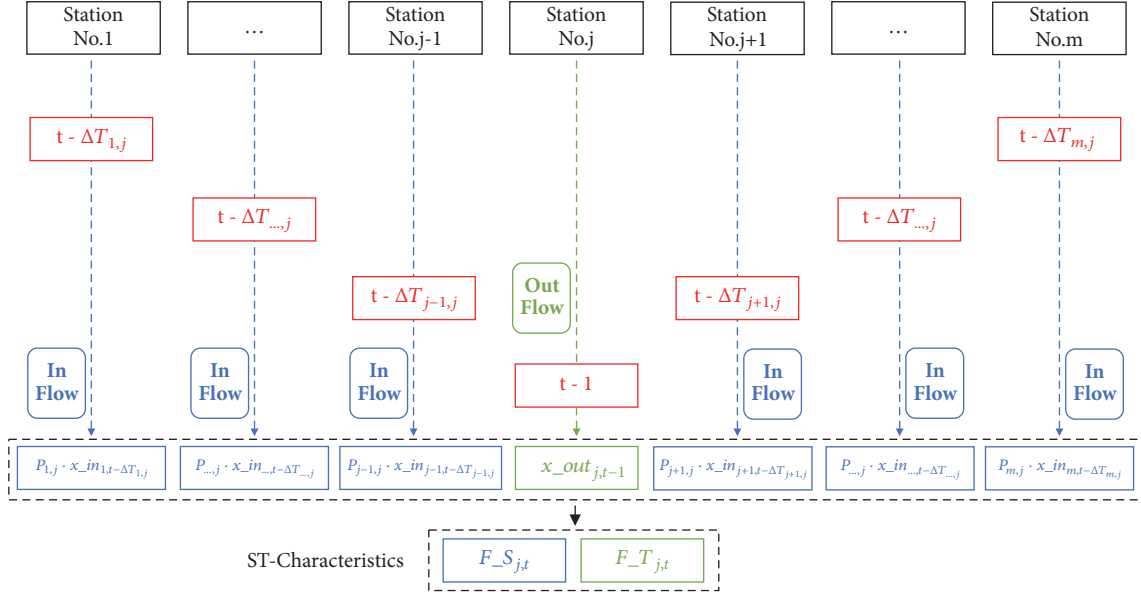


FIGURE 1: Design of the extraction of spatio-temporal features.

$$\Delta T_{k,j} = \overline{\Delta t_i}, \Delta t_i = i.dtime - i.otime \quad (5)$$

$$(i \in M, i.dtime \in T^*, i.ostation = k, i.dstation = j)$$

$$T^* = \{t + qw\}, \quad q = 0, 1, 2, \dots, n \quad (6)$$

where  $TCM_t$  is the TCM matrix in time  $t$ .  $\Delta T_{k,j}$  is the average of time cost between  $k$  and  $j$  in historical time series, where  $k$  is origin station and  $j$  is destination station.  $\Delta t_i$  is the time cost between two stations in record  $i$ .  $i$  is a card record in database  $M$ , which has been defined in Eq.(1).  $T^*$  is the historical time series of time  $t$ .  $w$  is a week and  $q$  is the number of weeks.

In the analysis, it is found that passenger flow varies by people's routine cycle, i.e., from Monday to Sunday. For example, the passenger flow in Thursday is similar to the one in last Thursday not yesterday. Therefore, to promote the extraction,  $\Delta T_{k,j}$  in time  $t$  represents the average time cost in historical time series of time  $t$ . In Eq. (6),  $T^*$  is the historical time series, which consists of time  $t$  and the same period in several weeks ago of it. This method is also used in the calculation of spatial correlation matrix.

**3.2.2. Spatial Correlation Matrix.** To forecast the exit passenger flow at station  $j$  in time  $t$ , passengers from station  $k$  in time  $t - \Delta T_{k,j}$  have to be considered. The entrance passenger flow of station  $k$  in time  $t - \Delta T_{k,j}$  ( $x.in_{k,t-\Delta T_{k,j}}$ ) is available. However, time  $t$  has not happened, so the proportion of passengers in  $x.in_{k,t-\Delta T_{k,j}}$ , which set off to station  $j$ , is unavailable. To solve this contradiction, spatial factor is introduced in this paper. Spatial factor  $P_{k,j}$  is the historical average probability of passengers between station  $k$  and  $j$  in entrance passenger flow. When forecasting the exit passenger flow at station  $j$  in time  $t$ , the spatial influence from station  $k$  can be calculated by multiplying the spatial factor  $P_{k,j}$

and entrance passenger flow  $x.in_{k,t-\Delta T_{k,j}}$ . The spatial factors between all stations constitute the spatial correlation matrix. There is an SCM matrix in each time, because the factors vary according to the time. Suppose there are  $m$  stations in the rail transit system; then the size of SCM matrix is  $m \times m$ , which is denoted by

$$SCM_t = \begin{pmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,m} \\ P_{2,1} & P_{2,2} & \dots & P_{2,m} \\ \dots & \dots & P_{k,j} & \dots \\ P_{m,1} & P_{m,2} & \dots & P_{m,m} \end{pmatrix} \quad (7)$$

$$c_{k,j,t-\Delta T_{k,j}} = \sum_{i \in M} \{1 \mid i.otime = t - \Delta T_{k,j}, i.ostation = k, i.dstation = j\} \quad (8)$$

$$P_{k,j} = \left( \frac{c_{k,j,t-\Delta T_{k,j}}}{x.in_{k,t-\Delta T_{k,j}}} \right), \quad t \in T^* \quad (9)$$

$$T^* = \{t + qw\}, \quad q = 0, 1, 2, \dots, n \quad (10)$$

where  $SCM_t$  is the SCM matrix of time  $t$ .  $P_{k,j}$  is the spatial factor, where  $k$  is origin station and  $j$  is destination station.  $\Delta T_{k,j}$  is the time cost from  $k$  to  $j$ .  $c_{k,j,t-\Delta T_{k,j}}$  is the number of passengers from  $k$  to  $j$ , whose origin time is  $t - \Delta T_{k,j}$ .  $x.in_{k,t-\Delta T_{k,j}}$  is the entrance passenger flow of station  $k$  in time  $t - \Delta T_{k,j}$ .  $T^*$ ,  $q$ , and  $w$  have been defined in Eq. (6).

**3.2.3. Extraction of Spatio-Temporal Features.** The structure of extraction method is shown in Figure 1. To forecast the exit passenger flow at station  $j$  in time  $t$ , the temporal features is the exit passenger flow at station  $j$  in time  $t - 1$ . The spatial features are gathered by calculating the number of passengers,

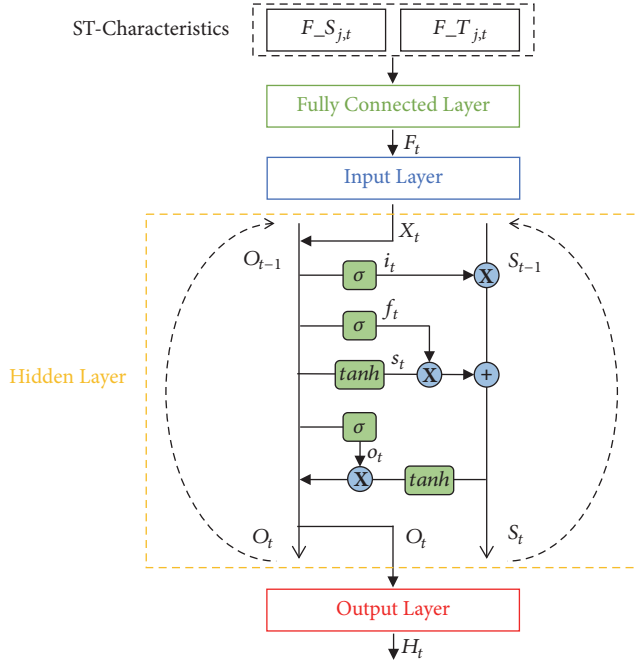


FIGURE 2: Structure of ST-LSTM network.

who will arrive at station  $j$  in time  $t$  and depart from other stations. The function of extraction can be set as

$$F-T_{j,t} = x\_out_{j,t-1} \quad (11)$$

$$F-S_{j,t} = \sum_{k \in N} P_{k,j} \cdot x\_in_{k,t-\Delta T_{k,j}} \quad (12)$$

where  $F-T_{j,t}$  is the temporal features of station  $j$  in time  $t$ .  $x\_out_{j,t-1}$  is the exit passenger flow of station  $j$  in time  $t-1$ .  $F-S_{j,t}$  is the spatial features of station  $j$  in time  $t$ .  $N$  is the set of stations and  $k$  is a station in it.  $P_{k,j}$  is the spatial factor, which has been defined in Eq. (9).  $x\_in_{k,t-\Delta T_{k,j}}$  is the entrance passenger flow of station  $k$  in time  $t-\Delta T_{k,j}$ .

**3.3. Structure of ST-LSTM Network.** Based on LSTM network, a fully connected layer is added to combine temporal features and spatial features in ST-LSTM network. The model will acquire the best mode of combination through the training.

The structure of ST-LSTM network is shown in Figure 2. The input of the model is spatio-temporal features  $F-T_{j,t}$  and  $F-S_{j,t}$ , and the output is the forecast result  $H_t$ . There are four layers in this model, namely, fully connected layer, input layer, hidden layer, and output layer. The fully connected layer combines the features at first and conveys the result  $F_t$  to the input layer. The input of hidden layer  $X_t$  is calculated through the input layer. The hidden layer has three gates, namely, input gate  $i_t$ , forget gate  $f_t$ , and output gate  $o_t$ . Moreover, the state of the hidden layer is indicated by  $S_t$ . The inputs of every gate are  $X_t$  and the previous state  $S_{t-1}$ . The blue points in Figure 2 are confluences, which stand for multiplications, and dashed lines are the transmitting of the previous state. Based on the

information flow, the structure of ST-LSTM network can be summarized as

$$F_t = W_{fs}S_{j,t} + W_{ft}T_{j,t} + b_f \quad (13)$$

$$X_t = W_{hx}F_t + W_{hh}O_{t-1} + b_h \quad (14)$$

$$i_t = \sigma(W^{(i)}X_t + U^{(i)}S_{t-1}) \quad (15)$$

$$f_t = \sigma(W^{(f)}X_t + U^{(f)}S_{t-1}) \quad (16)$$

$$o_t = \sigma(W^{(o)}X_t + U^{(o)}S_{t-1}) \quad (17)$$

$$s_t = \tanh(W^{(c)}X_t + U^{(c)}S_{t-1}) \quad (18)$$

$$S_t = f_t \circ S_{t-1} + i_t \circ s_t \quad (19)$$

$$O_t = o_t \circ \tanh(S_t) \quad (20)$$

$$H_t = W_{oh}O_t + b_y \quad (21)$$

where  $F_t$ ,  $X_t$ ,  $O_t$ , and  $H_t$  are the output of different layers.  $i_t$ ,  $f_t$ ,  $o_t$ , and  $s_t$  are the intermediate variables of the hidden layer.  $S_t$  is the state of the hidden layer.  $W_{fs}$ ,  $W_{ft}$ ,  $W_{hx}$ ,  $W_{hh}$ ,  $W_{oh}$ ,  $W^{(i)}$ ,  $W^{(f)}$ ,  $W^{(o)}$ ,  $W^{(c)}$ ,  $U^{(i)}$ ,  $U^{(f)}$ ,  $U^{(o)}$ , and  $U^{(c)}$  are weight matrices.  $b_f$ ,  $b_h$ , and  $b_y$  are bias vectors and  $\sigma$  is sigmoid function.

The cost function is activated after the forecast through the training. The proposed model is improved by reducing the output of cost function, which can be set as

$$loss = \|\hat{x}_{out,j,t} - x_{out,j,t}\|_2^2 \quad (22)$$

where  $\hat{x}_{out,j,t}$  is the forecast of station  $j$  in time  $t$  and  $x_{out,j,t}$  is the actual output.

**3.4. Training Algorithm.** The training algorithm contains two aspects. One is the extraction of spatio-temporal features, and the other is the training of ST-LSTM network. The key point of training is minimizing the output of cost function by adjusting the weight matrices and bias vectors. The training procedure can be stated as follows.

**Step 1. Obtaining the Inputs and Labels.** Capture the temporal features and the spatial features in each time  $t$ , which are the input of model. Collect the exit passenger flow in each time  $t+1$  as the labels.

**Step 2. Initialization of the ST-LSTM Network.** Initialize the weight matrices and bias vectors, including  $W_{fs}$ ,  $W_{ft}$ ,  $W_{hx}$ ,  $W_{hh}$ ,  $W_{oh}$ ,  $W^{(i)}$ ,  $W^{(f)}$ ,  $W^{(o)}$ ,  $W^{(c)}$ ,  $U^{(i)}$ ,  $U^{(f)}$ ,  $U^{(o)}$ ,  $U^{(c)}$ ,  $b_f$ ,  $b_h$ , and  $b_y$ .

**Step 3. Fine-tuning the Whole Network.** Fine-tune the whole network by adjusting the weight matrices and bias vectors in order to minimize the output of cost function. The process will be stopped until the output meets the qualification or the time of training reaches the limit.

TABLE 1: Performance of four models.

Model	ME	MAE	RMSE	MRE	Operation time
SARIMA	360.70	51.33	82.37	29.63%	75 min
PSO-SVR	153.45	18.89	27.48	21.46%	<b>8 min</b>
LSTM	183.59	22.82	32.29	26.45%	37 min
ST-LSTM	<b>145.72</b>	<b>18.42</b>	<b>25.89</b>	<b>20.77%</b>	38 min

## 4. Experiment

Based on the data of Chongqing rail transit, four models are contained in the experiment, namely Seasonal ARIMA (SARIMA) [23], Support Vector Regression Model combined with Particle Swarm Optimization (PSO-SVR) [24], LSTM network [25], and the proposed ST-LSTM network. The target of forecast is exit passenger flow with a frequency of 10 min. The four models will be trained and tested on 100 stations. The details of each model are as follows.

(1) SARIMA: The seasonal period 'S' is 100, due to the operating time being from 6.20 am to 23.00 pm (100×10 min). After the processing method in [23], ARIMA (2,1,0) × (0,1,1)100 is finally used.

(2) PSO-SVR: The time period is 100 (100×10 min per day), and the limit of parameter combination of SVR is from [10, 0.08] to [500, 0.3]. The final parameter combination will be selected by PSO in the training.

(3) LSTM network: The number of units is 10 and the time step is 100 (100×10 min per day).

(4) ST-LSTM network: The number of units is 10 and the time step is 100 (100×10 min per day).

*4.1. Data Description.* The data of card records are provided by Chongqing City Transportation Development & Investment Group Co., Ltd. Compared with other targets, such as Origin-Destination (OD) volume, exit passenger flow is more accurate and has less missing data. Therefore, we calculate the exit passenger flow from 01 March 2017 to 31 March 2017 based on the dataset. There are more than 46 million card records. After processing, 600 thousand data are calculated.

*4.2. Evaluation.* We use several criteria to compare the performance of four models. Maximum error (ME) and mean absolute error (MAE) are used to measure the accuracy of models. Root mean square error (RMSE) is sensitive to the stability of models. Mean relative error (MRE) is the most suitable to compare the performance of four models. The definitions of criteria are

$$ME = \max_{1 \leq i \leq n} \{|\hat{\varphi}_i - \varphi_i|\} \quad (23)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{\varphi}_i - \varphi_i| \quad (24)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\varphi}_i - \varphi_i)^2} \quad (25)$$

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{\varphi}_i - \varphi_i}{\varphi_i} \right| \quad (26)$$

where  $\hat{\varphi}_i$  is the forecast data, while  $\varphi_i$  is the measured data.

*4.3. Training and Testing.* 5-fold cross validation is used to evaluate the models. In 5-fold cross validation, the data is divided into 5 subsets. Each subset is a testing set, and the rest of data is the training set. The experiments are repeated 5 times for each station. After the experiments, the performance of four models was collected. The experiments are conducted under a desktop computer with Intel i7 3.20 GHZ CPU and 16 GB memory.

*4.4. Experiment Result.* The experiment results of different algorithms are shown in Table 1 and the operation time is averaged on all stations. Compared with SARIMA, PSO-SVR, and LSTM network, the proposed ST-LSTM network achieved a better performance. From the view of ME and MAE, ST-LSTM network is more accurate than the other models. Moreover, from the view of RMSE, ST-LSTM network has a better stability. Therefore, the proposed ST-LSTM network is more suitable for the short-term forecast of rail transit.

## 5. Analysis of Result

When the models have been tested on 100 stations of Chongqing rail transit, we find that ST-LSTM network achieves a higher accuracy than the other models. However, the performance of ST-LSTM network fluctuates on different stations, which are shown in Table 2. Therefore, we analyze the experimental results based on the field investigation. The stations in Table 2 are sorted in descending order by passenger volume. Due to the lack of space, Table 2 just exhibits the performance on stations of top-10 and bottom-10 passenger volume.

*5.1. Base Volume.* In our research, we find that base volume is one of the influence factors of the forecast. The performance of two stations is chosen to shown in Figure 3, which are station No.321 and station No. 334 of Chongqing rail transit. Both of them are located in the residential district of Chongqing. However, station No. 334 only attains 5% of the base volume of station No.321 monthly. In the test, the MRE on station No.321 is 13.52%, while the MRE on station No.334 is 26.58%. We use these two different performances as samples to show the influence of base volume. The research suggests that the stations with higher base volume usually have more prominent regional features. As a result, passenger flows of these stations have stronger regularity and are more insensitive to the emergent factors. Therefore, short-term forecast on stations with low base volume is one of the difficulties in rail transit forecast.



TABLE 2: Forecast performance of ST-LSTM network on different stations.

Station	ME	MAE	RMSE	MRE
321	366.96	71.08	99.12	13.52%
315	365.72	52.45	73.13	16.90%
318	280.70	32.45	47.31	11.13%
114	270.84	44.66	60.63	13.76%
322	215.52	30.41	41.12	12.10%
110	252.81	37.09	50.87	12.90%
606	238.99	33.44	48.46	11.97%
212	204.32	39.95	52.62	12.78%
108	215.70	32.76	47.10	13.94%
123	223.77	61.18	81.89	36.58%
...	...	...	...	...
335	69.81	6.58	10.10	29.34%
221	43.85	6.83	9.07	28.39%
305	58.25	9.55	12.93	32.25%
620	67.57	6.46	9.59	27.07%
621	45.15	6.69	8.91	30.47%
224	31.58	5.99	7.98	32.93%
219	36.12	5.40	7.40	31.86%
619	35.96	4.77	6.56	29.32%
118	23.63	4.41	5.72	29.88%
625	49.00	5.47	8.04	35.59%

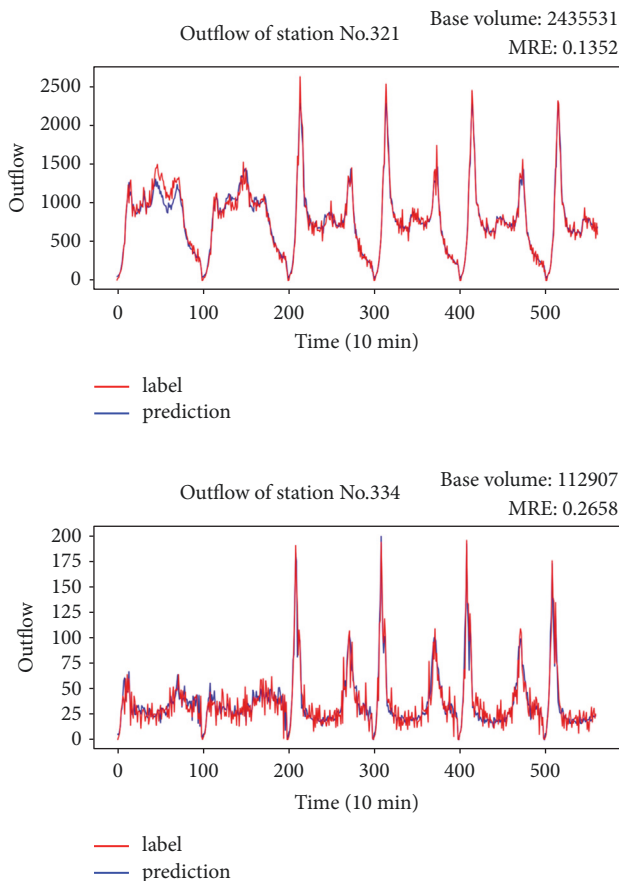


FIGURE 3: Influence of base volume in forecast.

5.2. *Randomness.* Except for the base volume, we discover that randomness is another influence factor of the forecast. As shown in Figure 4, the base volume of station No. 323 and station No.123 are both around 900 thousand monthly. However, the performance of forecast on two stations is quite different. In the test, the MRE on station No.323 is 16.73%, while the MRE on station No.123 is 38.37%. This phenomenon occurs on a few special stations, such as station No.123, which is located in the university town of Chongqing. Compared with the commuters, the undergraduates have more choice on the travel time. So the passenger flow of stations, which next to universities, has stronger randomness than others. Similarly, the passenger flow of stations, which next to railway stations or airports, is related to the flight schedule. Therefore, the randomness from the environment cannot be neglected on several stations. Short-term forecast on these stations is one of the difficulties in rail transit forecast.

## 6. Conclusion and Future Work

Short-term forecast for rail transit is an essential issue in ITS. We propose the ST-LSTM network, which combines the temporal features and spatial features. To extract spatial features, TCM matrix and SCM matrix are integrated into the method. Compared with other models, the proposed model is more suitable for rail transit forecast.

This study researches on prediction of exit passenger flow, but a model which also includes entrance passenger flow is more significant for the management. In addition, except the rail transit, ITS also contains bus system and taxi system. The

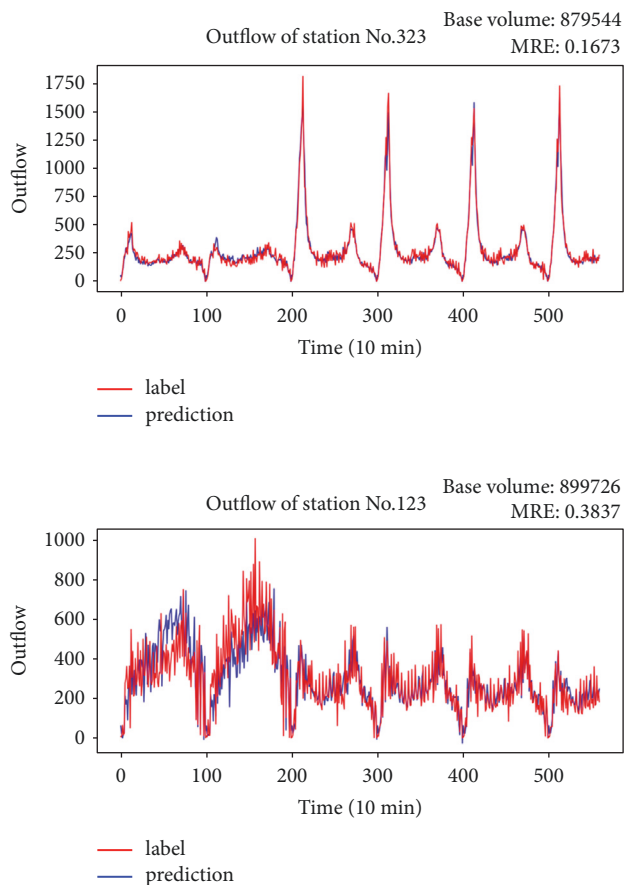


FIGURE 4: Influence of randomness in forecast.

correlation between different public transportation is worth consideration. In the future, we will try to forecast other targets of rail transit and then consider the relation among different transportation. Finally, a comprehensive system for rail transit will be built to output a more accurate result of short-term forecast.

## Data Availability

The data used in the experiments are freely available at [https://drive.google.com/open?id=1RuH080U\\_9PHdh9B9VoOjNur-ODWnHAYaf](https://drive.google.com/open?id=1RuH080U_9PHdh9B9VoOjNur-ODWnHAYaf) and more data of Chongqing rail transit are available by contacting the corresponding author.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

## Acknowledgments

This work is supported by the Central Universities under Grant 106112017CDJXY090001 and Chongqing Major Science and Technology Projects of Artificial Intelligence under Grant CSTC2017RGZN-ZDYF0150.

## References

- [1] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: A deep learning approach for Short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.
- [2] X. Chen, Z. Wei, X. Liu, Y. Cai, Z. Li, and F. Zhao, "Spatiotemporal variable and parameter selection using sparse hybrid genetic algorithm for traffic flow forecasting," *International Journal of Distributed Sensor Networks*, vol. 13, no. 6, pp. 1–14, 2017.
- [3] Q. Liu, Y. Cai, H. Jiang, X. Chen, and J. Lu, "Traffic state spatial-temporal characteristic analysis and short-term forecasting based on manifold similarity," *IEEE Access*, vol. 6, pp. 9690–9702, 2017.
- [4] K. Farokhi Sadabadi, M. Hamed, and A. Haghani, "Evaluating moving average techniques in short-term travel time prediction using an AVI data set," in *Transportation Research Board 89th Annual Meeting*, vol. 2010.
- [5] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.
- [6] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.
- [7] X. Fei, C. C. Lu, and K. Liu, "A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1306–1318, 2011.
- [8] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.
- [9] L. Li, X. Chen, Z. Li, and L. Zhang, "Freeway travel-time estimation based on temporal-spatial queueing model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1536–1541, 2013.
- [10] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3-4, pp. 187–210, 2003.
- [11] Y. Wei and M. Chen, "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 148–162, 2012.
- [12] M. Levin and Y. D. Tsao, "On forecasting freeway occupancies and volumes (abridgment)," *Transportation Research Record*, vol. 773, pp. 47–49, 1980.
- [13] A. Rosenblad, "J. J. Faraway: Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models," *Computational Statistics*, vol. 24, no. 2, pp. 369–370, 2009.
- [14] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.
- [15] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 125, no. 6, pp. 515–523, 2004.
- [16] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 1–11, 1984.
- [17] H. Liu, H. van Zuylen, H. van Lint, and M. Salomons, "Predicting urban arterial travel time with state-space neural networks and Kalman filters," *Transportation Research Record*, no. 1968, pp. 99–108, 2006.

- [18] P. Lingras, S. Sharma, and M. Zhong, "Prediction of recreational travel using genetically designed regression and time-delay neural network models," *Transportation Research Record*, vol. 13, no. 1, pp. 435–446, 2002.
- [19] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.
- [20] L. Shen, *Freeway Travel Time Estimation and Prediction using Dynamic Neural Networks (Ph. D. Dissertation)*, Florida International University, 2008.
- [21] X. Zeng and Y. Zhang, "Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling," *Computer-Aided Civil and Infrastructure Engineering*, vol. 28, no. 5, pp. 359–371, 2013.
- [22] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [23] S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *European Transport Research Review*, vol. 7, no. 3, 2015.
- [24] W. Hu, L. Yan, K. Liu, and H. Wang, "A short-term traffic flow forecasting method based on the hybrid PSO-SVR," *Neural Processing Letters*, vol. 43, no. 1, pp. 155–172, 2016.
- [25] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 47–60, 2018.