

Journal of Advanced Transportation

Innovative Methods for Data Informed Multimodal Transport Flow and System Analysis 2021

Lead Guest Editor: Zhiyuan Liu

Guest Editors: Inhi Kim, Xinyuan Chen, and Di Huang





**Innovative Methods for Data Informed
Multimodal Transport Flow and System
Analysis 2021**

Journal of Advanced Transportation

**Innovative Methods for Data Informed
Multimodal Transport Flow and System
Analysis 2021**

Lead Guest Editor: Zhiyuan Liu





Guest Editors: Inhi Kim, Xinyuan Chen, and Di
Huang



Copyright © 2021 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.

Associate Editors

Juan C. Cano , Spain
Steven I. Chien , USA
Antonio Comi , Italy
Zhi-Chun Li, China
Jinjun Tang , China

Academic Editors

Kun An, China
Shriniwas Arkatkar, India
José M. Armingol , Spain
Socrates Basbas , Greece
Francesco Bella , Italy
Abdelaziz Bensrhair, France
Hui Bi, China
María Calderon, Spain
Tiziana Campisi , Italy
Giulio E. Cantarella , Italy
Maria Castro , Spain
Mei Chen , USA
Maria Vittoria Corazza , Italy
Andrea D'Ariano, Italy
Stefano De Luca , Italy
Rocío De Oña , Spain
Luigi Dell'Olio , Spain
Cédric Demonceaux , France
Sunder Lall Dhingra, India
Roberta Di Pace , Italy
Dilum Dissanayake , United Kingdom
Jing Dong , USA
Yuchuan Du , China
Juan-Antonio Escareno, France
Domokos Esztergár-Kiss , Hungary
Saber Fallah , United Kingdom
Gianfranco Fancello , Italy
Zhixiang Fang , China
Francesco Galante , Italy
Yuan Gao , China
Laura Garach, Spain
Indrajit Ghosh , India
Rosa G. González-Ramírez, Chile
Ren-Yong Guo , China



Yanyong Guo , China
Jérôme Ha#rri, France
Hocine Imine, France
Umar Iqbal , Canada
Rui Jiang , China
Peter J. Jin, USA
Sheng Jin , China
Victor L. Knoop , The Netherlands
Eduardo Lalla , The Netherlands
Michela Le Pira , Italy
Jaeyoung Lee , USA
Seungjae Lee, Republic of Korea
Ruimin Li , China
Zhenning Li , China
Christian Liebchen , Germany
Tao Liu, China
Chung-Cheng Lu , Taiwan
Filomena Mauriello , Italy
Luis Miranda-Moreno, Canada
Rakesh Mishra, United Kingdom
Tomio Miwa , Japan
Andrea Monteriù , Italy
Sara Moridpour , Australia
Giuseppe Musolino , Italy
Jose E. Naranjo , Spain
Mehdi Nourinejad , Canada
Eneko Osaba , Spain
Dongjoo Park , Republic of Korea
Luca Pugi , Italy
Alessandro Severino , Italy
Nirajan Shiwakoti , Australia
Michele D. Simoni, Sweden
Ziqi Song , USA
Amanda Stathopoulos , USA
Daxin Tian , China
Alejandro Tirachini, Chile
Long Truong , Australia
Avinash Unnikrishnan , USA
Pascal Vasseur , France
Antonino Vitetta , Italy
S. Travis Waller, Australia
Bohui Wang, China
Jianbin Xin , China



Hongtai Yang , China
Vincent F. Yu , Taiwan
Mustafa Zeybek, Turkey
Jing Zhao, China
Ming Zhong , China
Yajie Zou , China




Contents

Stochastic Programming of Sustainable Waste Cooking Oil for Biodiesel Supply Chain under Uncertainty

Nana Geng , Qihong Fu, and Yixiang Sun 

Research Article (18 pages), Article ID 5335625, Volume 2021 (2021)

An Online Map Matching Algorithm Based on Second-Order Hidden Markov Model

Xiao Fu , Jiayu Zhang , and Yue Zhang 


Research Article (12 pages), Article ID 9993860, Volume 2021 (2021)

Multidepot Recycling Vehicle Routing Problem with Resource Sharing and Time Window Assignment

Yong Wang , Xiuwen Wang , Xiangyang Guan , and Jinjun Tang 

Research Article (21 pages), Article ID 2327504, Volume 2021 (2021)

Forecasting Civil Aviation Incident Rate in China Using a Combined Prediction Model

Yixiang Sun and Nana Geng 

Research Article (9 pages), Article ID 5521526, Volume 2021 (2021)

Research Article

Stochastic Programming of Sustainable Waste Cooking Oil for Biodiesel Supply Chain under Uncertainty

Nana Geng ¹, Qihong Fu,¹ and Yixiang Sun ²

¹Nanjing University of Posts and Telecommunications, Nanjing 210023, China

²Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Correspondence should be addressed to Yixiang Sun; sunyixiang@nuaa.edu.cn

Received 13 May 2021; Revised 4 July 2021; Accepted 7 July 2021; Published 20 July 2021

Academic Editor: xinyuan chen

Copyright © 2021 Nana Geng 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.

As an important emission reduction source for the transportation industry, biofuel has received strong support from the Chinese government. However, the development of the biofuel industry is still struggling. The high degree of uncertainty makes the development of the industry face huge challenges. Kitchen waste, as a biodiesel raw material with a large yield, has good development prospects. Reuse of kitchen waste can solve public health and safety problems. This paper proposes a two-stage stochastic programming model under supply disturbance to optimize the supply chain from the perspective of contract. Then current three main flow directions of kitchen waste are analysed and the reasonable price for biodiesel operators to purchase is determined. By signing contracts with the biodiesel operators, restaurant is guaranteed and encouraged to provide a certain percentage of kitchen waste to meet the demand for biodiesel production. Using actual case in the Yangtze River Delta region, the performance of the stochastic programming model under disturbance was compared. Through the sensitivity analysis of different parameters, this paper determines the influence of its supply chain network design and expected total system cost. Through the optimization of the waste cooking oil (WCO) for biodiesel supply chain, this paper can effectively improve the efficiency of the supply chain, reduce system costs, increase the profits of biofuel operators, and promote the sustainable development of the biofuel industry.

1. Introduction

In China, with the expansion of the bioenergy development planning in recent years, the amount of biodiesel needed in transport fuels is projected to increase. Moreover, the current biodiesel blend ratio of 5% in China is likely to increase to 15% or more in the near future, which will also create more biodiesel demand. To meet the growing demand of the biodiesel industry, society must provide more sustainable material to produce biodiesel. The Chinese government highly supports the development of the biodiesel industry with kitchen waste as the raw material. The development policy of the biodiesel industry (2015) clearly states that kitchen waste should become the main raw material of biodiesel, and its supply chain needs effective design to improve the overall operational efficiency. However, some studies have shown that few restaurants provide

kitchen waste to biodiesel operators due to uncertainty of the implementation of kitchen waste production in most restaurants and the lack of understanding. Therefore, this paper will focus on putting forward strategies to increase the supply of kitchen waste in restaurants and reduce the cost of the WCO-biodiesel supply chain.

The paper firstly defines and describes the supply chain problems. Previous studies have shown that providing a financial benefit to restaurants is a viable way to encourage restaurants to provide kitchen waste [1, 2]. Through signing a contract, ensure that restaurants sell a certain percentage of kitchen waste at a certain purchase price, which can effectively solve the problem of the uncertainty in the supply of raw materials for biodiesel [3]. Based on the above two models, the paper reconstructs the stochastic programming model under the supply disturbance. In this paper, biodiesel operators provide restaurants with the price of unit kitchen

waste by signing contracts to achieve a stable supply source, and the price is determined by the restaurant's allocation decision model. The supply uncertainty of kitchen waste is an important factor in the design of supply chain network. Therefore, in the process of contract pricing and supply chain model construction, this paper also considers the uncertainty of the total supply of kitchen waste into the supply chain stochastic model construction, so as to enhance the antidisturbance of supply chain network. Based on the above two models, the paper reconstructed stochastic programming model of the supply chain. At last, we choose a case study and do the sensitivity analysis from transportation cost, pretreatment rate, kitchen waste residual value, and penalty fee. The research on the supply chain network in the biodiesel collection stage can facilitate the transportation of kitchen waste from restaurants to biodiesel refineries for biodiesel production. The structure of this paper is as follows. Section 1 analyses the research background and purpose of this paper. Section 2 reviews the relevant literature. In Section 3, the network optimization of the restaurant kitchen waste supply chain is described. In Section 4, the basic stochastic programming model, allocation decision model, and reconstructed stochastic programming model of the restaurant kitchen waste supply chain under supply disturbance are presented. Section 5 discusses the acquisition of data and discussion of results for the case study. Finally, in Section 6, the research results of this paper are summarized.

2. Literature Review

2.1. Biodiesel Supply Chain Uncertainty. In the supply chain, the supply is uncertain due to the capacity limitation of internal enterprises (endogenous factors) and the demand change of external market (exogenous factors). Uncertainty comes from all stages and activities in the biodiesel supply chain, and its type and degree are all different [4]. At present, a large number of literature studies reported that the problems in production, procurement, inventory management, coordination, and other aspects of the biofuel supply chain are caused by supply uncertainty. Zahraee et al. [5] considered the cost-effectiveness problem of the biofuel supply chain under the uncertainty of supply and built a deterministic planning and scheduling model aiming at minimizing the cost. Geng et al. [3] considered the problem of two-level supply chain composed of farmers and producers in the environment of uncertain supply and proposed a cooperation mechanism to improve the profit of each member in the supply chain, but it was limited to the case that market supply and demand fluctuated less. Giarola et al. [6] considered the uncertainty of biomass supply and carbon emission quota trading plan, aimed at maximizing the net present value of supply chain and minimizing carbon emissions, and established a multi-period and multilevel mixed-integer linear programming model, so as to solve the problems of raw material allocation, production technology selection, and plant location of upstream supply chain of bioethanol. Osmani and Zhang [7] established a two-stage stochastic programming model

to maximize profits and minimize carbon emissions and carried out a case study on biomass supply, biofuel demand, and price uncertainty in four Midwest states of the United States using wood fiber as raw material for biofuel supply chain. Biomass supply uncertainty is one of the most important uncertainty factors in the procurement process. As for agricultural waste, on the one hand, biomass depends on planting and harvesting operations and requires a relatively fixed growth cycle, so the supply of biomass presents obvious seasonal characteristics [8]. On the other hand, under the joint action of farmers' planting willingness, weather conditions, soil conditions, and other factors, biomass yield in different planting or harvesting periods and locations also presents certain differences [8, 9]. For waste cooking oil and other municipal wastes, on the one hand, biomass production depends on the dining consumption habits of urban residents and economic development, and so on, and there is a spatial-temporal difference. On the other hand, the actual amount of biomass collected by refining plants is also subject to competition from the illegal production industry chain of "gutter oil," which further aggravates the uncertainty of the amount of biomass. Nguyen and Chen [10] proposed a mathematical model to tackle the supplier selection and operation planning problem in biomass supply chains to help decision-makers facing uncertainty of biomass feedstock supply. Hu and Feng [11] model a supply chain of service requirement and supply and demand uncertainty with revenue sharing contract and derive the buyer's optimal ordering policy and the supplier's optimal supply policy. Lin et al. [12] analysed equilibrium solutions for the cooperative supply chain across different channel structures under supply uncertainty.

2.2. Mathematical Modelling Methods for Uncertain Problems. In recent years, stochastic programming methods are mostly used to deal with supply uncertainty in the supply chain [13]. Stochastic programming is one of the emerging methods to deal with uncertain problems. The stochastic programming method was developed by Danzig and Beare, aiming to model random variables by using a set of discrete scenarios with known probabilities [14–17]. The demand assumption in stochastic planning is a random variable that obeys a uniform distribution, a normal distribution, or a Poisson distribution. Then, an optimal solution is found through stochastic programming to minimize (or maximize) the expected value of the objective function in the supply chain [18]. Due to the characteristics of supply chain design, two-stage stochastic planning has been widely used in problems related to supply chain management [19]. In the first stage, the strategic or long-term decision of supplier selection (such as the number and combination of suppliers) should be made before the realization of random variables. When the random variables are realized, tactical and operational decisions such as order allocation, inventory, production, and transportation will be made in the second stage. When the model is constructed, the uncertainty problem is usually transformed into a deterministic mathematical

programming method, which helps to minimize the impact of uncertainty factors. Dantzig et al. [20] proposed this concept and pointed out that stochastic programming can be divided into two types: two-stage stochastic programming model and multistage stochastic programming model. At present, two-stage stochastic programming is more common. Gupta and Maranas [21] pointed out that the two-order stochastic programming in uncertain supply chain design can be divided into two models, namely, Wait and See model and Here and Now model. An et al. [22] studied the reliable P -median location problem using a two-stage robust model. Ghodrtnama et al. [23] conducted robust and fuzzy target planning for the new multitarget hub location allocation problem. Stochastic programming has been gradually applied to various supply chain optimization and management problems.

2.3. Sustainability of Biodiesel Supply Chain. Sustainable supply chain design and optimization have become an emerging method, which tries to take environmental, economical, and social decisions into full consideration [24]. Geng and Sun [25] summarized the literature related to biodiesel supply chain optimization research. Economic sustainability means the most important objective of the biodiesel supply chain is to produce biodiesel in an economically viable manner [26]. Liu et al. [27] studied a mixed-integer linear programming model for optimizing economic, energy, and environmental objectives in a biofuel supply chain network design problem. The energy objective is measured by the fossil energy input per megajoule of biofuel production. Early work tends to focus on some environmental aspects of the engineering process, such as waste management and net heat consumption [28]. Giarola et al. [29] studied a biofuel supply chain network design model with the aim of cost and CO₂-eqv. minimization. Social sustainability reflects as the development of the bioenergy industry is likely to create new employment opportunities and bring greater economic vitality in rural areas [30–33].

In general, the optimization of biodiesel supply chain is extremely important in the development system of the bioenergy industry. Scholars have done extensive research on theories and methods of biodiesel supply chain optimization, especially in the areas of biodiesel supply chain facility location. However, the theory of supply chain optimization using kitchen waste as raw material in view of China's national conditions is relatively insufficient. There are few papers on both the tactical design of sustainable supply chains and the optimization of biodiesel modelling under uncertain conditions. This paper introduces the strategy of price contract between restaurants and biofuel operators to guarantee restaurants to supply appropriate kitchen waste to biodiesel refineries. Through the construction of a two-stage stochastic programming model, the disturbance of kitchen waste supply was integrated into the design of biodiesel supply chain network. The negotiated purchase price of kitchen waste and the structure of biodiesel supply chain network were determined to ensure the normal operation of biodiesel refineries under disturbance.

3. Problem Description

WCO for biodiesel supply chain is a three-level supply chain including kitchen waste supply point, pretreatment facility, and the demand point (biodiesel refinery). The first level of kitchen waste supply point is the restaurants. The second level is the kitchen waste pretreatment facility, which is used for pretreating kitchen waste. It is built by the biodiesel operator according to the economic cost and environmental impact. The biodiesel operators can sell the excess kitchen waste at the pretreatment facility and buy the kitchen waste when the kitchen waste shortage occurs according to the demand of biodiesel and the supply of kitchen waste. The third level is biodiesel refinery, where all waste cooking oil processed by the pretreatment facility is transported to the refinery for biodiesel production. The kitchen waste is provided by the restaurants and transported by truck to the kitchen waste pretreatment facility, where the treated kitchen waste is transported to the biodiesel refinery, as shown in Figure 1.

Biodiesel production from kitchen waste faces more uncertainties in the future such as feedstock supply, biodiesel demand, and changing regulations and policies. According to statistics in the past 10 years, the total kitchen waste has increased by 1%–8% in different proportions [34]. The government has introduced a series of measures to reduce the amount of kitchen waste. Kitchen waste will continue to have obvious uncertainty in the future. The purpose of this paper is to establish a stochastic programming model considering the contract establishment and provide reliable solutions for designing the whole kitchen waste supply chain under the potential supply disturbances in the future. The specific problems to be solved by the stochastic programming model include the determination of the purchase price of the kitchen waste in the contract signed by the biodiesel operator and the restaurants; the location of the infrastructure supporting this biodiesel supply chain system; and whether biodiesel derived from kitchen waste can be part of sustainable energy solutions that are economically viable and environmentally acceptable.

4. Model Construction

4.1. Two-Stage Stochastic Programming Model. The objective of WCO for biodiesel supply chain stochastic programming model under supply disturbance is to minimize the expected total system cost and carbon emission cost by satisfying some constraints. This paper assumes that the kitchen waste supply by restaurants depends on two factors: (a) the proportion of kitchen waste supplied to biodiesel refineries and (b) the total supply of kitchen waste. Biodiesel refineries offer a group of prices per unit kitchen waste of restaurants and promise to buy the kitchen waste they provide. Under this commitment, the biodiesel refiners buy the kitchen waste at that price and then the restaurants supply a percentage of kitchen waste. As the purchase price increases, the percentage may increase. After the kitchen waste is collected, it is treated by a pretreatment facility established by the biodiesel refinery and transported to the biodiesel refinery.

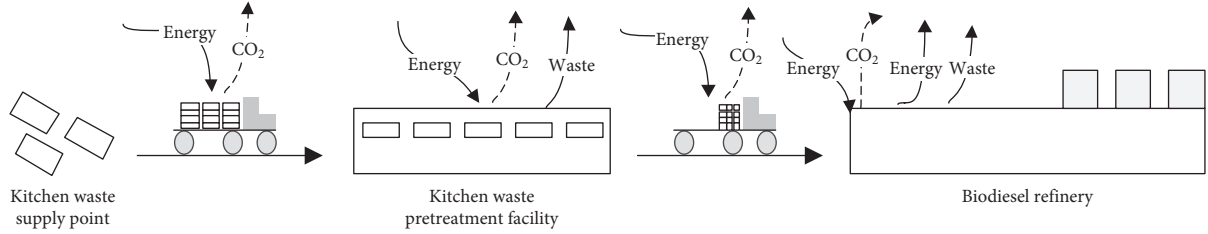


FIGURE 1: Waste cooking oil for biodiesel supply chain.

Pretreatment facilities pretreated kitchen waste from different kitchens and transported it to biodiesel refineries. Other assumptions include the following: (a) location of the pretreatment facility is determined by the biodiesel operator and selected from a set of known candidate sites; (b) biodiesel refineries need to process and convert the pretreated kitchen waste into biodiesel; (c) location of biodiesel refineries and demand for kitchen waste are deterministic; and (d) if the kitchen waste from the restaurants cannot meet the needs of the biodiesel refinery, the biodiesel operator needs to obtain it from an external supplier and there will be penalty charges for purchasing the unit's external kitchen waste. On the other hand, if the biodiesel operator has a surplus after meeting all the demand, it can sell the kitchen waste to make up for the loss of the excess raw materials.

In this paper, a scenario-based two-stage stochastic programming model is constructed. In the first stage, the location of the pretreatment facility and purchase price of the kitchen waste provided to the restaurants are determined. Then, based on the purchase price offered by the biodiesel operator, each restaurant decides how much kitchen waste allocates to biodiesel refineries for production. The total amount of kitchen waste was provided by each kitchen, which is influenced by random events such as the size of a city's population, consumption habits, and policy guidance. Supply uncertainty is addressed through a limited set of scenarios. When one scenario occurs, the total supply of the kitchen becomes known. In the second stage, the biodiesel refinery makes some decisions based on the calculated supply when the supply ratio and total supply are known. The decisions include the amount of kitchen waste transported from the restaurants to the pretreatment facility, the amount of kitchen waste transported from the pretreatment facility to the biodiesel refinery, the amount of kitchen waste purchased externally by each biodiesel

operator, and the amount of excess kitchen waste to be sold. Under the standard two-stage stochastic programming model, the first-stage decision must be made before the actual system uncertainty is realized. After random events occur, decisions are made in the second stage. The goal of a typical two-stage programming model is to make decisions by minimizing the cost of the first stage and randomizing the expected cost of the second stage. A two-stage stochastic programming based on scenario divided uncertain spaces into countable scenarios, and each scenario is provided with a corresponding probability. Although scenario-based two-stage stochastic programming sacrifices the solution's optimality, it ensures that the second-stage decision is made based on the occurrence of uncertainty. It is a good approximation of the total solution. Table 1 lists the symbols used in the model construction.

The two-stage stochastic programming model is expressed as follows:

$$F_1 = \text{Min} \sum_{j \in J} f_j \times Y_j + E[\psi(Y_j, \pi, \Phi_i^s)], \quad (1)$$

$$F_2 = \text{Min} \sum_{j \in J} eh_j \times Y_j + E[\chi(Y_j, \pi, \Phi_i^s)], \quad (2)$$

$$\sum_{j \in J} M^c \times Y_j \geq \sum_{k \in K} \frac{D_k}{\mu}, \quad (3)$$

$$h \leq \pi \leq c, \quad (4)$$

$$Y_j \in \{0, 1\}, \quad \forall j \in J. \quad (5)$$

Specific to a given specific scenario:

$$\begin{aligned} \psi(s) = \psi(Y_j, \pi, \Phi_i^s) = & \text{Min} \sum_{i \in I} \sum_{j \in J} c^{f,d} \times q_{ij}^s \times d_{ij} + \sum_{j \in J} \sum_{k \in K} c^{w,d} \times q_{jk}^s \times d_{jk} \\ & + \sum_{j \in J} \sum_{k \in K} \text{CPR} \times q_{jk}^s + \sum_{s \in S} \sum_{i \in I} \Phi_i^s \times \pi \times \Gamma_i(\pi) + \sum_{k \in K} c \times Q_j^s - \sum_{i \in I} h \times W_i^s, \end{aligned} \quad (6)$$

$$\begin{aligned} \chi(s) = \chi(Y_j, \pi, \Phi_i^s) = & \text{Min} \sum_{i \in I} \sum_{j \in J} \text{EHV} \times q_{ij}^s + \sum_{i \in I} \sum_{j \in J} \text{ECT} \times q_{ij}^s \times d_{ij} \\ & + \sum_{j \in J} \sum_{k \in K} \text{ECT} \times X_{jk}^s \times d_{jk} + \sum_{j \in J} \sum_{k \in K} \text{EBD} \times q_{jk}^s, \end{aligned} \quad (7)$$

TABLE 1: Description of symbols used in the base model.

Set	
I	Set of kitchen waste supply points i
J	Set of potential pretreatment facility locations j
K	Set of biodiesel refinery locations k
S	Set of uncertain scenarios s
Parameter	
$\Gamma_i(\pi)$	Function of the kitchen waste amount of supply point i provided for the biodiesel operator according to the price π
d_{ij}	Distance from supply point i to pretreatment facility j
d_{jk}	Distance from the pretreatment facility j to the biodiesel refinery k
Φ_i^s	Total amount of annual kitchen waste supply in each supply point i under scenario s
f_j	Fixed construction cost of the pretreatment facility j
P^s	Probability of scenario s
C	Penalty fees for shortage demand of biodiesel
H	Residual value of excess kitchen waste
D_k	Demand for biodiesel at the biodiesel refinery k
c^{fd}	Transport costs for kitchen waste by truck
c^{wd}	Transport costs for waste cooking oil by truck
EHV	Carbon emissions from unit kitchen waste collection
ECT	Carbon emissions from transportation by truck
M^c	Maximum processing capacity of the pretreatment facility
eh_j	Carbon emissions from the construction of pretreatment facility j
EBD	Carbon emissions from pretreatment facilities deal with unit kitchen waste
CPR	Processing cost from unit kitchen waste at the pretreatment facility
μ	Pretreatment rate of unit kitchen waste at pretreatment facility
α	The weight
CEP	Trading price from unit carbon emissions
Decision variables	
q_{ij}^s	Quantity of kitchen waste from supply point i to pretreatment facility j in scenario s
q_{jk}^s	Quantity of waste cooking oil from pretreatment facility j to biodiesel refinery k in scenario s
Y_j	1 when location j is used to build a pretreatment facility, 0 otherwise
Q_j^s	Quantity of outsourcing kitchen waste at the pretreatment facility j in the situation s
W_i^s	Quantity of excess kitchen waste at supply point i in scenario s
π	Negotiated purchase price for unit kitchen waste

$$\begin{aligned}
F = \text{Min } \alpha & \left(\sum_{i \in I} \sum_{j \in J} c^{fd} \times q_{ij}^s \times d_{ij} + \sum_{j \in J} \sum_{k \in K} c^{wd} \times q_{jk}^s \times d_{jk} + \sum_{j \in J} \sum_{k \in K} \text{CPR} \times q_{jk}^s \right. \\
& \left. + \sum_{s \in S} \sum_{i \in I} \Phi_i^s \times \pi \times \Gamma_i(\pi) + \sum_{j \in J} c \times Q_j^s - \sum_{i \in I} h \times W_i^s \right) + (1 - \alpha) \left(\sum_{i \in I} \sum_{j \in J} \text{EHV} \times q_{ij}^s \right. \\
& \left. + \sum_{i \in I} \sum_{j \in J} \text{ECT} \times q_{ij}^s \times d_{ij} + \sum_{j \in J} \sum_{k \in K} \text{ECT} \times X_{jk}^s \times d_{jk} + \sum_{j \in J} \sum_{k \in K} \text{EBD} \times q_{jk}^s \right) \times \text{CEP}.
\end{aligned} \tag{8}$$

The objective functions F_1 and F_2 minimize the expected total cost and carbon emissions of the system, respectively. The objective function F_1 includes the cost of the first stage and the expected cost of the second stage. Operators E_ψ represent mathematical expectations about random parameters $\psi(s)$. In objective function (1), the first-stage cost package is the fixed cost of the preprocessing facility construction, which is constant with the implementation of random parameters. The cost of the second stage depends on the scenario that is in formula (6): the transport cost of the biodiesel kitchen waste, the cost of the kitchen waste pretreatment facility, the purchase cost of the raw material kitchen waste, and the possible fuel shortage loss cost. Objective function (2) also includes the first stage cost and

the second stage expected cost. Operators E_χ represent mathematical expectations about random parameters $\chi(s)$. The first stage of carbon emissions includes fixed carbon emissions from the construction of the pretreatment facility, the value of which is constant with the implementation of the random parameters. The second stage of carbon emissions is scenario-dependent in formula (7), including kitchen waste collection carbon emission, pretreatment carbon emission, and transport carbon emission. It is assumed that the cost function of raw material purchase has a piecewise linear relationship with the purchase quantity. For the calculation of carbon emissions, this paper mainly refers to the relevant carbon emission factor method. These factors are mainly calculated by quoting scholars' literature and

some extrapolated data. Carbon emissions used in different processes are linear functions. The other constraint set is as follows:

$$\sum_{i \in I} q_{ij}^s \leq M^c \times Y_j, \quad \forall j \in J, \quad (9)$$

$$\mu \times \left(\sum_{i \in I} q_{ij}^s + Q_j^s \right) = \sum_{k \in K} q_{jk}^s, \quad \forall j \in J, \quad (10)$$

$$\sum_{j \in J} q_{ij}^s + W_i^s = \Gamma_i(\pi) \times \Phi_i^s, \quad \forall i \in I, \quad (11)$$

$$\sum_{j \in J} q_{jk}^s \geq D_k, \quad \forall k \in K, \quad (12)$$

$$q_{ij}^s, q_{jk}^s, W_i^s, Q_k^s \geq 0. \quad (13)$$

Constraint (3) means that the total capacity of the pretreatment facility must be greater than the total demand of the system because all kitchen waste need transport passes through the pretreatment facility. Constraint (4) defines the lower and upper limits of unit kitchen waste purchase price. The purchase price of kitchen waste must be greater than the residual value and must be less than the penalty cost. If the kitchen waste purchase price is lower than the residual value, the biodiesel operator only needs to make a profit by selling the kitchen waste. Conversely, if the purchase price is higher than the unit penalty cost, the biodiesel operator can meet all demand by purchasing from an external supplier. Constraint (5) is a binary variable. Objective function (6) represents the economic cost throughout the whole supply chain. Objective function (7) represents carbon emission throughout the whole supply chain. Objective function (8) is the total cost after taking unit carbon emissions trading price into account. Constraint (9) is the capacity constraints of each pretreatment facility. Constraint (10) is the balance between the inflow and outflow of the pretreatment facility. Constraint (11) is the supply constraint that ensures that all supply is transported to biodiesel refineries. Constraint (12) biodiesel refineries meet a demand greater or equal to the total biodiesel demand. Since the objective function and the constraint condition both contain $\Phi_i^s \times \Gamma_i(\pi)$, the problems ψ and χ change with different implementations of the random variable Φ_i^s . For each first-stage decision, the problems ψ and χ are feasible for all scenarios. Constraint (4) ensures that the capacity of kitchen waste transported through the pretreatment facility is sufficient so that all the pretreatment facility locations are feasible in the first stage. Biodiesel refineries can be outsourced to meet demand, so that the demand of each biodiesel refinery can always be outsourced regardless of the kitchen waste acquisition price and supply situation. Therefore, the properties of the mathematical model are largely dependent on $\Gamma_i(\pi)$.

4.2. Stochastic Programming Model Reconstruction

4.2.1. Allocation Decision Model for Restaurants. In order to determine the function Γ introduced in the mathematical formula, it is necessary to take the restaurant's decision into consideration. A model which determines the relationship between the price and the supply proportion of kitchen waste is needed. In this paper, we followed the methods of Uster and Memişoğlu [35] and Memişoğlu [36] to determine the relationship between the price and the supply proportion of kitchen waste. To form a hypothesis, restaurant makes a supply decision totally based on its expected profit. As a consequence, the kitchen uses the biodiesel operator if the expected profit is higher than the kitchen's current expected profit which is supplied to at least one other vendor. Suppose that restaurant i provides kitchen waste R_i to vendor r . The following symbols to construct the allocation decision model of the kitchen are shown in Table 2.

For each restaurant and vendor, the following equation must be followed:

$$\text{Sup}_{ir} \times \text{Pep}_r - \text{Ckp}_{ir} = (\text{Sup}_{ir} \times \text{Bkp}_{ir}) - \text{Ckp}_{ir}. \quad (14)$$

The left and right sides of equation (14), respectively, represent the expected unit profit (Yuan/year) that the restaurant provides to all vendors r and the biodiesel operators. A kitchen will supply kitchen waste to a biodiesel operator only if its expected profit is greater than its expected profit from supplying at least one of the other vendors. For kitchen i , there is $|R_i|$ vendor demands, and therefore, there is R_i balance price between the biodiesel operators, each represented by Bkp_{ir} , where r belongs to R_i . These prices can be determined by resolving (14) Bkp_{ir} , which is a supply-demand balance price that can persuade restaurants to supply kitchen waste to biodiesel operators rather than to other vendors.

4.2.2. Reconstruction of Stochastic Programming Model of WCO for Biodiesel Supply Chain under Supply Disturbance.

The optimal purchase price is one of the equilibrium prices of the restaurant's supply and demand. Thus, the set of continuous price values can be reduced to a set of discrete price points since each restaurant has a finite supply and demand equilibrium price. Let us define the set ξ , which represents all the supply and demand equilibrium prices $\xi = \cup_{i,r} \rho_{ir}$, $\forall i \in I, r \in R_i$ for all the kitchens, and a new set $F \subseteq \xi$, representing all the possible different prices. P_f is assumed to be a binary decision variable, $f \in F$; if the price f is selected, then $P_f = 1$; otherwise, it is 0 and π_f is the relevant price value. ω_{if} means that when the price f is provided, the proportion of allocation ω_{if} to the biodiesel operator determined by a kitchen i is equal to $\Gamma_i(\pi_f)$ which can be calculated relatively easily. For each scenario, the cost of the biodiesel operator to purchase kitchen waste is a function π_f .

TABLE 2: Description of the symbols used in the allocation decision model.

Symbol	Description
Sup_{ir}	Quantity that i expects to provide to vendor r for kitchen waste purchase (tons/year)
Pep_r	Expected purchase price given by vendor r (Yuan/ton)
Ckp_{ir}	Fees of restaurants i providing kitchen waste to vendor r (Yuan/year)
Sup_i	Quantity expected to be supplied by the restaurant i to the biodiesel operator (tons/year)
Bkp_{ir}	The equilibrium price between supply and demand of restaurant i and vendor r (Yuan/ton)
Ckp_i	Fees for the provision of kitchen waste by restaurant i to the biodiesel operator (Yuan/year)

The probability p^s of each scenario is assumed to be known. Therefore, the purchase cost item can be extracted from the problems ψ and χ and can be placed separately in the

objective functions (1) and (2). According to these modifications, the model proposed in Section 4.1 is reconstructed as follows:

$$F_1 = \text{Min} \sum_{j \in J} f_j \times Y_j + \sum_{i \in I} \sum_{s \in S} \sum_{f \in F} \pi_f \times p^s \times P_f \times w_{if} \times \Phi_i^s + E[\Delta(Y_j, P_f, \Phi_i^s)], \quad (15)$$

$$F_2 = \text{Min} \sum_{j \in J} eh_j \times Y_j + \sum_{i \in I} \sum_{s \in S} \sum_{f \in F} \text{EHV} \times p^s \times P_f \times w_{if} \times \Phi_i^s + E[\lambda(Y_j, P_f, \Phi_i^s)]. \quad (16)$$

Subject to (3), (5) and the following:

$$\sum_{f \in F} P_f = 1, \quad (17)$$

$$P_f \in \{0, 1\}, \quad \forall f \in F. \quad (18)$$

For a given particular implementation, scenario s ,

$$\begin{aligned} \Delta(Y_j, P_f, \Phi_i^s) &= \text{Min} \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} (c^{f,d} \times q_{ij}^s \times d_{ij} + c^{w,d} \times q_{jk}^s \times d_{jk}) + \sum_{j \in J} \sum_{k \in K} \text{CPR} \times q_{jk}^s \\ &\quad + \sum_{j \in J} c \times Q_j^s - \sum_{i \in I} h \times W_i^s, \\ \lambda(Y_j, P_f, \Phi_i^s) &= \text{Min} \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} \text{ECT} \times (q_{ij}^s \times d_{ij} + q_{jk}^s \times d_{jk}) + \sum_{j \in J} \sum_{k \in K} \text{EBD} \times q_{jk}^s, \\ F &= \text{Min} \alpha \times \left(\sum_{j \in J} f_j \times Y_j + \sum_{i \in I} \sum_{s \in S} \sum_{f \in F} \pi_f \times p^s \times P_f \times w_{if} \times \Phi_i^s + \sum_{i \in I} \sum_{j \in J} c^{f,d} \times q_{ij}^s \times d_{ij} \right. \\ &\quad \left. + \sum_{j \in J} \sum_{k \in K} c^{w,d} \times q_{jk}^s \times d_{jk} + \sum_{j \in J} \sum_{k \in K} \text{CPR} \times q_{jk}^s + \sum_{j \in J} c \times Q_j^s - \sum_{i \in I} h \times W_i^s \right) + (1 - \alpha) \\ &\quad \times \left(\sum_{j \in J} eh_j \times Y_j + \sum_{i \in I} \sum_{s \in S} \sum_{f \in F} \text{EHV} \times p^s \times P_f \times w_{if} \times \Phi_i^s + \sum_{i \in I} \sum_{j \in J} \text{ECT} \times q_{ij}^s \times d_{ij} \right. \\ &\quad \left. + \sum_{j \in J} \sum_{k \in K} \text{ECT} \times q_{jk}^s \times d_{jk} + \sum_{j \in J} \sum_{k \in K} \text{EBD} \times q_{jk}^s + \sum_{j \in J} eh_j Y_j \right) \times \mu. \end{aligned} \quad (19)$$

Subject to (9), (10), (12), and the following:

$$\sum_{j \in J} X_{ij}^s + W_i^s = \sum_{f \in F} P_f \times \Phi_i^s \times w_{if}, \quad \forall i \in I. \quad (20)$$

In the reconstructing model, the first and second formulas of objective function (15) represent the total fixed cost and the expected acquisition cost, respectively. The first and second formulas in (16) represent carbon emissions during construction and collection, respectively. Constraint (4) no longer needs to be removed from the entire model, while constraints (17) and (18) are added to the new model, which ensure that only one of the supply and demand equilibrium prices is selected. At the same time, the overall problem with this new model becomes linear. Although the introduction of new binary variables w_{if} into this model increases the number of decision variables, it also reduces the pricing decision to a limited set of choices.

4.2.3. Solutions. Aiming at the economic and environmental objectives involved in the model, this paper introduces the environmental cost (carbon emission trading price), converts the environmental target into the economic cost target, and refers to the weight coefficient between the environmental target and the economic target value [37], so as to transform the multiobjective problem into a single-objective problem. Thus, the single-objective supply chain optimization model can be solved by MATLAB programming software. Combined with the solution idea of two-stage stochastic programming model, this paper introduces situational variables to transform the optimization model under certain conditions into stochastic programming model under uncertain conditions. The specific solution idea of the two stages is as shown in Figure 2 [36]:

Step 1: make a first-stage decision to determine the construction status of each pretreatment facility

Step 2: calculate the cost of the first stage

Step 3: at the beginning of the second phase, realize all the uncertain supplies

Step 4: at the end of the second stage, see the realization of uncertainty and the decision of the first stage, and make the second stage decision, namely, supply price and transportation volume

Step 5: calculate the scenario cost of the second stage

Step 6: calculate the expected total cost

In this article, according to different circumstances, supply points calculated the average supply of decision variables in the first phase of Y_j , whether to choose pre-processor facilities with value of 1, then Y_j value will not be affected by changes in supply, will also receive a preliminary objective function value, and corresponds to a suitable supply chain network structure. In order to get the optimal allocation decision, the decision of the second stage is needed. In this stage, a subproblem is generated according to each different supply situation. In each subproblem, the objective function is no longer a decision about the

construction of pretreatment facilities, but a decision about the distribution of food and kitchen waste, that is, it determines q_{ij}^s and q_{jk}^s . Based on the sequential decision of location and distribution problems based on two-stage stochastic programming, the optimal design scheme of food waste supply chain network can be obtained.

5. Case Analysis

5.1. Disturbance Factors. The disturbance factor in the supply chain is the total supply of kitchen waste, and the specific calculation process is as follows: first, the supply quantity is estimated under the basic scenario. The case analysis in this section adopts the real data of the Yangtze River Delta region. It is assumed that the kitchens on the supply point of the kitchen waste in each city are clustered at the central position of the cities in the Yangtze River Delta. Then the geographic location and coordinates of the city centre are obtained through GIS as the location of the restaurants in the supply point. The calculation of the total supply of kitchen waste is estimated by the following equation [38]:

$$M_i = \kappa_i \times \bar{M}, \quad (21)$$

where κ_i is the population of city i and \bar{M} represents the annual production of kitchen waste per capita in China. In this paper, the statistical value of 2019, namely, 0.18 kg/day person, is used to get the total quantity of kitchen waste supply in each city. However, within the scope of China, although Shanghai, Nanjing, Suzhou, and other cities have carried out the first practice of garbage recycling and achieved preliminary results, most of the recycling objects are only for the kitchen waste produced by catering enterprises and large canning rooms, while the recycling cost of household waste is relatively high. Some studies have shown that kitchen waste from enterprises and large dining room takes only about 25% of the total current kitchen waste output; therefore, the proportion of restaurant waste that can be recycled for the firms in current Yangtze normal situation all is set to 25%. We concluded that the number of kitchen waste supply is used as a normal situation in this case in Table 3.

5.1.1. Determination of Kitchen Waste Supply Scenario. It can be seen from the data statistics in the previous section that people's increasing consumption in catering has brought about a rapid growth of kitchen waste. The government has introduced a series of measures to reduce the amount of kitchen waste. The Yangtze River Delta is also trying to introduce some policies to reduce the generation of kitchen waste, for example, "Empty Plate Campaign" and so on to encourage the moderate consumption. Considering the reduction policy on the influence of the amount of kitchen waste in this case study, we divided 10 classes according to 0.5% of the total recursive difference, respectively. Ten kinds of scenario have the same probability. Supply and geographical distribution of kitchen waste under standard circumstances are shown in Figure 3.

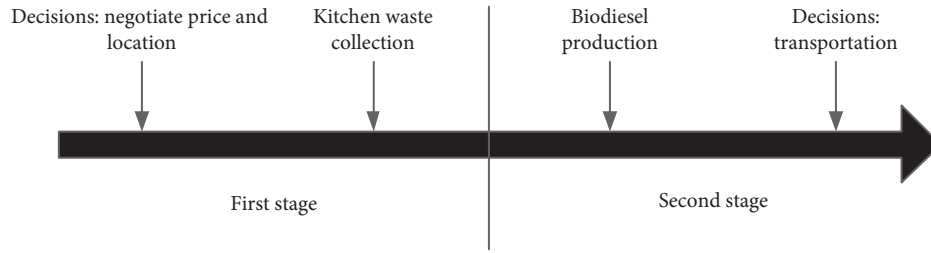


FIGURE 2: Solution idea of the two-stage stochastic programming model of WCO for biodiesel supply chain.

TABLE 3: Total supply of recyclable kitchen waste in the Yangtze River Delta.

Supply point	Quantity of food waste (tons)
Shanghai	396657
Hefei	127960.1
Ma'anshan	37155.23
Hangzhou	117533.6
Ningbo	95889.95
Jiaxing	57175.03
Huzhou	43329.95
Shaoxing	72772.1
Zhoushan	16003.03
Wenzhou	133643.4
Jinhua	78028.1
Quzhou	41983.9
Taizhou (Zhejiang Province)	98069.28
Lishui	43629.08
Nanjing	134946.7
Wuxi	106765.2
Xuzhou	141719.7
Changzhou	77130.73
Suzhou	174174.5
Nantong	119862.5
Lianyungang	73113.95
Huai'an	79694.63
Yancheng	118644.6
Yangzhou	73541.28
Zhenjiang	52089.95
Taizhou (Jiangsu province)	76190.63
Suqian	79545.08

5.2. Deterministic Factors

5.2.1. Proportion and Quantity Supplied to Biodiesel Refineries. In order to determine the proportion and quantity of kitchen waste supplied to biodiesel refineries, three directions of kitchen waste in Jiangsu Province in recent years should be determined first. According to the investigation, they are, respectively, gutter oil production, animal husbandry, and organic fertilizer production, and the vendors are represented by r . According to the survey data, the flow proportion of total kitchen waste in the Yangtze River Delta was determined, and the supply proportion of kitchen waste in each city was calculated. Then, according to the total 2018 kitchen waste in Yangtze River Delta, calculate the supply to the above three parties Sup_{ir} , in order to determine the kitchen waste purchase price of each demand point r , and this paper takes questionnaire investigation and gets the purchase price of all the demand point in 2016, 2017,

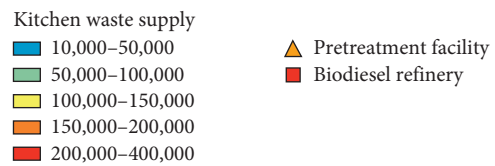
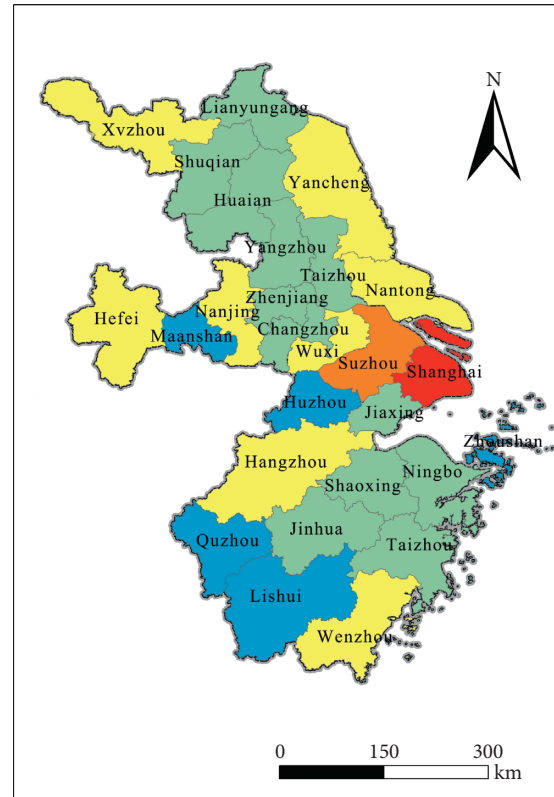


FIGURE 3: Supply of kitchen waste in the Yangtze River Delta under standard scenario.

and 2018. Then we determine the average value to the purchase price. For restaurants supply costs Ckp_{ir} , the survey found when a restaurant supplies kitchen waste for cooking oil production and organic fertilizer production manufacturers, and kitchens need to do a series of garbage classification which is simple to handle, so the cost is relatively a bit higher which is about 50 Yuan/ton. While provided to livestock feed manufacturers, restaurants processing fee is very low, only artificial collection and handling are needed, so the cost is relatively low. Here we assumed that the kitchen waste supply fee Ckp_{ir} is roughly the same as that of supply fee for gutter oil production, which is

estimated to be 50 Yuan/ton. Then, the above parameters were substituted into equation (14) to find the supply-demand balance price Bkp_{ir} of kitchen waste supplied to biodiesel operators. These figures were corresponding to the break-even price (π_f) in the model proposed in Section 4.

5.2.2. Kitchen Waste Demand Point and Candidate Facility Locations. Figure 4 shows the locations of candidate pretreatment facilities. After preliminary investigation, 27 prefecture-level cities can be listed as candidate cities for the construction of pretreatment facilities in the Yangtze River Delta. The geographic centre of each city is selected as the location coordinates of candidate pretreatment facilities and used in the case study.

The optimization of WCO for biodiesel supply chain aims at the process of biodiesel operators purchasing kitchen waste, so the location of the demand point here is the location of the demand point of kitchen waste, that is, the location of biodiesel refinery. Biodiesel production statistics website provides the list of the biological diesel oil refinery, and this paper chooses five large- and medium-sized biodiesel refining company according to the preliminary research as shown in Figure 4. They are Shanghai Jinshan Biological Diesel Co., Ltd., Hangzhou Xiaoshan Yuanhua Energy Technology Co., Ltd., Changzhou Yueda Carter New Energy Co., Ltd., and Jiangsu Clean Environment Co., Ltd. In order to determine the total demand of kitchen waste, this paper uses the company's annual report to estimate the demand for biodiesel of each company. According to the current biodiesel processing rate, calculations of total demand for kitchen waste are about 530,947 tons.

5.2.3. Other Parameters. There are other parameters in this paper, as shown in Table 4.

5.3. Result Analysis. In order to verify the model, the paper uses the actual data of the Yangtze River Delta region to conduct a case study. We first analyze the basic setting results of the standard parameter value, then change some parameter values, and analyze their effects on supply chain network design, total cost, and negotiated purchase price.

5.3.1. Analysis of Facility Location under Stochastic Programming. The above basic setting parameters are used here to obtain the optimal pretreatment facility location, as shown in Figure 5. Eight pretreatment facilities have been opened around the Yangtze River Delta. Four pretreatment facilities in the northwest and the middle of the Yangtze River Delta serve biodiesel refineries in Jiangsu and Anhui provinces. Another four pretreatment facilities opened in the south and east parts of the Yangtze River Delta serve biodiesel refineries in Zhejiang Province and Shanghai City.

The optimization results of the model show that the estimated total system cost is about 303.4096 million Yuan. Among them, more than half of the total system cost (about 46.8%) is the fixed cost for the construction of pretreatment facilities and the purchase cost for buying kitchen waste.

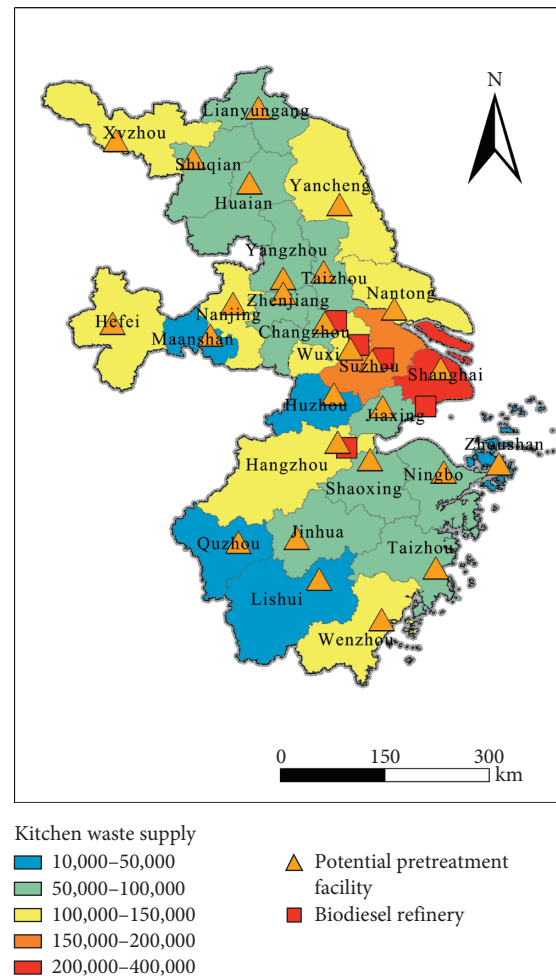


FIGURE 4: Location of alternative points of pretreatment facilities.

Most of the remaining expected cost is for transportation. The optimization results show that it is more economical to transport the kitchen waste from the middle part of the Yangtze River Delta where the supply is relatively high, to the middle part of the Yangtze River Delta where most of the demand occurs. It is better than to offer higher prices for the restaurant in the north and south of the Yangtze River Delta to provide a greater proportion of the kitchen waste to the biodiesel refineries. There is a trade-off between the purchase price of kitchen waste and the logistics cost in the system. The results of this section show that it is more economical to set a low purchase price for restaurants in the central part of the Yangtze River Delta to increase the proportion of kitchen waste supply than to transport kitchen waste supply from the north and south of the Yangtze River Delta to meet the demand.

5.3.2. Analysis of the Negotiated Purchase Price under the Base Scenario. Using the model proposed in Section 4, the optimal supply price is 348.2 Yuan/ton, and the supply ratio is shown in Table 5. At present, the purchase price of kitchen waste in the market is about 500 Yuan/ton. Therefore, biodiesel refineries under contract can not only guarantee

TABLE 4: Other parameters.

Other parameters	Parameter value
The conversion factor of kitchen waste to waste cooking oil	7%
Preprocessing cost for unit kitchen waste	15 Yuan/ton [39]
Penalty fees for shortage demand of biodiesel	700 Yuan/ton [40]
Disposal price of excess kitchen waste	60 Yuan/ton [41]
The distance transportation cost of unit kitchen waste	0.20 Yuan/ton/km [27]
The distance transportation cost per unit of waste oil	0.25 Yuan/ton/km [42]
The carbon emissions of unit kitchen waste collection	5.6 kg CO ₂ eqv./ton [43]
The carbon emission of unit kitchen waste pretreatment	12.6 kg CO ₂ eqv./ton [44]
Carbon emissions from transportation	0.1215 kg CO ₂ eqv./ton [43]

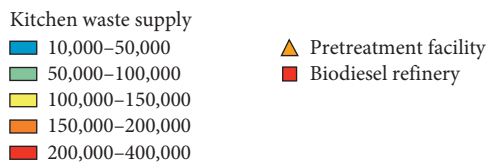
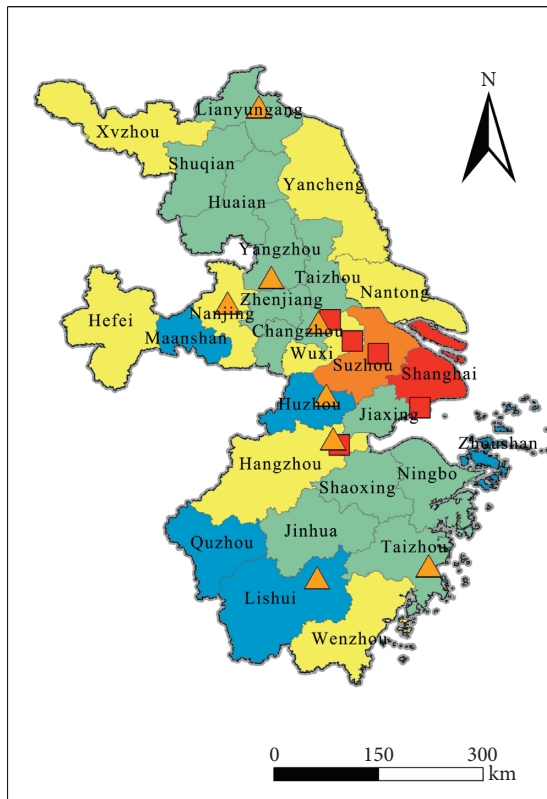


FIGURE 5: Optimization results of biodiesel supply chain in the collection stage under basic setting scenario.

the supply of kitchen waste, but also reduce the purchasing cost. At the same time, according to the calculation results, restaurants in Shanghai have the largest supply ratio according to the optimal price, with a supply ratio of 21.3%. This ratio obviously does not meet the national goal of vigorously developing the kitchen waste recycling for biodiesel production. At the optimal price offered to restaurants (348.2 Yuan/ton), the Yangtze River Delta provided a total of 515,637.8 tons of kitchen waste, which was lower than the total demand of 530,947 tons. In fact, in order to have a total

TABLE 5: Proportion of kitchen waste supply points in the Yangtze River Delta.

Supply point	Supply ratio
Shanghai	0.212979563
Hefei	0.204015776
Maanshan	0.19505199
Hangzhou	0.204015776
Ningbo	0.19505199
Jiaxing	0.186088204
Huzhou	0.177124417
Shaoxing	0.19505199
Zhoushan	0.177124417
Wenzhou	0.19505199
Jinhua	0.186088204
Quzhou	0.177124417
Taizhou (Zhejiang Province)	0.186088204
Lishui	0.177124417
Nanjing	0.204015776
Wuxi	0.19505199
Xuzhou (Jiangsu Province)	0.186088204
Changzhou	0.19505199
Suzhou	0.19505199
Nantong	0.186088204
Lianyungang	0.177124417
Huaian	0.177124417
Yancheng	0.177124417
Yangzhou	0.186088204
Zhenjiang	0.186088204
Taizhou	0.186088204
Suqian	0.177124417

expected supply of about 530,947 tons of kitchen waste, the price should be set at 350 Yuan/ton.

5.3.3. *Sensitivity Analysis.* This section mainly analyses the difference in pricing decisions and supply chain network structure by changing the value of parameters. This is done by changing the value of one input parameter in the base scenario and keeping the others at their standard values, to generate four scenarios as shown in Table 6. In the first scenario (S_1), the unit kitchen waste transportation cost is changed. In the second scenario (S_2), the pretreatment rate of kitchen waste is changed. In the third scenario (S_3), the residual value price per unit of food waste is changed, and finally, in the fourth scenario (S_4), the unit penalty fee is changed when demand is insufficient [36].

TABLE 6: Setting of sensitivity analysis.

Scenario	S_1	S_2	S_3	S_4
Transportation cost (Yuan/ton/km)	0.14–0.26	0.2	0.2	0.2
Pretreatment rate of kitchen waste (%)	0.07	0.055–0.085	0.07	0.07
Kitchen waste residual value (Yuan/ton)	60	60	60	0–120
Penalty fee (Yuan/ton)	700	700	400–1,000	700

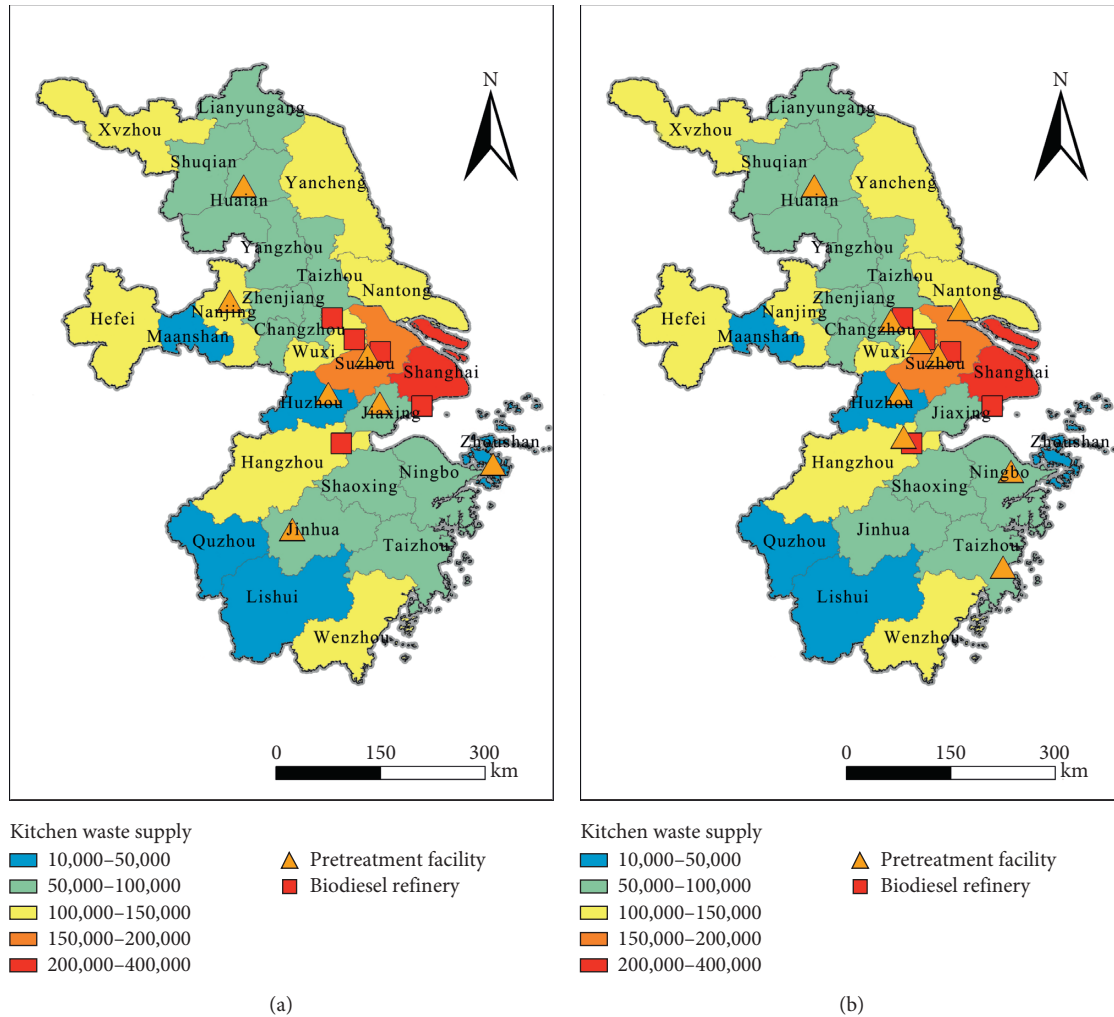


FIGURE 6: Network structure sensitivity analysis of penalty fee. (a) Penalty fee of 400 Yuan/ton. (b) Penalty fee of 1000 Yuan/ton.

(1) *Sensitivity Analysis of Transportation Costs S_1* . This scenario analyses different unit transportation cost systems ranging from 0.14 Yuan/ton-km to 0.26 Yuan/ton-km. As can be seen from Figure 6(a), for unit transportation cost ranging from 0.14 Yuan/ton-km to 0.24 Yuan/ton-km, the optimal purchase price is determined to be 348.2 Yuan/ton, which is the same as the basic scenario result. However, for a higher unit transport charge, that is, 0.26 Yuan/ton-km, the best purchase price offered to kitchens dropped to 325.5 Yuan/ton km. The main reason behind the change is that as unit transport costs increase, it becomes less economical for restaurants in distant locations to meet the biodiesel refineries demand. In other words, it becomes more economical for biodiesel refiners to choose to outsource rather

than pay high logistics costs to meet demand. As a result, the price of food waste will be lower and the expected supply will be reduced. Pretreatment facility decisions are also affected when unit transportation costs change.

Figures 7(a) and 7(b) show the network structure when the unit transportation cost is 0.14 Yuan/ton-km and 0.26 Yuan/ton-km.

For lower unit transportation cost, we can see that the supply chain system has fewer preprocessing facilities. The increase in the number of pretreatment facilities (from 7 to 9) is in line with the increase in unit transport costs. That means biodiesel refineries need to open pretreatment facilities in more dispersed locations to reduce the increase in transport costs. Although restaurants are less able to meet

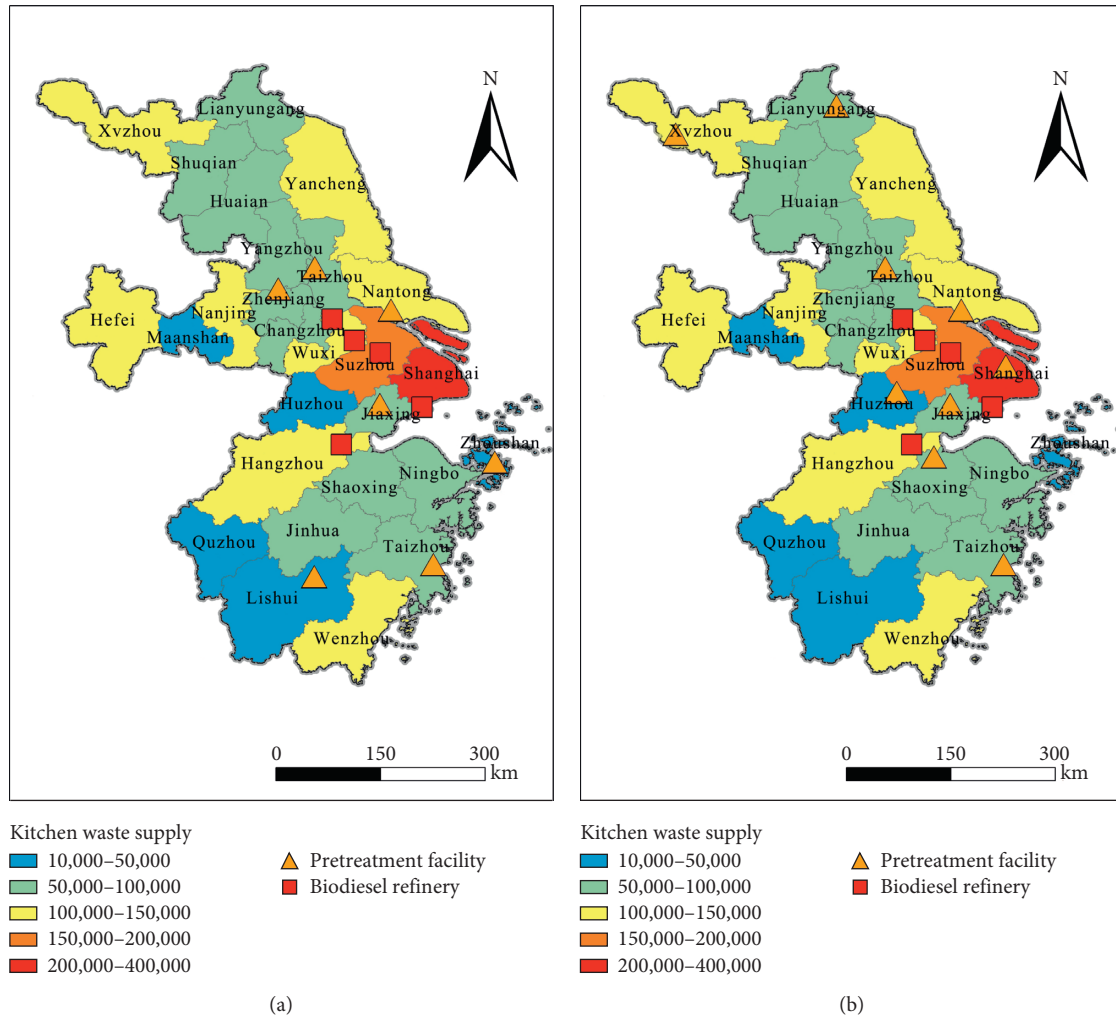


FIGURE 7: Network structure sensitivity analysis of transportation cost. (a) Transport cost of 0.14 Yuan/ton-km. (b) Transport cost of 0.26 Yuan/ton-km.

demand due to lower prices for kitchen waste, more pretreatment facilities have been opened to increase the supply of pretreated kitchen waste for transporting to biodiesel refineries. As shown in Figure 8(a), the estimated total system cost gradually increases as the unit transportation cost increases. When the unit transportation cost is from 0.14 Yuan/ton-km to 0.26 Yuan/ton-km, the expected total system cost will increase by about 7%. That indicates the logistics cost has some impact on the expected total system cost.

(2) *Sensitivity Analysis of Kitchen Waste Pretreatment Rate S_2 .* This scenario is mainly to analyse the sensitivity of kitchen waste pretreatment rate. Considering different kitchen waste pretreatment rates ranging from 5.5% to 8.5%, we observed the change in negotiated purchase price and network structure. The results show that, with the change in the pretreatment rate of kitchen waste, the purchase price provided to kitchens does not change basically (except for a small increase in 1 case), as shown in Figure 9(b). For all the kitchen waste pretreatment rates considered, the price remains the same as the basic scenario result, that is, 348.2

Yuan/ton. This indicates that the current range of kitchen waste pretreatment rate has no significant impact on the negotiated purchase price of kitchen waste. However, with the change of kitchen waste pretreatment rate, the structure of supply chain network has changed greatly. The paper found that, with the increase of kitchen waste pretreatment rate, pretreatment facilities opened decreased from 9 to 5, because increase in the rate of pretreatment compared with previous less eat hutch garbage can meet the demand, so biodiesel operators choose to reduce the number of facilities built pretreatment, in order to reduce building and operating costs. In addition, it can be observed from Figures 10(a) and 10(b) that after the pretreatment rate is improved, the location of pretreatment changes from the original area with small population and small food waste supply to the area with large supply and large population, which improves the overall supply chain efficiency. Figure 8(b) shows the relationship between the discount factor for economies of scale and the expected total system cost. Obviously, with the increase of the kitchen waste pretreatment rate, because the same amount of raw materials

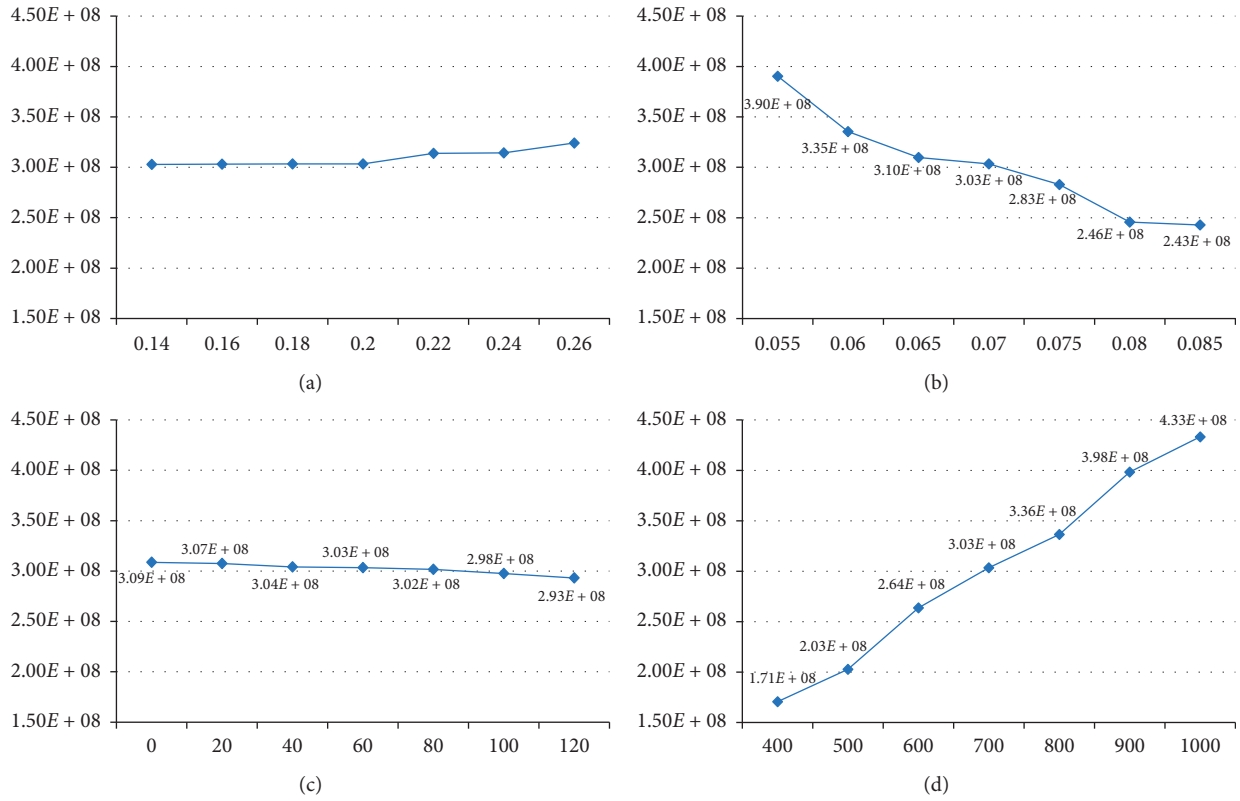


FIGURE 8: Influence of parameter values on the expected total system cost. (a) Transportation cost (Yuan/ton-km). (b) Pretreatment rate of kitchen waste. (c) Residual value of kitchen waste (Yuan/ton). (d) Penalty cost (Yuan/ton).

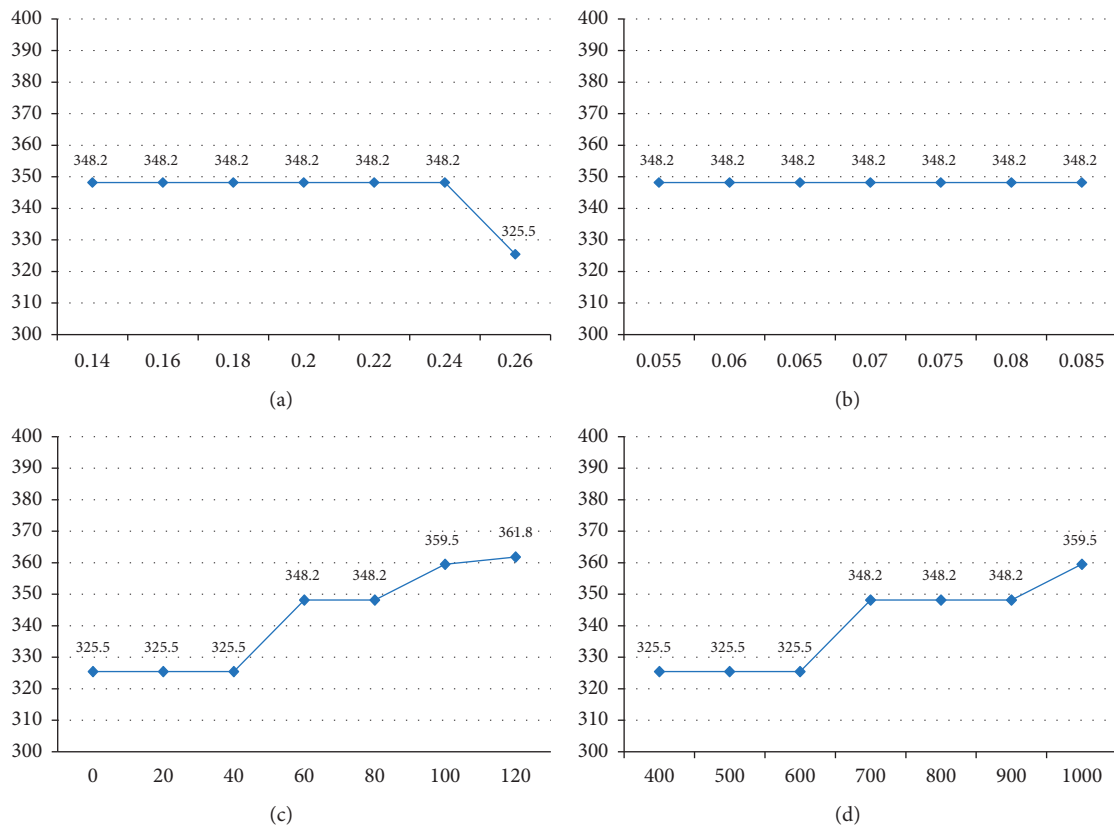


FIGURE 9: Influence of parameter values on price. (a) Transportation costs (Yuan/ton-km). (b) Pretreatment rate of kitchen waste. (c) Residual value of kitchen waste (Yuan/ton). (d) Penalty fee (Yuan/ton).

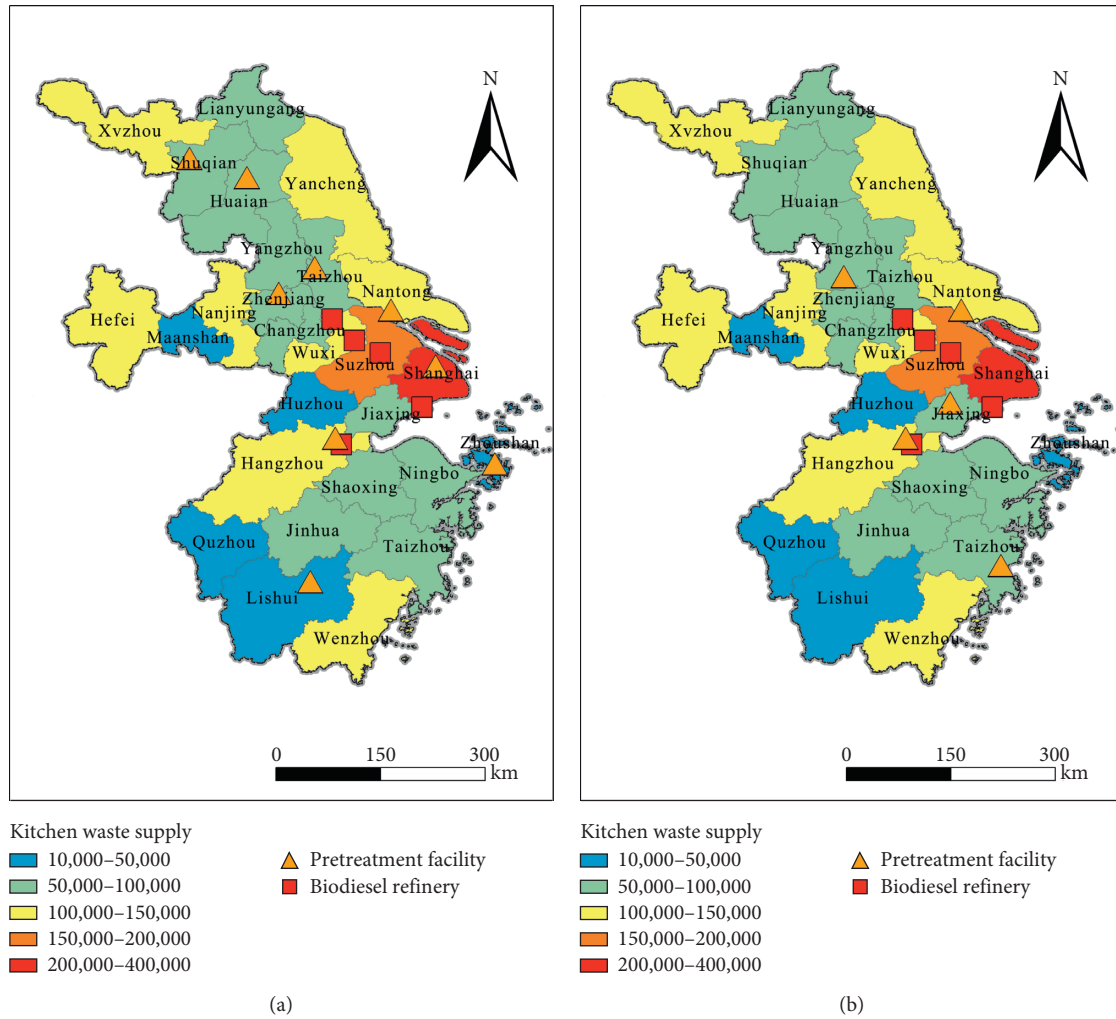


FIGURE 10: Network structure sensitivity analysis of kitchen waste pretreatment rate. (a) Pretreatment rate of 5.5%. (b) Transportation rate of 8.5%.

can produce more products, the expected total system cost is reduced, and the reduction is very significant. For example, when the increase in the kitchen waste pretreatment rate is doubled from its baseline pretreatment rate, the expected total system cost is reduced by nearly 25%. Therefore, it can be considered that the pretreatment rate of kitchen waste has a significant impact on the expected total system cost.

(3) *Sensitivity Analysis of Kitchen Waste Residual Value S_3 .* This scenario mainly analyses the influence of unit surplus kitchen waste residual value. Considering different residual values, ranging from 0 Yuan/ton to 120 Yuan/ton, the paper observes the impact on the first phase decision and the expected total system cost. When the residual value per unit is 0 Yuan/ton, the biodiesel refinery will not generate any revenue by selling excess kitchen waste. Figure 9(c) shows how the negotiated purchase price varies for different residual values. When the kitchen waste residual value is 0 Yuan/ton, 20 Yuan/ton, and 40 Yuan/ton, the negotiated purchase price is 325.5 Yuan/ton, which is lower than the price of basic settings. When the unit salvage value price is between 40 Yuan/ton and 80 Yuan/ton, the optimal

purchase price is 348.2 Yuan/ton. On the other hand, when the unit residual value price is 120 Yuan/ton, the purchase price of kitchen waste rises to 361.8 Yuan/ton. Therefore, the results show that as the unit residual value price increases, the negotiated price offered by biodiesel refineries to kitchens increases. The main reason for this result is that, with the rise in unit residual value prices, the income from the supply of excess kitchen waste increases. This incentivizes biodiesel operators to offer higher negotiated price to kitchens since the loss of excess kitchen waste can be compensated through residual values. Figures 11(a) and 11(b) show the optimal pretreatment facility decision when unit residual value is 0 Yuan/ton and 120 Yuan/ton, respectively. With the change of unit residual value, it was observed that the total number of pretreatment facility decisions (all are eight) and geographical distribution did not change, so it could be considered that unit residual value price had no significant influence on the decision of pretreatment facility. The relationship between unit residual value and the expected total system cost is shown in Figure 8(c). As the residual value goes up, the expected

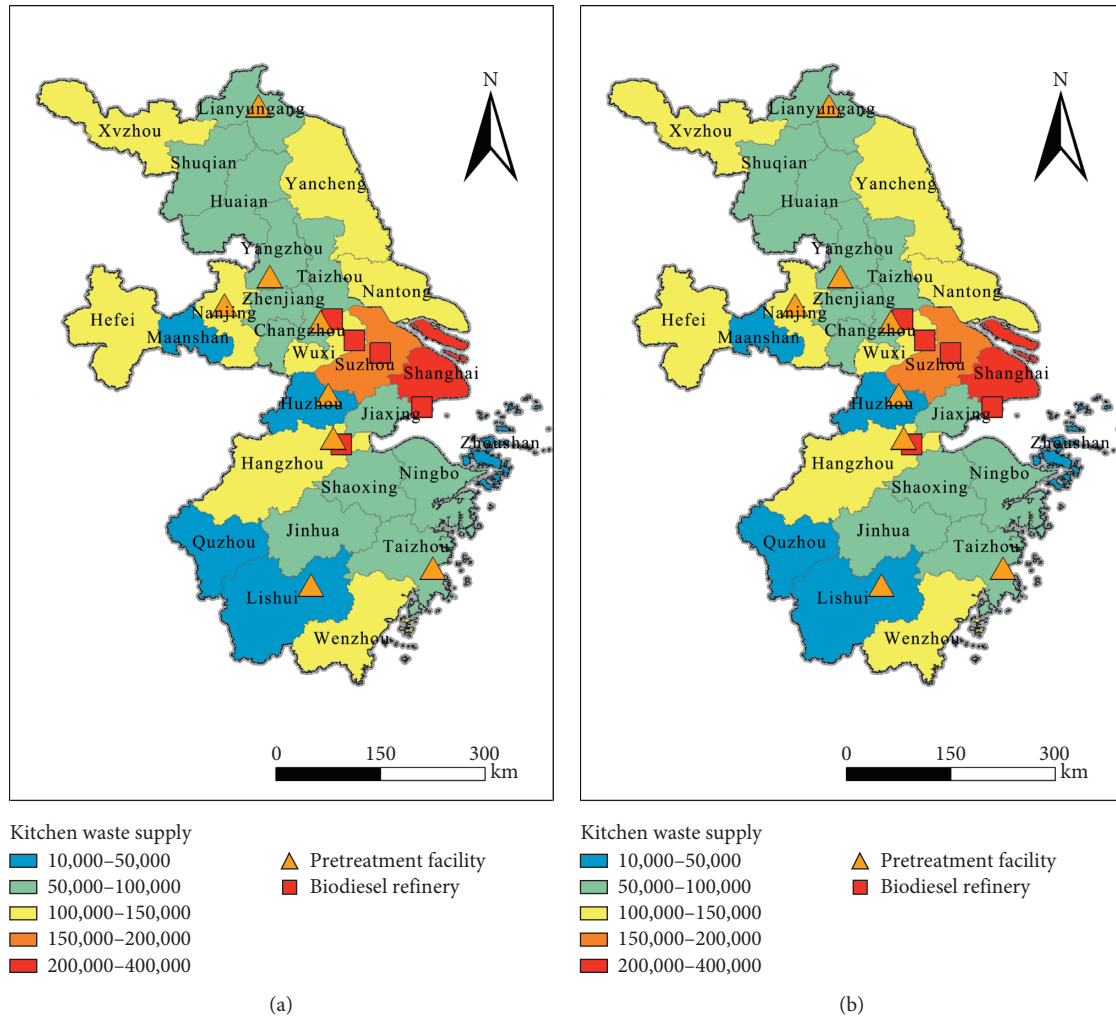


FIGURE 11: Network structure sensitivity analysis of kitchen waste residual value. (a) Residual value of 0 Yuan/ton. (b) Residual value of 120 Yuan/ton.

reduction in total system cost is not significant. For example, when the unit residual value price increases by a factor of two from its nominal value, the expected total system cost decreases by only 3.4%.

(4) *Sensitivity Analysis of Penalty Fee S_4* . This scenario takes into account variations in negotiated price, pretreatment facilities location, and expected total system costs at different unit penalty fee in the range from 400 Yuan/ton to 1000 Yuan/ton.

Figure 9(d) shows how the negotiated price varies with the changes in penalty fee. With the increase of unit penalty fee, the purchase price of kitchen waste increases sharply. When the unit penalty fee is 400 Yuan/ton, 500 Yuan/ton, and 600 Yuan/ton, the purchase price of kitchen waste is 325.2 Yuan/ton, which is lower than the optimal purchase price obtained in the basic scenario. However, when the unit penalty fee is 1000 Yuan/ton, the best purchase price offered to restaurants is increased to 359.5 Yuan/ton. This is because for a low penalty price, the logistics cost is greater than the penalty cost, and the biodiesel operator would rather choose other vendors to buy a certain amount of kitchen waste and

pay the penalty, rather than spend more money to meet the needs of individual restaurants. With the increase in unit penalty fees, biodiesel operators find it too costly to obtain biomass from other sources; therefore, they choose restaurants and offer a higher purchase price to capture more of the expected kitchen waste. There is a trade-off between the purchase price of kitchen waste and the punish fee.

Low purchase prices lead to low supply of kitchens, which in turn affects decisions at preprocessing facilities. Figure 6(a) and 6(b) show the location of the optimal pretreatment facility when the unit penalty cost is 400 Yuan/ton and 1000 Yuan/ton, respectively. When the unit penalty fee is 400 Yuan/ton, 7 pretreatment facilities need to be opened. With low expected supply and penalty costs, biodiesel operators can outsource their demand, and thus the demand for pretreatment facilities is reducing for. As the penalty fee increases, especially when the penalty fee increased to 1000 Yuan/ton, 9 pretreatment facilities need to be opened, and the orange area in the central part of the Yangtze River Delta pretreatment facilities increased by 2; they are Changzhou and Nantong. The two increased

pretreatment facilities in the region of the position are closer to area with high supply. The reason is the area with high supply kitchens open pretreatment facilities become more economic with the increase in the penalty fee. In this way, the refinery can get more kitchen waste supply and reduce the punishment cost. The relationship between unit penalty costs and expected total system costs is shown in Figure 8(d). The results show that as the unit penalty cost increases, the expected total system cost also increases in a very significant way. When the unit penalty cost increases by two times from its nominal value, the expected total system cost increases by 39%. At the same time, the curve can be understood as follows: When biodiesel operators are in insufficient demand, the punishment cost increases obviously. If the operators want to reduce total penalty cost, the out-of-stock rate needs to be decreased. So, the operators have to increase prices and purchase more raw materials on the basis of the original plant ensuring meeting production demand and reducing the punishment cost caused by shortage.

Figures 8 and 9, respectively, show the impact of these parameters on the price and expected total system cost.

6. Conclusion

In this paper, a stochastic programming model is proposed to optimize the WCO for biodiesel supply chain. The model proposed a system solution, which is a contract signed by the biodiesel operator and the restaurant to determine the purchase price and ensure the supply of kitchen waste. In order to incorporate this solution into the supply chain model, this paper presents an allocation decision model, which includes the relationship between the price of kitchen waste and the supply ratio. Then, a two-stage linear stochastic programming model was reconstructed by combining the restaurant's allocation decision model with the original supply chain model. At the same time, the influence of the kitchen waste supply uncertainty was considered, and the biodiesel supply chain network under disturbance was designed. This paper uses the actual data of the Yangtze River Delta region to conduct a case study to test the feasibility of the model. The results of the research in the basic scenario show that when the optimal purchase price provided by the biodiesel operator is set at 350 Yuan/ton, the restaurant can guarantee the amount of kitchen waste required by the biodiesel refinery, and the price is lower than the market price. By sensitivity analysis, the results show that the unit penalty fee and the residual value of kitchen waste have significant effects on the purchase price. The unit transportation cost will affect the network structure of supply chain. The variation of unit penalty fee and kitchen waste pretreatment rate has a significant impact on the network structure and the expected total cost of the system.

Data Availability

All data generated or analysed during this study are included within this article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Nanjing University of Posts and Telecommunications Research Start-Up Fund (NY219168), the Nature Foundation Incubation Fund of Nanjing University of Posts and Telecommunications (NY220214), the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (TJZ220011), and the Jiangsu Innovation Program for Graduate Education (KYLX15_0312).

References

- [1] S. Ishak and A. Kamari, "Biodiesel from black soldier fly larvae grown on restaurant kitchen waste," *Environmental Chemistry Letters*, vol. 17, no. 2, pp. 1143–1150, 2019.
- [2] S. Khan, M. Raza, A. Nosheen, R. Naz, S. M. U. Shah, and M. N. Hassan, "Quality comparison of biodiesel produced from waste cooking oil of restaurant and domestic kitchen," *International Journal of Green Energy*, vol. 17, no. 1, pp. 94–100, 2020.
- [3] N. Geng, Y. Zhang, and Y. Sun, "A coordinating strategy for biofuel supply chain under disturbance using revenue sharing contract approach," *PROMET-Traffic&Transportation*, vol. 30, no. 2, pp. 195–204, 2018.
- [4] L. V. Snyder, "Facility location under uncertainty: a review," *IIE Transactions*, vol. 38, no. 7, pp. 547–564, 2006.
- [5] S. M. Zahraee, N. Shiwakoti, P. Stasinopoulos, and S. Peter, "Biomass supply chain environmental and socio-economic analysis: 40-years comprehensive review of methods, decision issues, sustainability challenges, and the way forward," *Biomass and Bioenergy*, vol. 142, Article ID 105777, 2020.
- [6] S. Giarola, A. Zamboni, and F. Bezzo, "Environmentally conscious capacity planning and technology selection for bioethanol supply chains," *Renewable Energy*, vol. 43, pp. 61–72, 2012.
- [7] A. Osmani and J. Zhang, "Stochastic optimization of a multi-feedstock lignocellulosic-based bioethanol supply chain under multiple uncertainties," *Energy*, vol. 59, pp. 157–172, 2013.
- [8] B. Sharma, R. G. Ingalls, C. L. Jones, and A. Khanchi, "Biomass supply chain design and analysis: basis, overview, modeling, challenges, and future," *Renewable and Sustainable Energy Reviews*, vol. 24, no. 10, pp. 608–627, 2013.
- [9] M. Langholtz, E. Webb, B. L. Preston et al., "Climate risk management for the U.S. cellulosic biofuels supply chain," *Climate Risk Management*, vol. 3, pp. 96–115, 2014.
- [10] D. H. Nguyen and H. Chen, "Supplier selection and operation planning in biomass supply chains with supply uncertainty," *Computers & Chemical Engineering*, vol. 118, pp. 103–117, 2018.
- [11] B. Hu and Y. Feng, "Optimization and coordination of supply chain with revenue sharing contracts and service requirement under supply and demand uncertainty," *International Journal of Production Economics*, vol. 183, pp. 185–193, 2017.
- [12] F. Lin, X. Qin, X. Pu, W. Zhu, and X. Zhuo, "Effects of in-house production on channel structures in a co-opetitive supply chain under supply uncertainty," *Omega*, vol. 103, Article ID 102426, 2021.

- [13] K. Govindan, M. Fattahi, and E. Keyvanshokoh, "Supply chain network design under uncertainty: a comprehensive review and future research directions," *European Journal of Operational Research*, vol. 263, no. 1, pp. 108–141, 2017.
- [14] M. W. Carter and C. C. Price, *Operations Research: A Practical Introduction*, CRC Press, Boca Raton, FL, USA, 2017.
- [15] D. Huang, X. Chen, Z. Liu, C. Lyu, S. Wang, and X. Chen, "A static bike repositioning model in a hub-and-spoke network framework," *Transportation Research Part E: Logistics and Transportation Review*, vol. 141, Article ID 102031, 2020.
- [16] D. Huang, J. Xing, Z. Liu, and Q. An, "A multi-stage stochastic optimization approach to the stop-skipping and bus lane reservation schemes," *Transportmetrica A: Transport Science*, vol. 17, no. 4, pp. 1272–1304, 2021.
- [17] D. Huang, Y. Gu, S. Wang, Z. Liu, and W. Zhang, "A two-phase optimization model for the demand-responsive customized bus network design," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 1–21, 2020.
- [18] H. H. Turan, M. N. Serarslan, and N. Kasap, "A fuzzy stochastic model for telecommunications bandwidth brokers under probabilistic QoS measures," *Applied Mathematical Modelling*, vol. 38, no. 1, pp. 12–27, 2014.
- [19] B. Stefansdottir and M. Grunow, "Selecting new product designs and processing technologies under uncertainty: two-stage stochastic model and application to a food supply chain," *International Journal of Production Economics*, vol. 201, pp. 89–101, 2018.
- [20] G. Dantzig, A. Orden, and P. Wolfe, "The generalized simplex method for minimizing a linear form under linear inequality restraints," *Pacific Journal of Mathematics*, vol. 5, no. 2, pp. 183–195, 1955.
- [21] A. Gupta and C. D. Maranas, "Managing demand uncertainty in supply chain planning," *Computers & Chemical Engineering*, vol. 27, no. 8–9, pp. 1219–1227, 2003.
- [22] Y. An, B. Zeng, Y. Zhang, and L. Zhao, "Reliable p-median facility location problem: two-stage robust models and algorithms," *Transportation Research Part B: Methodological*, vol. 64, pp. 54–72, 2014.
- [23] A. Ghodrattnama, R. Tavakkoli-Moghaddam, and A. Azaron, "Robust and fuzzy goal programming optimization approaches for a novel multi-objective hub location-allocation problem: a supply chain overview," *Applied Soft Computing*, vol. 37, pp. 255–276, 2015.
- [24] F. Zhou, Y. Xu, C. Tang et al., "Review on the development of food waste management policy in China," *Low-Carbon Economy*, vol. 9, no. 2, p. 10, 2020.
- [25] N. Geng and Y. Sun, "Multiobjective optimization of sustainable WCO for biodiesel supply chain network design," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 6640358, 16 pages, 2021.
- [26] M. Eskandarpour, P. Dejax, J. Miemczyk, and O. Péton, "Sustainable supply chain network design: an optimization-oriented review," *Omega*, vol. 54, pp. 11–32, 2015.
- [27] Z. Liu, T. Qiu, and B. Chen, "A study of the LCA based biofuel supply chain multi-objective optimization model with multi-conversion paths in China," *Applied Energy*, vol. 126, pp. 221–234, 2014.
- [28] Y. Ge, L. Li, and L. Yun, "Modeling and economic optimization of cellulosic biofuel supply chain considering multiple conversion pathways," *Applied Energy*, vol. 281, Article ID 116059, 2021.
- [29] S. Giarola, A. Zamboni, and F. Bezzo, "Spatially explicit multi-objective optimisation for design and planning of hybrid first and second generation biorefineries," *Computers & Chemical Engineering*, vol. 35, no. 9, pp. 1782–1797, 2011.
- [30] Y. Wang and H. Li, "Research on prediction model of civil aviation accident symptomatic based on grey neural network," *China Safety Science Journal*, vol. 22, no. 3, pp. 10–15, 2012.
- [31] E. Gao, T. Sowlati, and S. Akhtari, "Profit allocation in collaborative bioenergy and biofuel supply chains," *Energy*, vol. 188, Article ID 116013, 2019.
- [32] G. Gosalbe, B. L. Montastruc, S. Negny et al., "Optimal design and planning of biomass-to-biofuel supply chain considering economic dimension under strategic and tactical levels: a case study in Ethiopia," *Computer Aided Chemical Engineering*, vol. 48, pp. 1111–1116, 2020.
- [33] G. Baudry, "How the cap limit for food-crop-based biofuels may affect France's stakeholders by 2030? a range-based multi-actor multi-criteria analysis," *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 291–308, 2018.
- [34] X. Qin, N. Li, G. X. Li et al., "Current status and progress of food waste recycling technology," *Coal & Chemical Industry*, vol. 38, no. 7, pp. 35–40, 2015.
- [35] H. Uster and G. Memişoğlu, "Biomass logistics network design under price-based supply and yield uncertainty," *Transportation Science*, vol. 52, no. 2, pp. 474–492, 2017.
- [36] G. Memisoglu, "Bio-energy logistics network design under price-based supply and yield uncertainty," Doctoral dissertation, Texas A & M University, College Station, TX, USA, 2014.
- [37] F. Xie, "Modeling sustainability in renewable energy supply chain systems," Dissertation, Clemson University, Clemson, SC, USA, 2015.
- [38] Y. Zhang and Y. Jiang, "Robust optimization on sustainable biodiesel supply chain produced from waste cooking oil under price uncertainty," *Waste Management*, vol. 60, pp. 329–339, 2017.
- [39] Y. Qi, Z. Wang, and Y. Li, "Cost analysis of food waste collection and transportation in Shanghai," *Environmental Sanitation Engineering*, vol. 16, no. 3, pp. 47–49, 2008.
- [40] Y. Jiang, *Robust Optimization of Sustainable Biofuel Supply Chain under Uncertain Conditions*, Southeast University, Nanjing, China, 2016.
- [41] P. Xu and X. Mu, "Application and development of kitchen waste in energy production," *Modern Chemical Industry*, vol. 36, no. 5, pp. 12–16, 2016.
- [42] R. Chen, J. Jiang, R. Huang et al., "Overview of the recovery and reuse of gutter oil," *Science and Technology Innovation and Application*, vol. 211, no. 9, pp. 7–9, 2017.
- [43] H. Chen, J. Liu, H. Zhong et al., "Carbon emission reduction potential analysis of different treatment modes of kitchen waste," *Chinese Environmental Science*, vol. 33, no. 11, pp. 2102–2106, 2013.
- [44] T. R. P. Ramos, M. I. Gomes, and A. P. Barbosa-Póvoa, "Planning waste cooking oil collection systems," *Waste Management*, vol. 33, no. 8, pp. 1691–1703, 2013.

Research Article

An Online Map Matching Algorithm Based on Second-Order Hidden Markov Model

Xiao Fu ¹, Jiaxu Zhang ², and Yue Zhang ³

¹Jiangsu Key Laboratory of Urban ITS, Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, School of Transportation, Southeast University, Nanjing, China

²State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China

³School of Geographic Sciences, East China Normal University, Shanghai, China

Correspondence should be addressed to Xiao Fu; fuxiao@seu.edu.cn

Received 5 April 2021; Revised 19 June 2021; Accepted 9 July 2021; Published 19 July 2021

Academic Editor: Tomio Miwa

Copyright © 2021 Xiao Fu 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.

Map matching is a key preprocess of trajectory data which recently have become a major data source for various transport applications and location-based services. In this paper, an online map matching algorithm based on the second-order hidden Markov model (HMM) is proposed for processing trajectory data in complex urban road networks such as parallel road segments and various road intersections. Several factors such as driver's travel preference, network topology, road level, and vehicle heading are well considered. An extended Viterbi algorithm and a self-adaptive sliding window mechanism are adopted to solve the map matching problem efficiently. To demonstrate the effectiveness of the proposed algorithm, a case study is carried out using a massive taxi trajectory dataset in Nanjing, China. Case study results show that the accuracy of the proposed algorithm outperforms the baseline algorithm built on the first-order HMM in various testing experiments.

1. Introduction

With the development of positioning and wireless communication technologies, floating car data (e.g., trajectories of taxis) have become a major data source for many applications such as location-based services, intelligent transportation systems, and transport policy appraisals [1–5]. The errors of positioning data collected by global positioning system (GPS) equipment on floating vehicles are inevitable and could come from satellite, transmission process, and receiver [6]. Map matching is the process of matching GPS data with errors onto the road network in order to eliminate the impact of errors and maximize the effectiveness of data. In practical applications, a map matching algorithm plays a vital role, for example, travel time prediction based on floating car data, which needs to match GPS points to the corresponding road segment accurately. Therefore, the map matching algorithm is the basis for the large-scale application of floating car data.

Existing map matching algorithms can be divided into four categories based on the technique they adopted [7]: geometric technique [8–10], topological technique [11–14], probability statistics technique [15], and integration of multiple technologies [16–18]. The algorithms with geometric technique utilize geometric information of GPS point and road network (e.g., distance, angle and shape) without considering the topology of the road network. These algorithms show high efficiency of map matching, but the accuracy is low when matching low-precision GPS data to complex road networks. With regard to topological technique, both geometric factors and road topology are considered. To some extent, topological technique improves the matching accuracy but is still vulnerable to the influence of low-frequency sampling interval and large sampling noise. The probability statistics technique sets an ellipse or rectangle confidence area for each GPS point, thus we can obtain the probability according to the distance between the GPS point and the position in confidence area. Optimal matching

paths are determined according to values of the probability. Compared to the geometric technique and topological technique, the probability statistics technique is relatively more complex and difficult to implement, and shows low time efficiency. By combining geometric, topological, and probability factors, advanced techniques, such as Kalman filter [19], Bayesian filter [20], fuzzy logic model [21], multi-hypothesis tree [18], and hidden Markov model (HMM) [22], can effectively improve the map matching accuracy and achieve online incremental matching.

Of the advanced techniques, HMM has become popular in map matching studies. HMM is a prevailing paradigm of network-based dynamics modeling, which well suits the process of finding the most suitable matching point (i.e., hidden state) to each GPS point (i.e., observed state) on the road network in map matching problem. Existing map matching algorithms based on HMM can be categorized into two categories [20]: offline algorithms and online algorithms (refer to Table 1).

Offline HMM map matching algorithms are applied using historical data, batching the whole input trajectory to find the optimal matching path in the road network [23–28]. Whole trajectories enable offline algorithms to take account of the relationship between the front and the back points to achieve higher accuracy. Offline algorithms show robustness to the reduction of sampling rate, but the computation efficiency is low. Online algorithms estimate the current segment immediately after obtaining GPS data, and this kind of algorithm can be used for providing online services such as real-time navigation and trajectory monitoring. Because of the unavailability of future points, online algorithms are more complicated and require higher computation demand for real-time applications. Most studies utilize the sliding window mechanism with fixed window size to realize online matching [29, 30]. As the number of GPS points increases, the points in sliding window change dynamically. However, under the condition of low data quality or complex road network, small window leads to a significant decrease in matching accuracy while large window brings a significant decrease in computation efficiency. A few online map matching algorithms adopt variable sliding windows, but it requires a lot of extra computation [31]. Considering these, in this study, we proposed self-adaptive sliding windows to realize online map matching based on HMM, which promises accuracy and efficiency at the same time.

HMM builds on the stochastic processes of observation and state transition. In the map matching context, two probabilities are important: observation probability and transition probability. Observation probability is usually obtained by the Gaussian distribution of great-circle distance between GPS points and candidate points. In the literature, several factors have been considered in calculating observation probability. For instance, unsupervised HMM [25] considers the location of Antenna when matching mobile phone data. Other studies, e.g., Quick Matching [26], Multistage Matching [27], and SnapNet [29] consider more factors including the speed constraint, road level, and vehicle heading. With regard to transition probability calculation, to consider temporal relationship of different points, some

factors such as speed constraint and free-flow travel time are considered in several studies [23, 25, 28, 30, 31]. To consider spatial relationship, some factors are included such as the difference between great-circle distance and route distance [24, 28, 29, 31], difference between vehicle's heading change and road segments' heading change [27], and same road priority [29]. Based on the analysis of advantages of each algorithm, this study is a pioneering endeavour devoted to comprehensively considering various factors in online map matching, i.e., road level, driver's travel preference, vehicle heading, and network topology (same/adjacent road priority).

To the best of our knowledge, almost all map matching algorithms based on HMM adopt first-order HMM. The basic hypothesis of first-order HMM is that the observation probability is only related to the current state while the transition probability is only related to the previous state. Because the moving of a vehicle is a continuous process, there is a complex space-time relationship between the current state and the previous states. There is no doubt that first-order HMM over-simplifies several practical systems. Recently, Salnikov et al. [32] explored possibilities to enrich the system description and exploited empirical pathway information by means of second-order Markov models. Experiments show that the higher-order model is more effective than the first-order model in dealing with space-time continuum. Therefore, a need is likely to exist for solving the map matching problem using higher-order (e.g., second-order) HMM to achieve better map matching results.

Along the line of previous online studies, this study proposes a new map matching algorithm based on the HMM technique. The proposed algorithm extends the previous studies in the following aspects: firstly, the proposed novel map matching algorithm is on the basis of second-order HMM, which can better consider the space-time relationship among different states. It can be effectively applied to complex urban road network with parallel segments using low-frequency sampling GPS data. Secondly, the proposed algorithm comprehensively considers driver's travel preference towards road segments, road level, vehicle heading, and network topology when calculating the probability matrix of second-order HMM in order to improve the matching accuracy. Thirdly, the proposed algorithm introduces a self-adaptive sliding window mechanism. Compared to the conventional fixed window size mechanism, the introduced mechanism using a self-adaptive window size can significantly improve the map matching accuracy and has a reasonable computational performance.

In summary, the contributions of this work are threefold:

- (i) An online map matching algorithm based on the second-order hidden Markov model (HMM) is proposed, which can better consider the spatial-temporal relationship among different states and large perception fields.
- (ii) The proposed algorithm comprehensively considers driver's travel preference, road level, vehicle heading, and network topology when calculating the

TABLE 1: Comparison of HMM-based map matching studies.

Category	Model	Sliding window	Factors of observation probability	Factors of transition probability	Order of HMM
Offline	Interactive-voting matching [23]	×	×	Speed constraint	First order
	HMM matching [24]	×	×	Distance difference	
	Unsupervised HMM [25]	×	Antenna location	Speed constraint	
	Quick matching [26]	×	Speed constraint	×	
	Multistage matching [27]	×	Vehicle heading	Heading change difference	
	Driver path preference-based HMM [28]	×	×	Distance difference; speed constraint; driver's travel preference	
Online	SnapNet [29]	Fixed size	Vehicle heading; road level	Distance difference; network topology (same road priority)	Second order
	Spatial and temporal matching [30]	Fixed size	×	Speed constraint	
	Route choice HMM [31]	Variable size	×	Distance difference; free-flow travel time	
	This study	Self-adaptive size	Vehicle heading; road level; driver's travel preference	Distance difference; network topology (same/adjacent road priority)	

probability matrix of second-order HMM to improve the matching accuracy.

- (iii) Experiments on real-world dataset show that with the help of the self-adaptive sliding window mechanism and an extended Viterbi algorithm, our second-order HMM-based model can reach a high accuracy while ensuring efficiency.

The rest of this paper is organized as follows: in the next section, we state the problem of map matching. After the problem statement, an online map matching algorithm is proposed based on second-order HMM. A case study is carried out using a large taxi trajectory dataset in Nanjing, China, to test the validity of the algorithm under various road conditions. Finally, we conclude this study and discuss directions for further research.

2. Problem Statement

Vehicle trajectory data are a series of GPS points recorded in chronological order. Each GPS point indicates longitude and latitude, vehicle speed, timestamp, etc. Because the errors of data collected by GPS equipment are inevitable, map matching is a key process before using the vehicle trajectory data. It is a process of matching GPS data onto the road segments and obtaining the continuous and specific locations of vehicles on the road. The concepts used in this study are listed as follows:

GPS Point. A GPS point g_t is a record indicating the longitude, latitude, timestamp, and velocity of the vehicle.

GPS Trajectory. A GPS trajectory T is a series of GPS points. A T is showed as: $g_1 \rightarrow g_2 \rightarrow \dots \rightarrow g_n$.

Road Network. Road network $G(V, E)$ is a directed graph where V is the set of vertexes and E is the set of edges.

Road Segment. A road segment e is a directed edge in road network with length, road level, start vertex, and end vertex.

Candidate Point. The candidate point c_t^n is the n th candidate point matched with GPS point g_t on the road network.

Route. A route R is a sequence of road segments that matched best to a GPS trajectory T ; each road segment belongs to the edge set E of road network $G(V, E)$. R is showed as: $e_1 \rightarrow e_2 \rightarrow \dots \rightarrow e_n$.

With the above concepts, the map matching problem solved in this study can be defined as follows: find the candidate points $c_t^1, c_t^2, \dots, c_t^n$ on each road segment e corresponding to GPS point g_t . Select the most likely candidate points sequence for GPS trajectories T , and connect the matched road segments on network G to get route R .

3. Second-Order HMM Map Matching

3.1. Data Preprocessing. Generally, there are a lot of “redundancy” and “incompleteness” in floating vehicle GPS data, which may be caused by devices or road environments (e.g., stopping in or passing through tunnels). In order to ensure the efficiency and accuracy of map matching, we first need to preprocess the GPS data, including the removal of redundant data and the interpolation of missing data.

For the currently received data point g_t , calculate the great-circle distance [24] of g_t and g_{t-1} (denoted as $D_{t-1,t}$); if $D_{t-1,t}$ is less than a predefined lower bound, the current point g_t is omitted and not matched. If $D_{t-1,t}$ is greater than an upper bound, the two points will be interpolated linearly.

With the data preprocessing, the redundant GPS data points can be effectively eliminated to avoid unnecessary matching. At the same time, interpolation of two points with too large intervals helps to process low-frequency GPS data.

3.2. *Candidate Point Selection.* For the currently received data point g_t , we search for its candidate points (refer to Figure 1(a)) with the following steps:

Step 1: using the R-tree index, the road segments within a predefined error circle or nearest to the point g_t are selected as road segment candidates [13, 17].

Step 2: vertically project the point g_t on the candidate road segments, and the projection point c_t^i is a candidate point for g_t . If the projection point falls outside the segment, choose the closer vertex of the segment as c_t^i . As shown in Figure 1(a), the candidate points for g_t are $c_t^1, c_t^2, \dots, c_t^5$. The distances from g_t to the candidate points are denoted as $d_t^1, d_t^2, \dots, d_t^5$, respectively.

3.3. *Observation Probability.* In the first-order HMM, the observation probability is used to measure the probability of getting some kinds of observed value in a hidden state [33]. The map matching algorithms based on HMM usually regard the GPS point g_t as the observation value of state t , and the actual position of g_t as the hidden value of state t . The observation probability is modeled using a Gaussian distribution for GPS trajectories. The first-order HMM observation probability in this paper is obtained as

$$P(g_t|c_t^i) = \frac{1}{\sqrt{2\pi}\sigma_t} e^{-0.5(\tau \cdot \rho \cdot d_t^i / \sigma_t)^2}, \quad (1)$$

where $P(g_t|c_t^i)$ is the observation probability of the candidate point c_t^i on g_t , d_t^i is the great-circle distance between g_t and the candidate point c_t^i , σ_t is the standard deviation of a Gaussian random variable that corresponds to the average great-circle distance between g_t and its candidate points. τ is a weight given on vehicle heading, which is related to the road direction angle α_{road} and the trajectory direction angle α_{GPS} :

$$\tau = v + \frac{e^{|\alpha_{\text{road}} - \alpha_{\text{GPS}}|}}{e^{2/\pi}}. \quad (2)$$

In equation (2), the road direction angle α_{road} is the direction angle of the two vertexes of a segment. The trajectory direction angle α_{GPS} indicates the direction angle of the last GPS point and the current GPS point. Because of the bidirectional property of the road, there are two results of $|\alpha_{\text{road}} - \alpha_{\text{GPS}}|$, and the smaller value of the two results should be used. v is a parameter which can be estimated with real data.

ρ is a weight reflecting the effect of road including road level (denoted as r level) and driver's travel preference for the road segment (denoted as p level):

$$\rho = 1 - \mu(r \text{ level} + p \text{ level}), \quad (3)$$

where μ is a parameter to be estimated. In this study, r level is within $[0, 5]$. A high r level indicates a high level of road. The value of p level is also ranging from 0 to 5. Considering driver's travel experience as a sigmoid curve [34], p level can be derived as

$$p \text{ level} = \frac{5}{1 + e^{-\omega + \omega'}}, \quad (4)$$

where ω is the actual number of times drivers pass the road segment in a certain time period, and ω' is a predefined expected number.

In this way, the observation probability can be obtained. By using vehicle heading weight τ and road weight ρ , we can consider road level, driver's travel preference, and the heading of the floating vehicle at that time, which are significant in online map matching with limited information. Take Figure 1(b) as an example to illustrate the merit of road weight ρ . The current GPS point g_t is located in the middle of two parallel road segments. The distances from g_t to c_t^1 and c_t^2 are the same. In conventional map matching methods, c_t^1 or c_t^2 is selected randomly as the real position of vehicle. However, if road level and travel preference are taken into account using our proposed method, we can consider c_t^1 as the real position of vehicle. It can be seen that without subsequent GPS points, we must make full use of the information provided by existing GPS points and road network in order to improve the matching accuracy.

Figure 1(c) shows the merits of incorporating vehicle heading weight τ . The GPS point g_{t+1} is located near the intersection, which is close to the candidate point c_{t+1}^1 and c_{t+1}^2 , and the distance d_{t+1}^1 is the same as d_{t+1}^2 . Connecting g_t and g_{t+1} , the vehicle heading weight between the connecting line and the two segments is τ_1 and τ_2 . Considering the impact of vehicle heading weight, c_{t+1}^2 has a greater probability of observation, and we can suppose that c_{t+1}^2 is the real position of the vehicle at time $t + 1$.

3.4. *Transition Probability.* In the first-order HMM, the transition probability measures the transition from one hidden state to another [33]. The map matching algorithm based on HMM uses the transition probability to measure the probability of moving from a candidate point c_{t-1}^j at time $t-1$ to a candidate point c_t^i at time t [29]. The formula for calculating the transition probability of the first-order HMM in this paper is given as Equation (5):

$$P(c_t^i|c_{t-1}^j) = \begin{cases} p_{\text{same}} \frac{1}{\beta} e^{-s_t/\beta}, & c_t^j \text{ and } c_{t-1}^i \text{ are on the same/adjacent road segments,} \\ (1 - p_{\text{same}}) \frac{1}{\beta} e^{-s_t/\beta}, & \text{otherwise,} \end{cases} \quad (5)$$

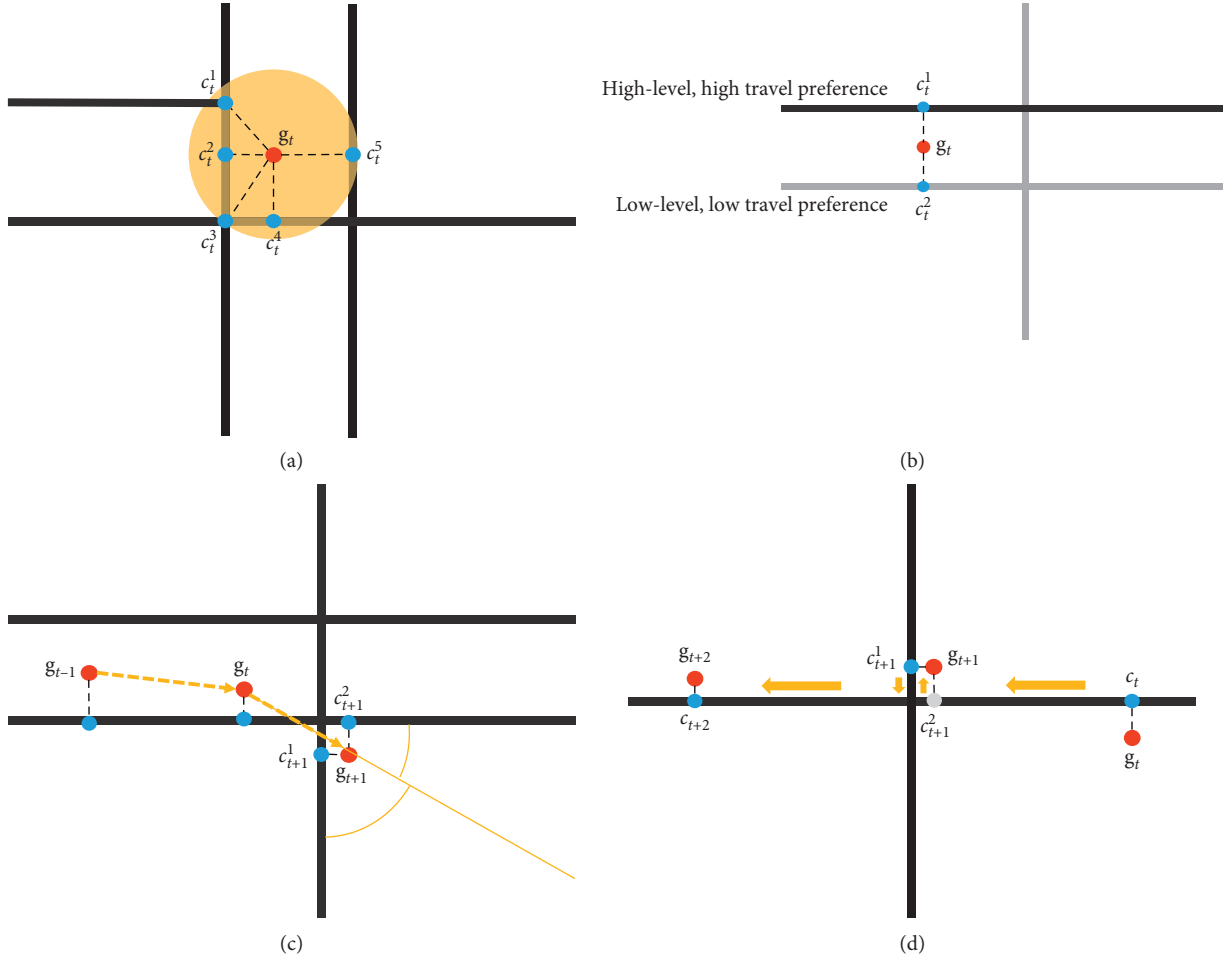


FIGURE 1: Illustration of merits of the proposed map matching method. (a) Candidate point selection. (b) Impact of road weight (ρ). (c) Impact of vehicle heading weight (τ). (d) Incorrect match with first-order HMM.

where $P(c_t^j | c_{t-1}^i)$ is the transition probability from candidate point c_{t-1}^i to candidate point c_t^j . p_{same} (>0.5) is a parameter. With equation (5), we can get the transition probability with explicit consideration of network topology (i.e., considering if c_{t-1}^i and c_t^j are on the same or adjacent road segments). In this way, the topological relation of road segments is taken into account. β is the mean of s_t . s_t is the difference between the great-circle distance from g_{t-1} to g_t (denoted as $\text{dist}(g_{t-1}, g_t)$) and the route length from c_{t-1}^i to c_t^j (denoted as $\text{routeDist}(c_{t-1}^i, c_t^j)$):

$$s_t = \left| \text{dist}(g_{t-1}, g_t) - \text{routeDist}(c_{t-1}^i, c_t^j) \right|. \quad (6)$$

3.5. Self-Adaptive Sliding Window and Second-Order Probability. Existing first-order HMM online map matching algorithms usually only focus on one single GPS point, considering its local geometric relation and road topology, which results in the precision of online map matching algorithm far behind the second-order map matching algorithm.

Figure 1(d) shows an example that the conventional first-order HMM online map matching results in an incorrect match. Obviously, from GPS point g_t to g_{t+2} , the vehicle does not turn and the correct matching path should be $c_t \rightarrow c_{t+1}^2 \rightarrow c_{t+2}$. However, in the process of the first-order HMM online incremental matching, an incorrect matching result is $c_t \rightarrow c_{t+1}^1 \rightarrow c_{t+2}$. The reason for this error is that the first-order HMM only considers the observation probability of a single point and the transition probability between two points. However, the measurement of transition probability should be on a larger scale. The real location of the current GPS point is not just related to the previous point, but to multiple previous points.

The higher-order HMM is an extension of the first-order HMM [35]. The basic assumption of the higher-order HMM is that the current state is not only related to one previous state but also to multiple previous states. In some cases, the second-order HMM is more consistent with the real situation, such as natural language processing, speech recognition, and so on [36, 37]. For the map matching problem, because the vehicle movement is continuous, the real position of the current point is not only related to the previous

point but also to the trajectory formed by two or more points. Therefore, the higher-order HMM is somewhat more suitable for map matching than the traditional first-order HMM. Analogous to human eyes observing things, we should first pay attention to the characteristics of things as a whole. For example, in Figure 1(d), the connection from g_t to g_{t+2} is approximately a straight line, so the GPS point g_{t+1} is more likely to be matched to c_{t+1}^2 than c_{t+1}^1 . To overcome the matching errors which may be resulting from the first-order HMM and to improve the accuracy of online map matching, in this study, we extend the first-order HMM map matching to a second-order one. Compared to the first-order HMM, the difficulties in using second-order HMM lie in the design of the probability matrix and how to improve the computational efficiency.

In the applications such as real-time navigation and travel time estimation, online map matching is necessary. The existing HMM map matching algorithms usually use the sliding window to realize online matching. Denote the sliding window size as w (i.e., number of GPS points). If the window overflows after the current point g_t entering the window, the first point in the window g_{t-w} is removed, and the matching result of g_{t-w} point will be finally determined. As the new point continues to join, matching results within the window may be changed continuously. The introduction of the sliding window makes online map matching possible, but it is difficult to determine the window size w . If w is too large, the matching speed will be too slow to meet the real-time performance requirement. If w is too small, the matching accuracy will be compromised. To solve this problem, a self-adaptive sliding window is proposed in this study.

In this study, we consider different sizes of self-adaptive sliding window. By calculating the average value of GPS points positioning error in the current window, sliding windows of different sizes are automatically selected to adapt to the current GPS positioning error, which can improve the accuracy of the online map matching as much as possible. The average value of GPS points positioning error (denoted as E_{ave}) can be obtained as

$$E_{ave} = \frac{\sum_{n=t-w+1}^{n=t} \text{dist}(g_n, c_n)}{w}, \quad (7)$$

where c_n is the candidate point which is matched to g_n .

The observation probability of the second-order HMM $P(g_{t-1}, g_t | c_{t-1}^i, c_t^j)$ can be obtained from the first-order HMM:

$$P(g_{t-1}, g_t | c_{t-1}^i, c_t^j) = P(c_t^j | c_{t-1}^i) \cdot P(g_{t-1} | c_{t-1}^i) \cdot P(g_t | c_t^j). \quad (8)$$

Define the second-order HMM state transition probability (denoted as $P(c_t^i | c_{t-2}^j, c_{t-1}^k)$) as

$$P(c_t^i | c_{t-2}^j, c_{t-1}^k) = \frac{1}{\lambda} e^{-k_t/\lambda}, \quad (9)$$

where λ is the mean of k_t . k_t is the difference between the great-circle distance from g_{t-1} to g_{t+1} and the route length from c_{t-1}^j to c_{t+1}^i :

$$k_t = \left| \sum_{n=t-2}^{n=t-1} \text{dist}(g_n^j, g_{n+1}^j) - \sum_{n=t-2}^{n=t-1} \text{routeDist}(c_n^j, c_{n+1}^i) \right|. \quad (10)$$

The second-order transition probability describes the state transition between three consecutive candidate points, that is, the actual position of the current GPS point is related to the previous two points. In this way, the strong assumption of the first-order HMM is relaxed and the accuracy of map matching is improved. In fact, we can continue to extend the proposed method to the third-order HMM and define appropriate observation and transition probabilities to improve accuracy. However, the third-order HMM will make the calculation process more complicated, which is not conducive to online map matching.

3.6. Extended Viterbi Algorithm. In the previous sections, we introduce the second-order HMM to solve the map matching problem. Although we use the sliding window mechanism to reduce the computational complexity of matching a single GPS point, the algorithm complexity of traversing the second-order HMM is still $O(n^w)$. Traversal search seriously affects the online performance of the matching algorithm. Thus, some dynamic programming algorithms should be used to reduce the complexity.

The objective function of second-order HMM dynamic programming is defined as

$$\max \prod_{n=t-w+3}^{n=t} (P(c_n^i | c_{n-2}^j, c_{n-1}^k) \times P(g_{n-2}, g_{n-1} | c_{n-2}^j, c_{n-1}^k)). \quad (11)$$

Viterbi algorithm is an efficient dynamic programming algorithm, which can effectively avoid repeated searches of path and quickly achieve the optimal solution. It is widely used to solve the first-order HMM. For solving the second-order HMM with a complexity of $O(n^2)$, we extend the traditional Viterbi algorithm [38] using an order reduction process as follows:

Step 1: order reduction

In the second-order HMM, $P(g_{t-1}, g_t | c_{t-1}^i, c_t^j)$ is regarded as the observation probability, which is equivalent to the observation probability of a single candidate point in the first-order HMM. Equation (8) shows that the observation probability of the second-order HMM is the product of the observation probability of two consecutive candidates in the first-order HMM and the state transition probability. Thus, the order of the second-order HMM can be reduced by using equation (8) (refer to Figure 2). If the second-order HMM has two layers, each layer has m and n nodes, respectively, the second-order HMM can be reduced to one layer with $m \times n$ nodes.

Step 2: recursive tracing

After Step1, we can use the traditional Viterbi algorithm for iterative calculation to solve the second-order HMM in the following process (refer to Figure 2):

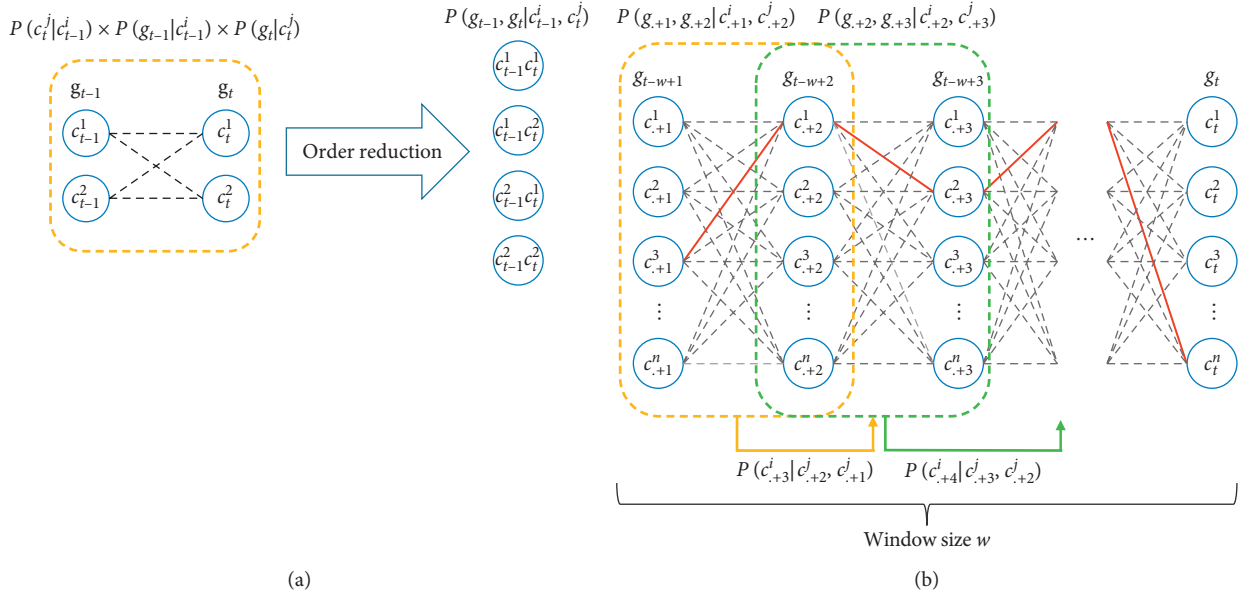


FIGURE 2: Illustration of the extended Viterbi algorithm. (a) Order reduction process. (b) Extended Viterbi algorithm.

- Starting from the first layer's nodes, the observation probability of each layer's nodes after reduction and the transition probability between adjacent two layers' nodes are calculated.
- Calculate the maximum total probability of each node from the second layer to the last layer. Save maximum total probability and precursor node of each node.
- Select the node with the highest total probability in the last layer, and go back to its precursor node until the first layer.

With the above steps, we can find the optimal matching path $(c_{t-w+1}^i, c_{t-w+2}^j, \dots, c_t^k)$ in the sliding window.

4. Case Study

In this section, we make sensitivity analyses of the parameters involved in the algorithm, and use real data to show the merits of the proposed second-order HMM map matching algorithm.

4.1. Data Preparation and Evaluation Metric. We used the road network data of Qinhuai District in Nanjing, China, including 6901 sections and 4647 nodes. Taxi GPS data with 30 s sampling interval collected in September 2016 were used, including 500 trajectories for 20 taxis. We manually match these trajectories to the road network as the ground truth. In order to verify the effectiveness of the algorithm under extreme conditions and reflect the advantages of the proposed algorithm, we resampled the original data and added the random noise of Gaussian distribution. The resampling intervals are 60 s to 300 s. The Gaussian noises with a standard deviation of 10 m to 80 m (convert to degrees) were added to the longitude and latitude.

Evaluation metric is defined as follows: first, we find the common matching sequence X (the sequence that matched correctly) between the matched output route M and the real trajectory T . Based on this sequence, the precision and the recall of the map matching result (denoted as pcs and rc , respectively) can be calculated as

$$pcs = \frac{X}{M}, \quad (12)$$

$$rc = \frac{X}{T}, \quad (13)$$

where pcs is defined as the ratio of the length of matched sequence X and the total length of the matched trajectory M . rc is defined as the ratio between the length of the matched sequence X and the total length of the real trajectory T . In this study, F_1 -score, which is widely used to evaluate the performance of classification models and prediction models [39], is adopted to evaluate the proposed model:

$$F_1 - \text{score} = \frac{2 \cdot pcs \cdot rc}{pcs + rc}. \quad (14)$$

5. Results

Effects of different parameters on map matching accuracy are investigated in this study. In the proposed model, there are three parameters to be estimated, i.e., μ , ν , and p_{same} . According to previous studies, the approximate range of the three parameters can be obtained. Figure 3 shows the impact of different parameter values on F_1 -score and Table 2 shows the optimal parameter values. It can be seen that, when the road weight μ is around 0.02, the vehicle heading weight ν is around 0.6, and the same/adjacent road priority p_{same} is around 0.6, and their impact on the final performance becomes optimal and stable.

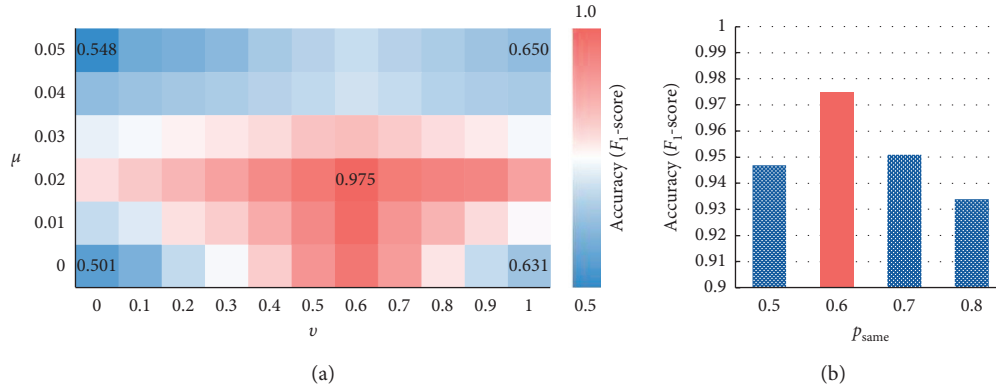


FIGURE 3: Sensitivity analyses of model parameters. (a) The effect of μ and ν . (b) The effect of p_{same} .

TABLE 2: Optimal parameter values.

Parameter	Optimal value	Range	Impact
μ	0.02	0–0.05	Control the effect of road weight on results
ν	0.6	0–1	Control the effect of vehicle heading weight on results
p_{same}	0.6	0.5–0.8	Control the effect of same/adjacent road priority on results

Figure 4(a) shows the effect of window size w on the accuracy of map matching. It can be seen that when $w=3$, the value of F_1 -score increases significantly. The reason is that when the size of the sliding window is larger than 3, the second-order HMM comes into play. Under different standard deviations of noise (SDNs), when the sliding window size increases from 3 to 10, the matching accuracy remains unchanged. However, as the sliding window's size increases, the computation time of matching a single GPS point increases rapidly. Thus, the optimal self-adaptive sliding window sizes are 3, 4, and 5.

Figure 4(b) shows the effects of the sample interval and the random SDN on accuracy of map matching. With the increase in the sampling interval and SDN, the F_1 -score decreases. It can be seen from Figure 4(b) that when the sampling interval is between 30 s and 90 s and the SDN ranges from 0 to 30 m, the F_1 -score is kept above 0.9.

With the map matching algorithm proposed in this paper, various factors (i.e., road level, driver's travel preference, vehicle heading, and network topology) are considered. Figure 5 shows some map matching results in complex urban road network environment. From Figure 5(a), it can be seen that the first-order HMM map matching algorithm may bring about mismatch when it deals with parallel road segments. Under the constraints of topological relations, the second-order HMM algorithm gives a greater transition probability to the segment, which is adjacent to the previous segment to effectively reduce errors. When the GPS points are located near the road intersection, the first-order HMM algorithm may match the GPS points to the section that intersects with the current road. The second-order HMM and sliding window can help solve this problem. The second-order transition probability can effectively avoid the detour of matching trajectory at the intersection and improve the accuracy of map matching. Figure 5(b) shows an overview of map matching result in the

central area of Nanjing, where the road network is dense and complex. The proposed algorithm is found well performed on parallel segments and intersections. This is because the second-order HMM model has a wider field of view, and our method considers a variety of factors, which is helpful for map matching in complex conditions.

Figure 6(a) compares the accuracy of the proposed second-order HMM map matching algorithm with the accuracy of our baseline (the first-order HMM map matching algorithm) at different sample intervals without adding random noise. It can be seen that the F_1 -score of the proposed algorithm is higher than that of the first-order HMM. With the increase of the sampling interval, the advantages of the proposed algorithm become obvious. Taking the 300 seconds sampling interval as an example, the distance between two GPS points is about 2500 meters considering the average speed of 30 km/h on urban roads. In this situation, the position correlation between two consecutive GPS points is very low. The traditional first-order HMM algorithm only considers the transition probability between two points, so the error tends to be very large. Our proposed algorithm integrates several factors such as road level and driver's travel preference, and the second-order transition probability can match GPS trajectory on a larger scale, so it shows higher accuracy (F_1 -score is about 0.67).

Figure 6(b) compares the accuracy of the proposed second-order HMM map matching algorithm with our baseline (the first-order HMM algorithm) at different SDNs with 30 s sample interval. The map matching accuracy of the proposed algorithm is always higher than that of the first-order algorithm. The reason is that the conventional first-order HMM algorithm only considers the difference between the great-circle distance and route distance when calculating the observation probability of candidate points. When the positioning error of GPS point increases and the road network is dense, matching errors are numerous. In

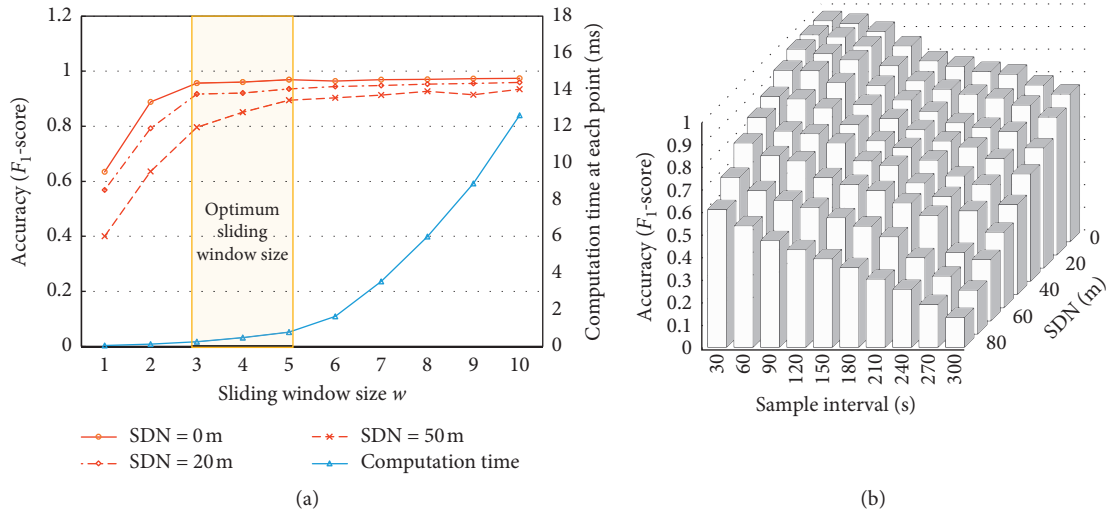


FIGURE 4: Effects of sliding window size and sampling on map matching accuracy. (a) Effects of sliding window size. (b) Effects of sample interval and SDN.

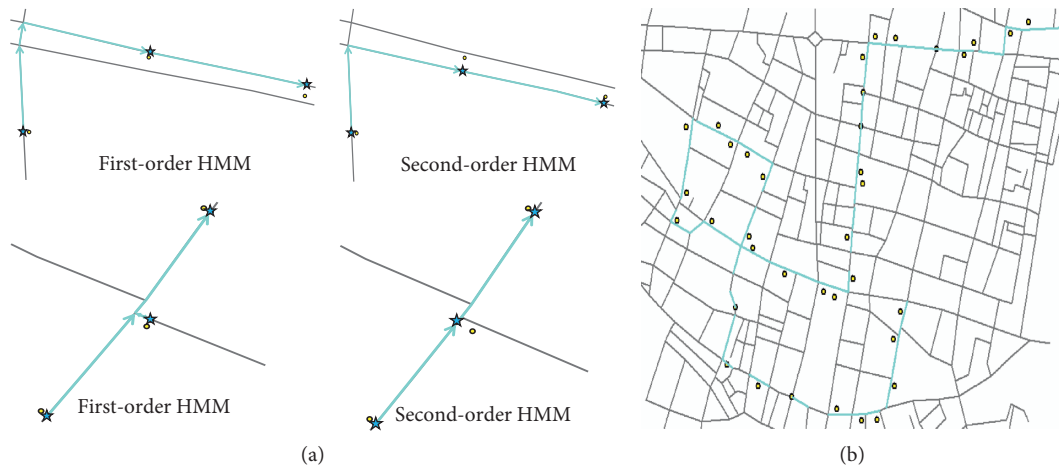


FIGURE 5: Demonstration of several map matching cases. (a) Map matching results at parallel segments and road intersections. (b) An overview of map matching results.

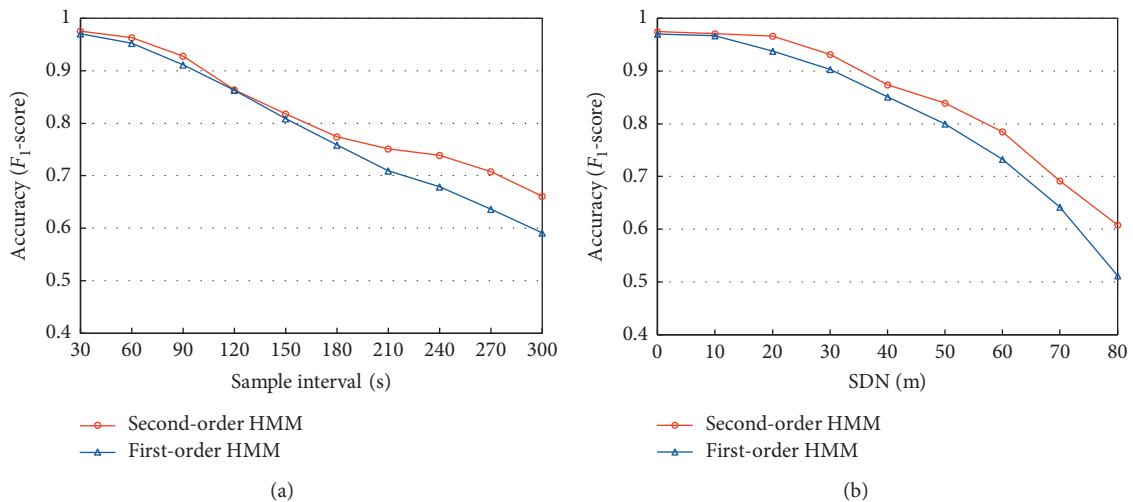


FIGURE 6: Analysis of map matching accuracy. (a) Accuracy at different sampling intervals. (b) Accuracy at different SDNs.

TABLE 3: Comparison of the accuracy (F_1 - score) with some state-of-the-art methods.

Method	Accuracy
HMM-DPP [28]	0.910
SnapNet [29]	0.909
This study	0.975

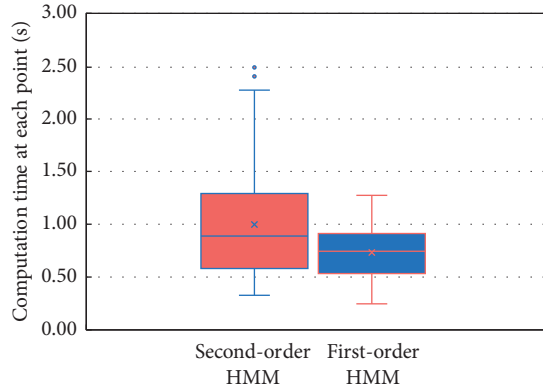


FIGURE 7: Comparison of map matching efficiency between the first-order and second-order HMM.

practice, the GPS positioning error is significant in city centre with dense high-rise buildings. As the proposed second-order HMM algorithm excelled conventional algorithms in accuracy (0.6 compared to 0.5 when SDN equals 80 m), the proposed algorithm can be adopted to achieve high accuracy of map matching in the whole city.

When comparing with the state-of-the-art methods that are most relevant to our proposed method on the condition of raw GPS data, the results in Table 3 show that our second-order HMM method performs well with regard to accuracy. Figure 7 compares the efficiency of the proposed second-order HMM map matching algorithm with the conventional first-order HMM algorithm. For the first-order HMM algorithm, the sliding window size is set to 5. It can be seen from Figure 7 that the computation time at each point using the second-order HMM algorithm is slightly longer than when using the first-order HMM algorithm, and the average computation time is less than 1 s. In the process of self-adaptation of the sliding window size, a small number of outliers appear. For example, using the second-order HMM algorithm, there are a few points whose the computation time is longer than 2 seconds. However, in this example, the overall matching efficiency is close to the first-order HMM map matching, which can meet the requirements of online map matching. Moreover, compared to the first-order HMM, the second-order HMM can better consider the spatial-temporal relationship among different states and larger perception fields, which can get remarkable accuracy under complex conditions.

6. Conclusions

Accurate and efficient matching of GPS data onto road network is the basis and prerequisite for conducting traffic

flow analysis and providing location-based service. An online map matching algorithm based on the second-order HMM is presented in this paper. Various factors (i.e., road level, driver's travel preference, vehicle heading, and network topology) are explicitly considered in the algorithm, which effectively improve the accuracy of map matching in complex urban road network environment. An extended Viterbi algorithm is adopted to solve the map matching problem efficiently. A self-adaptive sliding window mechanism is proposed to adjust window size on a real-time basis and ensures high accuracy.

We tested the proposed algorithm using real road network and massive taxi GPS data collected in Nanjing, China. The proposed map matching approach was found to outperform state-of-the-art algorithms built on the first-order HMM in various testing environments. Sliding window with self-adaptive size is shown to be an effective method for online incremental map matching. Some typical types of mismatching can be avoided in complex urban road network environment such as parallel road segments and various road intersections. The map matching accuracy of the proposed algorithm is demonstrated to be higher than that of the conventional first-order HMM algorithm. The efficiency of the proposed algorithm is close to the first-order HMM map matching algorithm, which can meet the requirements of online map matching. Therefore, the proposed algorithm is applicable in real-time navigation, trajectory monitoring, traffic flow analysis, and other related fields.

To solve the map matching problem, there are some other solutions such as considering driving direction and turning behaviour. The consideration of users with heterogeneous activity/travel behaviour is suggested as another interesting extension of the proposed method, potentially improving the accuracy of map matching [31, 40]. In the case study, the proposed algorithm is tested using a single processor. How to incorporate the parallel computing technologies into the proposed algorithm with a large number of trajectories needs further investigation [41]. Besides, the comparison of the advantages and disadvantages of the second-order-HMM-based method and other advanced map matching algorithms can also be the focus of future research.

Data Availability

The GPS data used to support the findings of this study have not been made available because of the confidentiality agreement.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work described in this paper was jointly supported by the National Key Research and Development Program of China (2018YFB1600900), National Natural Science

Foundation of China (71601045), and “Zhishan” Scholars Programs of Southeast University.

References

- [1] W. Wong and S. C. Wong, “Network topological effects on the macroscopic bureau of public roads function,” *Transportmetrica*, vol. 12, pp. 272–296, 2015.
- [2] J. Bao, P. Liu, X. Qin, and H. Zhou, “Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data,” *Accident Analysis & Prevention*, vol. 120, pp. 281–294, 2018.
- [3] W. Huang, W. Jia, J. Guo et al., “Real-time prediction of seasonal heteroscedasticity in vehicular traffic flow series,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3170–3180, 2018.
- [4] Z. He, Y. Lv, L. Lu, and W. Guan, “Constructing spatio-temporal speed contour diagrams: using rectangular or non-rectangular parallelogram cells?” *Transportmetrica B: Transport Dynamics*, vol. 7, no. 1, pp. 44–60, 2019.
- [5] B. Y. Chen, H. Yuan, Q. Li, S.-L. Shaw, W. H. K. Lam, and X. Chen, “Spatiotemporal data model for network time geographic analysis in the era of big data,” *International Journal of Geographical Information Science*, vol. 30, no. 6, pp. 1041–1071, 2016.
- [6] M. Bierlaire, J. Chen, and J. Newman, “A probabilistic map matching method for smartphone GPS data,” *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 78–98, 2013.
- [7] Q. Luo, J. Auld, and V. Sokolov, “Addressing some issues of map-matching for large-scale, high-frequency GPS data sets,” in *Proceedings of 95th Annual Meeting of the Transportation Research Board*, Washington, D.C., USA, January 2016.
- [8] C. E. White, D. Bernstein, and A. L. Kornhauser, “Some map matching algorithms for personal navigation assistants,” *Transportation Research Part C: Emerging Technologies*, vol. 8, no. 1–6, pp. 91–108, 2000.
- [9] G. Taylor, G. Blewitt, D. Steup, S. Corbett, and A. Car, “Road reduction filtering for GPS-GIS navigation,” *Transactions in GIS*, vol. 5, no. 3, pp. 193–207, 2001.
- [10] J. Schweizer, S. Bernardi, and F. Rupi, “Map-matching algorithm applied to bicycle global positioning system traces in Bologna,” *IET Intelligent Transport Systems*, vol. 10, no. 4, pp. 244–250, 2016.
- [11] N. R. Velaga, M. A. Quddus, and A. L. Bristow, “Improving the performance of a topological map-matching algorithm through error detection and correction,” in *Proceedings of 91th Annual Meeting of the Transportation Research Board*, Washington, D.C., USA, January 2012.
- [12] B. Y. Chen, H. Yuan, Q. Li, W. H. K. Lam, S.-L. Shaw, and K. Yan, “Map-matching algorithm for large-scale low-frequency floating car data,” *International Journal of Geographical Information Science*, vol. 28, no. 1, pp. 22–38, 2014.
- [13] N. Tradišauskas, J. Juhl, H. Lahrmann, and C. S. Jensen, “Map matching for intelligent speed adaptation,” *IET Intelligent Transport Systems*, vol. 3, no. 1, pp. 57–66, 2009.
- [14] Z. c. He, S. Xi-wei, L. j. Zhuang, and P. I. Nie, “On-line map-matching framework for floating car data with low sampling rate in urban road networks,” *IET Intelligent Transport Systems*, vol. 7, no. 4, pp. 404–414, 2013.
- [15] L. Zheng, X. Liu, and B. Yi, “Dynamic weighted real-time map matching algorithm considering spatio-temporal property,” *Journal of Computer Applications*, vol. 37, no. 8, pp. 2381–2386, 2017.
- [16] X. Liu, K. Liu, M. Li, and F. Lu, “A ST-CRF map-matching method for low-frequency floating car data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 5, pp. 1241–1254, 2017.
- [17] Y.-J. Gong, E. Chen, X. Zhang, L. M. Ni, and J. Zhang, “AntMapper: an ant colony-based map matching approach for trajectory-based applications,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 390–401, 2018.
- [18] K. Zhang, S. Liu, Y. Dong, D. Wang, Y. Zhang, and L. Miao, “Vehicle positioning system with multi-hypothesis map matching and robust feedback,” *IET Intelligent Transport Systems*, vol. 11, no. 10, pp. 649–658, 2017.
- [19] C. Takenga, T. Peng, and K. Kyamakya, “Post-processing of fingerprint localization using kalman filter and map-matching techniques,” in *Proceedings of 9th International Conference on Advanced Communication Technology*, Gangwon, Korea (South), February 2007.
- [20] S. Taguchi, S. Koide, and T. Yoshimura, “Online map matching with route prediction,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 338–347, 2019.
- [21] M. Ren and H. A. Karimi, “A fuzzy logic map matching for wheelchair navigation,” *GPS Solutions*, vol. 16, no. 3, pp. 273–282, 2012.
- [22] Z. Chen and R. C. Qiu, “Prediction of channel state for cognitive radio using higher-order hidden Markov model,” in *Proceedings of the IEEE SoutheastCon 2010 (SoutheastCon)*, March 2010.
- [23] J. Yuan, Y. Zheng, C. Zhang, X. Xie, and G.-Z. Sun, “An interactive-voting based map matching algorithm,” in *Proceedings of the 2010 Eleventh International Conference on Mobile Data Management*, May 2010.
- [24] P. Newson and J. Krumm, “Hidden Markov map matching through noise and sparseness,” in *Proceedings of ACM SIGSPATIAL*, pp. 336–343, Seattle, WA, USA, November 2009.
- [25] L. Bonnetain, A. Furno, J. Krug, and N.-E. E. Faouz, “Can we map-match individual cellular network signaling trajectories in urban environments? Data-driven study,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 7, 2019.
- [26] R. Song, W. Lu, W. Sun, Y. Huang, and C. Chen, “Quick map matching using multi-core CPUs,” in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems-SIGSPATIAL’12*, pp. 605–608, Redondo Beach California, November 2012.
- [27] M. M. Atia, A. R. Hilal, C. Stellings et al., “A low-cost lane-determination system using GNSS/IMU fusion and HMM-based multistage map matching,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, pp. 1–11, 2017.
- [28] C. Song, X. Yan, N. Stephen, and A. A. Khan, “Hidden Markov model and driver path preference for floating car trajectory map matching,” *IET Intelligent Transport Systems*, vol. 12, no. 10, pp. 1433–1441, 2018.
- [29] R. Mohamed, H. Aly, and M. Youssef, “Accurate real-time map matching for challenging environments,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 847–857, 2017.
- [30] Y. Lou, C. Zhang, Y. Zheng, W. Wang, and Y. Huang, “Map-matching for low-sampling-rate GPS trajectories,” in *Proceedings of the ACM-GIS*, pp. 352–361, Seattle, WA, USA, November 2009.

- [31] G. R. Jagadeesh and T. Srikanthan, "Online map-matching of noisy and sparse location data with hidden Markov and route choice models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, pp. 1–12, 2017.
- [32] V. Salnikov, M. T. Schaub, and R. Lambiotte, "Using higher-order Markov models to reveal flow-based communities in networks," *Scientific Reports*, vol. 6, Article ID 23194, 2016.
- [33] S. R. Eddy, "What is a hidden Markov model?" *Nature Biotechnology*, vol. 22, no. 10, pp. 1315–1316, 2004.
- [34] N. Leibowitz, B. Baum, G. Enden, and A. Karniel, "The exponential learning equation as a function of successful trials results in sigmoid performance," *Journal of Mathematical Psychology*, vol. 54, no. 3, pp. 338–340, 2010.
- [35] F. Ye and Y. Wang, "Research advancement of high-order hidden Markov model," *Advances in Mathematics*, vol. 2, no. 2, pp. 845–848, 2014.
- [36] T. Lee, F. Zheng, W. Wu, and D. Chen, "The hidden Markov model of Co-articulation and its application to the continuous speech recognition," *Journal of Electronics (China)*, vol. 17, no. 3, pp. 242–247, 2000.
- [37] M. Gales, S. Watanabe, and E. Fosler-Lussier, "Structured discriminative models for speech recognition: an overview," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 70–81, 2012.
- [38] Y. He, "Extended Viterbi algorithm for second order hidden Markov process," in *Proceedings of the 9th International Conference on IEEE*, IEEE, Rome, Italy, November 1988.
- [39] E. M. Voorhees, "Variations in relevance judgments and the measurement of retrieval effectiveness," *Information Processing & Management*, vol. 36, no. 5, pp. 697–716, 2000.
- [40] X. Fu and W. H. K. Lam, "Modelling joint activity-travel pattern scheduling problem in multi-modal transit networks," *Transportation*, vol. 45, no. 1, pp. 23–49, 2018.
- [41] Q. Li, T. Zhang, and Y. Yu, "Using cloud computing to process intensive floating car data for urban traffic surveillance," *International Journal of Geographical Information Science*, vol. 25, no. 8, pp. 1303–1322, 2011.

Research Article

Multidepot Recycling Vehicle Routing Problem with Resource Sharing and Time Window Assignment

Yong Wang ¹, Xiuwen Wang ¹, Xiangyang Guan ², and Jinjun Tang ³

¹School of Economics and Management, Chongqing Jiaotong University, Chongqing 400074, China

²Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA

³Smart Transport Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China

Correspondence should be addressed to Jinjun Tang; jinjuntang@csu.edu.cn

Received 8 April 2021; Revised 19 April 2021; Accepted 3 May 2021; Published 17 May 2021

Academic Editor: xinyuan chen

Copyright © 2021 Yong Wang 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 study aims to provide tactical and operational decisions in multidepot recycling logistics networks with consideration of resource sharing (RS) and time window assignment (TWA) strategies. The RS strategy contributes to efficient resource allocation and utilization among recycling centers (RCs). The TWA strategy involves assigning time windows to customers to enhance the operational efficiency of logistics networks. A biobjective mathematical model is established to minimize the total operating cost and number of vehicles for solving the multidepot recycling vehicle routing problem with RS and TWA (MRVRPRSTWA). A hybrid heuristic algorithm including 3D k -means clustering algorithm and nondominated sorting genetic algorithm- (NSGA-) II (NSGA-II) is designed. The 3D k -means clustering algorithm groups customers into clusters on the basis of their spatial and temporal distances to reduce the computational complexity in optimizing the multidepot logistics networks. In comparison with NSGA algorithm, the NSGA-II algorithm incorporates an elitist strategy, which can improve the computational speed and robustness. In this study, the performance of the NSGA-II algorithm is compared with the other two algorithms. Results show that the proposed algorithm is superior in solving MRVRPRSTWA. The proposed model and algorithm are applied to an empirical case study in Chongqing City, China, to test their applicability in real logistics operations. Four different scenarios regarding whether the RS and TWA strategies are included or not are developed to test the efficacy of the proposed methods. The results indicate that the RS and TWA strategies can optimize the recycling services and resource allocation and utilization and enhance the operational efficiency, thus promoting the sustainable development of the logistics industry.

1. Introduction

Sustainability is a constant topic in the logistics industry and research [1, 2]. In reverse logistics, recycling waste products can save resources and facilitate sustainable development [3]. Recycling vehicle routing problem (RVRP), one part of reverse logistics, often appears in logistics operations. For example, empty beer bottles are recycled at convenience stores. An increasing number of enterprises have begun to consider recycling products to improve the utilization of resources due to economic, legal, and social concerns [4, 5]. Therefore, the growing demand for recycling services will stimulate the development of reverse logistics. The design,

planning, and operation of sustainable reverse logistics systems pose a challenge to logistics organizations [6].

The low transportation resource utilization and irrational transport operations of logistics facilities still appear and need to be addressed [7, 8]. Resource sharing (RS) strategies have been proposed to optimize resource utilization and reduce operating cost in multidepot logistics networks [9]. Customers are reassigned to appropriate logistics facilities through these strategies to avoid unreasonable routing via customer information sharing [10]. Transportation resources are shared within and between logistics facilities to improve transportation resource utilization. Specifically, vehicles can be used to serve customers

of a single logistics facility or customers of multiple logistics facilities many times when the time windows of customers allow [11]. In this study, the RS strategy is developed in the multidepot RVRP with time window (MRVRPTW) to optimize the resource utilization.

In traditional vehicle routing problems with time window (VRPTW), the service time windows for customers are obtained by the logistics service providers after negotiation with customers or are determined by customers [12, 13]. Logistics service providers often recycle cargoes from customers at the same time every day [14]. In real life, the time windows for most customers are viewed as soft time windows, that is, they can be violated by paying certain penalties [15, 16]. Most logistics service providers try to recycle cargoes within the predetermined time windows in logistics operations to reduce the penalty cost [17]. However, the vehicles cannot provide recycling services for customers in the expected time windows due to long transportation distance and unreasonable service time windows for logistics facilities [18, 19]. Therefore, a set of candidate time windows are assigned to customers in this study, and recycling vehicle routes are designed on the basis of time window assignment (TWA) to optimize the operational efficiency and minimize the operating cost in multidepot RVRP with RS and TWA (MRVRPRSTWA).

In this study, the RS and TWA strategies are adopted to improve resource utilization, optimize resource allocation, and enhance the operational efficiency of multidepot recycling logistics networks. A bi-objective mathematical model is presented to obtain the minimum operating cost and the number of vehicles (TNV) in a multidepot logistics network with RS and TWA. A hybrid heuristic algorithm, including a 3D k -means clustering algorithm and non-dominated sorting genetic algorithm- (NSGA-) II (NSGA-II), is devised to optimize the recycling routes. The 3D k -means clustering algorithm groups customers on the basis of their spatial and temporal distances to reduce the computational complexity. NSGA-II is used to assign time windows to customers and find the optimal solutions in the MRVRPRSTWA strategy. A real-world case study is performed to verify the applicability of MRVRPRSTWA in logistics operations. The results confirm the capability of the RS and TWA strategies in optimizing multidepot logistics networks.

The rest of this paper is organized as follows. In Section 2, relevant literature is reviewed. In Section 3, MRVRPRSTWA is described in detail. In Section 4, a multiobjective mathematical model is formulated to minimize the total operating cost and TNV. In Section 5, a hybrid heuristic algorithm is introduced to solve MRVRPRSTWA. In Section 6, a real-world case is presented to verify the applicability of the proposed method. In Section 7, the conclusions are drawn, and future research directions are proposed.

2. Literature Review

With the increasing attention on environmental protection, recycling and related problems have received great concern.

The RVRP with time window (RVRPTW) as the main form of reverse logistics has been widely studied to reduce the operating cost in recent years [20]. The multidepot RVRPTW (MRVRPTW) is developed to coordinate the recycling services among logistics facilities [21], and the RS strategy can be considered to optimize the resource utilization in MRVRPTW (MRVRPRSTW) [22]. Candidate time windows are assigned to customers via the TWA strategy to improve operational efficiency compared with the traditional time window in MRVRPRSTW [23]. MRVRPRSTW with TWA (MRVRPRSTWA) investigates the RS and TWA strategies to optimize the resource allocation and operational efficiency in multidepot logistics networks.

Increasing interest in recycled products and materials has led to a growing concern in recycling logistics [24, 25]. El korch and Millet [26] investigated a framework, which used to generate and assess the reverse logistics channel structure, and proposed an environment-friendly structure with high economic benefits to optimize recycling services. Chan et al. [27] proposed a framework of reverse logistics on the basis of the value of reversed products, cost issue, and legislation perspective to optimize reverse logistics. Bai and Sarkis [28] introduced a flexible framework for reverse logistics and a performance measurement of third-party reverse logistics service provider model on the basis of a novel neighborhood rough set method to set the foundation for future research of reverse logistics. Roghanian and Pazhoheshfar [29] presented a probabilistic mixed liner programming model and genetic algorithm to minimize the total cost in reverse logistics. Shaik and Abdul-Kader [30] proposed a multi-criteria performance measurement model considering product lifecycle stages, strategies, capabilities, processes, perspectives, and measures to verify and enhance the enterprise's performance in reverse logistics.

The RS strategy, as a major issue, has drawn increasing attention on the enhanced resource utilization and reduced operating cost [31–33]. An et al. [34] constructed a network DEA model to calculate the cost of logistics network before and after RS and show the performance of the RS strategy. Sun et al. [35] analyzed and evaluated the effect of the RS strategy on alleviating or solving the traffic and environmental problems in megacities from five perspectives of resource, environment, convenience, economy, and governance. Quintero-Araujo et al. [36] investigated different collaborative scenarios on the basis of the RS strategy to obtain a lower logistics cost. The RS strategy is developed to minimize TNV and the total operating cost for improving the efficiency and sustainability of logistics networks [37]. Molina et al. [38] established a mathematical model that aims at maximizing the number of served customers and minimizing the total cost in the case of RS. Wang et al. [39] studied the RS strategy in multiple service periods and multidepot to enhance resource efficiency and refine the resource configuration.

Clustering algorithms are always used to group customers before determining vehicle routes in MDVRPTW to simplify the complexity of calculation [40, 41]. Min et al. [42] proposed the maximum-minimum distance clustering method to cluster customers in split-delivery VRP for the

better performance of the algorithm. Ge et al. [43] added the service radius and load expansion factors to the clustering algorithm to avoid the vehicle overloading in MDVRPTW. Fan et al. [44] introduced a clustering algorithm based on the temporal-spatial distance to reduce the computational complexity and enhance the quality of initial solution in MDVRPTW. Liu et al. [45] investigated the clustering algorithm that groups customers in terms of the minimum distance and maximum demand to improve the efficiency of solving MDVRPTW.

Several algorithms are studied to improve the quality of solution in MRVRPTW [46–48]. Aras et al. [21] established two mixed-integer linear programming models and proposed a Tabu-search-based heuristic algorithm to solve MRVRPTW. Kim et al. [49] developed a two-stage solution procedure to minimize the total distance and cost in MRVRPTW. Liu and He [50] proposed a clustering-based multiple ant colony system algorithm to minimize the total cost for enhancing the route compactness in MRVRPTW. Ramos and Oliveira [51] developed a mathematical model to minimize the variable costs and attain equity, and designed a heuristic algorithm to address MRVRPTW. Liao [4] established the genetic mixed-integer nonlinear programming model to maximize total profit and designed a hybrid genetic algorithm to solve MRVRPTW. Govindan et al. [52] proposed a mathematical model and a fuzzy solution method to minimize the total cost in MRVRPTW.

TWA vehicle routing problem (TWAVRP) differs from the traditional VRPTW, where the service time windows are determined by customers. In TWAVRP, candidate time windows are assigned to customers to improve the operational efficiency of logistics networks [53, 54]. Subramanyam et al. [55] introduced a scenario decomposition algorithm to solve TWAVRP for minimizing the expected routing cost of logistics facility. Neves-Moreira et al. [56] presented a two-stage stochastic optimization problem of TWAVRP and addressed it by using a fix-and-optimize-based meta-heuristic. Jalilvand et al. [57] developed a stochastic model and a progressive hedging algorithm to minimize the routing cost. The results show that the efficiency of the proposed method has good performance in addressing TWAVRP. A mathematics model is proposed to enhance the operational efficiency and reduce the total cost, and a hybrid heuristic algorithm is used for TWAVRP [58].

The aforementioned literature covers many aspects of MRVRPRSTWA but has the following limitations: (1) The RS strategy is rarely considered between and within logistics facilities. (2) The TWA strategy is insufficiently discussed in multidepot logistics network optimization. (3) An efficient mathematical model for optimizing TNV in the presence of RS strategy is lacking. (4) The traditional evolutionary algorithms have limited applicability when solving the MRVRPRSTWA.

The contributions of this paper are as follows: (1) The RS strategy is incorporated in the MRVRPRSTWA to maximize resource utilization. (2) The TWA strategy is developed in the MRVRPRSTWA to improve the operational efficiency. (3) A bi-objective integer programming model is proposed to minimize the total operating cost and TNV in multidepot

logistics networks with shared resources. (4) A hybrid heuristic algorithm, including 3D k -means clustering and NSGA-II, is developed to solve MRVRPRSTWA.

3. Problem Statement

The RS and TWA strategies are proposed in MRVRPRSTWA to optimize the recycling service of the multidepot logistics network. The multidepot logistics network is composed of multiple RCs and customers. Long-distance and crisscross trips often exist in the initial non-optimal logistics network (Figure 1). The transportation resources are shared among logistics facilities, and customers with irrational time windows are reassigned new time windows to adjust optimized recycling routes through the TWA strategy. The comparison of the multidepot logistics network before and after optimization is shown in Figures 1 and 2.

In Figure 1, each RC operates independently in the nonoptimal logistics network. Several customers are served by the farther RC rather than the closer one. For example, customer 21 is closer to RC3, while it is served by RC1, resulting in long-distance and crisscross transportation. In addition, RCs can only provide on-time service for two-thirds of customers, and the time windows for the remaining third of customers cannot be met. Relevant measures must be taken to enhance the operational efficiency and optimize the configuration of logistics networks. In Figure 2, centralized transportation is performed among RCs by a fleet of semitrailer trucks and is used to merge and transship goods between RCs, considering changes in customer demand assignment. Resources are shared among RCs, and customers are reassigned to the corresponding RC based on geographical locations and time windows. Irrational transportation phenomenon is reduced, and vehicles are shared among RCs. For example, due to the difference of service time between different routes, V2 is shared between RC1 and RC2, which first serve the customers of RC2 and then the customers of RC1, and thus the vehicle utilization rate is improved and the logistics operating cost is reduced. The violations of time windows are decreased when the TWA strategy is adopted. The RS and TWA strategies call for improving the utilization of transportation resources and operational efficiency and optimizing multidepot logistics networks.

This study assumes that the centralized transportation cost (TC) among logistics facilities is \$22 per unit time, and the recycling cost from logistics facilities to customers is \$20 per unit time. The penalty cost for earliness and delay service is \$20 per unit time. The assignment cost (AC) from the initial time window to assigned time (AT) window is \$5 per unit time. The maintenance cost (MC) of each semitrailer truck is \$300, and that of each vehicle is \$200. The AT, TNV, the number of semitrailer trucks (TNS), and the total cost are compared and listed in Table 1.

In Table 1, the total cost of reverse logistics network is decreased from \$4640 to \$4170. In addition, TNV is reduced from nine to five when sharing resources among logistics facilities. When the TWA strategy was applied, the penalty

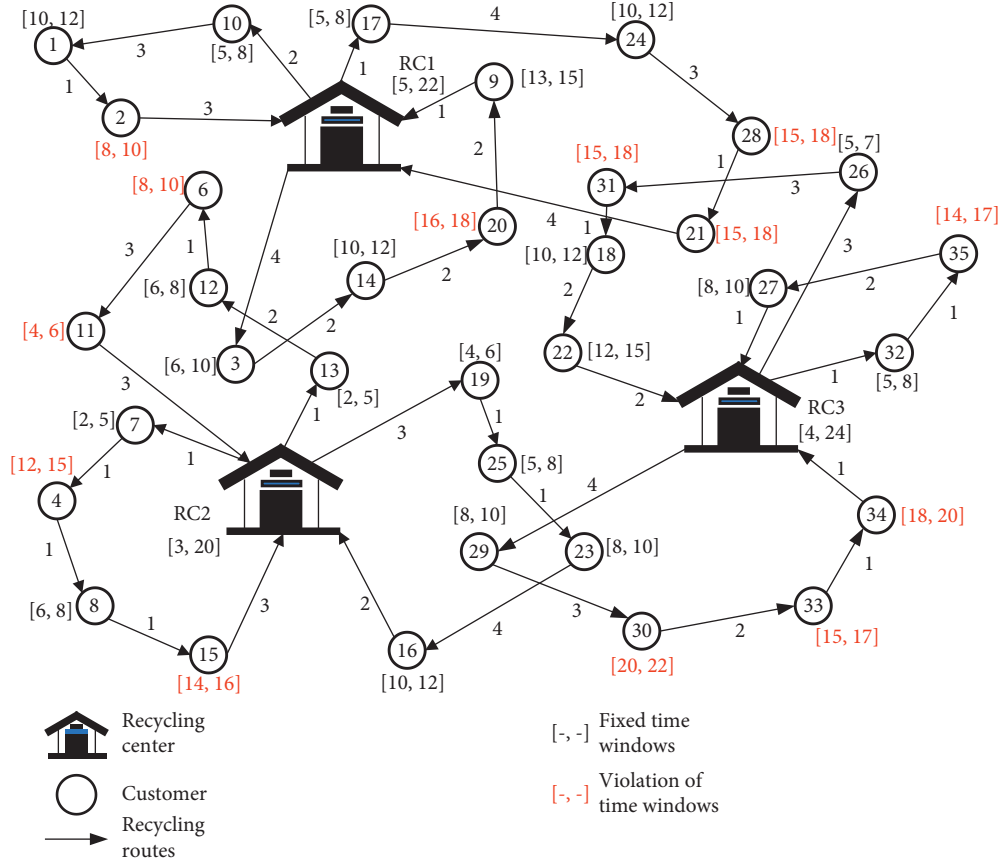


FIGURE 1: Logistics network without the RS and TWA strategies.

cost decreases by 63%. The RS and TWA strategies can improve the resource utilization and operational efficiency and optimize multi-depot logistics networks.

4. Model Formulation for MRVRPTWA

4.1. Related Definitions and Variables. In this section, the related notations and definitions on sets, parameters, and variables are shown in Table 2.

4.2. Mathematical Model. A bi-objective optimization model for MRVRPRSTWA is formulated to minimize the total operating cost in equation (1) and minimize TNV in equation (2).

$$\text{Min TC} = \text{TC}_1 + \text{TC}_2 + \text{TC}_3, \quad (1)$$

$$\text{Min MNV} = \sum_{v \in V} \min \left\{ \sum_{i \in P} \sum_{j \in C} x_{ijv}, 1 \right\}. \quad (2)$$

TC_1 contains three components in equation (3): $\sum_{m \in P} \sum_{n \in P} \sum_{s \in S} (U_s \times D_{mn} \times W \times y_{mns})$ represents the centralized TCs by semitrailer trucks among logistics facilities. $\max \left\{ \sum_{n \in P} (t_{q_{mn}} / Q_s) \right\} \times (M_s / B)$ is the MC of semitrailer trucks. $\sum_{m \in P} \text{FC}_m$ represents the fixed cost of logistics facilities.

$$\begin{aligned} \text{TC}_1 = & \sum_{m \in P} \sum_{n \in P} \sum_{s \in S} (U_s \times D_{mn} \times W \times y_{mns}) \\ & + \max_{m \in P} \left\{ \sum_{n \in P} \frac{t_{q_{mn}}}{Q_s} \right\} \times \frac{M_s}{B} + \sum_{m \in P} \text{FL}_m. \end{aligned} \quad (3)$$

TC_2 contains two components in equation (4): $\sum_{i \in \text{PUC}} \sum_{j \in \text{PUC}} \sum_{v \in V} (U_v \times D_{ij} \times W \times x_{ijv})$ is the TC by vehicles from logistics facilities to customers. $\sum_{v \in V} \min \left\{ \sum_{i \in P} \sum_{j \in C} x_{ijv}, 1 \right\} \times (M_v / B)$ represents the MC of vehicles.

$$\begin{aligned} \text{TC}_2 = & \sum_{i \in \text{PUC}} \sum_{j \in \text{PUC}} \sum_{v \in V} (U_v \times D_{ij} \times W \times x_{ijv}) \\ & + \sum_{v \in V} \min \left\{ \sum_{i \in P} \sum_{j \in C} x_{ijv}, 1 \right\} \times \frac{M_v}{B}. \end{aligned} \quad (4)$$

TC_3 contains five components in equation (5): $\sum_{i \in \text{PUC}_f} \sum_{j \in \text{C}_f} \sum_{v \in V} \sum_{k \in \text{O}_v} x_{ijv} \times \mu_e \times [\max\{e_i - at_{iv}^k, 0\}]$ and $\sum_{i \in \text{PUC}_a} \sum_{j \in \text{C}_a} \sum_{v \in V} \sum_{k \in \text{O}_v} x_{ijv} \times \mu_e \times [\max\{\alpha_i - at_{iv}^k, 0\}]$ represent the penalty cost caused by vehicles that provide earlier service than time windows of customers. $\sum_{i \in \text{PUC}_a} \sum_{j \in \text{C}_a} \sum_{v \in V} \sum_{k \in \text{O}_v} x_{ijv} \times \mu_d \times [\max\{at_{iv}^k - l_i, 0\}]$ and $\sum_{i \in \text{PUC}_a} \sum_{j \in \text{C}_a} \sum_{v \in V} \sum_{k \in \text{O}_v} x_{ijv} \times \mu_d \times [\max\{at_{iv}^k - \beta_i, 0\}]$ represent the

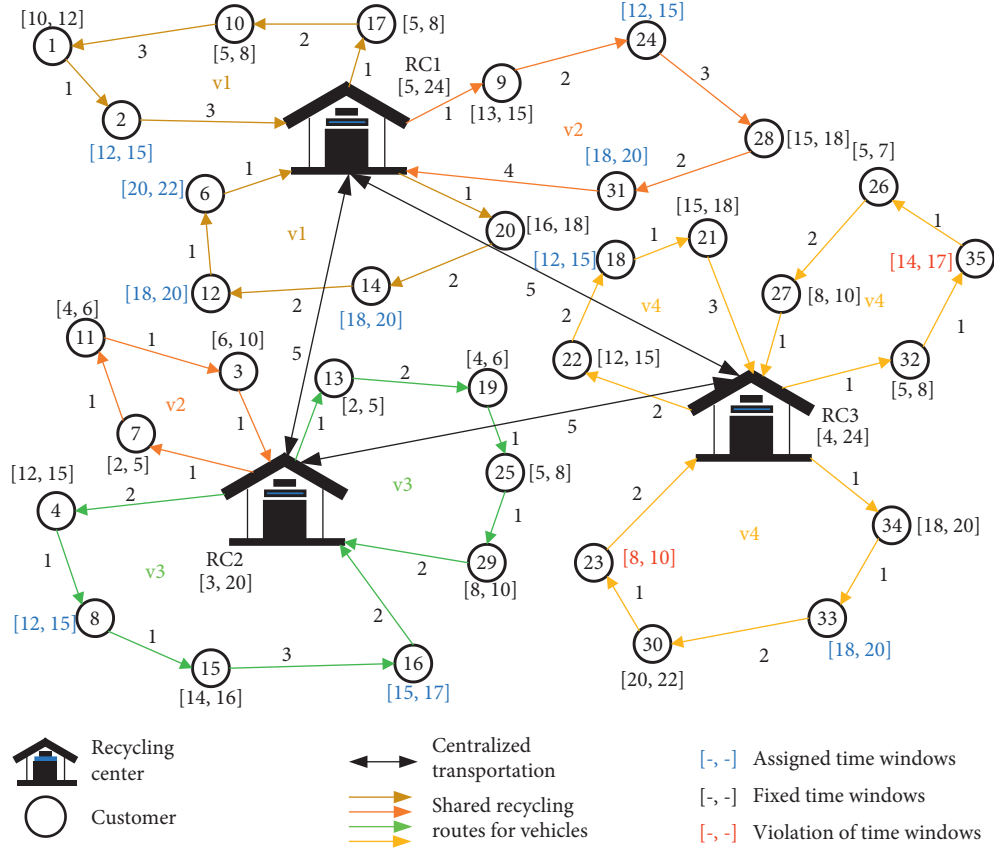


FIGURE 2: Logistics network with the RS and TWA strategies.

TABLE 1: Comparison of before and after RS and TWA.

Case	TC (\$)	Recycling cost (\$)	Penalty cost (\$)	AT	AC (\$)	TNV	TNS	MC (\$)	Total cost (\$)
Nonoptimal logistics network	—	1760	1080	—	—	9	—	1800	4640
Optimal logistics with RS and TWA	990	1400	400	56	280	4	1	1100	4170

penalty cost caused by vehicles that provide delayed service than time windows of customers. $\sum_{i \in C_a} \mu_a \times [\min\{|e'_i - \alpha_i|, |\beta_i - l'_i|\}]$ denotes the cost of TWA.

$$\begin{aligned}
 TC_3 = & \sum_{i \in PUC_f} \sum_{j \in C_f} \sum_{v \in V} \sum_{k \in O_v} x_{ijv} \times \mu_e \times [\max\{e_i - at_{iv}^k, 0\}] \\
 & + \sum_{i \in PUC_f} \sum_{j \in C_f} \sum_{v \in V} \sum_{k \in O_v} x_{ijv} \times \mu_d \times [\max\{at_{iv}^k - l_i, 0\}] \\
 & + \sum_{i \in PUC_a} \sum_{j \in C_a} \sum_{v \in V} \sum_{k \in O_v} x_{ijv} \times \mu_e \times [\max\{\alpha_i - at_{iv}^k, 0\}] \\
 & + \sum_{i \in PUC_a} \sum_{j \in C_a} \sum_{v \in V} \sum_{k \in O_v} x_{ijv} \times \mu_d \times [\max\{at_{iv}^k - \beta_i, 0\}] \\
 & + \sum_{i \in C_a} \mu_a \times [\min\{|e'_i - \alpha_i|, |\beta_i - l'_i|\}].
 \end{aligned} \tag{5}$$

TABLE 2: Notations and definitions in the MRVRPRSTWA.

Set definition	
P	Set of recycling centers (RCs), $P = \{p p = 1, 2, 3, \dots, m\}$ and m is the total number of RCs
C	Set of all recycling customers, $C = \{c c = 1, 2, 3, \dots, h\}$ and h is the total number of customers
C_f	Set of customers with fixed time windows
C_a	Set of customers with assigned time windows
V	Set of vehicles for recycling, $V = \{v v = 1, 2, 3, \dots, b\}$ and b is the total number of vehicles
S	Set of semitrailer trucks used between RCs, $S = \{s s = 1, 2, 3, \dots, w\}$ and w is the total number of vehicles
O_v	Maximum times for executing recycling routes of vehicle v in a working period, $v \in V$
Parameters	
q_i	Recycling demand quantity of customer i , $i \in C$
tq_{mn}	Transportation quantity from logistics facility m to n , $m, n \in P$
Q_s	Maximum capacity of semitrailer truck s , $s \in S$
Q_v	Maximum capacity of vehicle v , $v \in V$
Q_m	Maximum capacity of logistics facility m , $m \in P$
U_s	Usage cost of semitrailer truck s , $s \in S$ (unit: dollar/km)
U_v	Usage cost of vehicle v , $v \in V$ (unit: dollar/km)
D_{ij}	Distance from logistics facility or customer i to j , $i, j \in P \cup C$, $i \neq j$ (unit: km)
D_{mn}	Distance from logistics facility m to logistics facility n , $m, n \in P$ (unit: km)
W	Number of working days in one planning period
B	Number of planning periods in one year
M_s	Annual MC of semitrailer truck s , $s \in S$
M_v	Annual MC of vehicle v , $v \in V$
FL_m	Fixed cost of logistics facility m , $m \in P$
$[e_p, t_i]$	Fixed time window of customer i , $i \in C_f$
$[e^p, i, l^p, i]$	Expected time window of customer i , $i \in C_a$
$[\alpha_p, \beta_i]$	Time window assigned to customer i , $i \in C_a$
$[E_m, L_m]$	Service time window of logistics facility m , $m \in P$
μ_e	Penalty cost for early arrival per unit time
μ_d	Penalty cost for delayed arrival per unit time
μ_a	Cost coefficient of the customer's time window from the expected time window to the assigned time window per unit time
dtk_{vm}	Departure time of the k th route of vehicle V from logistics facility m , $v \in V$, $m \in P$, $k \in O_v$
atk_{iv}	Arrival time of the k th route of vehicle v at node i , $v \in V$, $i \in P \cup C$, $k \in O_v$
t_{ijv}	Travel time of vehicle v between entities i and j , $i, j \in P \cup C$, $v \in V$
$ N_s $	Total number of logistic facilities served by semitrailer truck s , $s \in S$
$ N_v $	Total number of customers served by vehicle v , $v \in V$
BN	Big number
Decision variables	
x_{ijv}	If vehicle v travels from logistics facility or customer i to j , then $x_{ijv} = 1$; otherwise, $x_{ijv} = 0$, $i, j \in P \cup C$, $v \in V$
zk_v	If vehicle v has the k th route, then $zk_v = 1$; otherwise, $zk_v = 0$, $m \in P$, $v \in V$, and $k \in O_v$
ω_{ivm}	If vehicle v departs from logistics facility m to served customer i , $\omega_{ivm} = 1$; otherwise, $\omega_{ivm} = 0$, $i \in C$, $v \in V$, and $m \in P$
y_{mns}	If semitrailer truck s transports between logistics facility m and n , $y_{mns} = 1$; otherwise, $y_{mns} = 0$, $m, n \in P$, and $s \in S$
τ_{imn}	If customer i 's logistic facility changes from logistics facility m to n after optimization, $\tau_{imn} = 1$; otherwise, $\tau_{imn} = 0$, $i \in C$, and $m, n \in P$
g_{ms}	If semitrailer truck s departs from logistics facility m , $g_{ms} = 1$; otherwise, $g_{ms} = 0$, $s \in S$, and $m \in P$

Subject to

$$\sum_{m \in P} y_{mns} = 1, \quad \forall n \in P, s \in S, m \neq n, \quad (6)$$

$$\sum_{n \in P} y_{nms} - \sum_{n \in P} y_{nfs} = 0, \quad \forall m, f \in P, m \neq f, s \in S, \quad (7)$$

$$\sum_{m \in P} y_{mns} = 1, \quad \forall n \in P, s \in S, m \neq n, \quad (8)$$

$$\sum_{m \in P} \sum_{n \in P} y_{mns} \times g_{ps} \leq |N_s| - 1, \quad \forall p \in P, s \in S, p \neq m, p \neq n, \quad (9)$$

$$tq_{mn} = \sum_{i \in C} \tau_{imn} \times q_i, \quad \forall m, n \in P, \quad (10)$$

$$\sum_{m \in P} \sum_{n \in P} tq_{mn} \times y_{mns} \leq Q_s, \quad \forall s \in S, \quad (11)$$

$$\sum_{i \in P} x_{ijv} = 1, \quad \forall j \in C, v \in V, \quad (12)$$

$$\sum_{i \in C} \sum_{j \in C} x_{ijv} = 1, \quad \forall v \in V, \quad (13)$$

$$\sum_{j \in P} x_{ijv} - \sum_{j \in P} x_{jfv} = 0, \quad \forall i, f \in P \cup C, i \neq f, v \in V, \quad (14)$$

$$\sum_{i \in P} x_{jiv} = 1, \quad \forall j \in C, v \in V, \quad (15)$$

$$\sum_{i \in C} \sum_{j \in C} x_{ijv} \leq |N_v| - 1, \quad \forall v \in V, \quad (16)$$

$$\sum_{j \in C} q_j \times x_{ijv} \times z_v^k \leq Q_v, \quad \forall i \in P \cup C, v \in V, k \in O_v, \quad (17)$$

$$\sum_{j \in C} \sum_{v \in V} q_j \times \omega_{ijvm} \leq Q_m, \quad \forall i \in P \cup C, m \in P, \quad (18)$$

$$dt_{vm}^k + t_{miv} - BN(1 - x_{miv}) \leq at_{iv}^k, \quad \forall m \in P, i \in C, v \in V, k \in O_v, \quad (19)$$

$$dt_{vm}^k + t_{miv} + BN(1 - x_{miv}) \geq at_{iv}^k, \quad \forall m \in P, i \in C, v \in V, k \in O_v, \quad (20)$$

$$at_{iv}^k + t_{ijv} - BN(1 - x_{ijv}) \leq at_{jv}^k, \quad \forall i \in C, j \in C \cup P, v \in V, k \in O_v, \quad (21)$$

$$at_{iv}^k + t_{ijv} + BN(1 - x_{ijv}) \geq at_{jv}^k, \quad \forall i \in C, j \in C \cup P, v \in V, k \in O_v, \quad (22)$$

$$E_m \times x_{miv} \times z_v^k \leq dt_{vm}^k \leq L_m \times x_{miv} \times z_v^k, \quad \forall m \in P, i \in C, v \in V, k \in O_v, \quad (23)$$

$$E_m \times x_{miv} \times z_v^k \leq at_{mv}^k \leq L_m \times x_{miv} \times z_v^k, \quad \forall m \in P, i \in C, v \in V, k \in O_v, \quad (24)$$

$$at_{mv}^k + t_{miv} - M(1 - z_v^{k+1}) \leq dt_{vm}^{k+1}, \quad \forall m, n \in P, v \in V, k \in O_v, \quad (25)$$

$$x_{ijv} = \{0, 1\}, \quad \forall i, j \in P \cup C, v \in V, \quad (26)$$

$$z_v^k = \{0, 1\}, \quad \forall v \in V, k \in O_v, \quad (27)$$

$$\omega_{ivm} = \{0, 1\}, \quad \forall i \in C, v \in V, m \in P, \quad (28)$$

$$y_{mns} = \{0, 1\}, \quad \forall m, n \in P, \forall s \in S, \quad (29)$$

$$\tau_{imn} = \{0, 1\}, \quad \forall i \in C, \forall m, n \in P, m \neq n, \quad (30)$$

$$g_{ms} = \{0, 1\}, \quad \forall m \in P, s \in S. \quad (31)$$

Constraint (6) ensures that each semitrailer truck originally departs from logistics facilities. Constraint (7) is the flow conservation on each logistics facility. Constraint (8) ensures that each semitrailer truck finally returns to logistics facilities. Constraint (9) is used to avoid subtours for semitrailer trucks. Constraint (10) calculates the transportation quantity among logistics facilities. Constraint (11) stipulates that each semitrailer truck has enough capacity to satisfy the transportation quantity of the assigned logistics facilities. Constraint (12) ensures that each vehicle departs from logistics facilities only. Constraint (13) ensures that each customer can be served once by one vehicle. Constraint (14) is the flow conservation on each customer. Constraint (15) ensures that each vehicle finally returns to logistics facilities. Constraint (16) is used to avoid subtours for vehicles. Constraint (17) represents that each vehicle has enough capacity to satisfy the total demand of assigned customers. Constraint (18) ensures that the total demand of customers served by a logistics facility should be beyond its capacity. Constraints (19) and (20) guarantee the continuous departure time of vehicles at logistics facilities. Constraints (21) and (22) guarantee the continuous arrival time of vehicles at customers. Constraints (23) and (24) guarantee the departure and return times of vehicles at logistics facilities. Constraint (25) ensures the continuous departure time of shared routes of each vehicle. Constraints (26)–(31) indicate the binary restrictions on decision variables.

5. Solution Procedure

The hybrid heuristic algorithm is widely applied to solve multiobjective optimization problems [59–61]. This algorithm, including the 3D k -means clustering and NSGA-II algorithms, is proposed to solve the MRVRPRSTWA. The 3D k -means clustering algorithm is often utilized to construct clusters according to the spatial and temporal distances of data [53]. The 3D k -means clustering algorithm can effectively reduce the computational complexity of multi-depot logistics networks [37]. The NSGA-II can assign appropriate time windows to customers and obtain Pareto optimal solutions in MRVRPRSTWA. The elitist strategy in NSGA-II can effectively avoid the loss of the best individual and enhance the computational speed and robustness. The

elitist strategy is composed of the genetic operations (i.e., selection, crossover, and mutation) and the fast non-dominated sorting and crowding distance. The parameters utilized in the flow structure are defined as follows: i is the number of RCs and clustering centers, R is set as the present number of optimization runs, $Rmax$ represents the maximum number of optimization runs, N is set as the present number of generations, and $Nmax$ is set as the maximum number of generations. The flow structure of the hybrid heuristic algorithm is shown in Figure 3.

The optimization procedure of MRVRPRSTWA is as follows:

Step 1: 3D k -means clustering. Set the number of RCs as the number of clusters and the data from RCs as the initial data of each cluster. Calculate the spatial and temporal distances between cluster centers and customers. Assign each customer to the nearest cluster center.

Step 2: the NSGA-II algorithm is used to establish pickup vehicle routes. The initial parameters are set and the initial population is randomly generated.

Step 3: genetic operation, including selection, partial-mapped crossover (PMC), and mutation is performed to generate offspring population. The parent and offspring population are combined.

Step 4: the new population is selected by performing nondominated sorting and calculating the crowding distance of each individual.

Step 5: determine whether the internal termination is satisfied. If the termination condition is unsatisfied, then select customers that accepted the TWA strategy and have time window violations, and assign the appropriate time windows to them; otherwise, return to Step 3.

Step 6: calculate the total cost and TNV with the adoption of the TWA strategy by NSGA-II and determine whether the termination condition is satisfied. If the termination condition is unsatisfied, then return to Step 1 to confirm whether the clustering results need to be adjusted; otherwise, terminate the algorithm procedure and find the optimal solution.

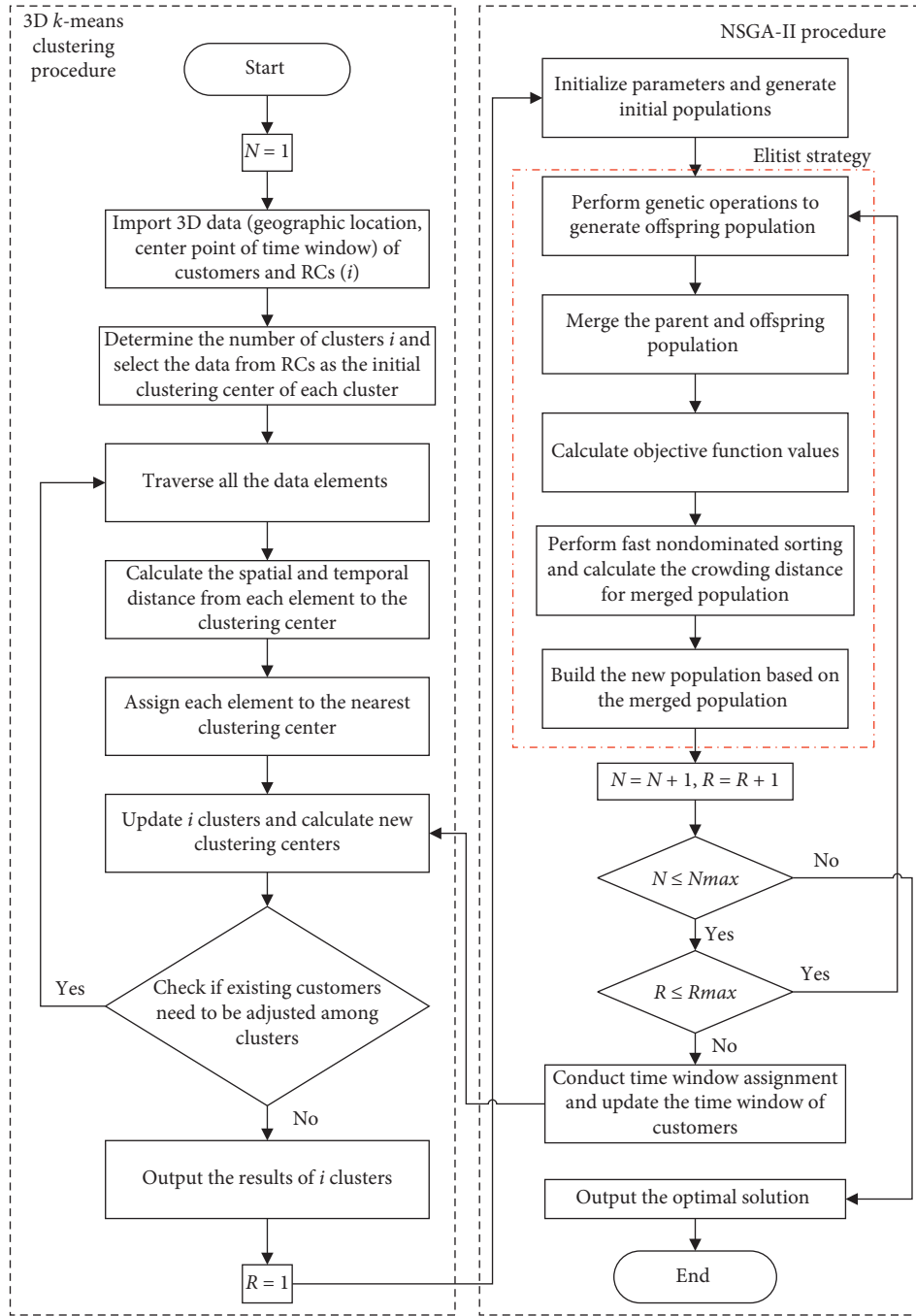


FIGURE 3: Algorithm flow structure for MRVRPRSTWA.

5.1. 3D *k*-Means Clustering. The clustering algorithm is always utilized to construct clusters to each depot in multi-depot VRP to reduce computational complexity [62, 63]. The traditional *k*-means clustering algorithm groups customers on the basis of the spatial distances between customers and clustering centers [64, 65]. The 3D *k*-means clustering algorithm, a variant of the traditional *k*-means clustering algorithm, constructs clusters according to the spatial and temporal distances among clustering centers and customers. In this study, the 3D *k*-means clustering algorithm is applied to optimize MRVRPRSTWA. The spatial and temporal

distances based on geographic coordinates and time windows of customers and logistics facilities can be calculated by

$$\beta_1 + \beta_2 = 1, \quad (32)$$

$$\text{dis} = \beta_1 \left(|x_m - x_p| + |y_m - y_p| \right) + \beta_2 |z_m - z_p|. \quad (33)$$

In equation (32), β_1 and β_2 represent the proportion of spatial and temporal distances, respectively. In equation (33), (x_m, y_m, z_m) expresses the geographic location and center point of time windows of logistics facility. (x_p, y_p, z_p)

expresses the geographic location and center point of time windows of customer. The 3D k -means clustering algorithm procedure is shown in Algorithm 1.

In Algorithm 1, the process of clustering can be divided into several components: First, the geographic locations and center point of the time windows of RCs and customers are used to construct the initial 3D data elements. Second, set i on the basis of the number of RCs as the number of clusters and the data from RCs as the initial clustering center of each cluster. Third, traverse data elements and calculate the spatial and temporal distances from each element to every clustering center. Fourth, assign each element to the nearest clustering center, and elements assigned to the same clustering center are placed in one cluster. Fifth, update i clusters, and continue to execute this procedure till no existing customers need to be adjusted among clusters; then, i clusters, including RCs and their served customers, are reported. The clustering results provide the initial population for multidepot vehicle routing optimization.

5.2. NSGA-II. Heuristic algorithms have obvious performances in solving multidepot logistics networks [66]. The NSGA-II is an enhanced algorithm based on the NSGA to solve multiobjective evolutionary problems [67, 68]. NSGA-II ranks different individuals via nondominated sorting algorithm and crowding distance comparison operators to force individuals to converge toward and diffuse along the Pareto front. The main procedures of NSGA-II are summarized below.

5.2.1. Selection. The selection operation involves choosing certain individuals every time from the parent population with selection probability sp , and the best individual will be inserted into the offspring population. The selection operation will continue until the offspring population size reaches a predetermined size. The specific steps can be described as follows: First, individuals are chosen from the parent population with selection probability sp at random. Then, the individual with the best fitness values is selected and added into the offspring population. The above-mentioned procedures are repeated until the offspring population reaches a predetermined size.

5.2.2. PMC. PMC is used as the operator of the crossover operation [69]. The specific procedure of PMC is shown in Figure 4, and the main steps of PMC are as follows: First, randomly select the beginning and ending positions of a gene sequence of a pair of chromosomes. Second, the two sets of genes selected on the two chromosomes are exchanged separately. Third, conduct conflict detection and establish a mapping relationship on the exchanged genes. Fourth, repeat this step until no duplicate genes are present on the same chromosome.

The mapping relationship is illustrated in Figure 4. For example, a mapping relationship of 3-1 and 5-6 genes is presented. Two genes 3 in offspring 1 are selected in the second step, and the unselected gene is transformed into 1.

Two genes 5 in offspring 1 are selected in the third step, and the unselected gene is transformed into 6. Finally, two children with no conflict genes are generated.

5.2.3. Mutation. Mutation operation aims to change the values of a gene at a certain locus of chromosome to generate a new chromosome [70]. Polynomial mutation is used in this work to prevent local convergence in the course of evolution. The main steps of polynomial mutation are shown in Figure 5.

In Figure 5, a chromosome is randomly selected from the parent generation in terms of the mutation probability pc . Then, three genes on the parental chromosome are randomly selected, and the exchange and regenerated operation are conducted on the three genes. If duplicate genes appear, then they should be removed and regenerated.

5.2.4. Nondominated Sorting and Crowding Distance Assignment. Several definitions are provided to describe the procedures of nondominated sorting and crowding distance. N represents the set of population, and it can be divided into several subsets (i.e., A_1, A_2, A_3, \dots). The subsets can be indicated as follows: A_1 is the set of nondominated population of N , A_2 is the set of nondominated population of $N - A_1$, A_3 is the set of nondominated population of $N - (A_1 \cup A_2)$, and so on. In each individual m , let $id(m)$ be the index of the front to which m belongs and $cd(m)$ be the crowding distance of m . The crowding distance can be calculated in equation (34) as follows:

$$cd(m) = \frac{f_1(m_1) - f_1(m_2)}{f_1^{\max} - f_1^{\min}} + \frac{f_2(m_3) - f_2(m_4)}{f_2^{\max} - f_2^{\min}}, \quad (34)$$

where m_1 and m_2 are the closest populations of m in the same front on the basis of f_1 ; m_3 and m_4 are the closest populations of m in the same front on the basis of f_2 ; and f_1^{\max} and f_1^{\min} are the maximum and minimum values of f_1 , respectively. In addition, if n is the boundary individual, then the crowding distance of n is infinity, namely, $cd(n) = \infty$. If the $id(m)$ is smaller than the $id(m')$ or $id(m)$ is equal to the $id(m')$, and the $cd(m)$ is larger than the $cd(m')$, then we can say m is preferred to m' . The next generation is selected through the above-mentioned method, which is nondominated sorting and crowding distance [71].

5.2.5. TWA Strategy. The TWA strategy is designed in the NSGA-II algorithm to assign candidate time windows to customers with improved operational efficiency. The main procedures are presented as follows:

Step 1: customers that accepted the TWA strategy with violation of time windows are selected.

Step 2: assume that the time window for customer A is $[E_A, L_A]$, and the actual service time of customer A is i hours earlier (or delayed) than E_A . If $E_A - i$ (or $L_A + i$) is between the candidate time window $[E' A, L' A]$, then calculate the TWA cost, and compare it with the penalty cost.

Input: Number of RCs, geographic coordinates, and time windows of customers and RCs
 Output: Clusters, including RCs and their assigned customers

- (1) Import the geographic locations and center point of the time windows of RCs and customers as initial data.
- (2) Establish 3D vectors on the basis of the initial data and construct 3D data elements.
- (3) Set i as the number of clusters on the basis of the number of RCs.
- (4) Select the data from RCs as the initial clustering center of each cluster.
- (5) Traverse all the data elements.
- (6) Calculate the spatial and temporal distances from each element to the clustering center.
- (7) Assign each element to the nearest clustering center.
- (8) Update i clusters, and calculate new clustering centers.
- (9) **If** existing customers need to be adjusted among clusters
- (10) Then return to step 5;
- (11) **Else**
- (12) End if
- (13) End for
- (14) Output the results of i clusters.

ALGORITHM 1: 3D k -means clustering algorithm.

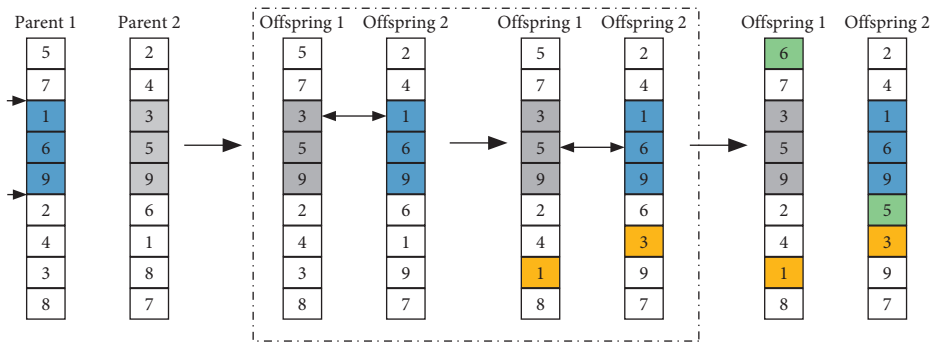


FIGURE 4: Procedure of PMC.

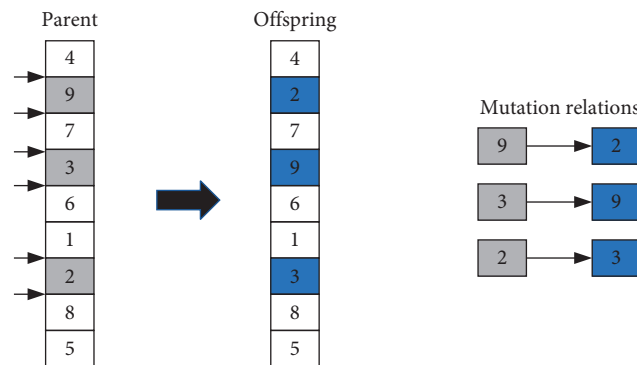


FIGURE 5: Procedure of mutation.

Step 3: if the TWA cost is lower than the penalty cost, then $[E' A, L' A]$ will assign to customer A ; otherwise, keep the initial time window of customer A .

Therefore, the appropriate time windows are assigned to the corresponding customers. The process of NSGA-II is described in Algorithm 2.

6. Implementation and Analysis

6.1. Algorithm Comparison. In this section, benchmark instances are performed to test the quality of NSGA-II in MRVVRPRSTWA for comparing its results with those received by multiobjective evolutionary algorithm (MOEA) [72] and multiobjective particle swarm optimization

Nondominated Sorting Algorithm-II (NSGA-II)

Input: pop_{size} , nodes, R , $Rmax$, $Nmax$, sp , cp , and mp

Output: Pareto front optimal solutions

- (1) Initialize parameters
- (2) # set the population size (pop_{size}), number of customers (nodes), number of generation (R), maximum number of generations ($Rmax$), maximum number of runs ($Nmax$), selection probability (sp), crossover probability (cp), and mutation probability (mp)
- (3) For $N = 1:Nmax$
- (4) For $R = 1:Rmax$
- (5) Generate the initial population with size pop_{size}
- (6) Objective function evaluation
- (7) # compute the objective function to minimize the total operating cost and TNV
- (8) Divide pop_{size} into nondominance front and calculate the crowding distance of each individual
- (9) For $i = 1: R$
- (10) Implement selection, PMC, and mutation operations to generate offspring population
- (11) Combine parent and offspring population, perform nondominated sorting and calculate the crowding distance of each individual
- (12) Build the new population on the basis of the parent and offspring populations following partial order
- (13) end
- (14) end
- (15) For $n = 1: nodes$
- (16) Select the customers that accept time window and have time window violation
- (17) Select the appropriate time window from candidate time windows to assign to customer
- (18) Calculate the TWA cost and compare it with the penalty cost
- (19) Assign the time window to the customer or keep the expected time window of customer
- (20) end
- (21) Find the Pareto front optimal solution
- (22) end

ALGORITHM 2: NSGA-II algorithm operation.

(MOPSO) [73]. The multidepot VRPTW (MDVRPTW) datasets obtained from the database of the NEO research group (<https://neo.lcc.uma.es/vrp/vrp-instances/multiple-depot-vrp-with-time-windows-instances/>) are illustrated in Table 3.

Table 3 presents the relevant characteristics of 20 benchmarks. The parameters for NSGA-II and MOEA are as follows: population size $pop_{size} = 150$, selection probability $sp = 0.6$, crossover probability $cp = 0.9$, and mutation probability $mp = 0.1$. The parameters for MOPSO are as follows: maximum iteration number $max_it = 200$, inertia weight $\omega = 0.9$, personal confidences $pc = 2$, and social learning confidences $gc = 3$. The results of the three algorithms, which contain total cost (cost), TNV (vehicle), and computation time (Ct), are compared in Table 4.

Table 4 exhibits an adequate difference among the optimal results of those three algorithms on the basis of the values of t -test and p -value. The NSGA-II algorithm is superior to MOEA and MOPSO in obtaining the optimal cost in MRVRPRSTWA. The costs of MOEA and MOPSO are higher than that of NSGA-II in 20 benchmark instances. For instance, the average cost of NSGA-II is \$15479. However, the costs of MOEA and MOPSO are \$20019 and \$17520, respectively. The minimum number of vehicles calculated by NSGA-II is nine compared to TNV of MOEA and MOPSO. The NSGA-II outperforms MOEA and MOPSO in terms of the minimum computation time. The average computation time of NSGA-II is 237s, which is lower than the computation times of MOEA and MOPSO.

Therefore, the NSGA-II algorithm can achieve better performance in solving MRVRPRSTWA compared to MOEA and MOPSO.

6.2. Data Source. A real reverse logistics network in Chongqing, China, is used to verify the applicability and feasibility of the proposed solution in MRVRPRSTWA. The logistics network has five RCs (i.e., RC1, RC2, RC3, RC4, and RC5) and 183 customers. The spatial distribution of logistics network is shown in Figure 6. The RCs and customers are marked as stars and rhombuses, respectively. Rhombuses in various colors represent customers served by different RCs. The initial vehicle routes are shown in Table 5.

In Table 5, customers are served by 21 vehicle routes among five RCs. Each RC is operated independently, which results in the waste of resources. For example, three customers (i.e., P180, P177, and P171) belonging to RC1 and two customers (i.e., P75 and P65) belonging to RC5 are served by a vehicle, respectively. The irrational vehicle routes should be decreased to enhance the operational efficiency of the multidepot logistics network.

6.3. Relevant Parameter Setting. The objective of MRVRPRSTWA is to obtain the lowest operating cost and TNV in the multidepot reverse logistics network. The related parameters used in the optimization model and the hybrid heuristic algorithm initialization are shown in Table 6. The time window of RC1 is [6, 18]. RC2 and RC3 have the same

TABLE 3: Relevant settings of data instances.

Instances	Datasets	Number of depots	Number of customers	Number of candidate time windows	Vehicle capacity
1	MDPR1	4	48	2	200
2	MDPR2	4	96	2	195
3	MDPR3	4	144	3	190
4	MDPR4	4	192	4	185
5	MDPR5	4	240	4	180
6	MDPR6	4	288	5	175
7	MDPR7	6	72	2	200
8	MDPR8	6	144	3	190
9	MDPR9	6	216	4	180
10	MDPR10	6	288	5	170
11	MDPR11	4	48	2	200
12	MDPR12	4	96	2	195
13	MDPR13	4	144	3	190
14	MDPR14	4	192	4	185
15	MDPR15	4	240	4	180
16	MDPR16	4	288	5	175
17	MDPR17	6	72	2	200
18	MDPR18	6	144	3	190
19	MDPR19	6	216	4	180
20	MDPR20	6	288	5	170

TABLE 4: Comparison of the results of the three algorithms on different instances.

Instances	NSGA-II			MOEA			MOPSO		
	Cost (\$)	Vehicle	Ct (s)	Cost (\$)	Vehicle	Ct (s)	Cost (\$)	Vehicle	Ct (s)
1	8450	3	184	12742	7	189	10689	5	179
2	9617	4	195	14906	9	221	12837	7	204
3	13716	8	210	16137	10	211	15067	9	215
4	15764	10	225	18940	13	243	19923	14	231
5	15829	10	231	21070	15	238	18041	12	244
6	17032	11	251	21412	15	267	19338	13	247
7	16952	9	250	20373	12	255	17325	9	261
8	20048	12	254	23601	15	248	18550	10	269
9	17968	10	266	25526	17	261	22484	14	268
10	19119	11	278	23500	15	287	19454	11	288
11	8335	3	207	14720	9	210	8673	3	217
12	10598	5	216	15840	10	231	14772	9	231
13	12743	7	224	19165	13	242	17094	11	233
14	15789	10	230	20110	14	257	18053	12	246
15	15929	10	237	21171	15	241	17119	11	251
16	18074	12	241	20260	14	246	19231	13	265
17	15932	8	251	21234	13	279	20173	12	271
18	18158	10	257	19411	11	264	19367	11	279
19	19998	12	262	24482	16	298	21461	13	278
20	19537	11	268	25779	17	317	20742	12	288
Average	15479	9	237	20019	13	250	17520	11	248
<i>t</i> -test				-12.955			-5.529		
<i>p</i> -value				3.51502E-11			1.24E-05		

time window [8, 20]. The time window of RC4 and RC5 is the same as [10, 22]. The candidate time windows are constructed on the basis of the real situation of customers that accepted TWA. Hence, four candidate time windows are constructed: [8, 10], [10, 11], [12, 15], and [16, 18].

6.4. Optimization Results. Customers are clustered on the basis of the 3D *k*-means clustering algorithm to reduce the computational complexity in the multidepot reverse logistics

network. Customers are reassigned to the RC with the nearest spatial and temporal distances. The results of the 3D *k*-means clustering algorithm are shown in Table 7 and Figure 7.

Table 7 illustrates that 183 customers are reassigned to the corresponding logistics facilities and grouped into five clusters. RC1 is allocated 28 customers, RC2 is allocated 28 customers, RC3 is allocated 42 customers, and RC4 and RC5 are allocated 42 and 43 customers, respectively. The spatial and temporal distances of the customers are shown in

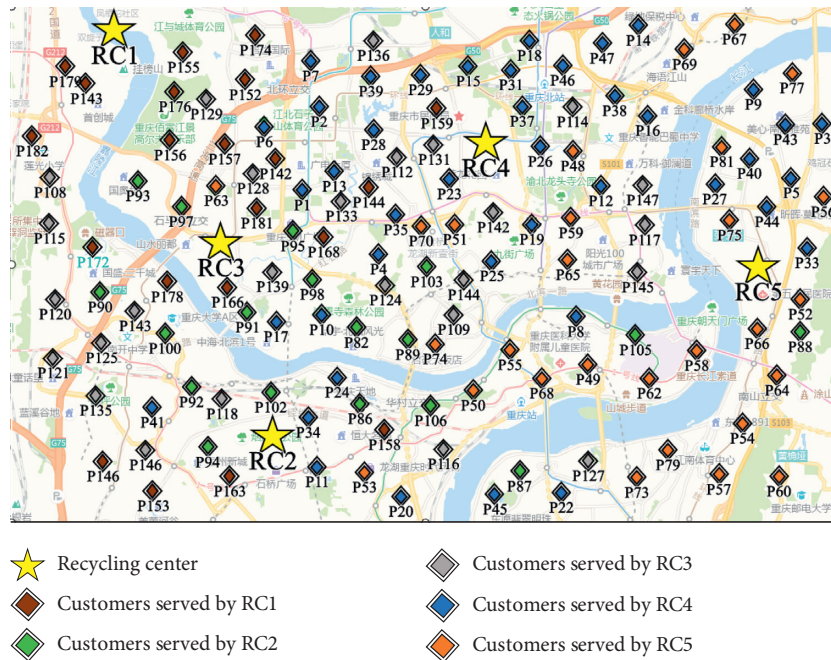


FIGURE 6: Spatial distribution of RCs and customers.

TABLE 5: Initial vehicle routes before optimization.

Facilities	Routes
RC1	RC1 → P175 → P174 → P167 → P157 → P163 → P172 → P169 → P176 → P162 → P179 → P155 → RC1
	RC1 → P180 → P177 → P171 → RC1
	RC1 → P150 → P159 → P152 → P165 → P168 → P161 → P170 → P154 → P181 → P158 → RC1
	RC1 → P182 → P164 → P156 → P183 → P160 → P153 → P166 → P151 → P178 → P173 → RC1
RC2	RC2 → P98 → P107 → P95 → P97 → P99 → P93 → P94 → P84 → P85 → P104 → P103 → RC2
	RC2 → P82 → P100 → P89 → P90 → P101 → P96 → P88 → P105 → P87 → P102 → P92 → RC2
	RC2 → P91 → P86 → P106 → P83 → RC2
RC3	RC3 → P122 → P121 → P113 → P114 → P115 → P108 → P131 → RC3
	RC3 → P124 → P149 → P112 → P147 → P140 → P144 → P141 → RC3
	RC3 → P126 → P143 → P120 → P145 → P123 → P125 → P128 → P139 → RC3
	RC3 → P148 → P137 → P110 → P138 → P118 → P134 → P136 → P142 → P132 → P129 → RC3
RC4	RC4 → P135 → P146 → P133 → P109 → P111 → P116 → P117 → P119 → P130 → P127 → RC3
	RC4 → P22 → P16 → P40 → P47 → R42
	RC4 → P14 → P5 → P9 → P3 → P36 → P33 → P18 → P12 → P30 → P11 → P4 → P35 → P6 → RC4
	RC4 → P15 → P7 → P39 → P28 → P37 → P38 → P23 → RC4
RC5	RC4 → P2 → P26 → P10 → P24 → P17 → P8 → P45 → P20 → P29 → P19 → P21 → RC4
	RC4 → P13 → P1 → P32 → P25 → P34 → P41 → P44 → P43 → P27 → P46 → P31 → RC4
	RC5 → P52 → P49 → P58 → P66 → P79 → P57 → P68 → P50 → P53 → P51 → P55 → P71 → RC5
	RC5 → P59 → P74 → P76 → P62 → P61 → P70 → P78 → P60 → P80 → P54 → P67 → P56 → RC5
	RC5 → P64 → P81 → P77 → P48 → P69 → P73 → P72 → P63 → RC5
	RC5 → P75 → P65 → RC5

Figure 7, wherein x represents the longitude of coordinate, y indicates the latitude of coordinate, and z denotes the mid value of the time window. The optimized vehicle routes with the RS and TWA strategies are shown in Table 8.

Table 8 exhibits that seven vehicles are shared on different routes to serve customers in the multidepot reverse logistics network. For example, vehicle 1 is shared in three routes. One route (e.g., RC1 → P128 → P138 → P7 → P69 → P78 → P15 → P36 → P42 → RC1) departs from RC1.

Meanwhile, two routes (e.g., RC3 → P1 → P122 → P50 → P95 → P109 → P51 → P41 → P129 → P13 → P166 → P158 → RC3, RC3 → P98 → P124 → P86 → P112 → P82 → P136 → P28 → P137 → P113 → P104 → RC3) depart from RC3. Vehicle sharing can improve the resource utilization and operational efficiency of logistics facilities and promote sustainable development among logistics networks. The comparison of results before and after optimization is shown in Table 9 and Figure 8.

TABLE 6: Parameter values utilized in the computational experiments.

Notation	Definition
β_1	Coefficient of spatial distance
β_2	Coefficient of temporal distance
Q_s	Maximum capacity of the semitrailer truck
Q_v	Maximum capacity of vehicle
U_s	Usage cost of the semitrailer truck (unit: dollar/km)
U_v	Usage cost of vehicle v (unit: dollar/km)
M_s	Annual MC of the semitrailer truck
M_v	Annual MC of the vehicle
μ_e	Penalty cost for early arrival per unit time
μ_d	Penalty cost for delayed arrival per unit time
μ_a	AC per unit time
W	Number of working days in one planning period
B	Number of planning periods in one year
FL_1	Fixed cost of RC1
FL_2	Fixed cost of RC2
FL_3	Fixed cost of RC3
FL_4	Fixed cost of RC4
FL_5	Fixed cost of RC5
Pop_{size}	Population size
N_{max}	Maximum number of generations runs
R_{max}	Maximum number of runs
sp	Selection probability
cp	Crossover probability
mp	Mutation probability

TABLE 7: 3D k -means clustering results in MRVRPRSTWA.

Facilities	RC1	RC2	RC3	RC4	RC5
			P1 P2 P4 P6 P10 P13	P8 P14 P16 P19 P20	P3 P5 P9 P12 P18 P27
			P26 P28 P29 P35 P41	P21 P22 P23 P31 P37	P30 P52 P54 P56 P57
	P7 P15 P39 P42 P48	P11 P17 P24 P25 P32	P50 P51 P60 P74 P76	P38 P45 P46 P47 P55	P58 P64 P65 P75 P77
	P69 P72 P73 P78 P93	P34 P53 P91 P92 P97	P82 P83 P86 P87 P95	P59 P63 P68 P80 P84	P79 P81 P88 P90 P96
	P99 P110 P118 P138	P102 P133 P135 P146	P98 P103 P104 P106	P85 P89 P94 P114 P123	P100 P101 P105 P108
Customers	P130 P131 P177 P180	P148 P160 P183 P115	P107 P109 P112 P113	P134 P139 P141 P142	P111 P125 P127 P140
	P182 P33 P36 P40 P43	P116 P117 P119 P120	P121 P122 P124 P129	P144 P145 P147 P151	P143 P150 P152 P157
	P44 P49 P126 P128	P61 P62 P70 P66 P67	P136 P137 P149 P153	P155 P161 P165 P168	P159 P162 P163 P167
	P132	P71	P154 P156 P158 P164	P170 P171 P173 P179	P169 P172 P174 P175
			P166	P181	P176 P178

In Table 9 and Figure 8, the total cost is reduced from \$38025 to \$21373 in the initial and optimized reverse logistics network on the basis of the RS and TWA strategies in the MRVRPRSTWA. In addition, the TC is decreased from \$4675 to \$1307. The sum of the penalty cost and AC after optimization is lower than that before optimization through the TWA strategy. The total number of vehicles is decreased to seven when the vehicles are shared among five RCs.

6.5. *Analysis and Discussion.* Four cases are considered to verify the applicability of the proposed method in MRVRPRSTWA. In Case 1, each logistics facility operates independently. In Case 2, the RS strategy is considered among logistics facilities. In Case 3, the TWA strategy is adopted among logistics facilities. In Case 4, the RS and TWA strategies are adopted among logistics facilities. The calculation results of the four scenarios are shown in Table 10 and Figure 9.

In Table 10 and Figure 9, the total cost, violated time, and TNV are significantly reduced in the optimized logistics network with the RS and TWA strategies. TNV in Case 2 is lower than that in Case 1, which shares vehicles among reverse logistics facilities. The violated time is significantly reduced when the TWA strategy is adopted among logistics facilities in Case 3 compared to that of Case 1. The total cost, violated time, and TNV of Case 4 have the minimum value of \$21373, 87 min, and 7, respectively. The total cost, violated time, and TNV are compared in Figure 9. Therefore, the proposed method in MRVRPRSTWA can effectively reallocate resources and improve the operational efficiency.

6.6. *Implications.* The RS and TWA strategies are considered to optimize recycling service in the multidepot reverse logistics network. The implications obtained from this work are described as follows:

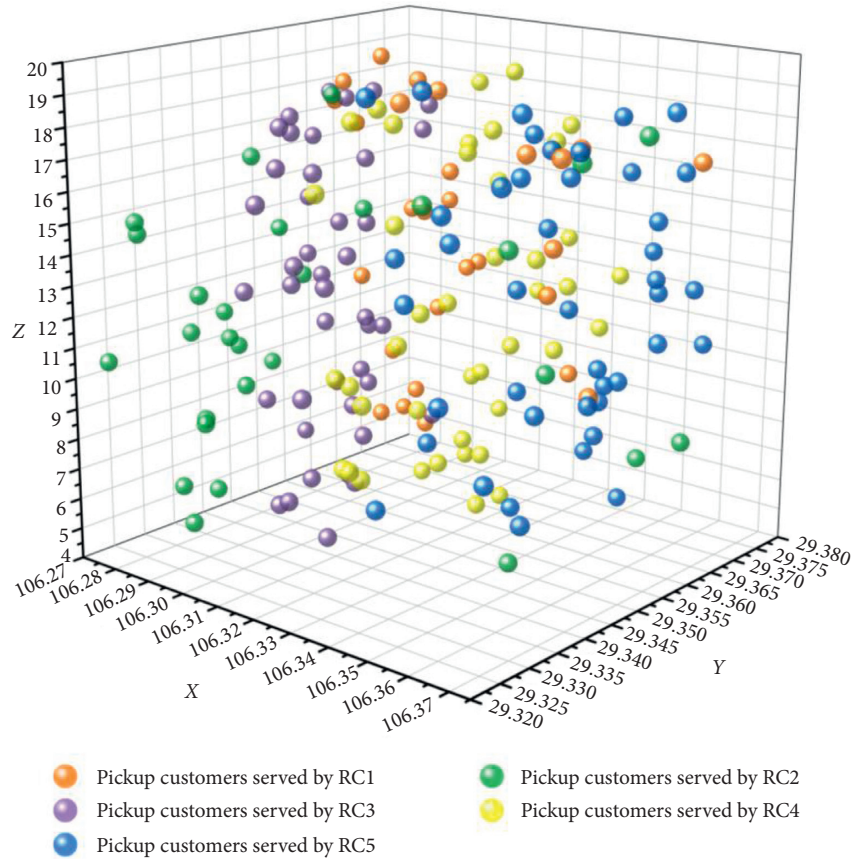


FIGURE 7: Customer clustering results in MRVRPRSTWA.

TABLE 8: Vehicle routes with the RS and TWA strategies.

Vehicles	Facilities	Routes
Vehicle 1	RC1	RC1 → P128 → P138 → P7 → P69 → P78 → P15 → P36 → P42 → RC1
	RC3	RC3 → P1 → P122 → P50 → P95 → P109 → P51 → P41 → P129 → P13 → P166 → P158 → RC3
	RC3	RC3 → P98 → P124 → P86 → P112 → P82 → P136 → P28 → P137 → P113 → P104 → RC3
Vehicle 2	RC2	RC2 → P133 → P11 → P160 → P17 → P71 → P115 → P34 → P148 → P120 → P32 → P91 → RC2
	RC2	RC2 → P102 → P24 → P25 → P146 → P119 → P183 → P70 → P9 → P97 → RC2
	RC2	RC2 → P61 → P62 → P135 → P116 → P53 → P117 → P66 → P67 → RC2
Vehicle 3	RC3	RC3 → P106 → P83 → P87 → P29 → P103 → P10 → P76 → P74 → P4 → P60 → P2 → P107 → RC3
	RC4	RC4 → P31 → P141 → P59 → P68 → P147 → P45 → P151 → P55 → P38 → P123 → P94 → RC4
	RC5	RC5 → P154 → P26 → P121 → P35 → P176 → P175 → P90 → P179 → P172 → RC5
Vehicle 4	RC4	RC4 → P14 → P114 → P37 → P84 → P85 → P134 → P173 → P156 → P181 → P178 → P168 → RC4
	RC4	RC4 → P63 → P21 → P20 → P145 → P19 → P6 → P89 → P22 → P47 → P161 → RC4
	RC4	RC4 → P144 → P165 → P8 → P46 → P139 → P142 → P16 → P80 → P155 → P23 → RC4
Vehicle 5	RC5	RC5 → P150 → P88 → P152 → P57 → P100 → P125 → P162 → P56 → P5 → P96 → P159 → RC5
	RC5	RC5 → P108 → P65 → P54 → P81 → P30 → P77 → P75 → P101 → P111 → RC5
Vehicle 6	RC1	RC1 → P39 → P48 → P93 → P180 → P131 → P44 → P118 → P177 → P110 → RC1
	RC1	RC1 → P72 → P99 → P73 → P182 → P132 → P126 → P49 → P33 → P130 → P43 → P40 → RC1
Vehicle 7	RC5	RC5 → P3 → P169 → P140 → P9 → P153 → P167 → P164 → P52 → P171 → P170 → P174 → RC5
	RC5	RC5 → P105 → P79 → P18 → P12 → P58 → P157 → P163 → P64 → P143 → P27 → P127 → RC5

TABLE 9: Comparison of the results before and after optimization.

Scenarios	Transportation cost (\$)	Penalty cost (\$)	Assignment cost (\$)	Fixed cost (\$)	Number of semitrailer trucks	Number of vehicles	Total cost (\$)
Before optimization	4675	6000	—	6350	—	21	38025
After optimization	1307	1566	2150	6350	2	7	21373

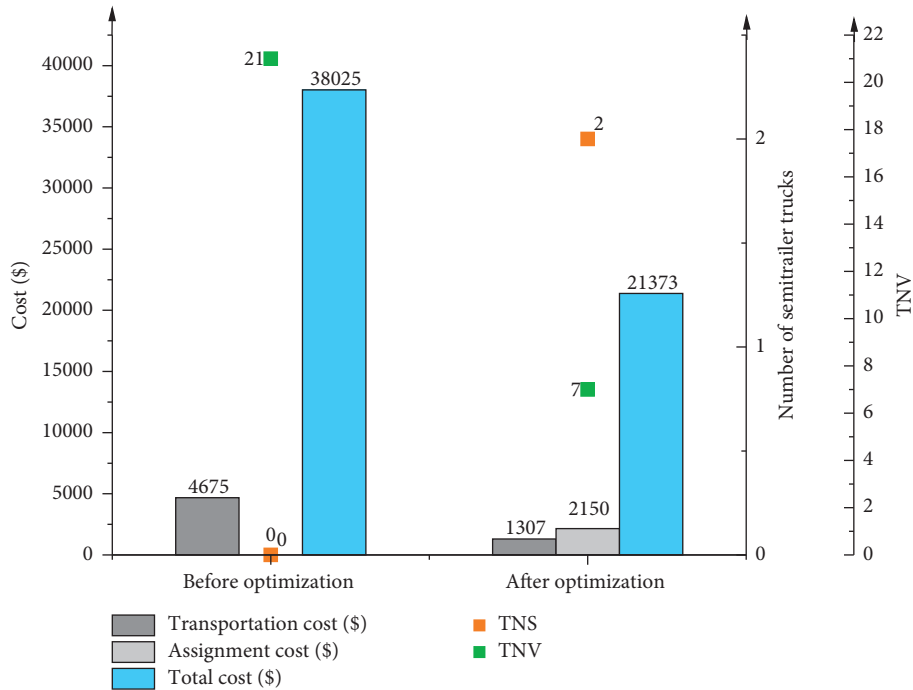


FIGURE 8: Comparison of the results before and after optimization.

TABLE 10: Comparison of the optimization results of the four scenarios.

Scenarios	The total cost (\$)	Violated time (min)	Assigned time (min)	The number of semitrailer trucks	The number of vehicles
Case 1	38025	333	—	—	21
Case 2	27939	188	—	2	12
Case 3	32616	107	247	2	17
Case 4	21373	87	215	2	7

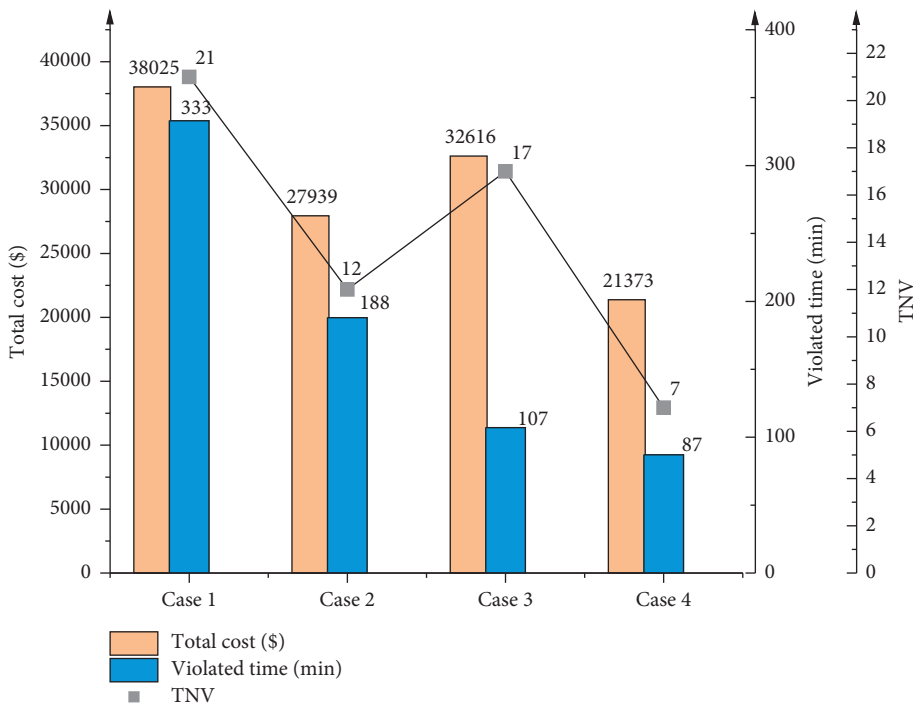


FIGURE 9: Comparison of the total cost, violated time, and TNV.

- (1) The RS strategy contributes to the resource utilization of reverse logistics network and promotes the sustainable development of the intelligent logistics system. Vehicles often serve customers on one route in a working period, thereby resulting in low vehicle utilization. Vehicles serve customers on several routes in a working period based on satisfying customer service time window with the adoption of the RS strategy. This strategy shares the resources within and between logistics facilities to obtain the maximum resource utilization in MRVRPRSTWA.
- (2) The TWA strategy can enhance the operational efficiency of logistics networks. The candidate time windows of customers with high penalty costs on account of the violation of time windows are assigned to corresponding customers through the TWA strategy to achieve the maximum operational efficiency in MRVRPRSTWA. The TWA strategy is of great theoretical and practical significance to the sustainable development of reverse logistics networks.
- (3) The RS and TWA strategies in reverse logistics network can reduce the negative impact of logistics enterprises on the environment, optimize resource allocation, maximize benefits, promote the development of green logistics, and provide related development decisions for logistics enterprises and government departments. The established mathematical model and the hybrid heuristic algorithm can also provide a reliable reference for the applicability and effectiveness of the RS and TWA strategies in addressing MRVRPRSTWA.

7. Conclusions

This work studies the RS and TWA strategies in multidepot recycling logistics networks. The RS strategy can effectively improve resource utilization and optimize resource configuration in logistics networks. A bi-objective optimization model is proposed for MRVRPRSTWA with the minimum operating cost and TNV. A hybrid heuristic algorithm, which includes 3D *k*-means clustering algorithm and NSGA-II algorithm, is also designed to address MRVRPRSTWA. The 3D *k*-means clustering algorithm is used to cluster customers in the first step of the hybrid heuristic algorithm for reducing the computational complexity of multidepot vehicle routing problem. Next, the NSGA-II algorithm is used to find the optimal vehicle routes and assign candidate time windows to customers.

A real-world case study in Chongqing City, China, is analyzed to verify the performance of the proposed mathematical model and the hybrid heuristic algorithm in MRVRPRSTWA. The computational results demonstrate that the total operating cost and TNV can be decreased through the proposed method. The total operating cost is reduced from \$38025 to \$21373, and TNV is minimized from 21 to 7. The comparison of the performance of three algorithms shows that the NSGA-II has a better

performance in solving MRVRPRSTWA than MOEA and MOPSO. Four different scenarios with and without RS and TWA strategies are presented to demonstrate the applicability of the proposed method. The results of four scenarios indicate that the operational efficiency of the logistics network can be improved via the RS and TWA strategies.

In this study, the RS and TWA strategies are encouraged to promote the sustainable development of multidepot logistics network. On the one hand, the RS strategy among logistics facilities can optimize the resource configuration and enhance the resource utilization. On the other hand, the TWA strategy can improve the operational efficiency of logistics network by assigning candidate time windows to customers. Although only the RS and TWA strategies can improve the operational efficiency of logistics network to a certain extent, the greatest benefits can be obtained by simultaneously adopting both, which is demonstrated in section 6.5. The RS and TWA strategies can be used as references when policy-makers make decisions for logistics development.

Many research questions can be studied in the future on the basis of this work. (1) Delivery activity can be considered on the basis of the MRVRPRSTWA to construct pickup and delivery logistics network. (2) The proposed algorithm can be combined with exact algorithms to enhance the accuracy and efficiency of computation. (3) Collaboration among logistics facilities and profit allocation mechanism can be investigated on the MRVRPRSTWA. (4) The environmental impact in the MRVRPRSTWA can be explored in the future.

Data Availability

The vehicle routing data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 71871035 and 41977337), Humanity and Social Science Youth Foundation of Ministry of Education of China (Grant No. 18YJC630189), Key Science and Technology Research Project of Chongqing Municipal Education Commission (Grant No. KJZD-K202000702), Key Project of Human Social Science of Chongqing Municipal Education Commission (Grant No. 20SKGH079), Social Science Foundation of Chongqing of China (Grant No. 2019YBGL054), and Chongqing Graduate Tutor Team Construction Project (Grant No. JDDSTD2019008). This research was also supported by 2018 Chongqing Liuchuang Plan Innovation Project (Grant No. cx2018111).

References

- [1] T. R. P. Ramos, M. I. Gomes, and A. P. Barbosa-Póvoa, "Planning a sustainable reverse logistics system: balancing costs with environmental and social concerns," *Omega*, vol. 48, pp. 60–74, 2014.
- [2] M. Mojtahedi, A. M. Fathollahi-Fard, R. Tavakkoli-Moghaddam, and S. Newton, "Sustainable vehicle routing problem for coordinated solid waste management," *Journal of Industrial Information Integration*, vol. 23, p. 100220, 2021.
- [3] Z. Xiao, J. Sun, W. Shu, and T. Wang, "Location-allocation problem of reverse logistics for end-of-life vehicles based on the measurement of carbon emissions," *Computers & Industrial Engineering*, vol. 127, pp. 169–181, 2019.
- [4] T.-Y. Liao, "Reverse logistics network design for product recovery and remanufacturing," *Applied Mathematical Modelling*, vol. 60, pp. 145–163, 2018.
- [5] M. Rahimi and V. Ghezavati, "Sustainable multi-period reverse logistics network design and planning under uncertainty utilizing conditional value at risk (CVaR) for recycling construction and demolition waste," *Journal of Cleaner Production*, vol. 172, pp. 1567–1581, 2018.
- [6] J. Trochu, A. Chaabane, and M. Ouhimmou, "A two-stage stochastic optimization model for reverse logistics network design under dynamic suppliers' locations," *Waste Management*, vol. 95, pp. 569–583, 2019.
- [7] Z. Yang, L. Guo, and Z. Yang, "Emergency logistics for wildfire suppression based on forecasted disaster evolution," *Annals of Operations Research*, vol. 283, no. 1-2, pp. 917–937, 2017.
- [8] Y. Wang, Q. Li, X. Guan, M. Xu, Y. Liu, and H. Wang, "Two-echelon collaborative multi-depot multi-period vehicle routing problem," *Expert Systems with Applications*, vol. 167, p. 114201, 2021.
- [9] J. Li, Y. Li, and P. M. Pardalos, "Multi-depot vehicle routing problem with time windows under shared depot resources," *Journal of Combinatorial Optimization*, vol. 31, no. 2, pp. 515–532, 2016.
- [10] Y. Wang, J. Zhang, K. Assogba, Y. Liu, M. Xu, and Y. Wang, "Collaboration and transportation resource sharing in multiple centers vehicle routing optimization with delivery and pickup," *Knowledge-Based Systems*, vol. 160, pp. 296–310, 2018.
- [11] J. Li, R. Wang, T. Li, Z. Lu, and P. M. Pardalos, "Benefit analysis of shared depot resources for multi-depot vehicle routing problem with fuel consumption," *Transportation Research Part D: Transport and Environment*, vol. 59, pp. 417–432, 2018.
- [12] Y.-J. Gong, J. Zhang, O. Liu, R.-Z. Huang, H. S.-H. Chung, and Y.-H. Shi, "Optimizing the vehicle routing problem with time windows: a discrete particle swarm optimization approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 2, pp. 254–267, 2012.
- [13] İ. Küçüköğlü and N. Öztürk, "A differential evolution approach for the vehicle routing problem with backhauls and time windows," *Journal of Advanced Transportation*, vol. 48, no. 8, pp. 942–956, 2014.
- [14] D. Taş, O. Jabali, and T. Van Woensel, "A vehicle routing problem with flexible time windows," *Computers & Operations Research*, vol. 52, pp. 39–54, 2014.
- [15] Z. Miao, F. Yang, K. Fu, and D. Xu, "Transshipment service through crossdocks with both soft and hard time windows," *Annals of Operations Research*, vol. 192, no. 1, pp. 21–47, 2010.
- [16] S. Iqbal, M. Kaykobad, and M. S. Rahman, "Solving the multi-objective vehicle routing problem with soft time windows with the help of bees," *Swarm and Evolutionary Computation*, vol. 24, pp. 50–64, 2015.
- [17] S. Mouthuy, F. Massen, Y. Deville, and P. Van Hentenryck, "A multistage very large-scale neighborhood search for the vehicle routing problem with soft time windows," *Transportation Science*, vol. 49, no. 2, pp. 223–238, 2015.
- [18] R. Spliet and G. Desaulniers, "The discrete time window assignment vehicle routing problem," *European Journal of Operational Research*, vol. 244, no. 2, pp. 379–391, 2015.
- [19] S. Martins, M. Ostermeier, P. Amorim, A. Hübner, and B. Almada-Lobo, "Product-oriented time window assignment for a multi-compartment vehicle routing problem," *European Journal of Operational Research*, vol. 276, no. 3, pp. 893–909, 2019.
- [20] A. Mihi Ramírez, "Product return and logistics knowledge: influence on performance of the firm," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 6, pp. 1137–1151, 2012.
- [21] N. Aras, D. Aksen, and M. Tuğrul Tekin, "Selective multi-depot vehicle routing problem with pricing," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 5, pp. 866–884, 2011.
- [22] P. He, S. Zhang, and C. He, "Impacts of logistics resource sharing on B2C E-commerce companies and customers," *Electronic Commerce Research and Applications*, vol. 34, p. 100820, 2019.
- [23] R. Spliet and A. F. Gabor, "The time window assignment vehicle routing problem," *Transportation Science*, vol. 49, no. 4, pp. 721–731, 2015.
- [24] W. Wilcox, P. A. Horvath, S. E. Griffis, and C. W. Autry, "A Markov model of liquidity effects in reverse logistics processes: the effects of random volume and passage," *International Journal of Production Economics*, vol. 129, no. 1, pp. 86–101, 2011.
- [25] M. M. Veiga, "Analysis of efficiency of waste reverse logistics for recycling," *Waste Management & Research*, vol. 31, no. 10, suppl, pp. 26–34, 2013.
- [26] A. El korchi and D. Millet, "Designing a sustainable reverse logistics channel: the 18 generic structures framework," *Journal of Cleaner Production*, vol. 19, no. 6-7, pp. 588–597, 2011.
- [27] F. T. S. Chan, H. K. Chan, and V. Jain, "A framework of reverse logistics for the automobile industry," *International Journal of Production Research*, vol. 50, no. 5, pp. 1318–1331, 2012.
- [28] C. Bai and J. Sarkis, "Flexibility in reverse logistics: a framework and evaluation approach," *Journal of Cleaner Production*, vol. 47, pp. 306–318, 2013.
- [29] 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.
- [30] M. N. Shaik and W. Abdul-Kader, "A hybrid multiple criteria decision making approach for measuring comprehensive performance of reverse logistics enterprises," *Computers & Industrial Engineering*, vol. 123, pp. 9–25, 2018.
- [31] X. Wang, H. Wang, X. Zhang, Z. Feng, and J. Wu, "Stochastic seat allocation models for passenger rail transportation under customer choice," *Transportation Research Part E: Logistics and Transportation Review*, vol. 96, pp. 95–112, 2016.
- [32] Y. Wang, K. Assogba, J. Fan, M. Xu, Y. Liu, and H. Wang, "Multi-depot green vehicle routing problem with shared transportation resource: integration of time-dependent speed

- and piecewise penalty cost,” *Journal of Cleaner Production*, vol. 232, pp. 12–29, 2019.
- [33] X. Xu, J. Hao, and Y. Zheng, “Multi-objective artificial bee colony algorithm for multi-stage resource leveling problem in sharing logistics network,” *Computers & Industrial Engineering*, vol. 142, p. 106338, 2020.
- [34] Q. An, Y. Wen, T. Ding, and Y. Li, “Resource sharing and payoff allocation in a three-stage system: integrating network DEA with the Shapley value method,” *Omega*, vol. 85, pp. 16–25, 2019.
- [35] L. Sun, S. Wang, S. Liu, L. Yao, W. Luo, and A. Shukla, “A complete research on the feasibility and adaptation of shared transportation in mega-cities—a case study in Beijing,” *Applied Energy*, vol. 230, pp. 1014–1033, 2018.
- [36] C. L. Quintero-Araujo, A. Gruler, A. A. Juan, and J. Faulin, “Using horizontal cooperation concepts in integrated routing and facility-location decisions,” *International Transactions in Operational Research*, vol. 26, no. 2, pp. 551–576, 2019.
- [37] Y. Wang, J. Zhang, X. Guan, M. Xu, Z. Wang, and H. Wang, “Collaborative multiple centers fresh logistics distribution network optimization with resource sharing and temperature control constraints,” *Expert Systems with Applications*, vol. 165, p. 113838, 2021.
- [38] J. C. Molina, J. L. Salmeron, I. Eguia, and J. Racero, “The heterogeneous vehicle routing problem with time windows and a limited number of resources,” *Engineering Applications of Artificial Intelligence*, vol. 94, p. 103745, 2020.
- [39] Y. Wang, Q. Li, X. Guan, J. Fan, Y. Liu, and H. Wang, “Collaboration and resource sharing in the multidepot multiperiod vehicle routing problem with pickups and deliveries,” *Sustainability*, vol. 12, no. 15, p. 5966, 2020.
- [40] D. Dechampai, L. Tanwanichkul, K. Sethanan, and R. Pitakaso, “A differential evolution algorithm for the capacitated VRP with flexibility of mixing pickup and delivery services and the maximum duration of a route in poultry industry,” *Journal of Intelligent Manufacturing*, vol. 28, no. 6, pp. 1357–1376, 2015.
- [41] H. Xu, P. Pu, and F. Duan, “A hybrid ant colony optimization for dynamic multidepot vehicle routing problem,” *Discrete Dynamics in Nature and Society*, vol. 2018, pp. 1–10, 2018.
- [42] J. N. Min, C. Jin, and L. J. Lu, “Maximum-minimum distance clustering method for split-delivery vehicle-routing problem: case studies and performance comparisons,” *Advances in Production Engineering & Management*, vol. 14, no. 1, pp. 125–135, 2019.
- [43] X. Ge, G. Xue, and P. Wen, “Proactive two-level dynamic distribution routing optimization based on historical data,” *Mathematical Problems in Engineering*, vol. 2018, Article ID 5191637, 15 pages, 2018.
- [44] H. Fan, Y. Zhang, P. Tian, Y. Lv, and H. Fan, “Time-dependent multi-depot green vehicle routing problem with time windows considering temporal-spatial distance,” *Computers & Operations Research*, vol. 129, p. 105211, 2021.
- [45] D. Liu, Z. Deng, X. Mao, Y. Yang, and E. I. Kaisar, “Two-Echelon vehicle-routing problem: optimization of autonomous delivery vehicle-assisted E-grocery distribution,” *IEEE Access*, vol. 8, pp. 108705–108719, 2020.
- [46] V. Hemmelmayr, K. F. Doerner, R. F. Hartl, and S. Rath, “A heuristic solution method for node routing based solid waste collection problems,” *Journal of Heuristics*, vol. 19, no. 2, pp. 129–156, 2011.
- [47] R. Alizadeh Foroutan, J. Rezaeian, and I. Mahdavi, “Green vehicle routing and scheduling problem with heterogeneous fleet including reverse logistics in the form of collecting returned goods,” *Applied Soft Computing*, vol. 94, p. 106462, 2020.
- [48] A. Parchami Afra and J. Behnamian, “Lagrangian heuristic algorithm for green multi-product production routing problem with reverse logistics and remanufacturing,” *Journal of Manufacturing Systems*, vol. 58, pp. 33–43, 2021.
- [49] H. Kim, J. Yang, and K.-D. Lee, “Reverse logistics using a multi-depot VRP approach for recycling end-of-life consumer electronic products in South Korea,” *International Journal of Sustainable Transportation*, vol. 5, no. 5, pp. 289–318, 2011.
- [50] J. Liu and Y. He, “A clustering-based multiple ant colony system for the waste collection vehicle routing problems,” in *Proceedings of 2012 Fifth International Symposium on Computational Intelligence and Design*, Washington, DC, USA, October 2012.
- [51] T. R. P. Ramos and R. C. Oliveira, “Delimitation of service areas in reverse logistics networks with multiple depots,” *Journal of the Operational Research Society*, vol. 62, no. 7, pp. 1198–1210, 2017.
- [52] K. Govindan, H. Mina, A. Esmaeili, and S. M. Gholami-Zanjani, “An integrated hybrid approach for circular supplier selection and closed loop supply chain network design under uncertainty,” *Journal of Cleaner Production*, vol. 242, p. 118317, 2020.
- [53] K. Dalmeijer and R. Spliet, “A branch-and-cut algorithm for the time window assignment vehicle routing problem,” *Computers & Operations Research*, vol. 89, pp. 140–152, 2018.
- [54] Y. Wang, S. Zhang, X. Guan et al., “Collaborative multi-depot logistics network design with time window assignment,” *Expert Systems with Applications*, vol. 140, p. 112910, 2020.
- [55] A. Subramanyam, A. Wang, and C. E. Gounaris, “A scenario decomposition algorithm for strategic time window assignment vehicle routing problems,” *Transportation Research Part B: Methodological*, vol. 117, pp. 296–317, 2018.
- [56] F. Neves-Moreira, D. Pereira da Silva, L. Guimarães, P. Amorim, and B. Almada-Lobo, “The time window assignment vehicle routing problem with product dependent deliveries,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 116, pp. 163–183, 2018.
- [57] M. Jalilvand, M. Bashiri, and E. Nikzad, “An effective Progressive Hedging algorithm for the two-layers time window assignment vehicle routing problem in a stochastic environment,” *Expert Systems with Applications*, vol. 165, p. 113877, 2021.
- [58] G. Xu, M. Xu, Y. Wang, Y. Liu, Q. Lv, and E. Osaba, “Collaborative multidepot petrol station replenishment problem with multicompartments and time window assignment,” *Journal of Advanced Transportation*, vol. 2020, Article ID 8843397, 22 pages, 2020.
- [59] D. K. Jana and B. Das, “A two-storage multi-item inventory model with hybrid number and nested price discount via hybrid heuristic algorithm,” *Annals of Operations Research*, vol. 248, no. 1-2, pp. 281–304, 2016.
- [60] Z. Xiong, X.-H. Li, J.-C. Liang, and L.-J. Li, “A multi-objective hybrid algorithm for optimization of grid structures,” *International Journal of Applied Mechanics*, vol. 10, no. 1, p. 1850009, 2018.
- [61] X. Sun, S. Guo, J. Guo, and B. Du, “A hybrid multi-objective evolutionary algorithm with heuristic adjustment strategies and variable neighborhood search for flexible job-shop scheduling problem considering flexible rest time,” *IEEE Access*, vol. 7, pp. 157003–157018, 2019.
- [62] J. Li, P. M. Pardalos, H. Sun, J. Pei, and Y. Zhang, “Iterated local search embedded adaptive neighborhood selection approach for the multi-depot vehicle routing problem with

- simultaneous deliveries and pickups,” *Expert Systems with Applications*, vol. 42, no. 7, pp. 3551–3561, 2015.
- [63] D. Cinar, K. Gakis, and P. M. Pardalos, “A 2-phase constructive algorithm for cumulative vehicle routing problems with limited duration,” *Expert Systems with Applications*, vol. 56, pp. 48–58, 2016.
- [64] S. Gao, Y. Wang, J. Cheng, Y. Inazumi, and Z. Tang, “Ant colony optimization with clustering for solving the dynamic location routing problem,” *Applied Mathematics and Computation*, vol. 285, pp. 149–173, 2016.
- [65] Y. Wang, Y. Yuan, K. Assogba et al., “Design and profit allocation in two-echelon heterogeneous cooperative logistics network optimization,” *Journal of Advanced Transportation*, vol. 2018, Article ID 4607493, 20 pages, 2018.
- [66] Y. Wang, X. Ma, Z. Li, Y. Liu, M. Xu, and Y. Wang, “Profit distribution in collaborative multiple centers vehicle routing problem,” *Journal of Cleaner Production*, vol. 144, pp. 203–219, 2017.
- [67] M. Rashidnejad, S. Ebrahimnejad, and J. Safari, “A bi-objective model of preventive maintenance planning in distributed systems considering vehicle routing problem,” *Computers & Industrial Engineering*, vol. 120, pp. 360–381, 2018.
- [68] Y. Wang, C. Li, X. Jin, Y. Xiang, and X. Li, “Multi-objective optimization of rolling schedule for tandem cold strip rolling based on NSGA-II,” *Journal of Manufacturing Processes*, vol. 60, pp. 257–267, 2020.
- [69] Y. Wang, Y. Yuan, X. Guan, H. Wang, Y. Liu, and M. Xu, “Collaborative mechanism for pickup and delivery problems with heterogeneous vehicles under time windows,” *Sustainability*, vol. 11, no. 12, 2019.
- [70] Y. Wang, Y. Yuan, X. Guan et al., “Collaborative two-echelon multicenter vehicle routing optimization based on state-space-time network representation,” *Journal of Cleaner Production*, vol. 258, p. 120590, 2020.
- [71] Y. Wang, S. Peng, X. Zhou, M. Mahmoudi, and L. Zhen, “Green logistics location-routing problem with eco-packages,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 143, p. 102118, 2020.
- [72] B. Li, J. Li, K. Tang, and X. Yao, “Many-objective evolutionary algorithms,” *ACM Computing Surveys*, vol. 48, no. 1, pp. 1–35, 2015.
- [73] A. Ameli, S. Bahrami, F. Khazaeli, and M.-R. Haghifam, “A multiobjective particle swarm optimization for sizing and placement of DGs from DG owner’s and distribution company’s viewpoints,” *IEEE Transactions on Power Delivery*, vol. 29, no. 4, pp. 1831–1840, 2014.

Research Article

Forecasting Civil Aviation Incident Rate in China Using a Combined Prediction Model

Yixiang Sun¹ and Nana Geng² 

¹Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

²Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Correspondence should be addressed to Nana Geng; shiyu0618@163.com

Received 25 February 2021; Revised 20 March 2021; Accepted 1 April 2021; Published 13 April 2021

Academic Editor: di huang

Copyright © 2021 Yixiang Sun and Nana Geng. 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 the rapid development of air transportation, the complexity, importance, and severity of civil aviation safety have gradually become prominent. It is essential to use various data to analyze and predict the level of aviation safety. This paper used a combined prediction model based on Induced Ordered Weighted Averaging (IOWA) operator to forecast the civil aviation incident rate. We compiled and calculated civil aviation incident data and total flight hours from 2008 to 2019 in China and took the civil aviation incident rate (incident numbers per ten thousand flight hours) as the prediction object. First, this paper used the nonlinear regression model, Grey Verhulst model, and Holt-Winters exponential smoothing model to forecast the civil aviation incident rate individually. Then, it used the smallest sum of squared errors as the principle to use a combined prediction model based on the IOWA operator. It can be seen from the experimental results that the prediction accuracy of the combined model is better than single models. Finally, this paper forecasted the civil aviation incident rate in 2020 and 2021. The results showed that the predicted rates are 0.524 and 0.551. Most notably the incident rate will increase significantly compared with 2019.

1. Introduction

Safety is the primary prerequisite for the development of civil aviation. In recent years, the safety level of civil aviation in China has been continuously improved. By the end of September 2020, China civil aviation achieved consecutive safe flight for 86.69 million flight hours and 121 months [1]. The civil aviation accident index in China is better than the world average. However, aviation safety should not be relaxed because accidents in civil aviation will often cause significant losses. The research of aviation accident data and incident data plays a vital role in aviation safety management. Data prediction with scientific algorithms is the key to accident prevention.

Worldwide travel restrictions related to the COVID-19 pandemic caused a sharp drop in the total number of flights. According to reports from Flightradar24, a global flight tracking website, the number of commercial flights in 2020 fell 42% from the previous year. Nevertheless, crashes of

large commercial airliners worldwide caused more deaths than the previous year. According to reports from the Dutch aviation consultancy To70, 299 people died in crashes of civil aviation large commercial airliners in 2020. The number of people who died in crashes of civil aviation large commercial airliners in 2019 is 257. The aviation operation is full of randomness and uncertainty. So it is essential to strengthen civil aviation data analysis and prediction.

Civil aircraft incident refers to the occurrence of events related to aircraft that do not constitute an accident but may affect safety during the aircraft's operation [2]. The incident rate generally refers to the number of incidents occurring in 10,000 flight hours. So we used the incident rate as the prediction object in this research. More accurate predictions of future civil aviation safety status are conducive to reducing risks and losses before accidents and major accidents occur. Effective prediction can reduce economic losses and has a particular significance for civil aviation safety management development. The civil aviation incident prediction

methods mainly include single predictions such as regression prediction, time series prediction, and Grey prediction and combination predictions of Grey Markov prediction and intelligent algorithm models such as SVM and BP. In actual prediction, different methods can be used to predict the same problem. Improving the forecasting accuracy of incident rate has an important practical significance for aviation practitioners to judge aviation safety trends and formulate correct safety strategies.

This paper used a combined prediction model for forecasting civil aviation incident rates in the next two years based on IOWA operator. First, based on the incident rate data from 2008 to 2019, we used the nonlinear regression model, Grey Verhulst model, and Holt-Winters exponential smoothing model to forecast the aviation accident rate individually. Then, we established a combined prediction model based on the IOWA operator with the smallest sum of squared errors as the principle. Finally, we forecasted the incident rate of civil aviation in China from 2020 to 2021.

2. Literature Review

At present, scholars' research on civil aviation safety data analysis mainly focuses on the following aspects:

- (1) The monitoring or recognition of a certain kind of event, risk analysis, or related model research. For example, Korsun and Poplavskii [3] estimated the measurement error of the aircraft angle of attack using satellite navigation data. Huang et al. [4] proposed a feature extraction method for distinguishing abnormal flight events. Sun and Han [5] proposed a method for analyzing Quick Access Recorder (QAR) overrun events and flight record data using a different test, which can solve the massive QAR Information with guiding significance extracted from flight data. Qi et al. [6] studied the high-risk areas that may trigger the QAR overrun event using the QAR record data and the golden section method. Sun and Meng [7] used the cluster analysis method to study aviation flight incidents' time distribution law and obtained a short period of high aviation incidents. Liu and He [8] studied the flight safety risk index based on the frequency and intensity of flight parameter overrun. Geng [9] used QAR data to analyze pilots' long-term performance and carried out quantitative evaluation methods for their technical flight level. Wang et al. [10] used flight QAR data to study heavy landing accidents and established a quantitative risk assessment model for heavy landing using a statistical modeling method. Jian et al. [11] presented a risk assessment method for civil aviation flight safety based on QAR overrun events.
- (2) Use relevant data for fault diagnosis and trend prediction. Some researchers have used a variety of forecasting models to forecast passenger traffic [12, 13]. Shan [14] used the Bayesian network to predict civil aviation unsafe events and made noise

diagnosis and high-risk identification for unsafe events. Dai [15] analyzed the management mode and relevant civil aviation flight quality monitoring measures and analyzed the civil aviation flight quality monitoring and early warning management. Wang et al. [16] used the Auto Regressive Moving Average (ARMA) model to predict tail collision events in domestic civil aviation transportation. Sun et al. [17] proposed a spatiotemporal analysis method of flight quality monitoring based on temporal Geographic Information System (GIS). Du [18] used exponential function to perform regression analysis on civil aviation accident rate data from 1993 to 2007 and conducted a correlation test. Liu et al. [19] used correlation analysis to choose five positively related influencing factors. They adopted air transport incidents' historical data and their influencing factors to establish air transportation incidents' cluster prediction model. Wang et al. [20] proposed a nonlinear regression prediction model of civil aviation incidents based on taking off and landing sorties. Bin et al. [21] proposed a combined prediction model based on GM (1,1) and Markov model to forecast the civil aviation incidents based on the incident data and operational aircraft data from 2003 to 2012. Wang and Li [22] analyzed civil aviation incidents' characteristics and proposed a series of grey neural network combined prediction models. Xiong et al. [23] used the Long Short-Term Memory (LSTM) neural network model to train and predict the bird strike incidents data. Chen et al. [24] proposed a combined prediction model based on Autoregressive Integrated Moving Average (ARIMA) and Back Propagation Neural Networks (BPNN) to forecast civil aviation incidents in other months. Liang and Li [25] used the combined model of ARIMA, Least Squares Support Vector Machines (LS-SVM), and BPNN to forecast and analyze the airline company's monthly flight incidents rate. Valdés et al. [26] developed five Bayesian models of increasing difficulty to predict and anticipate incidents. Subramanian and Rao [27] used LSTM networks to forecast incident data derived from the National Aeronautics and Space Administration's Aviation Safety Reporting System incident database. The forecast generated helps identify factors that contribute significantly to the trends seen in multiple categories of incidents and also provides insight into which categories of incidents are more (or less) likely to occur in the forecast period.

It can be seen from the literature review that econometric models have been widely used in civil aviation event prediction, such as the ARIMA model, exponential smoothing model, and regression model. The Grey prediction model has good applicability in the prediction of a civil aviation incident. Most of them focus is on optimizing GM (1, 1) model in many aspects and applying it to predicting the civil aviation incident. As intelligent algorithms

can solve traditional prediction models' limitations, more and more intelligent algorithm technologies have been applied to predicting civil aviation incidents, including BP algorithm, RBF algorithm, genetic algorithm, etc. In large data samples, intelligent algorithms have more advantages, and in the case of small data samples, traditional models such as the Grey model have good accuracy. In general, the combination prediction models have attracted more and more attention because of their better prediction accuracy.

3. Combined Prediction Model of Civil Aviation Incident Rate Based on IOWA Operator

The civil aviation industry's development is full of uncertainties and challenges, so predicting with a single model may not achieve good results. It is possible to consider combining multiple types of prediction methods appropriately to comprehensively utilize multiple models' information to improve the prediction's reliability and accuracy. It is the principle of the combined prediction model. The IOWA operator can give weight according to the prediction accuracy of each prediction model at each time and establish a combined prediction model with the sum of square error as the objective function [28–30]. IOWA operator theory is introduced as follows [31].

3.1. Ordered Weighted Averaging (OWA) Model. Assume $f_{owa}: R^m \rightarrow R$ is the function with m variables; thus,

$$f_{owa}(a_1, a_2, \dots, a_m) = \sum_{i=1}^m w_i b_i, \quad (1)$$

$W = (w_1, w_2, w_3, \dots, w_m)^T$ is the weight vector related to the function f_{owa} , where

$$\sum_{i=1}^m w_i = 1, \quad w_i \geq 0, \quad i = 1, 2, \dots, m. \quad (2)$$

Resort the sequence a_1, a_2, \dots, a_m from the largest to the smallest and get a new sequence; w_i is related to the order of a_1, a_2, \dots, a_m .

3.2. IOWA Operator. Assume there are m two-dimensional arrays, $\langle u_1, a_1 \rangle, \langle u_2, a_2 \rangle, \dots, \langle u_m, a_m \rangle$, and $f_{iowa}: R^m \rightarrow R$ is the function with m variables; thus,

$$f_{iowa}(\langle u_1, a_1 \rangle, \langle u_2, a_2 \rangle, \dots, \langle u_m, a_m \rangle) = \sum_{i=1}^m w_i a_{u\text{-index}(i)}. \quad (3)$$

Here, $u_1, u_2, u_3, \dots, u_m$ are the induced values and $u\text{-index}(i)$ is the subscript of the i -th number in a_1, a_2, \dots, a_m arranged in descending order. w_i is not related to the size and position of a_1, a_2, \dots, a_m , but it is related to the location of the induced value.

3.3. Combined Prediction Model Building Process. If m kinds of single prediction method are used to predict the observation value of an indicator sequence $\{y_t, t = 1, 2, \dots\}$, then the prediction accuracy is

$$u_{it} = \begin{cases} 1 - |(x_t - x_{it})/x_t|, & |(x_t - x_{it})/x_t| < 1; \\ 0, & |(x_t - x_{it})/x_t| \geq 1. \end{cases} \quad (4)$$

$i = 1, 2, \dots, m$ means the number of prediction method, $t = 1, 2, \dots, N$ means time, u_{it} means the prediction accuracy, and $u_{it} \in [0, 1]$, x_t is the actual value and x_{it} is the prediction value. Assume u_{it} is the induced value of x_{it} , and the prediction accuracy and the prediction value form a two-dimensional array $[u_{1t}, x_{1t}], [u_{2t}, x_{2t}], [u_{3t}, x_{3t}], \dots, [u_{it}, x_{it}]$.

Arrange $u_{1t}, u_{2t}, u_{3t}, \dots, u_{mt}$ from the largest to the smallest, and mark the i -th prediction accuracy as $u\text{-index}(it)$; thus, the combined prediction value is

$$f[(u_{1t}, x_{1t}), (u_{2t}, x_{2t}), \dots, (u_{mt}, x_{mt})] = \sum_{i=1}^m w_i x_{u\text{-index}(it)}. \quad (5)$$

Assume $e_{u\text{-index}(it)} = x_t - x_{u\text{-index}(it)}$. If the experimental standard is specified as the smallest sum of squared errors, the model can be rewritten as

$$\begin{aligned} \min S(L) &= \sum_{i=1}^m \sum_{j=1}^m w_i w_j \left(\sum_{t=1}^N e_{u\text{-index}(it)} e_{u\text{-index}(jt)} \right), \\ \text{s.t. } &\begin{cases} \sum_{i=1}^m w_i = 1, w_i \geq 0, & i = 1, 2, \dots, m. \end{cases} \end{aligned} \quad (6)$$

3.4. Evaluation of Prediction Model. Prediction accuracies are usually compared using the following four indicators:

Mean square error (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (\hat{x}_t - x_t)^2, \quad t = 1, 2, \dots, n. \quad (7)$$

Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{x}_t - x_t)^2}, \quad t = 1, 2, \dots, n. \quad (8)$$

Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |\hat{x}_t - x_t|, \quad t = 1, 2, \dots, n. \quad (9)$$

Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{x}_t - x_t}{x_t} \right|, \quad t = 1, 2, \dots, n. \quad (10)$$

In the above four evaluation index expressions, x_t is the actual value at time t , \hat{x}_t is the predicted value of the model at time t , and n is the number of forecast periods.

4. Forecasting Civil Aviation Incident Rate in China

4.1. Date Source. According to relevant data [32, 33], the necessary data of civil aviation incidents and civil aviation flight hours in China from 2008 to 2019 are selected.

It can be seen from Table 1 that the civil aviation incident rate showed a clear increasing trend before 2018, from 0.303 in 2008 to 0.554 in 2017, and the rate decreased from 2018. It shows that civil aviation development direction in China is shifting from rapid expansion to high-quality development. Civil aviation safety management, especially the analysis and prediction of accidents and incidents, is necessary for such a critical period.

4.2. Nonlinear Regression Model Prediction. According to relevant data [32, 33], the necessary data of civil aviation incidents in China, civil aviation flight hours, and civil aviation incidents per 10,000-hour rates from 2008 to 2019 are selected. A nonlinear regression model is a commonly used prediction method. Due to the randomness of civil aviation events, the number and frequency data are relatively discrete. The reasonable degree of linear regression is insufficient, but the reasonable degree of nonlinear regression models such as exponential function, power function, and polynomial is good. We use power function regression, exponential function regression, logarithmic function regression, polynomial regression, and other nonlinear regressions to establish the prediction models. Since the error is the smallest and the fitting degree is the highest, the power function regression is selected to establish the power function regression prediction model:

$$y = 2.8393 \times 10^{-287} \times x^{86.6185846}. \quad (11)$$

According to the above formula, the prediction result and error and accuracy are calculated, shown in Table 2. Comparison of power function regression forecast data and actual data is shown in Figure 1.

4.3. Grey Verhulst Model Prediction. The Grey Verhulst model combines the specific advantages of the grey prediction theory and the Verhulst model. Based on the Verhulst model's whitening equation, the difference equation's parameters and the differential equation have an excellent agreement through a series of mathematical derivation. Simultaneously, it applies the sample with insufficient information and little data and has good prediction feasibility.

TABLE 1: Civil aviation incident rate in China from 2008 to 2019.

Year	Incident rate	Year	Incident rate
2008	0.303	2014	0.424
2009	0.324	2015	0.463
2010	0.384	2016	0.547
2011	0.363	2017	0.554
2012	0.477	2018	0.492
2013	0.437	2019	0.463

TABLE 2: The prediction and fitting data of the power function regression model.

Years	Actual data	Forecast data	Error	Accuracy
2008	0.303	0.342	-0.039	87.13%
2009	0.324	0.357	-0.033	89.81%
2010	0.384	0.373	0.011	97.14%
2011	0.411	0.389	0.022	94.65%
2012	0.477	0.406	0.071	85.12%
2013	0.437	0.424	0.013	97.03%
2014	0.424	0.443	-0.019	95.52%
2015	0.463	0.462	0.001	99.78%
2016	0.547	0.483	0.064	88.30%
2017	0.554	0.504	0.05	90.97%
2018	0.492	0.526	-0.034	93.09%
2019	0.463	0.549	-0.086	81.43%

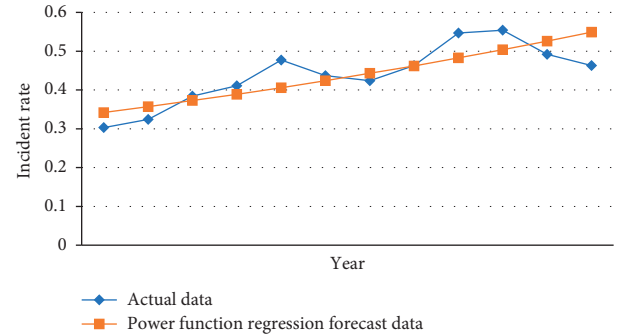


FIGURE 1: Comparison of power function regression forecast data and actual data.

Establish the nonnegative original sequence:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)), \quad (12)$$

and one-time accumulation sequence is

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)), \quad (13)$$

where

$$X^{(1)}(k) = \sum_{i=1}^k X^{(1)}(i), \quad k = 1, 2, \dots, n, \quad (14)$$

and background value sequence is

$$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)), \quad (15)$$

where

$$z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k-1) + x^{(1)}(k)), \quad k = 2, 3, \dots, n. \quad (16)$$

The basic expression form of the Grey Verhulst model is

$$X^{(0)}(k) + az^{(1)}(k) = b(z^{(1)}(k))^2. \quad (17)$$

The parameter estimation vector of the model is $\hat{v} = (a, b)^T$. Thus, the least squares estimation of $X^{(0)}(k) + az^{(1)}(k) = b(z^{(1)}(k))^2$ is $\hat{v} = (a, b)^T = (B^T B)^{-1} B^T Y$, where

$$B = \begin{bmatrix} -z^{(0)}(2)(z^{(0)}(2))^2 \\ -z^{(0)}(3)(z^{(0)}(3))^2 \\ \vdots \\ -z^{(0)}(n)(z^{(0)}(n))^2 \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}. \quad (18)$$

The whitening differential equation of the Grey Verhulst model is

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b(x^{(1)})^2. \quad (19)$$

The time response function can be obtained as

$$x^{(1)}(t+1) = \frac{ax^{(1)}(0)}{bx^{(1)}(0) + (a - bx^{(1)}(0))e^{at}}. \quad (20)$$

Discretize the albino differential equation and take the initial value condition: $\hat{x}^{(1)}(1) = x^{(1)}(1)$.

The time response sequence of the Grey Verhulst model can be obtained as

$$X^{(1)}(k+1) = \frac{ax^{(1)}(1)}{bx^{(1)}(1) + (a - bx^{(1)}(1))e^{ak}}, \quad k = 1, 2, \dots, n. \quad (21)$$

Finally, the predicted and fitted values of the Grey Verhulst model can be obtained by reduction of formula.

The parameters $a = -0.286$ and $b = -0.556$ can be obtained through calculation.

The time response sequence of the Grey Verhulst model is

$$\hat{x}^{(1)}(k+1) = \frac{-0.087}{-0.168 + -0.117e^{-0.286k}}, \quad k = 1, 2, \dots, n. \quad (22)$$

The prediction and fitting data of the Grey Verhulst model are shown in Table 3 and Figure 2.

4.4. Exponential Smoothing Model Prediction. The exponential smoothing model is an essential time series analysis method, which is also called the exponential weighted average method. The exponential smoothing model has been widely used in practice because of its characteristics, such as simple operation, convenient calculation, and excellent prediction performance.

TABLE 3: The prediction and fitting data of the Grey Verhulst model.

Years	Actual data	Forecast data	Error	Accuracy
2008	0.303	0.303	0	100%
2009	0.324	0.337	-0.013	96.14%
2010	0.384	0.369	-0.015	96.09%
2011	0.411	0.397	0.014	96.59%
2012	0.477	0.421	0.056	88.26%
2013	0.437	0.441	0.004	99.09%
2014	0.424	0.457	-0.033	92.78%
2015	0.463	0.47	-0.007	98.51%
2016	0.547	0.48	0.067	87.75%
2017	0.554	0.488	0.066	88.09%
2018	0.492	0.494	-0.002	99.60%
2019	0.463	0.499	-0.036	92.79%

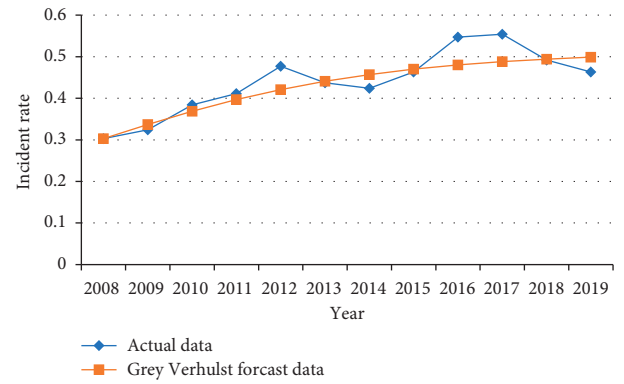


FIGURE 2: Comparison of Grey Verhulst model forecast data and actual data.

Holt-Winters Model was first proposed by Winters in 1960, and later, through the continuous improvement of Cipra, Romera, Hyndman, etc., the existing model formula was gradually formed. Basically, the Holt-Winters model is a cubic exponential smoothing approach, which is one of the standard models for time series analysis.

In the process of modeling the Holt-Winters model, the sample sequence is first decomposed into three parts: linear trend, trend increment, and seasonal change, and then the three components are estimated separately using the exponential smoothing method, and finally the extrapolation is established separately modeling and obtaining the extrapolated predicted value of the sample sequence.

The Holt-Winters nonseasonal model is suitable for the trend prediction problem without the influence of seasonal changes. The mathematical expression of the model is as follows:

$$\begin{cases} S_t = \alpha X_t + (1 - \alpha)(S_{t-1} - b_{t-1}), \\ b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}, \\ F_{t+m} = S_t + mb_t. \end{cases} \quad (23)$$

The Holt-Winters additive model is suitable for time series forecasting problems with seasonal additive changes. The mathematical expression of the model is as follows:

$$\begin{cases} S_t = \alpha(X_t - I_{t-L}) + (1 - \alpha)(S_{t-1} - b_{t-1}), \\ I_t = \beta(X_t - S_t) + (1 - \beta)I_{t-L}, \\ b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}, \\ F_{t+m} = S_t + mb_t + I_{t-L+m}. \end{cases} \quad (24)$$

The Holt-Winters multiplication model is suitable for time series forecasting problems with multiplicative seasonal changes. The mathematical expression of the model is as follows:

$$\begin{cases} S_t = \alpha \left(\frac{X_t}{I_{t-L}} \right) + (1 - \alpha)(S_{t-1} - b_{t-1}), \\ I_t = \beta \left(\frac{X_t}{S_t} \right) + (1 - \beta)I_{t-L}, \\ b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1}, \\ F_{t+m} = (S_t + mb_t)I_{t-L+m}. \end{cases} \quad (25)$$

In the formula, X_t is the actual value of the civil aviation incident rate; F_{t+m} is the predicted value of the civil aviation incident rate for period m ; F_{t+m} is the length of the season; α , β , γ are smoothing parameters, and the value range is a

closed interval of 0~1; b_t is trend component, and the initial value is 0; S_t is stable component at time t , and the initial value is $S_t \sum_{i=1}^L (X_i/L)$; I_t is the seasonal component at time t , and the initial value of I_t in the additive model is $X_t - S_t$. The initial value of I_t in the additive model is X_t/S_t .

In this paper, EViews software is used to establish a Holt-Winters Exponential Smoothing Model. The initial smoothing value is the default value of the system. Alpha, Beta, and Gamma values are automatically selected by EViews software to minimize the sum of squares between the actual value of the sequence and the sequence's smoothing value. For the Holt-Winters model, its parameter estimation is shown in Table 4.

At this point, the smoothing parameter α is 0.36000, β is 1.0000, γ is 0.0000, and the error sum of squares is 0.004171.

The prediction and fitting data of the Holt-Winters model are shown in Table 5 and Figure 3.

4.5. *Combination Prediction Model.* Establish a two-dimensional array:

$$[a_{1t}, x_{1t}], [a_{2t}, x_{2t}], [a_{3t}, x_{3t}], \dots, [a_{mt}, x_{mt}]. \quad (26)$$

Calculate the IOWA combination forecast value as follows:

$$\begin{aligned} f_L([a_{11}, x_{11}], [a_{21}, x_{21}], [a_{31}, x_{31}]) &= 0.342L_1 + 0.303L_2 + 0.320L_3, \\ f_L([a_{12}, x_{12}], [a_{22}, x_{22}], [a_{32}, x_{32}]) &= 0.357L_1 + 0.337L_2 + 0.316L_3. \end{aligned} \quad (27)$$

Then use LINGO software to get the weight of each individual forecast:

$$\begin{aligned} w1 &= 0.9999998, \\ w2 &= 1.42E - 08, \\ w3 &= 7.35E - 09. \end{aligned} \quad (28)$$

Constructing the combination prediction model with this weight, the calculated results are shown in Table 6 and Figure 4. The average prediction accuracy of the combined prediction model is 96.92%.

4.6. *Prediction Model Evaluation.* As shown from the result figures of the prediction model, both the Grey Verhulst model and the power function model are monotone curves, and their fitting curves are relatively stable. However, neither of these two models can reflect the fluctuation of data. So, when the incident rate fluctuates significantly in the years of 2012 and 2016, the error is relatively large. The exponential smoothing model curve can reflect the fluctuation of the actual curve. The incident rate is highly dependent on the recent data and has obvious trend change, but there is always a certain distance between the actual value and the predicted value. Combined prediction weights and averages the predicted values of each individual forecast at each time point,

and the result is highly correlated with the prediction accuracy at each time point. It is a dynamic combination of three single forecast models. Combination forecasting is more coincident with the factual data.

To compare and analyze the prediction effects of different models, we use the four indicators of *MSE*, *RMSE*, *MAE*, and *MAPE* to evaluate the models. Calculate the four evaluation indicators of the three single prediction models, and compare them with the combination prediction model. The comparison results are shown in Table 7.

It can be seen from the four error evaluation indexes that the error evaluation of the combined forecasting model is less than that of the single forecasting method. Because the combined prediction effectively integrates the sample information and selects the advantages of a single prediction, the fitting accuracy is also higher than that of the general combined prediction model. Due to the randomness of events, certain important factors may be ignored when using a particular forecasting method. Careful consideration of several forecasting methods and operators' use as a combined forecasting model will improve the forecasting accuracy.

4.7. *Forecast the Incident Rate.* Use the combination prediction model to predict the civil aviation incident rate of the

TABLE 4: The prediction and fitting data of the Grey Verhulst model.

Alpha	0.3600
Beta	1.0000
Gamma	0.0000
Sun of squared residuals	0.004171
Root mean squared error	0.018643

TABLE 5: The prediction and fitting data of the Holt-Winters model.

Years	Actual data	Forecast data	Error	Accuracy
2008	0.303	0.320	-0.017	94.39%
2009	0.324	0.316	0.008	97.53%
2010	0.384	0.363	0.021	94.53%
2011	0.411	0.424	-0.013	96.84%
2012	0.477	0.458	0.019	96.02%
2013	0.437	0.427	0.010	97.71%
2014	0.424	0.444	-0.020	95.28%
2015	0.463	0.496	-0.033	92.87%
2016	0.547	0.496	0.051	90.68%
2017	0.554	0.531	0.023	95.85%
2018	0.492	0.499	-0.007	98.58%
2019	0.463	0.486	-0.023	95.03%

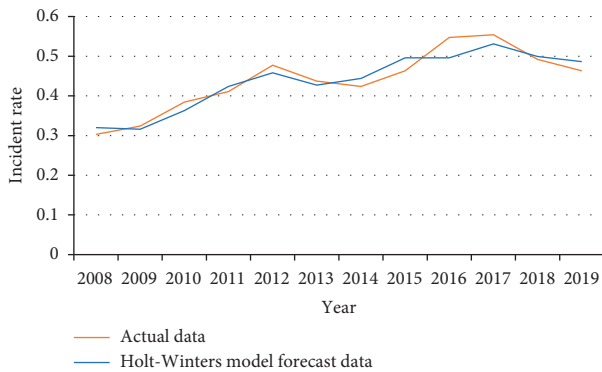


FIGURE 3: Comparison of Holt-Winters model forecast data and actual data.

TABLE 6: The prediction and fitting data of the combined prediction model.

Years	Actual data	Forecast data	Accuracy
2008	0.303	0.303	100.00%
2009	0.324	0.316	97.53%
2010	0.384	0.373	97.14%
2011	0.411	0.424	96.84%
2012	0.477	0.458	96.02%
2013	0.437	0.441	99.08%
2014	0.424	0.443	95.52%
2015	0.463	0.462	99.78%
2016	0.547	0.496	90.68%
2017	0.554	0.531	95.85%
2018	0.492	0.494	99.59%
2019	0.463	0.486	95.03%

year 2020 and 2021. Significantly, the combined prediction model based on IOWA Operator takes the arithmetic mean of all the weights as the weight of the forecast moments when

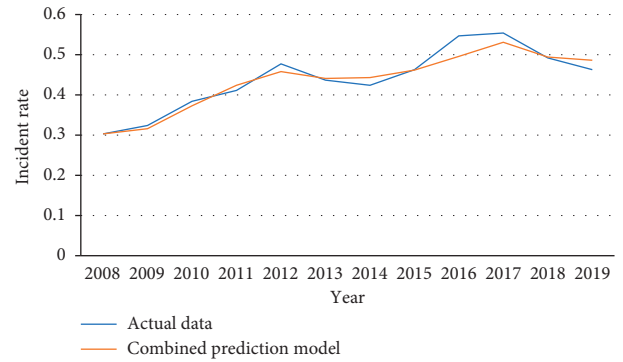


FIGURE 4: Comparison of combined prediction model forecast data and actual data.

TABLE 7: Comparison of combination prediction model and single prediction models.

Model	MSE	RMSE	MAE	MAPE
Power function regression	0.001995	0.044661	0.036917	0.083368
Grey Verhulst	0.001252	0.035384	0.026083	0.054683
Holt-Winters	0.000552	0.023495	0.020417	0.045578
Combined prediction	0.000396	0.019908	0.014500	0.030785

TABLE 8: Civil aviation incident rates of years 2020 and 2021.

Year	2020	2021
Combined prediction model	0.524	0.551

determining the future prediction weight coefficient. When determining each forecast model's weight during the forecast period, the simple average method is used to determine the three single forecast models' weights in the forecast period. After calculation, during the forecast period, the weights of the Power Function Regression model, Grey Verhulst model, and the Holt-Winters model are $w_1 = 0.1816, w_2 = 0.1819, w_3 = 0.6365$. The predicted value of the civil aviation incident rate in the next two years by the combined forecasting model based on the IOWA operator is calculated. The prediction results are shown in Table 8.

According to the combined model's prediction results, civil aviation incident rates in 2020 and 2021 will be 0.524 and 0.551, showing a rising trend. This prediction result is contrary to the civil aviation incident rate in China which has declined for two consecutive years from 2018 to 2019. Simultaneously, considered with the aviation industry's current situation under the influence of the COVID-19, aviation safety needs more concerns.

5. Conclusions

Based on the theoretical research on civil aviation incident rate forecasting, this paper used a combined prediction model based on IOWA operator to overcome the shortcomings of the single prediction model and applies it to the 2020-2021 civil aviation incident rate prediction. Using the incident rate instead of the number of incidents to predict

civil aviation's safety level can weaken incidents' randomness and have higher stability. It can reflect the safety level of civil aviation more accurately.

Regarding the reasons for the possible increase in incident rate, we hold that, on the one hand, the aviation industry is currently in a difficult period, and aviation safety management may be lax. On the other hand, due to the COVID-19 epidemic, many uncertain factors have been caused, which further leads to safety issues. COVID-19 has brought considerable changes to the aviation industry and brought new safety risks. The total number of flights will continue to be affected by worldwide travel restrictions. It has a significant impact on the flight crew and other aviation industry-related employees. Many employees were suspended from work. Aviation safety issues under the influence of the COVID-19 pandemic deserve more attention and further discussion. For example, after the aviation industry's recovery, flight attendants' "skill decline" may become a critical problem.

The aviation industry must take preventive measures to control new safety risks. We all hope to sweep away the haze of the past, and reducing the rate of civil aviation incidents becomes one of the goals. To effectively maintain aviation safety management, we should increase the investigation and management of potential safety hazards and strengthen the control of critical risks. In addition to strengthening flight attendants' skill training, it is equally essential for countries to resolve international conflicts and reduce potential risks peacefully.

Data Availability

All data generated or analyzed during this study are included within this article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Nanjing University of Posts and Telecommunications research start-up fund (NY219168), the Nature Foundation Incubation Fund of Nanjing University of Posts and Telecommunications (NY220214), the Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (TJZ220011), and the Jiangsu Innovation Program for Graduate Education (KYLX15_0312).

References

- [1] Information Office of the State Council, *China's Civil Aviation Development has made Six Achievements During the 13th Five Year Plan*, Information Office of the State Council, Beijing, China, 2020, <http://n.eastday.com/pnews/1603361709021070,2020-10-22/2020-12-10>.
- [2] MH/T (2001–2018), "Civil aircraft incident sign," 2018.
- [3] O. N. Korsun and B. K. Poplavskii, "Estimation of systematic errors of onboard measurement of angle of attack and sliding angle based on integration of data of satellite navigation system and identification of wind velocity," *Journal of Computer and Systems Sciences International*, vol. 50, no. 1, pp. 130–143, 2011.
- [4] Y. Huang, J. Wang, and B. Gu, "Feature extraction method of QAR data based on ball vector machine," *Informatization Research*, vol. 23, pp. 107–108, 2010.
- [5] R. Sun and W. Han, "Parameter characteristics analysis of flight overrun events based on difference test," *China Science and Technology of Work Safety*, vol. 10, pp. 10–11, 2011.
- [6] L. Qi, X. Shao, and H. Chi, "Flight operation risk diagnosis method for QAR overrun event," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 36, pp. 17–18, 2011.
- [7] R. Sun and L. Meng, "Study on time distribution of aviation flight incidents," *Traffic Information and Safety*, vol. 2, no. 31, pp. 83–87, 2013.
- [8] X. Liu and Y. He, "Flight safety quantitative evaluation model fram-fd based on FDR," *Journal of University of Electronic Science and Technology of China*, 2006, vol. 35, no. 1, 2013.
- [9] S. Geng, "Research on pilot flight technology evaluation method based on QAR data," *Tianjin: Civil Aviation University of China*, vol. 08, pp. 14–17, 2007.
- [10] L. Wang, R. Sun, and C. Wu, "Heavy landing risk quantitative evaluation model based on flight QAR data," *Chinese Journal of Safety Sciences*, vol. 24, no. 2, pp. 88–92, 2014.
- [11] W. Jian, Z. Xia, and J. Wang, "Civil aviation flight safety risk assessment method based on QAR overrun event," *Science and Technology Guide*, vol. 37, no. 11, pp. 101–108, 2019.
- [12] J. Zhou, X. Kuang, and Z. Chen, "A demand forecasting model for air passenger traffic in China: based on the stochastic frontier analysis model and model averaging," *Systems Engineering—Theory &*, vol. 40, no. 11, pp. 117–127, 2020.
- [13] J. Biay, A. Rurak, and W. K. J. Tomasz, "Statistical analysis of the volume of passenger air traffic in Europe in years 2004–2015," *Transportation Research Procedia*, vol. 35, pp. 72–79, 2018.
- [14] J. Shan, *Prediction Reasoning and Risk Identification of Civil Aviation Unsafe Event Data*, Beijing Jiaotong University, Beijing, China, 2016.
- [15] L. Dai, "Analysis of civil aviation flight quality monitoring and early warning management," *Dual Use Technologies and Products*, vol. 14, 2017.
- [16] L. Wang, X. Yang, and S. Gao, "Prediction of civil aviation tail scraping events based on time series model," *Safety and Environmental Engineering*, vol. 27, no. 2, pp. 216–220, 2020.
- [17] H. Sun, Y. Liu, and Y. Chen, "Spatial temporal analysis of QAR big data based on Temporal GIS," *Computer Engineering and Design*, vol. 39, no. 3, pp. 694–700, 2018.
- [18] Y. Du, "Regression analysis of the 10,000-hour rate of civil aviation incidents," *Journal of Civil Aviation Flight University of China*, vol. 1, pp. 42–44, 2010.
- [19] L. Liu, Y. Sun, and H. Zheng, "Prediction model of the air transport incidents of civil aviation based on set pair analysis," *Safety and Environmental Engineering*, vol. 20, no. 5, pp. 154–158, 2013.
- [20] Y. Wang, J. Zhu, and N. Ga, "Nonlinear regression prediction of civil aviation incidents based on taking off and landing sorties," *Journal of Wuhan University of Technology (Information and Management Engineering)*, vol. 38, no. 6, pp. 672–677, 2016.
- [21] W. Bin, "Realization of GMPC model in incident prediction," *Science and Technology in Western China*, vol. 13, no. 4, pp. 17–30, 2014.
- [22] Y. Wang and H. Li, "Research on prediction model of civil aviation accident symptomatic based on grey neural

- network,” *China Safety Science Journal*, vol. 22, no. 3, pp. 10–15, 2012.
- [23] M. Xiong, H. Wang, and Y. Xu, “General aviation safety research based on prediction of bird strike symptom,” *Systems Engineering and Electronics*, vol. 42, pp. 2033–2040, 2020.
- [24] F. Chen, D. Zhang, and W. Wei, “Integrated forecasting model for the civil aviation incidents rate based on the ARIMA and BP neural network,” *Journal of Safety and Environment*, vol. 6, pp. 2040–2047, 2019.
- [25] W. Liang and X. Li, “Pre-diction of air traffic accidents based on ARIMA, LS-SVM and BP neural network models,” *Safety and Environ-Mental Engineering*, vol. 25, no. 1, pp. 130–136, 2018.
- [26] R. M. A. Valdés, V. F. G. Comendador, and L. P. Sanz, “Prediction of aircraft safety incidents using Bayesian inference and hierarchical structures - ScienceDirect,” *Safety Science*, vol. 104, pp. 216–230, 2018.
- [27] S. V Subramanian and A. H Rao, “Deep-learning based time series forecasting of go-around incidents in the national airspace system,” in *Proceedings of the 2018 AIAA Modelling and Simulation Technologies Conference*, Grapevine, TX, USA, January 2018.
- [28] Q. Chen and H. Chen, “Optimal combination forecasting model based on two kinds of criteria of Iowa operator and its application,” *Mathematical Statistics and Management*, vol. 5, pp. 91–97, 2013.
- [29] H. Zhang, Q. Chen, and C. Xiong, “Optimal grey combination model based on Iowa operator and its application,” *Statistics and Decision*, vol. 32, no. 174, pp. 75–79, 2017.
- [30] D. Huang, X. Chen, Z. Liu, C. Lyu, S. Wang, and X. Chen, “A static bike repositioning model in a hub-and-spoke network framework,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 141, Article ID 102031, 2020.
- [31] H. Chen and C. Liu, “Combination forecasting method based on Iowa operator,” *Forecasting*, vol. 22, no. 6, pp. 61–65, 2003.
- [32] Civil Aviation Administration of China, *China Air Transport Development Report*, Civil Aviation Administration of China, Beijing, China, 2009.
- [33] Civil Aviation Administration of China, *Civil Aviation Industry Development Statistics Bulletin*, Civil Aviation Administration of China, Beijing, China, 2019.