Modelling for Sustainable Cities

Lead Guest Editor: Haoran Zhang
Guest Editors: Ryosuke Shibasaki, Xuan Song, and Jie Yan

Modelling for Sustainable Cities
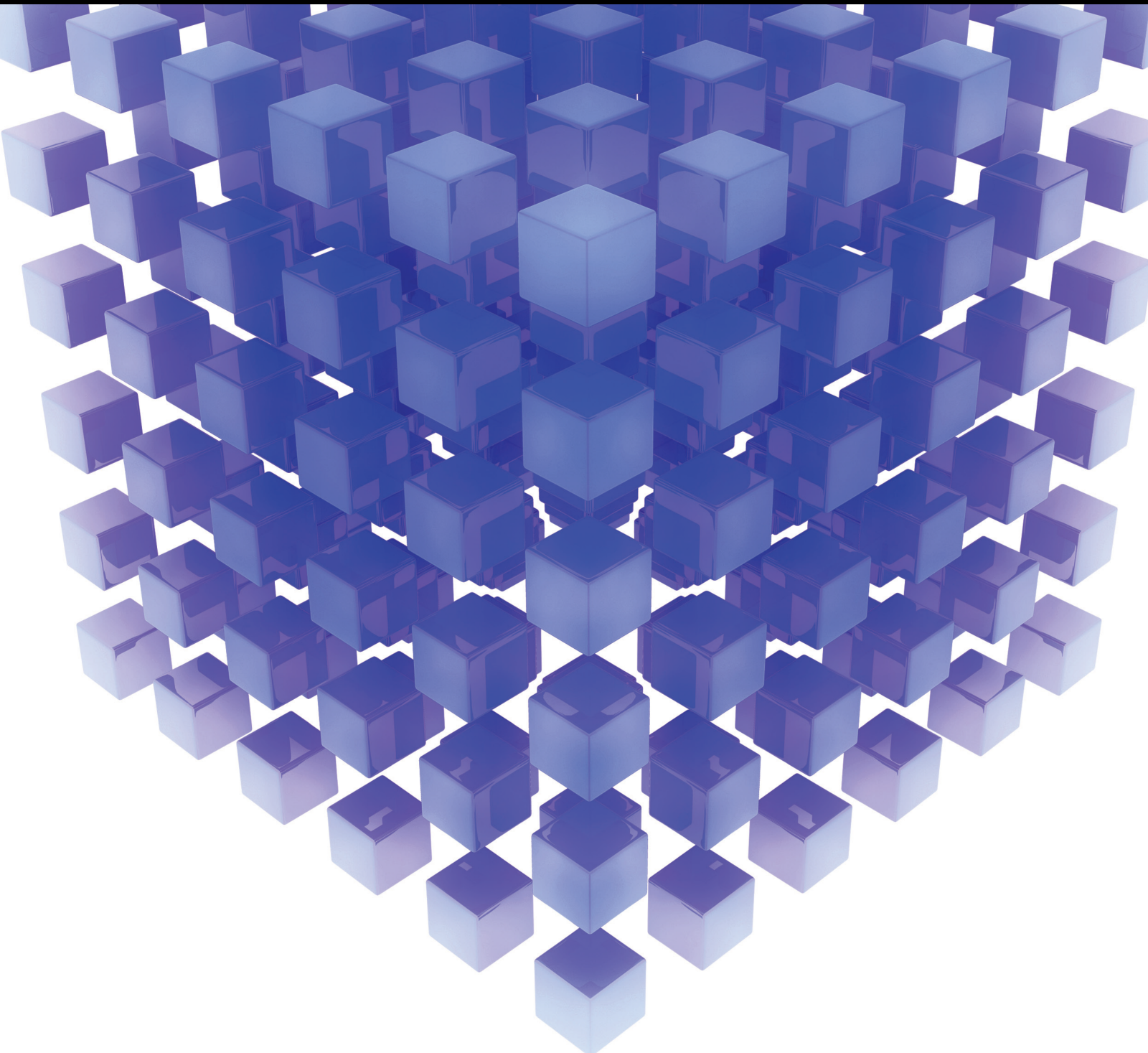
Modelling for Sustainable Cities

Lead Guest Editor: Haoran Zhang
Guest Editors: Ryosuke Shibasaki, Xuan Song, and Jie Yan

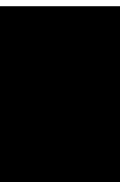Paolo Boscariol, Italy
Daniela Boso (iD), Italy
Guillermo Botella-Juan, Spain
Abdesselem Boulkroune (iD), Algeria
Boulaïd Boulkroune, Belgium
Fabio Bovenga (iD), Italy
Francesco Braghin (iD), Italy
Ricardo Branco, Portugal
Julien Bruchon (iD), France
Matteo Bruggi (iD), Italy
Michele Brun (iD), Italy
Maria Elena Bruni, Italy
Maria Angela Butturi (iD), Italy
Bartłomiej Błachowski (iD), Poland
Dhanamjayulu C (iD), India
Raquel Caballero-Águila (iD), Spain
Filippo Cacace (iD), Italy
Salvatore Caddemi (iD), Italy
Zuowei Cai (iD), China
Roberto Caldelli (iD), Italy
Francesco Cannizzaro (iD), Italy
Maosen Cao (iD), China
Ana Carpio, Spain
Rodrigo Carvajal (iD), Chile
Caterina Casavola, Italy
Sara Casciati, Italy
Federica Caselli (iD), Italy
Carmen Castillo (iD), Spain
Inmaculada T. Castro (iD), Spain
Miguel Castro (iD), Portugal
Giuseppe Catalanotti (iD), United Kingdom
Alberto Cavallo (iD), Italy
Gabriele Cazzulani (iD), Italy
Fatih Vehbi Celebi, Turkey
Miguel Cerrolaza (iD), Venezuela
Gregory Chagnon (iD), France
Ching-Ter Chang (iD), Taiwan
Kuei-Lun Chang (iD), Taiwan
Qing Chang (iD), USA
Xiaoheng Chang (iD), China
Prasenjit Chatterjee (iD), Lithuania
Kacem Chehdi, France
Peter N. Cheimets, USA
Chih-Chiang Chen (iD), Taiwan
He Chen (iD), China

Kebing Chen (iD), China
Mengxin Chen (iD), China
Shyi-Ming Chen (iD), Taiwan
Xizhong Chen (iD), Ireland
Xue-Bo Chen (iD), China
Zhiwen Chen (iD), China
Qiang Cheng, USA
Zeyang Cheng, China
Luca Chiapponi (iD), Italy
Francisco Chicano (iD), Spain
Tirivanhu Chinyoka (iD), South Africa
Adrian Chmielewski (iD), Poland
Seongim Choi (iD), USA
Gautam Choubey (iD), India
Hung-Yuan Chung (iD), Taiwan
Yusheng Ci, China
Simone Cinquemani (iD), Italy
Roberto G. Citarella (iD), Italy
Joaquim Ciurana (iD), Spain
John D. Clayton (iD), USA
Piero Colajanni (iD), Italy
Giuseppina Colicchio, Italy
Vassilios Constantoudis (iD), Greece
Enrico Conte, Italy
Alessandro Contento (iD), USA
Mario Cools (iD), Belgium
Gino Cortellessa, Italy
Carlo Cosentino (iD), Italy
Paolo Crippa (iD), Italy
Erik Cuevas (iD), Mexico
Guozeng Cui (iD), China
Mehmet Cunkas (iD), Turkey
Giuseppe D'Aniello (iD), Italy
Peter Dabnichki, Australia
Weizhong Dai (iD), USA
Zhifeng Dai (iD), China
Purushothaman Damodaran (iD), USA
Sergey Dashkovskiy, Germany
Adiel T. De Almeida-Filho (iD), Brazil
Fabio De Angelis (iD), Italy
Samuele De Bartolo (iD), Italy
Stefano De Miranda (iD), Italy
Filippo De Monte (iD), Italy

Francisco J. Martos, Spain
Elio Masciari, Italy
Paolo Massioni, France
Alessandro Mauro, Italy
Jonathan Mayo-Maldonado, Mexico
Pier Luigi Mazzeo, Italy
Laura Mazzola, Italy
Driss Mehdi, France
Zahid Mehmood, Pakistan
Roderick Melnik, Canada
Xiangyu Meng, USA
Jose Merodio, Spain
Alessio Merola, Italy
Mahmoud Mesbah, Iran
Luciano Mescia, Italy
Laurent Mevel, France
Constantine Michailides, Cyprus
Mariusz Michta, Poland
Prankul Middha, Norway
Aki Mikkola, Finland
Giovanni Minafò, Italy
Edmondo Minisci, United Kingdom
Hiroyuki Mino, Japan
Dimitrios Mitsotakis, New Zealand
Ardashir Mohammadzadeh, Iran
Francisco J. Montáns, Spain
Francesco Montefusco, Italy
Gisele Mophou, France
Rafael Morales, Spain
Marco Morandini, Italy
Javier Moreno-Valenzuela, Mexico
Simone Morganti, Italy
Caroline Mota, Brazil
Aziz Moukrim, France
Shen Mouquan, China
Dimitris Mourtzis, Greece
Emiliano Mucchi, Italy
Taseer Muhammad, Saudi Arabia
Ghulam Muhiuddin, Saudi Arabia
Amitava Mukherjee, India
Josefa Mula, Spain
Jose J. Muñoz, Spain
Giuseppe Muscolino, Italy
Marco Mussetta, Italy

Hariharan Muthusamy, India
Alessandro Naddeo, Italy
Raj Nandkeolyar, India
Keivan Navaie, United Kingdom
Soumya Nayak, India
Adrian Neagu, USA
Erivelton Geraldo Nepomuceno, Brazil
AMA Neves, Portugal
Ha Quang Thinh Ngo, Vietnam
Nhon Nguyen-Thanh, Singapore
Papakostas Nikolaos, Ireland
Jelena Nikolic, Serbia
Tatsushi Nishi, Japan
Shanzhou Niu, China
Ben T. Nohara, Japan
Mohammed Nouari, France
Mustapha Nourelfath, Canada
Kazem Nouri, Iran
Ciro Núñez-Gutiérrez, Mexico
Wlodzimierz Ogryczak, Poland
Roger Ohayon, France
Krzysztof Okarma, Poland
Mitsuhiro Okayasu, Japan
Murat Olgun, Turkey
Diego Oliva, Mexico
Alberto Olivares, Spain
Enrique Onieva, Spain
Calogero Orlando, Italy
Susana Ortega-Cisneros, Mexico
Sergio Ortobelli, Italy
Naohisa Otsuka, Japan
Sid Ahmed Ould Ahmed Mahmoud,
Saudi Arabia
Taoreed Owolabi, Nigeria
EUGENIA PETROPOULOU, Greece
Arturo Pagano, Italy
Madhumangal Pal, India
Pasquale Palumbo, Italy
Dragan Pamučar, Serbia
Weifeng Pan, China
Chandan Pandey, India
Rui Pang, United Kingdom
Jürgen Pannek, Germany
Elena Panteley, France
Achille Paolone, Italy

Zaoli Yang (iD), China
Jun Ye (iD), China
Min Ye (iD), China
Luis J. Yebra (iD), Spain
Peng-Yeng Yin (iD), Taiwan
Muhammad Haroon Yousaf (iD), Pakistan
Yuan Yuan, United Kingdom
Qin Yuming, China
Elena Zaitseva (iD), Slovakia
Arkadiusz Zak (iD), Poland
Mohammad Zakwan (iD), India
Ernesto Zambrano-Serrano (iD), Mexico
Francesco Zammori (iD), Italy
Jessica Zangari (iD), Italy
Rafal Zdunek (iD), Poland
Ibrahim Zeid, USA
Nianyin Zeng (iD), China
Junyong Zhai (iD), China
Hao Zhang (iD), China
Haopeng Zhang (iD), USA
Jian Zhang (iD), China
Kai Zhang, China
Lingfan Zhang (iD), China
Mingjie Zhang (iD), Norway
Qian Zhang (iD), China
Tianwei Zhang (iD), China
Tongqian Zhang (iD), China
Wenyu Zhang (iD), China
Xianming Zhang (iD), Australia
Xuping Zhang (iD), Denmark
Yinyan Zhang, China
Yifan Zhao (iD), United Kingdom
Debao Zhou, USA
Heng Zhou (iD), China
Jian G. Zhou (iD), United Kingdom
Junyong Zhou (iD), China
Xueqian Zhou (iD), United Kingdom
Zhe Zhou (iD), China
Wu-Le Zhu, China
Gaetano Zizzo (iD), Italy
Mingcheng Zuo, China

Contents

Research Article

Analysis of Individual's Activity Space Based on the Cellular Signaling Data

Shaofei Song (ID)**, Weifeng Li** (ID)**, Qing Yu** (ID)**, and Dongyuan Yang** (ID)

Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China

Correspondence should be addressed to Weifeng Li; liweifeng@tongji.edu.cn

In the overall planning of a city, it is important to formulate the reasonable structure of urban space which needs lots of research studies as strong supports. One of these supports is the relationship between the urban built environment and human behavior, and this has been of interest to the field of urban transportation planning. The essential element in this research field is the development of appropriate measures for individual's activity space based on the collected data. This study introduced a new dataset, the cellular signaling data (CSD), and corresponding measures to analyze the relationship between the urban built environment and individual's activity space. The CSD have more detailed time-space stamps of individual's activities compared with traditional surveys, questionnaires, and even call detailed record (CDR) data. The individual's activity space is defined based on the anchor point theory. The convex polygon approach was used to describe the geometrical shape of individual's activity space. The proposed methodology was verified with the CSD collected in Shanghai. The results show that the total number of the cellphone users investigated in this study can be categorized into three different groups with specific characteristics of activity spaces. The results may benefit for related urban agencies to implement customized policy for the purpose of transportation demand management.

1. Introduction

The urban strategic spatial policy has raised a number of topics, for example, during the outline stage of urban master planning, how to develop the urban space structure. We have known that what we need are not only the healthy modes of transportation but also a healthy space structure. The traffic patterns with the priority of bus cannot take the place of a reasonable urban space structure, which should also be the focus of strategic controlling. Under these circumstances, it is necessary to discuss the relationship between the urban built environment (BE) and individual's activity space. And this arises two related issues: how to discuss the activity space of the same kind of residents under different built environment and how to discuss the activity space of different kinds of residents under the same built environment.

The relationship between the urban built environment and individual's activity space has been of interest to the field of urban transportation planning [1–3]. For example, recent studies [4, 5] have focused on the impacts of land use and design policies on the usage of different transport modes, such as transit, walking, and bicycling. The results of such relationship then can be used by urban planners for the evaluation of proper policies to guide human travel activities. The essential element in this research field is the development of appropriate measures for individual's activity space based on the collected data. In other words, there are two keys in this research field: the data and the corresponding appropriate measures.

Previous studies on the link between the urban built environment and individual's activity space mostly rely on traditional traffic surveys and questionnaires and the corresponding measures including travel behavior patterns [6], human mobility patterns [7], and activity patterns [8]. However, the traditional traffic data collection is time consuming and expensive, and the results from traffic

surveys and questionnaires are specific for certain research area under certain time period. All the above shortcomings of the traditional data collection restrict the development of the temporal and spatial patterns of individual's activities.

Recently, the widely implemented intelligent devices, for example, GPS, GIS, the Internet, and especially the mobile phone, offer a great opportunity to investigate the detailed relationship between urban built environment and individual's activity space. There are generally two different kinds of data generated from mobile phones. One is the call detailed record (CDR) data, which are generated with the time and location information when an individual makes a call or sends a message. This dataset is currently used in some studies [9–12]. However, because individuals make calls or send messages randomly, the major drawback of the CDR data is that the data are not uniformly generated. The sample size of the CDR data may not be sufficient to analyze the temporal and spatial distribution of individuals' activity space. The other data generated from mobile phones are the cellular signaling data (CSD), which are generated not only when people make phone calls or send messages but also when the location of the device changes, e.g., from the coverage of the current base station to the adjacent station (please see the details in the data section). Thus, compared with CDR data, the CSD contain much more temporal and spatial records of individuals. Yet, the questions of what and how to use the CSD on the link between the urban built environment and individual's activity space remain largely unexplored.

2. Literature Review

Under the background of the big data, the research emphasis transferred from "based on OD" to "individual's regular activity pattern caused by the external environment effect." So, how to describe the individual's activity space effectively and comprehensively has become a necessity. And there are two main tasks for us to do: to describe the activity space and to mark the category attributes according to activity patterns.

The data collected from mobile phones are the emerging widely and used in current studies, such as the road travel speed [13], the acquisition of OD matrices [9, 14], and the traffic predication and path selection for urban road network [15]. Meanwhile, lots of work based on mobile phone data have been done for urban and traffic planning (like the decision support system [10]) and transportation construction and management [16]. Besides the cell phone data, traditional surveys have contributed to the research of travel behavior, like pedestrian behavior, driving behavior [6, 7, 17], and consumers' trip [18].

Some studies have attempted to figure out the human mobility patterns or to what extent can we predict human mobility patterns. By measuring different entropies of individuals' trajectories, the distribution of actual entropy was captured. And there was a 93% potential predictability (the distribution is narrowly peaked) in user mobility which conflicts with intuition that is the relative regularity of users who travel the most is higher the others [19]. The gyration

radius was calculated to interpret the user's characteristic distance, and its distribution follows the truncated power-law as the same as the travel step size. Furthermore, after removing the anisotropy and rescaling the trajectories, all human mobility collapsed into a same pattern and researchers found that individual trajectories can be characterized by the same gyration radius-independent two-dimensional probability distribution [11, 20]. For the verification of results from mobile phone data, studies find that trajectories estimated by models are similar to the real ones, and the radius of gyration is an appropriate way to present human mobility [12].

From the view of a large scale, human mobility patterns and the corresponding travel behavior form the individuals' activity space. Researchers have described the activity space from different aspects, most of which relied on the anchor point theory [21]. To obtain the human activity space, trips were reconstructed and the distribution of reconstructed starting time and duration of activities was discussed [22]. Another method is to study the spatial density and distribution of individuals' activity [23] or the intensity and entropy of activity [24]. To describe the geometrical shape of the activity space, the standard deviational ellipse (SDE) technique was used [25]. However, the SDE technique will overestimate the spatial spread. Now, in this paper, we found a convex polygon to solve the problem, and it can also show the direction as well as the SDE. Furthermore, geographers have tried to explain human activity space from the aspect of time geography [26, 27], and visualization of human mobility patterns was made in 2D and 3D dimension [28, 29]. Table 1 concludes the typical articles relating these issues with new data resources.

3. Data and Methods

3.1. Dataset. The data we used are cellular signaling data (CSD) which were recorded from September 1st, 2011, to September 30th, 2011, and the detailed format of the data is shown in Table 2. The column "MSID" is the only identification of the mobile phone user which is encrypted. The "Date Time" is the timestamp of the signal. And the combination of "LAC" and "CI" can identify the base station through which we can locate the mobile phone user (obtain the latitude and longitude coordinates).

The great advantage of cellular signaling data is that they not only contain CDR data but the location update data and the handover data. The location update data include normal location update, periodic location update, and IMSI attach (caused by cell phone power on). The handover data include the CDR data and the switch data between base stations. In more general words, these following activities will generate CSD: the cell phone powers on; the user makes a call or sends a message; and the user moves from the coverage of a base station into another one (Figure 1). So, we can obtain detailed information of an individual's trajectory and cell phone usage.

Two datasets were used in this study. The first dataset (D1) has a total of 6441389 logs, which were collected from 1500 people who are randomly in Shanghai. Figure 2 shows

TABLE 1: A summary of related studies.

Research topics	Studies	Measures	Data	Sample size
Human mobility patterns	Human trajectory	Marta et al. 2008	Distribution of displacement; distribution of gyration radius	CDR only	100,000 people 6 months + 206 people 2 weeks
	OD matrices	Iqbal et al. 2014	OD matrices; traffic microsimulation	CDR + video count	6.9 million users + 13 key location 3 days
	Mobility predictability	Song et al. 2010	Trajectory; entropy; predictability	CDR only	50,000 people 3 months
Travel behavior	Spatial behavior	Olle et al. 2014	Monthly variability; standard deviational ellipse	CDR only	1310 people one year
	Travel distance	Xu et al. 2015	Anchor point; travel distance	CDR only	1,219,198 people
Activity space	Spatial and temporal figure	Guo et al. 2013	Spatial distribution and density of individuals' activities	GPS/GSM + surveys	100 residents 1 week
	Activity patterns	Widhalm et al. 2015	Reconstruction of trips; distribution of restricted starting time and duration of activities	CDR + CSD + land use	600,000 people 2 weeks + 1 million 2 weeks
	Activity patterns	Cheng et al. 2013	Home-workplace location; activity patterns visualization; space-time activity density	CSD + surveys	12635 people 2 months

TABLE 2: Format of the cellular signaling data.

MSID	Date time	LAC	CI
912805633C598B3F2C9D03D8A4C4D93F	20110901000302	06150	12914
6478865FA8C6BAABAFADB8A5B31865AD	20110901000253	06204	12626
51E38D9D77BA1979C260994F9CF701F2	20110901000253	06324	62290

If a user "A" moves from the coverage of base station "X" into the coverage of base station "Y", this will generate a log in CSD
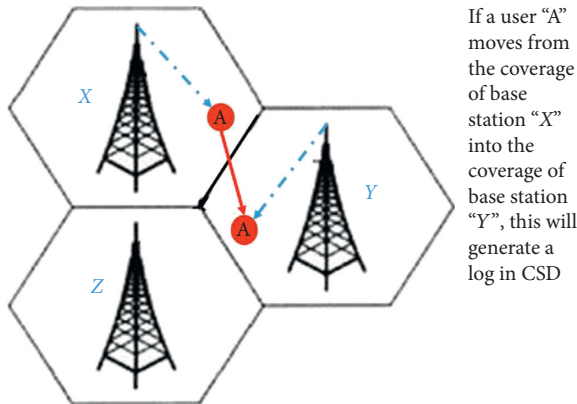
FIGURE 1: The generation of one kind of CSD.

the spatial distribution of the D1. The second dataset (D2) consists of 18844 people who lived in three communities (Jingan, Dahua, and Gucun) (see Figure 2(c)) along Shanghai Metro Line7. These 3 communities locate, respectively, around the inner, central, and outer ring, and others characteristics are shown in Table 3. We use the dataset D2 as comparison group to the dataset D1. A detailed description can be seen in the section of methodology.

3.2. Data Preprocessing. The raw CSD cannot be used directly in this study due to the data noise and the structure of the data. Thus, a three-stage framework was proposed in this study to preprocess the raw CSD as follows and a detailed description of the data preprocessing can be seen in our previous study [30]:

Step 1. Binning method: Because there are some overlaps in the coverage areas of two adjacent base transceiver station, frequency handover may occur as the mobile phone enters the overlaps of the serving cell and the adjacent cells. This may lead to the data noise. In order to clean the data noise, the binning method was used in this study to smooth the location information and reduce the volume of data. First, we distributed the chronologically sorted logs into bins of equal width (10 minutes). This threshold will mostly clean the noise and has little effect on data quality. Second, we replaced all the logs by one equivalent log. So, the location information was replaced by the weighted average of the original coordinates in the same bin.

Step 2. Raster data structure: There were totally 23,918 base transceiver stations which are distributed unevenly and irregularly throughout Shanghai. So, the data structure was not fit for temporal and spatial analysis. In this study, we constructed a raster to cover the city territory of Shanghai. Cells of the raster were delimited with latitude and longitude with fixed intervals. So, every cell could be determined by the coordinate of centroid (lon_c, lat_c) and the length of sides Δlon and Δlat, as shown in Figure 3. Given the radio coverage of the base transceiver station, the size of cells was 500 meters by 500 meters, and the base transceiver
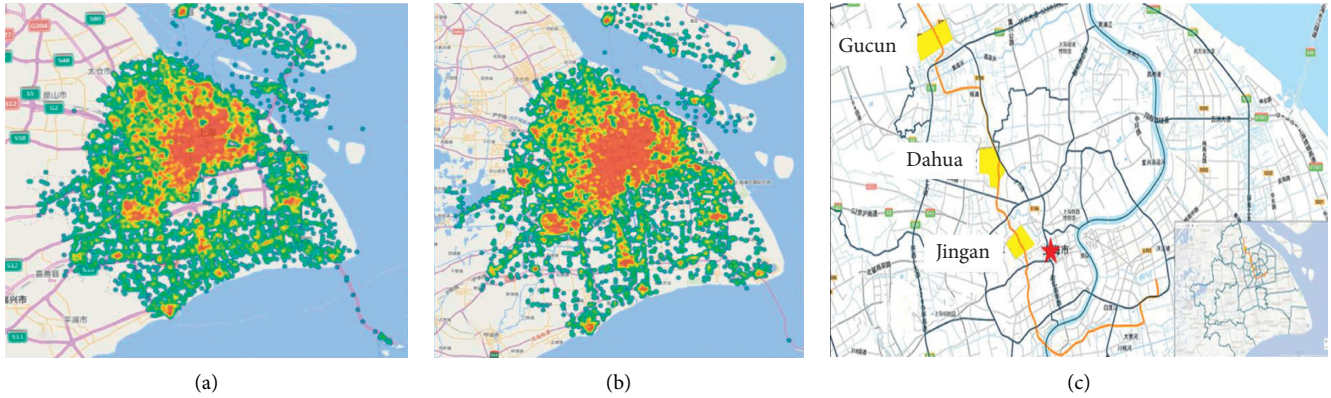
FIGURE 2: Coverage of two datasets: (a) dataset 1 (D1); (b) dataset 2 (D2); (c) 3 large residential districts.

TABLE 3: Characteristics of three communities.

	Jingan	Dahua	Gucun
Area (km^2)	2.56	3.44	4.5
Distance to city center (km)	3.16	7.46	16.71
Residents (number of people)	14901	2955	1363
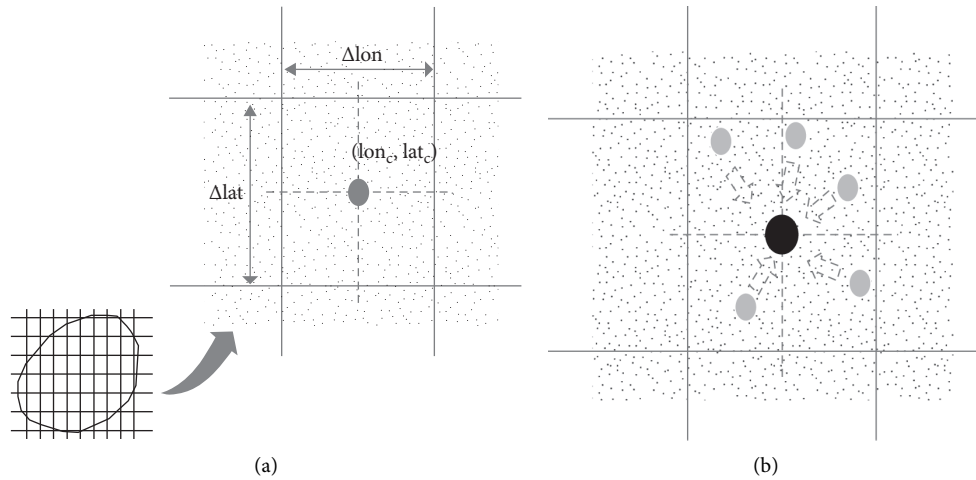
FIGURE 3: Illustration of the method to raster: (a) raster and cell; (b) transportation of BTS coordinates.

stations in the same cell were replaced by one equivalent base transceiver station at the cell centroid.

Step 3. Identification of activity points: We are more interested in the points where people stay rather than those points people just pass by. So, we defined those interested points where people continuously stay for more than 30 minutes as activity points.

3.3. Identification of Anchor Points. Anchor point theory was proposed by Golledge to explain the formation of the activity space: an individual in a specific space will first find the primary nodes (like home or workplace) and then secondary nodes and the roads connecting nodes will be recognized. These hierarchical roads and nodes form an individual's

activity space. The primary nodes are defined as anchor points to be distinguished from other activity points (secondary nodes). However, for the lack of individuals' personal properties in the data, we cannot recognize individuals' home-work places. So, here in this study, we set the rules for an anchor point "S" to follow:

(1) In one day, an individual stays at the activity point "S" for more than 4 hours

(2) During the 30 days of observation, (1) happens for more than 10 days, which means more than 40 hours totally

(3) The distance between anchor point "$S1$" and anchor point "$S2$" for one individual should be more than 3

3.4. Method to Describe Activity Space.

An important parameter of human activity space is its geometric shape. The widely used method [25, 32–34] is the standard deviational ellipse (SDE) technique; however, the SDE will definitely cause the overestimation of the activity space.

In this study, a convex polygon method was proposed to solve this problem. In the 30-day period, each individual had several activity points, representing several sets of coordinates in the map. We chose the outermost points to form a convex polygon by following the arrows (from 1 to 7) in Figure 4, and then we can calculate the area of the polygon. Furthermore, to represent the sprawl direction of activity space, we defined the major-minor axis ratio = major axis/ minor axis, where the major axis is the longest distance between the activity points (the length of AB in Figure 4) and the minor axis is the plus of longest distances from both sides to major axis (CD + EF in Figure 4). The specific method is as follows:

(1) The outermost point of the resident stay point is screened out, and the distance between the two is calculated, among which the longest distance AB is the long axis of the convex polygon.

(2) Starting from the leftmost point A, circle the stop points A into a convex polygon counterclockwise and calculate their area.

(3) The distance from the outermost stop point to the long axis AB is calculated. The sum of the upper maximum distance EF and the lower maximum distance CD is the length of the short axis, and the ratio of the long axis to the short axis is used as the major-minor axis ratio to describe the spatial orientation of residents.

4. Results and Discussion

4.1. Distribution of Anchor Points.

The proposed methodology was used to obtain the distribution of anchor points in D1. The results show that there are 803 people (54%) with only one anchor point (group 1, G1), 301 people (20%) with two anchor points (group 2, G2), and 20 (1%) people with 3 anchor points (group 3, G3). There are also 376 people who had no anchor point which mainly concluded the following parts: tourists or businessmen; old people who rarely used cell phone; and others with data missing. This group of people is very complicated, so in this paper, we will not take them into consideration.

Group 1 with 54% people means these people have one anchor point. However, this does not mean that these people all live and work at the same place. This group includes (1) people have no work, like old people and young children; (2) people have no fixed work place, like deliveryman and drivers, and during the work time, they move in a region; (3) people who work at home; and (4) people who have two
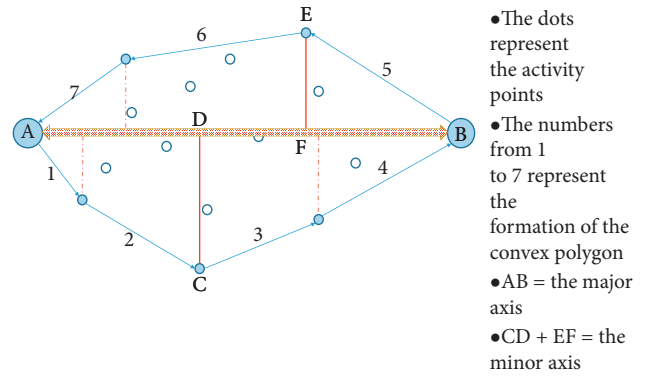
FIGURE 4: Illustration of the convex polygon.

- The dots represent the activity points
- The numbers from 1 to 7 represent the formation of the convex polygon
- AB = the major axis
- CD + EF = the minor axis

cellphones, and they leave a cell phone at home when they go to work.

Group 2 with 20% people means these people have tow anchor points. This group includes people who have fixed work place or have two home places. Group 3 (1%) means people with three regular activity place, like home, work place, or leisure and entertainment place.

The spatial distribution of anchor points of each group is shown in Figure 5. And samples from the 3 groups tell us the intuitive geometry and distribution of the activity space. Group 1 converges in 1 center. Group 2 distributes to 2 centers. And group 3 is more scattering and with a larger range.

As for the individuals in dataset D2, we considered them with one anchor point because their homes were in the three large residential communities along Line 7.

It can be seen from Figure 5 that the individual with one anchor point moves around the anchor point. The individual with two anchor points has a relatively flat and long activity space, and other activity points gather around the two anchor points. An individual with three anchor points has a triangular activity space.

4.2. The Area and Ratio of Activity Space.

After applying this convex polygon method to the three different groups in dataset D1, we got the corresponding distributions of area and major-minor axis ratio as follows:

(1) We found that distribution of the polygon's area (S) can be approximated by an exponential (Figure 6):

$$P(S) = \alpha \exp(-aS) + \beta \exp(-bS), \qquad (1)$$

with exponents $\alpha|G_1 = 0.075$, $\beta|G_1 = 0.02$, $a|G_1 = 0.032$, and $b|G_1 = 0.003$ and $\alpha|G_2 = 0.033$, $\beta|G_2 = 0.004$, $a|G_2 = 0.005$, and $b|G_2 = 0.00009$. For group 3, there were only 20 individuals, and it was not enough to find a proper distribution. However, still we can see the obvious variation tendency of area from group 1 to group 3: the probability of a large area gets bigger. And this is quite fit with the actual situation: people with more anchor points are more likely to own a larger activity space.

(a)   (b)   (c)   (d)
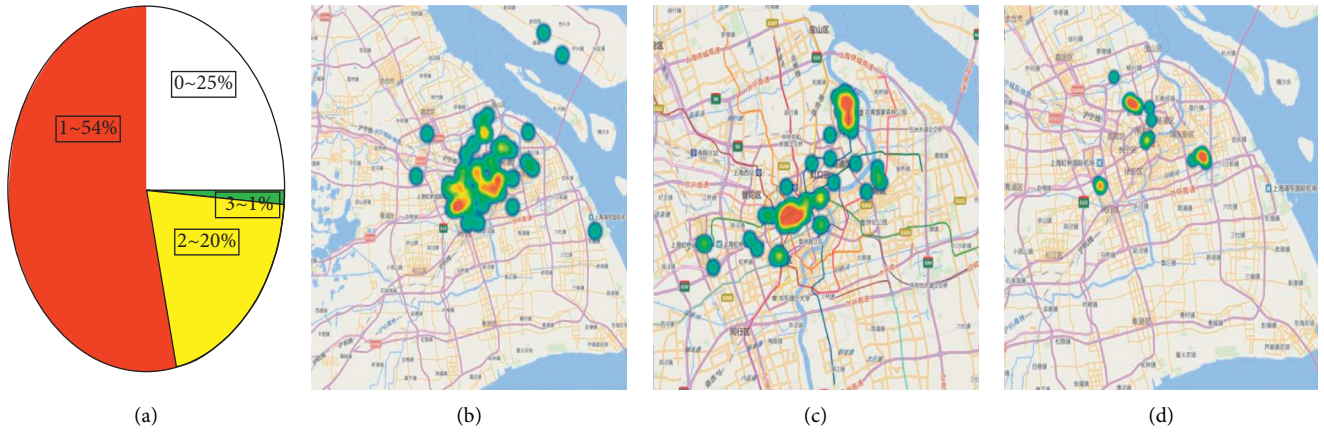
FIGURE 5: Distribution and samples with different number of anchor points; (a) distribution; (b) a sample with 1 anchor point; (c) a sample with 2 anchor points; (d) a sample with 3 anchor points.

Group 1 (with 1 anchor point)

$P(S) = 0.075e^{-0.032S} + 0.02e^{-0.003S}$

(a)

Group 2 (with 2 anchor points)

$P(S) = 0.033e^{-0.005S} + 0.004e^{-0.00009S}$

(b)

Group 3 (with 3 anchor points)

(c)

Group 1 (with 1 anchor point)

$P(R) = 0.26R^{1.2} \exp(-0.12R^{2.2})$

(d)

Group 2 (with 2 anchor point)

$P(R) = 0.26R^{1.2} \exp(-0.12R^{2.2})$

(e)

Group 3 (with 3 anchor point)

$P(R) = 0.27R^{2} \exp(-0.09R^{3})$

(f)
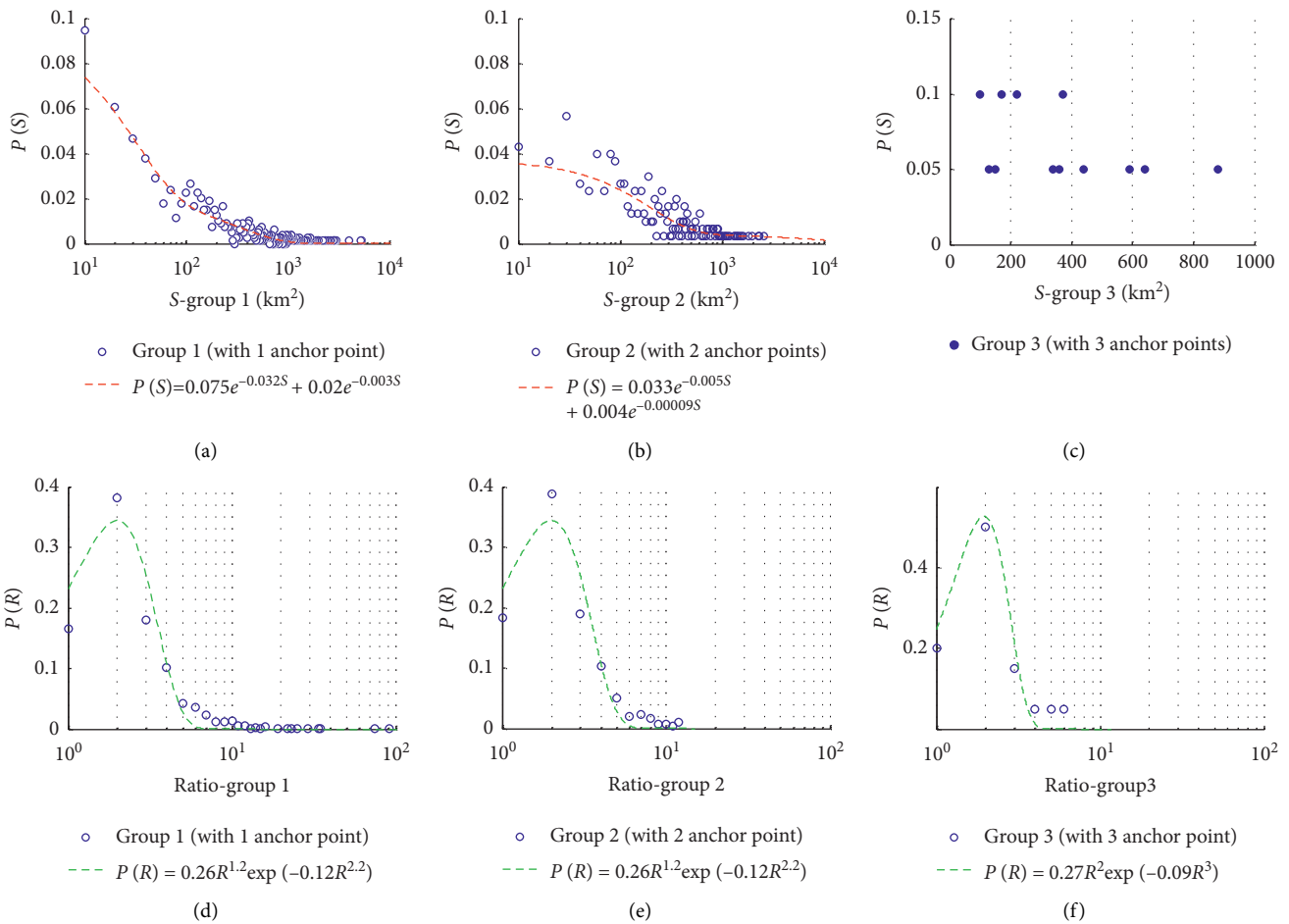
FIGURE 6: Distributions of area and major-minor axis ratio (D1).

(2) The major-minor axis ratio represents the ductility of the activity space. Intuitively, the shape of the individuals' activity space from group 2 should be more flat long than the shape from group 1, so the probability of a large ratio should be bigger for group 2. However, actually we found people in group 1 shared the same distribution of ratio with people in group 2, which can be well approximated by the Weibull model with 2 parameters (R means the major-minor axis ratio):
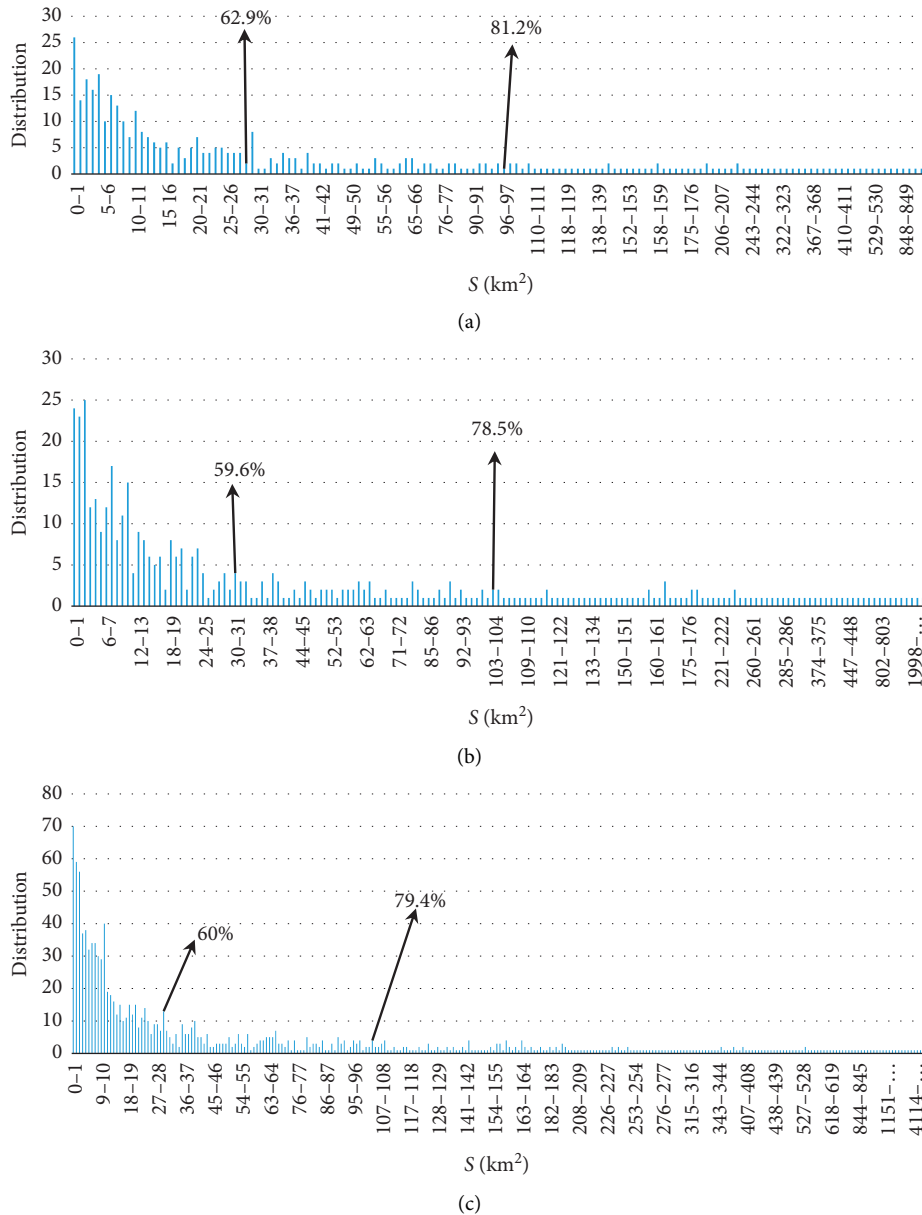
Figure 7: Distribution of the area of the samples: (a) group 1; (b) group 2; (c) the whole group.

$$P(R) = abR^{(b-1)} \exp\left(-aR^b\right), \qquad (2)$$

with exponents $a = 0.12$ and $b = 2.2$. This means group 1 and group 2 share the same potential of spatial anisotropy or sprawl direction. As for the group 3, the difference is the parameters: $a = 0.09$ and $b = 0.018$. This indicates a bigger probability in a round activity space. However, overall, the shape of most individuals' activity space is roundness (14%) or ellipse (57%, length-width ratio = 2 or 3); that is, a person in a society tends to travel in different directions rather than a single one.

The distribution of activity space size of the three groups of residents is very similar, almost obeying the distribution of long tail. Residents living within 30 square kilometers account for 62.9 percent and 59.6 percent, respectively, in group 1 and group 2 (Figure 7 shows distribution of the area of the samples). The vast majority of residents live within 100 square kilometers, that is, within 10 kilometers *10 kilometers, accounting for 81.2% and 78.5% in group 1 and group 2. For transportation planning and urban planning, the travel needs of residents within this scope must be considered in the planning stage. For example, in terms of subway planning, the influence of subway station setting on the residents around the station should not be considered only, but the travel needs of residents in the region should be considered. And how can residents have better travel experience. The same is true for road network planning. Besides expressway, how can other roads in the road network

$P(S) = 0.06e^{-0.016S} + 0.022e^{-0.004S}$

(a)

$P(R) = 0.26R^{1.3}\exp(-0.11R^{2.3})$

(b)
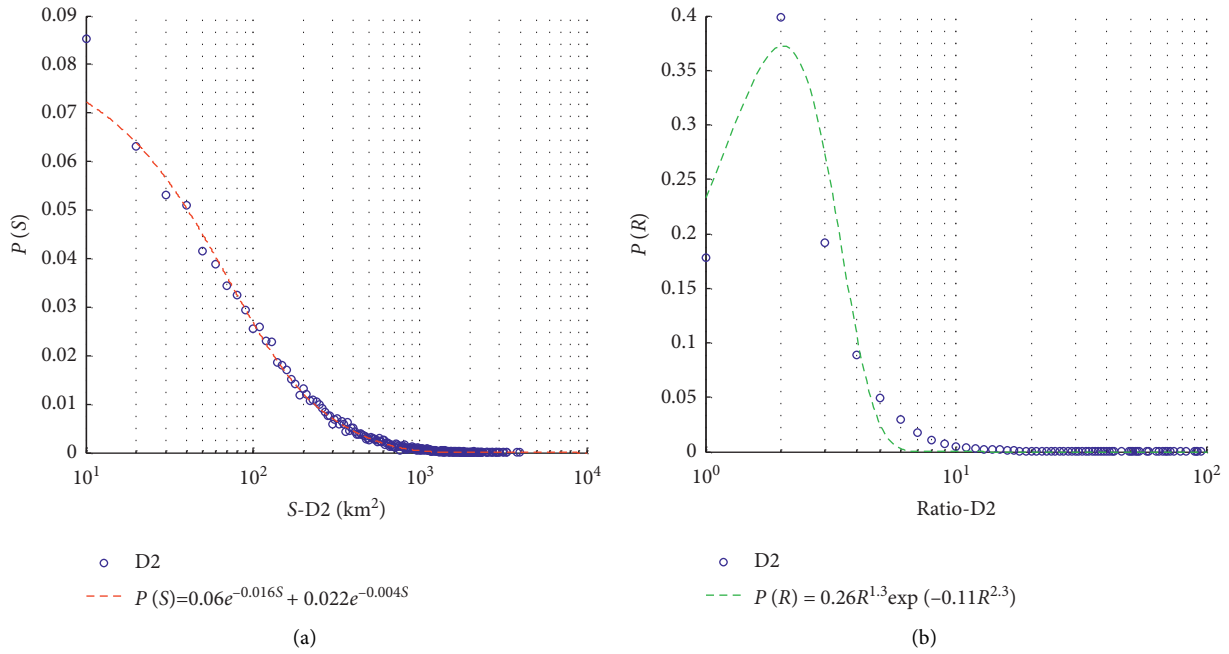
Figure 8: Distributions of area and major-minor axis ratio (D2).

(a)

(b)
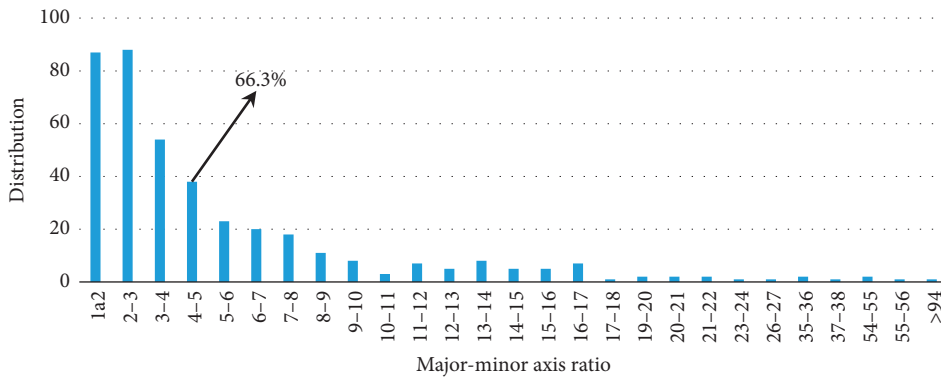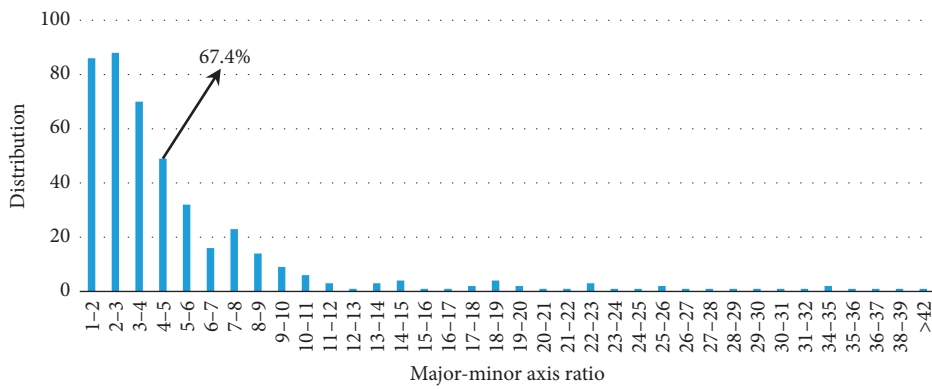
Figure 9: Continued.
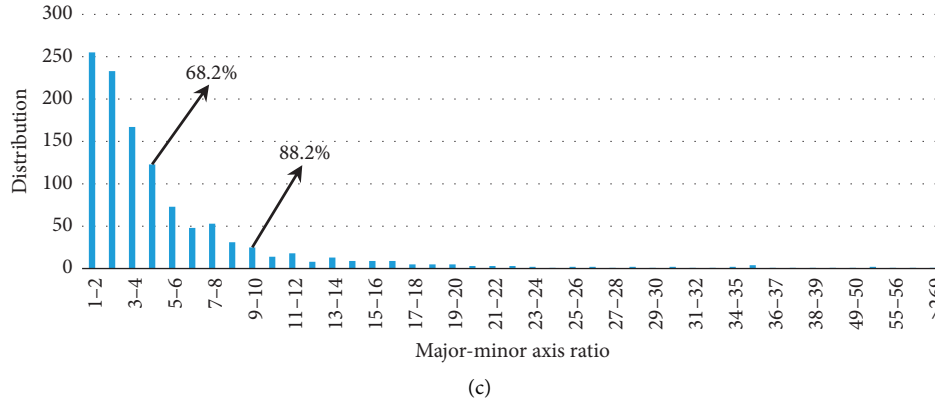
FIGURE 9: Distribution of the major-minor axis ratio: (a) group 1; (b) group 2; (c) the whole sample.

effectively provide sufficient support for residents' middle-distance travel.

As for the whole sample, the area of their activity range is also the same long tail distribution, with 60% of the population living within 30 kilometers and 79.4% living within 100 square kilometers.

As for the dataset D2, we use it as a comparison group to G1. Because we have known that the individuals from D2 live in the 3 communities long Line 7, this means they have one single anchor point: the home. The distribution of area and ratio also shows the consistency with G1: the exponential with close parameters $\alpha|D_2 = 0.06$, $\beta|D_2 = 0.022$, $a|G_1 = 0.016$, and $b|G_1 = 0.004$ and the Weibull model with close parameters $a = 0.11$ and $b = 2.3$. These parameters are extremely close (Figure 8).

This similarity between G1 and D2 means the consistency on the human activity space. So, we can say these people share the same characteristics in the spatial activity dimension. This also illustrates from sides the rationality of the rule to distinguish anchor points.

4.3. Directional Distribution of Individuals' Activity Space.
Similar to the distribution of activity space size of residents, the directional distribution of activity space has little difference. It can be seen in Figure 9 that the major-minor axis ratio of most individuals' activity space is within 5, respectively, 66.3% and 67.4% in group 1 and group 2. Very few residents have a major-minor axis ratio greater than 10. This shows the regularity of the activity space: the activity space of residents tends to extend in one direction with the major-minor axis ratio less than 5, and only a small number of residents have activity space that extends in the same direction and not in other directions.

This is of great research significance for transportation planning. First of all, individuals' daily travel presents a flat spatial structure, so in some traffic corridors and in the direction where some individual s travel actively, it is necessary to provide transportation support for the movement of these individuals, like subway or expressway construction. Second, there is more randomness to long-distance travel, which makes harder for transportation forecasting.
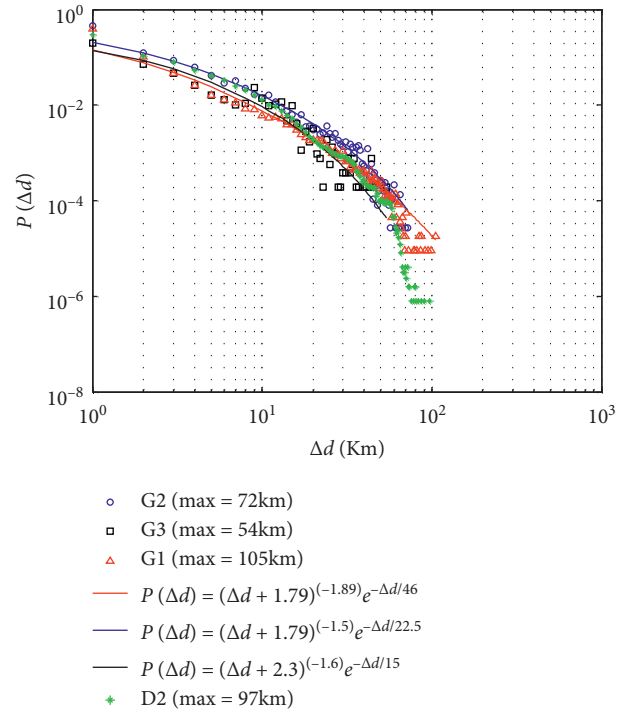
FIGURE 10: Distribution of travel distance.

4.4. The Distribution of Travel Distance.
As we know, people's travel distance or displacement can be well approximated by a truncated power-law distribution (11, 20):

$$P(\Delta d) = (\Delta d + \Delta d_0)^{-\beta} \exp\left(\frac{-\Delta d}{\kappa}\right), \quad (3)$$

where Δd is the displacement, β is the power-law exponent, and κ is the cut-off value. These values are different but close for different group of people (see in Figure 10).

Things should be different for the three groups separated by the number of anchor points. However, we found that all the three groups collapsed into almost the same distribution. The only different thing is the maximum distance: 105 km for G1, 72 km for G2, and 54 km for G3. And the maximum distance of dataset D2 is 97 km, which is close to G1. So, we

can say people usually travel in the same mobility pattern, except that people with fewer anchor points have to travel farther very occasionally.

5. Conclusions

Under the background of the big data, the relationship between urban built environment and individual's activity space is a necessity for urban strategic space policy. However, for the particularity of data environment, the representation of individual's activity space needs appropriate modification in the big data analysis. We need to extract effective indicators to describe individual's activity space based on spatial subordination relationships.

This study proposed a methodology to investigate the individuals' activity patterns based on the cellular signaling data (CSD). Compared with call detailed record (CDR) data, CSD contain more detailed information of an individual's activity trajectory. The anchor point theory was used in this study to investigate individuals' activity spaces due to the lack of personal properties and failing to identify the workplace. A convex polygon method was proposed to describe the shape of the activity space, with area to represent the size and the major-minor axis ratio to represent the ductility. The distributions of area and ratio can describe the human activity space comprehensively. For human mobility pattern, the important index—the travel distance—should also be discussed according to different groups. Individual activity space area is very important for traffic planning and urban planning. For previous research and the traditional traffic model, it is difficult to describe the individual activity space. In this study, mobile phone data can be used to calculate the activity space scope and form of residents, which can effectively describe the current situation of the city in the initial stage of urban traffic planning and provide strong data support for decision makers and urban planners.

The proposed methodology was tested with sample datasets collected in Shanghai. By setting a rule to recognize the anchor point, we separated the sample (D1) into 3 groups. Every group shows particular characteristics. Group 1 has the same distribution of major-minor axis ratio with group 2, but the probability with a large activity space is bigger for group 2. Group 3 tends to own a larger and more flat long activity space than the other two groups. People from D2 are very consistent with the people from group 1 in every aspect we mentioned in this paper. So, we can say these people belong to the same category from the perspective of activity space and human mobility.

This study is only the first step trying to use CSD to describe individuals' activity patterns. In future, a more scientific rule is needed to distinguish the anchor point with the normal activity point. The rule should include not only the time threshold but also the frequency of the contact between activity points. The present results of the study indicate that individuals with different number of anchor points have various mobility patterns and activity spaces. This will give us a better understanding of the individuals' intrinsic characteristics temporally and spatially in transportation area. Follow-up work will include the following: the sample size will be exaggerated to the whole Shanghai to study the morphology and spatial structure of the whole city; the relationship between the built environment and the living space is considered, and the differences of activity space of different residents under the same built environment and the differences of the same type of residents under different built environment are discussed.

Data Availability

The data used to support the findings of this study are restricted in order to protect patient privacy. Data are available from Shaofei Song (1610764@tongji.edu.cn) for researchers who meet the criteria for access to confidential data.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] D. Mi, "Sustainable metropolitan growth strategies: exploring the role of the built environment," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010.

[2] A. Z. Khan, "Towards sustainable built environment: understanding sustainability prospects in a metropolitan framework," *Journal of Research in Architecture & Planning*, vol. 11, pp. 31–55, 2011.

[3] F. T. Demdime, "Integrating public transport networks and built environmentThe Case of Addis Ababa and Experiences from Stockholm," Master's thesis, KTH-Royal Institute of Technology, Stockholm, Sweden, 2012.

[4] X. Wang, "How the built environment influences driving: insights from global positioning system data," Dissertations & Thesis-Gradworks, University of Michigan, Ann Arbor, MI, USA, 2010.

[5] A. Nasri and L. Zhang, "Assessing the impact of metropolitan-level, county-level, and local-level built environment on travel behavior: evidence from 19 U.S. Urban areas," *Journal of Urban Planning & Development*, vol. 141, no. 3, 2014.

[6] S. Handy, X. Cao, and P. Mokhtarian, "Correlation or causality between the built environment and travel behavior? Evidence from northern California," *Transportation Research Part D: Transport & Environment*, vol. 10, no. 6, pp. 427–444, 2005.

[7] X. Cao, S. L. Handy, and P. Mokhtarian, "The influences of the built environment and residential self-selection on pedestrian behavior: evidence from Austin, TX," *Transportation*, vol. 33, no. 1, pp. 1–20, 2006.

[8] P. Collins, Y. Al-Nakeeb, A. Nevill, and M. Lyons, "The impact of the built environment on young people's physical activity patterns: a suburban-rrural comparison using GPS," *International Journal of Environmental Research & Public Health*, vol. 9, no. 9, pp. 3030–3050, 2012.

[9] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin-destination matrices using mobile

phone call data," *Transportation Research Part C: Emerging Technologies*, vol. 40, no. 1, pp. 63–74, 2014.

[10] Q. Lin, "Research on decision support system of urban transport planning based on mobile data," in *Proceedings of the 5th China Intelligent Transportation Conference*, Beijing, China, 2009.

[11] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[12] S. Hotei, S. Secci, S. Sobolevsky, G. Pujolle, and C. Ratti, "Estimating real human trajectories through mobile phone data," in *Proceedings of the 2013 IEEE 14th International Conference on Mobile Data Management*, pp. 148–153, Milan, Italy, 2013.

[13] X. Zuo, Y. Zhang, and C. Feng, "A compute method of road travel speed based on mobile phone handover location," *Journal of Networks*, vol. 7, no. 10, pp. 1639–1645, 2012.

[14] J. Li, "Research on acquisition method of dynamic OD matrices based on mobile phone positioning data," Master's Thesis, Kunming University of Science and Technology, Kunming, China, 2013.

[15] H. Liu, "Traffic prediction and path selection for urban road networks based on data collected by mobile phones," Master's Thesis, Kunming University of Science and Technology, Kunming, China, 2014.

[16] Z. Guan, "Research on urban transport planning, designing, construction, management decision-making support based on mobile data," in *Proceedings of the Seventh Annual China Intelligent Transportation*, Changsu, China, 2013.

[17] P. Mokhtarian and X. Cao, "Examining the impacts of residential self-selection on travel behavior: a focus on methodologies," *Transportation Research Part B: Methodological*, vol. 42, no. 3, pp. 204–228, 2008.

[18] D. Wang and J. Zhang, "The analysis of consumer trip characteristics and spatial structure of commercial facilities in Shanghai," *City Planning Review*, vol. 25, no. 10, pp. 6–14, 2001.

[19] C. Song, Z. Qu, A. L. Barabási, and N. Blumm, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[20] G. D. Lorenzo and F. Calabrese, "Identifying human spatio-temporal activity patterns from mobile-phone traces," in *Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pp. 1069–1074, Washington, DC, USA, 2011.

[21] R. G. Golledge, "Spatial Behavior: A Geographic Perspective," *Economic Geography*, vol. 74, no. 1, p. 83, 1997.

[22] P. Widhalm, Y. Yang, M. Ul, S. Athavale, and M. C. González, "Discovering urban activity patterns in cell phone data," *Transportation*, vol. 42, pp. 597–623, 2015.

[23] W. Guo, Y. Zhang, Y. Chai, and Y. Shen, "Study on spatio-temporal figure of suburban residents based on gps data: a case study of Tiantongyuan and Yizhuang in Beijing," *Areal Research & Development*, vol. 32, no. 6, pp. 159–164, 2013.

[24] X. Cheng, "Research on travel characteristics and classification of urban residents based on mobile phone data," Doctoral Thesis, Tongji University, Shanghai, China, 2014.

[25] O. Järv, R. Ahas, and F. Witlox, "Understanding monthly variability in human activity spaces: a twelve-month study using mobile phone call detail records," *Transportation Research Part C: Emerging Technologies*, vol. 38, no. 1, pp. 122–135, 2014.

[26] Y. Zhao, Y. W. Chai, J. Chen, and M. Jun, "GIS-based analyzing method for spatial-temporal behavior data," *Geography and Geo-Information Science*, vol. 25, no. 5, pp. 1–5, 2009.

[27] Y. Chai and J. Shen, "Activity-based approach to human spatial behavior research," *Scientia Geographica Sinica*, vol. 28, no. 5, pp. 594–600, 2008.

[28] C. Kang, S. Gao, X. Lin, Y. Xiao, Y. Yuan, and Y. Liu, "Analyzing and geo-visualizing individual human mobility patterns using mobile call records," in *Proceedings of the 18th International Conference on Geoinformatics*, pp. 1–7, Beijing, China, 2010.

[29] X. Cheng, W. Li, F. Jia, and D. Yang, "Analyzing human activity patterns using cellular phone data: case study of Jinhe New Town in Shanghai, China," in *Proceedings of the Transportation Research Board 92nd Annual Meeting*, Washington, DC, USA, 2013.

[30] W. Li, X. Cheng, Z. Duan, D. Yang, and G. Guo, "A framework for spatial interaction analysis based on large-scale mobile phone data," *Computational Intelligence & Neuroscience*, vol. 2014, Article ID 363502, 11 pages, 2014.

[31] X. Li, "Research of Urban Residents Radius of Space Activities Based on Mobile Phone," Master's Thesis, Tongji University, Shanghai, China, 2014.

[32] R. N. Buliung and P. S. Kanaroglou, "A GIS toolkit for exploring geographies of household activity/travel behavior," *Journal of Transport Geography*, vol. 14, no. 1, pp. 35–51, 2006.

[33] N. Levine, "Spatial statistics and gis: software tools to quantify spatial patterns," *Journal of the American Planning Association*, vol. 62, no. 3, pp. 381–391, 1996.

[34] R. K. Rai, M. Balmer, M. Rieser, V. S. Vaze, S. Schönfelder, and K. W. Axhausen, "Capturing human activity spaces," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2021, no. 1, pp. 70–80, 2007.

Research Article

A Power Transformer Fault Diagnosis Method Based on Random Vector Functional-Link Neural Network

Qian Wang [iD],[1] **Shinan Wang** [iD],[1] **Rong Shi**,[2] **and Yong Li**[3]

[1]*School of Automation and Information Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710054, China*
[2]*State Grid Shaanxi Electric Power Company Economic Research Institute, Xi'an, Shaanxi 710065, China*
[3]*Trinity International Ltd., Chaoyang, Beijing 100022, China*

Correspondence should be addressed to Qian Wang; qianqian82@126.com

The random vector functional link (RVFL) network is suitable for solving nonlinear problems from transformer fault symptoms and different fault types due to its simple structure and strong generalization ability. However, the RVFL network has a disadvantage in that the network structure, and parameters are basically determined by experiences. In this paper, we proposed a method to improve the RVFL neural network algorithm by introducing the concept of hidden node sensitivity, classify each hidden layer node, and remove nodes with low sensitivity. The simplified network structure could avoid interfering nodes and improve the global search capability. The five characteristic gases produced by transformer faults are divided into two groups. A fault diagnosis model of three layers with four classifiers was built. We also investigated the effects of the number of hidden nodes and scale factors on RVFL network learning ability. Simulation results show that the number of implicit layer nodes has a large impact on the network model when the number of input dimensions is small. The network requires a higher number of implicit layer neurons and a smaller threshold range. The size of the scale factor has significant influence on the network model with larger input dimension. This paper describes the theoretical basis for parameter selection in RVFL neural networks. The theoretical basis for the selection of the number of hidden nodes, and the scale factor is derived. The importance of parameter selection for the improvement of diagnostic accuracy is verified through simulation experiments in transformer fault diagnosis.

1. Introduction

Urban construction and sustainable development have become the focus of modern social development. The sustainable development of a city is inseparable from the safe and stable operation of the urban power grid. An important part of city construction and planning is the power network. The power transformer is one of the most important components in the power system. A good fault diagnosis method for power transformers is very important to improve the reliability of transformers [1–7]. Dissolved gas analysis (DGA) method is an effective method for transformer fault diagnosis. DGA methods can be divided into two categories: conventional and intelligent methods, including critical gas, IEC ratio, Rogers ratio, Doernenburg ratio, etc., all of which detect faults by the amount of gas released in the transformer oil. Fault diagnosis in these methods is based on the decisions of human experts. They explain the different gas releases due to temperature and energy variations of the fault [8]. All of these methods are relatively simple to implement, but their percentage of fault detection is quite low and may not be sufficient to reliably determine the type of fault [9]. Poor fault diagnosis results prompted researchers to look for new techniques.

With the development of artificial intelligence, machine learning techniques are also widely used to solve complex classification and regression problems in power systems. Among them are artificial neural networks, Bayesian networks, and support vector machines. Using intelligent methods can achieve a considerably higher diagnostic accuracy than traditional similar methods. However, the complex and variable relationship between power

transformer fault symptoms and different fault types makes it difficult to establish simulation models for power transformer fault diagnosis and network training.

The Artificial Neural Network (ANN) diagnostic method has a high diagnostic capability and compensates for the shortcomings of the three-ratio method for determining mixed fault types [10–16]. However, the method suffers from overfitting and nonconvergence in the learning process [17]. Bayesian multicore learning method was proposed and applied in transient stability assessment, but the method requires large-scale training data and the threshold range is chosen empirically [18–21]. Least-squares support vector machine (LS-SVM) was applied for load prediction [22–28]. The main drawback is that the kernel functions are difficult to select and must satisfy the Mercer condition [29, 30]. The key parameters directly affect the generalization performance but lack a structured approach to achieve optimal parameter selection.

The random vector function link (RVFL) network is a single hidden layer feedforward neural network [21–33]. It is widely used in deep and transfer learning [34–37]. Also, it has good potential for handling large-scale data, fast dynamic modeling, and real-time data processing [38–40]. In recent years, RVFL network models have been used in classification and regression, but there are few studies involving the number of hidden nodes and scale factor values. Neural network models with few hidden nodes cannot guarantee modeling performance, but neural network models with more hidden nodes may be overfitted leading to poor generalization performance [41–45]. Over the past few years, some constructive solutions have received considerable attention. These solutions start with a smaller network and gradually generate hidden nodes and output weights until an acceptable learning performance is achieved [46, 47].

This paper examines common shortcomings in RVFL network data modeling and transformer fault diagnosis. Radbas activation functions and uniformly distributed unbiased network models are selected. This study explains the instability of the neural network model used in numerous literatures. The importance of randomly assigning ranges to implicit parameters in RVFL networks is illustrated by an arithmetic example that points out the validity and infeasibility of two-step generation of RVFL networks with randomly assigned inputs. Furthermore, modeling data on RVFL networks with fixed range settings is risky. The main contributions of this paper are summarized as follows:

Firstly, an improved RVFL network model is proposed to introduce the concept of hidden node sensitivity. Each hidden node is classified and the hidden nodes with low sensitivity are eliminated in Section 2. Secondly, a standardization method using characteristic inputs is proposed in Section 3. The proposed method of processing the input data ensures full utilization of the sample. Finally, the theoretical basis for the number of hidden layer nodes and the scale factor values of the RVFL neural network and the value rules are presented in Section 4. It is verified with a transformer fault diagnosis model.

2. The Improved Random Vector Functional Link Algorithm

2.1. The RVFL Algorithm. The feedforward neural network (FNN) is one of the most popular algorithms in small sample machine learning [48–51]. One of the most commonly used topologies is the single hidden layer feedforward neural network (SLFN). Neurons between adjacent layers (input layer to hidden layer and hidden layer to output layer) are interconnected, but there is no interconnection between neurons in nonadjacent layers. The algorithm has been widely applied in classification and regression problems.

The random vector functional link (RVFL) network is a single hidden layer feedforward neural network (SLFN) [52]. The red line in Figure 1 indicates the direct connection between the input neuron and the output neuron. The RVFL combines the advantages of random weights and function chains and can be used to solve classification, regression, and prediction problems in many fields. In addition, RVFL networks can be combined with other learning methods to generate hybrid algorithms.

The RVFL network is a random weighted neural network. Suppose the output $y \in R$, the RVFL neural network can be described as a weighted sum of the outputs of the nodes in the L hidden layer:

$$y = \sum_{m=1}^{L} \beta_m \mathbf{h}_m(\mathbf{x}; \mathbf{w}_m) = \beta^T \mathbf{h}(\mathbf{x}; \mathbf{w}_1, \ldots, \mathbf{w}_L). \quad (1)$$

Here β is the output weight, and the input is a d-dimensional real vector $x \in R_d$. The mth transform is parameterized by the vector $\mathbf{w}_m$. Here, $h: x \longrightarrow R$ is called a base. The sigmoid activation function is used in the simulation:

$$h_k(i, w) = f(x_k) = \exp\left\{-\left(\sum_{m=1}^{M} w_{m,k} i_m^{(n)}\right)^2\right\}. \quad (2)$$

At the beginning of the learning process, $w_1, \ldots, w_B$ are selected at first. In the RVFL neural network, parameters of the implicit function are randomly selected from a predefined probability distribution. Assuming that the basis function is continuous and stationary, the general approximation ability of RVFL can be guaranteed if L is large enough. By using this method, equation (1) can be transformed into a linear regression equation for coefficients $\beta = \beta_1, \beta_2, \ldots, \beta_L$. If the required function for N samples is provided, it is called a training set, $S = \{x_i, y_i\}$, $i = 1, 2, \ldots, N$.

$$\mathbf{H} = \begin{pmatrix} h_1(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_N) \end{pmatrix}. \quad (3)$$

The optimal parameter β can be obtained by using a standard regular least squares method

$$\beta^* = \arg\min_{\beta \in R^L} \frac{1}{2}\|\mathbf{H}\beta - \mathbf{Y}\|_2^2 + \frac{\lambda}{2}\|\beta\|_2^2. \quad (4)$$
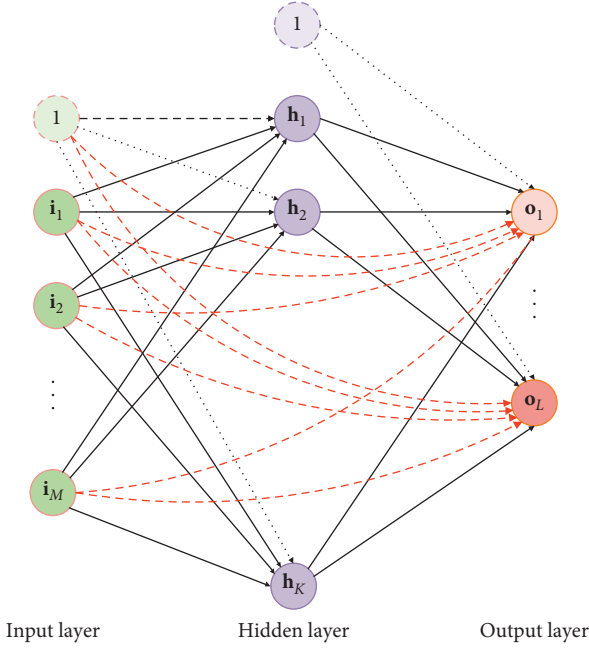
FIGURE 1: Schematic diagram of an RVFL network.

Let the gradient of $J(\beta)$ be zero, which is

$$\frac{\partial \mathbf{J}}{\partial \beta} = \mathbf{H}^{\mathbf{T}}\mathbf{H}\beta - \mathbf{H}^{\mathbf{T}}\mathbf{Y} + \lambda\beta = 0. \tag{5}$$

The optimal parameter can be written as

$$\beta^* = \left(\mathbf{H}^{\mathbf{T}}\mathbf{H} + \lambda\mathbf{I}\right)^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{Y}. \tag{6}$$

The parameter I is the identity matrix. The inverse of the $L \times L$ matrix affects the computational complexity of RVFL. In the case of $N \ll L$, for any $\lambda > 0$, it can be simplified by the following equation:

$$\left(\mathbf{H}^{\mathbf{T}}\mathbf{H} + \lambda\mathbf{I}\right)^{-1}\mathbf{H}^{\mathbf{T}} = \mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{H}^{\mathbf{T}} + \lambda\mathbf{I}\right)^{-1}. \tag{7}$$

By combining equations (6) and (7), the optimal weight β can be obtained.

$$\beta^* = \mathbf{H}^{\mathbf{T}}\left(\mathbf{H}\mathbf{H}^{\mathbf{T}} + \lambda\mathbf{I}\right)^{-1}\mathbf{Y}. \tag{8}$$

The RVFL algorithm is simple and has short training time, but it is also weak in computational complexity and generating ability. The concept of hidden node sensitivity is introduced, and each hidden node is classified according to its sensitivity. Finally, several hidden nodes with the lowest sensitivity are removed, which leads to a simplified network structure. This method has been applied to ELM [53–55]. This paper uses the same idea to optimize and improve RVFL.

2.2. Description of the Improved RVFL. The sensitivity of hidden nodes is the space distance between the output results before and after removing hidden nodes.

$$E(l, i) = \left\|f_{\text{bef}} - f_{\text{aft}}\right\| = \left|h_{li}\right|\left\|\beta_l\right\|. \tag{9}$$

If $E(l, i)$ is larger, the greater the sensitivity. $E(l, i)$ is only based on one sample, which is not enough to measure the real state. Therefore, it needs a large amount of samples, which enables it to fully show its change law and trends. As a result, the equation could be transformed to

$$E_{\text{avg}}(l) = \frac{1}{N}\sum_{i=1}^{N}\left|h_{li}\right|\left\|\beta_l\right\|. \tag{10}$$

$E_{\text{avg}}(l) \longrightarrow +\infty$ shows the sensitivity is the greatest. For different practical applications, the sensitivity of hidden nodes cannot be decided only by the value of $E(l, i)$ due to the uncertainty of its value. For instance, its sensitivity is very large in some programs, but very small in others. Therefore, a novel solution is proposed in this paper.

Given a training sample (x_i, y_i), $i = 1, \ldots, N$, the sensitivity based on each sample is $E(l, 1)$, $E(l, 2)$, $\ldots$, $E(l, N)$. Therefore, the sensitivity of the hidden node can be represented by a point Se, where Se $= (E(l, 1), E(l, 2), \ldots, E(l, N))$. In practical applications, the Se cannot be directly measured. So the Murkowski distance between Se and space original point O can be used when $q = 2$, which is the Euclidean distance.

$$d(\text{Se}, O) = \left(\sum_{i=1}^{N}\left|E(l, i) - O_i\right|^q\right)^{1/q} = \left(\sum_{i=1}^{N}\left|E(l, i)\right|^q\right)^{1/q}. \tag{11}$$

Se is similar to $E_{\text{avg}}(l)$. However, it is inconsistent with people's cognitive habits. Therefore, it can be transformed to

$$\text{Sen}(l) = e^{-\alpha/d(\text{Se}, O)} = e^{-\alpha\|\beta l\|_q^{-1}\left(\sum_{i=1}^{N}|h_{li}|^q\right)^{-(1/q)}}. \tag{12}$$

$\text{Sen}(l) \in (0, 1)$, α is the adjustment factor governing the sensitivity distribution. $\text{Sen}(l) \longrightarrow 1$ indicates higher sensitivity.

After the above detailed analysis, the description of improved RVFL is shown in Figure 2:

(1) According to the specific problem, the threshold parameters ε and ε_{Se} are initialized first, and then the SLFN with sufficient hidden nodes is established through the RVFL. Finally, the output weight β and the output matrix of the hidden node are obtained according to equation (8).

(2) For each input training sample, the sensitivity of each hidden node $E(s, i)$ is obtained using equation (9), and then the sensitivity $\text{Sen}(l)$ is calculated using equation (12).

(3) For current input samples x_{N+1}, the output of SLFNS y_{N+1} can be obtained at a given threshold ε. If the error of the actual output y'_{N+1} and the expected output y_{N+1} are satisfied with SSE $(y_{N+1} - y'_{N+1}) < \varepsilon$, it can remove some hidden nodes with lower sensitivity and go to step (4). Otherwise, it should be prepared for prediction or classification based on the next new sample.
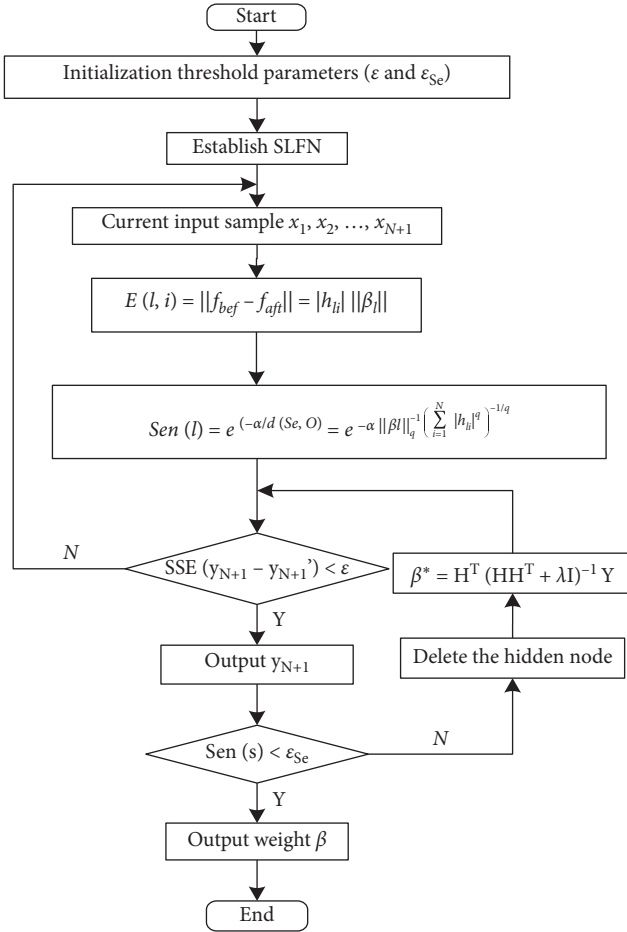
FIGURE 2: Hidden layer node improvements.

(4) For a given threshold ε_{Se}, if the sensitivity of the Sth hidden node satisfies $Sen(s) < \varepsilon_{Se}$, the hidden node can be deleted.

(5) After deleting the hidden node, update the output weight with formula (8), and continue to analyze whether there is a hidden node with low sensitivity, and go to step (3).

3. Design of Transformer Fault Diagnosis Model Based on Improved RVFL Neural Network

The fault mechanism of oil-immersed power transformers is rather complicated and there are many types of faults. In fact, no matter what the cause of the failure is, it can ultimately be attributed to two factors: thermal or electrical. Therefore, the transformer fault analysis can be divided into five states of operation: normal state, low energy discharge, high energy discharge, low temperature heat generation, and high temperature overheating. Transformer fault analysis is essentially a multi-classification problem. However, the RVFL-based classification algorithm involves only two classification problems. Therefore, it is necessary to convert multiple classifications into two classifications.

In the event of a power transformer failure, the transformer oil will decompose to produce many hydrocarbon molecules, including hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2). Therefore, five kinds of gases mentioned above are analyzed.

3.1. Fault Diagnosis Flow. After determining the feature variables and sample data, and building a transformer fault diagnosis model, the classification model can be learned on the training set according to Figure 3. In addition, four classifiers can be constructed. The diagnosis process is shown in Figure 3:

The four RVFL classifiers are constructed in the following steps. Firstly, the activation function is chosen as a Gaussian kernel function and a suitable number of nodes in the hidden layer are selected. Then weights and biases are calculated to adjust the inputs, while the input-output relationship of the diagnostic model is obtained based on the correlation vector and the correlation function. Finally, the training is completed and the test data set is tested.

The random vector function chain neural network (RVFL) is simple in structure and training. The nonlinear problem in transformer faults is reasonably solved with a very fast convergence rate using the strong learning ability and the principle of approximating the minimum value in the RVFL network model.

3.2. Diagnostic Model Based on RVFL Network. In this paper, a diagnostic model is built using a binomial tree approach to transform the five classification problems into a three-level, four-classification problem. Each classifier is binary and the model is shown in Figure 4. Each set of input data of the RVFL neural network model is five-dimensional, representing five characteristic gases. The output data are one-dimensional. The five-dimensional inputs correspond to a set of one-dimensional outputs. The output type is normal or one of the four fault types.

The RVFL network model has three layers. It contains four classifiers and five different running states. A classifier 1 (RVFL1) in the first layer is used to diagnose whether the power transformer is in a fault or normal state. Classifier 2 (RVFL2) located in the second layer is responsible for determining whether the specific fault of the transformer in the fault state is a thermal or electrical fault. The third layer has two classifiers. Classifier 3 (RVFL3) is used to distinguish between low-energy discharge and high-energy discharge faults of transformers with electrical faults. Classifier IV (RVFL4) is responsible for the diagnosis of specific faults of power transformers in the thermal fault state, such as low temperature heating faults or high temperature heating faults. The stage value of each classifier is set up, as shown in Table 1. It should be pointed out that each classifier is independent of each other when constructing the model. Since each binary classifier is independent of each other, they can perform their own classification without interfering with each other.

3.3. Selection of Characteristic Variables. Dissolved gas samples in transformer oil usually have a large variation and dispersion due to differences in voltage levels and
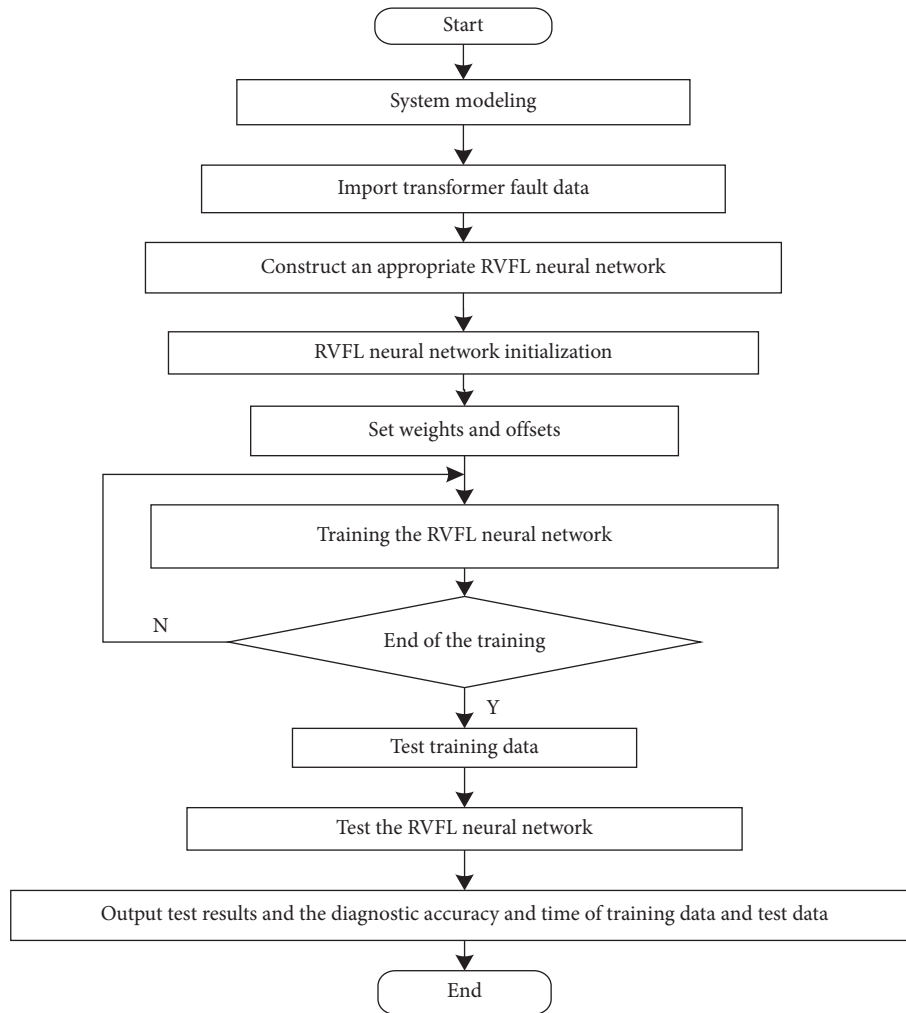
FIGURE 3: Flow chart of transformer fault diagnosis based on the RVFL network.
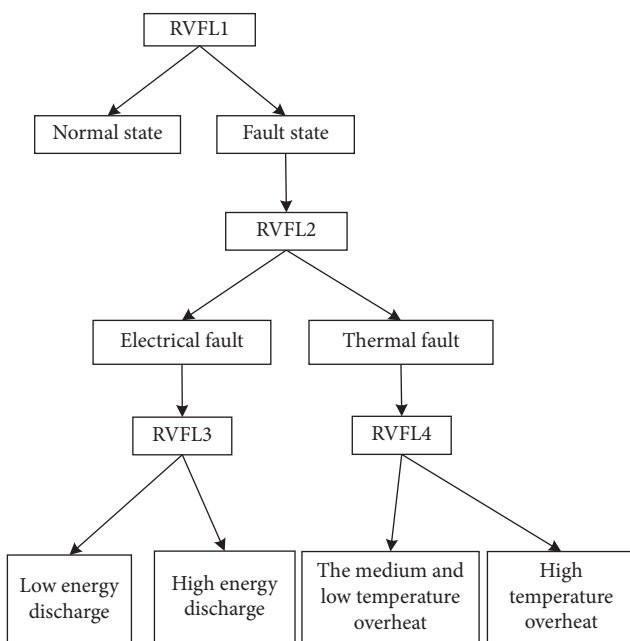
FIGURE 4: Fault diagnosis model of power transformer based on the RVFL.

transformer capacities. In some specific cases, the model is not sensitive to smaller inputs. If raw data are used directly as inputs, some important information will be missed. In order to improve the accuracy of diagnosis, different neural network models are built for different feature variables. In addition, more effective information needs to be fully utilized to reduce the effect of magnitude differences. It should be noted that the characteristic gas samples need to be normalized as inputs. Therefore, there are two scenarios for selecting the feature variables.

3.3.1. Characteristic Gas Content Ratio as a Characteristic Variable. The ratio of the separated gas content in the five transformer oils was selected as the input characteristic. Obviously, there are some differences in the order of magnitude of each group of gas content data compared to the original characteristic gas content. Using it directly as the input data of the simulation model is likely to complicate the network model structure or overfit the test results.

To investigate the diagnostic value and runtime of the test results, the training set data were tested again. The model was trained using these data and the accuracy of the

TABLE 1: Classifier value settings.

Classifier	Target value	
RVFL1	Normal state: 1	Fault target: 0
RVFL2	Electrical fault: 1	Thermal fault: 0
RVFL3	Low energy discharge: 1	High energy discharge: 0
RVFL4	Medium and low temperature overheat: 1	High temperature overheat: 0

calculated results was compared with the test set data. The input gas content was preprocessed according to the requirements of the model parameters and results. Data with larger dimensions were normalized based on the choice of scale factors. The significant effect of dimensional differences on the test results was eliminated.

The ratios of the five gases of H_2, CH_4, C_2H_6, C_2H_4, and C_2H_2 to the total gas content are selected as the characteristic variables. Because the ratio of the five characteristic gas contents is between 0 and 1, it does not need to be standardized and can be directly used as input data.

3.3.2. Standardized Processing of Characteristic Gas Content as a Characteristic Variable. There is a large difference between the data sets in terms of raw characteristic gas content. If the characteristic content is used directly as input data, it is not only complicated to operate but also the fault diagnosis accuracy is low. In this paper, the minimum distance obtained by normalized least squares method is used to normalize the raw characteristic gas content, which avoids the significant influence of input data with different dimensions on the simulation model.

$$X_{\text{new}} = C_{\min} + \frac{X - X_{\min}}{X_{\max} - X_{\min}} (C_{\max} - C_{\min}), \qquad (13)$$

where X_{new} is the gas content after standardization. $X_{\min}$ and $X_{\max}$ are the minimum and maximum value of the gas content, respectively. $C_{\min}$ and $C_{\max}$ are the upper and lower bounds of the gas content standardization, respectively.

Different neural network models are built for different feature variables to obtain different diagnostic results. In order to reduce the effect caused by the size difference, more effective information can be fully utilized. Assume that we have n samples, where $X_n = \{x_{n1}, x_{n2}, x_{n3}, x_{n4}, x_{n5}\}$. Five of them represent the input quantities of five gases, namely, hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) [56]. Each set of input data has five dimensions and the output data have one dimension. Depending on the five characteristic gases of the inputs, each output represents a fault state or a normal state of the transformer output.

The diagnostic results of the two models with feature variables as input data will be compared in the simulation. More effective information will be fully explored by analyzing the requirements and relationships of the relevant parameters of the RVFL network model under different input types.

3.4. Selection of Sample Data. In the establishment of any neural network model, in order to achieve a good fault

diagnosis effect, the most representative and universal sample data should be selected. When selecting the type of dissolved gas in the transformer oil, the sample data should contain all the data in the operating state.

In addition, try to satisfy each run with roughly the same proportion of input data. However, because of the influence of the transformer internal and external uncertainties, the real amount of dissolved gas in transformer span is greater than the training sample span. Hence, small sample training cannot meet the practical requirements of transformer fault diagnosis, and large sample size training sample spans a wide range of existence. The problem of sample dispersion will reduce the generalization ability of the neural network. Therefore, for samples with large amount of data and discrete order of magnitude of features, the input feature scaling is used to normalize the sample data.

When the network model is tested, it is also necessary to include every running state in the test data. The proportion is basically the same. The data volume of the training set and the test set are divided in a ratio of roughly 2 : 1. Each data set should contain transformer data in various operating states. And each run state has a roughly equal share of each data set.

4. Simulation Results

4.1. The Experimental Setup. To verify the effectiveness of the proposed algorithm, four independent fault detectors with no direct relationship between diagnostic results are used to identify four transformer states. The RVFL1, RVFL2, RVFL3, and RVFL4 classifiers are used to distinguish normal from fault, thermal from electrical fault, low-energy from high-energy discharge, and low-temperature from high-temperature fever, respectively. The simulation results and diagnostic accuracy α of each classifier will be introduced separately below (α = the correct number of test samples/the total number of test samples).

First, the scale factor values are determined for different sample data and input characteristics. Considering the effects of model structure and input parameters, the model produces more accurate results when diagnosing the training set. However, the model has higher error in the results when diagnosing the test set data. Therefore, it is necessary to preprocess the input data to ensure the validity of the experiment and the accuracy of the simulation results.

In this paper, the RVFL neural network normalizes the input feature data before diagnosis. Data standardization avoids large network diagnostic errors caused by large numerical differences between input data and output data. Secondly, network diagnostics are not always faster and more accurate when using a scaling range of [−1, 1], and optimal performance is not always guaranteed. Therefore, in

order to achieve the desired order of magnitude of input sample data in each dimension and to ensure the validity of RVFL neural network diagnosis, simulation was carried out to study the input characteristics in two different cases, that is, the content characteristic variables of the five characteristic gases as input data and the content characteristic variables of the characteristic gases as standardized input data. After that, the number of hidden layer neurons under different sample data and input characteristics was determined. The number of nodes in the hidden layer has an important influence on the accuracy of model diagnosis during the construction of the RVFL neural network. The experiment was performed by separately changing the characteristics of two different variables as input data for the number of hidden nodes. Observe the data model of transformer fault diagnosis with the same number of hidden layer nodes in different neural networks. Record the running time of training data and test data, and diagnose the difference in accuracy of test data set and test data set diagnosis.

As a case study, the sample data set contains 162 groups of content data of five characteristic gases of transformer during operation, namely, hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), and acetylene (C_2H_2) [18]. Among them, there are 64 groups of characteristic gas content data in normal operation state and 98 groups of data in fault operation state. In this simulation experiment, 162 samples of characteristic gas content extracted from 10 kV power transformer oil were provided by our cooperative enterprise.

4.2. Analysis of Test Results

4.2.1. RVFL1 Simulation and Test Results. In order to ensure the reliability of learning, 120 sets of data were randomly selected from 162 sample data as training sets. The remaining 42 sets of data serve as test sets. When the characteristic variable is the characteristic gas content, the value of the scaling factor S is adjusted to obtain the new simulation results, as shown in Figure 5.

It can be seen from Figure 5 that the expected purpose can be achieved by adjusting the scale of the two data sets on the premise that the testing process of the training set and the test set is not affected. The total amount of the training data set is appropriately expanded, so it can be seen that the scaling factor $S = 10$ is the optimal threshold for diagnostic accuracy and running time. Therefore, when the input data are the transformer gas content, the selected scale factor $S = 10$ is the best value of the existing data set, and its random range $[-S, S]$ is $[-10, 10]$.

When the characteristic variable is the characteristic gas content, adjust the value of N, select the optimal number of hidden layer nodes, and sort out selected partial results, as shown in Figure 6.

The observation of the test data shows that starting from $N = 1$, the diagnostic accuracy of the RVFL neural network for both the training and test sets grows rapidly in a stepwise manner as the number of hidden nodes increases. When the number of hidden nodes is $13 = N < 20$, the diagnostic
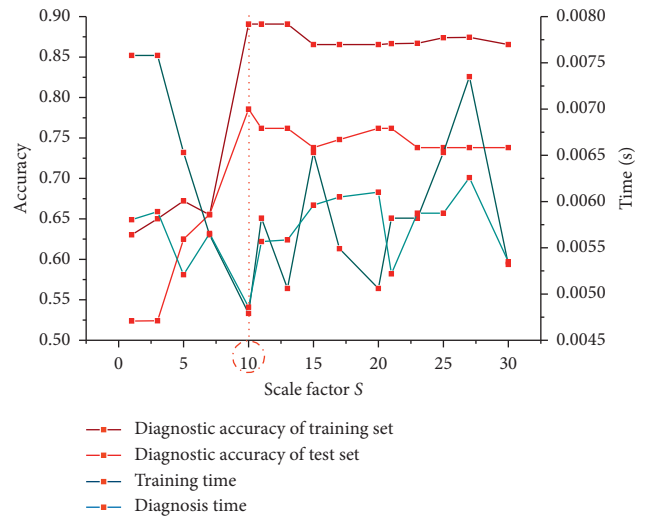
FIGURE 5: Selection of scale factor S of classifier RVFL1.
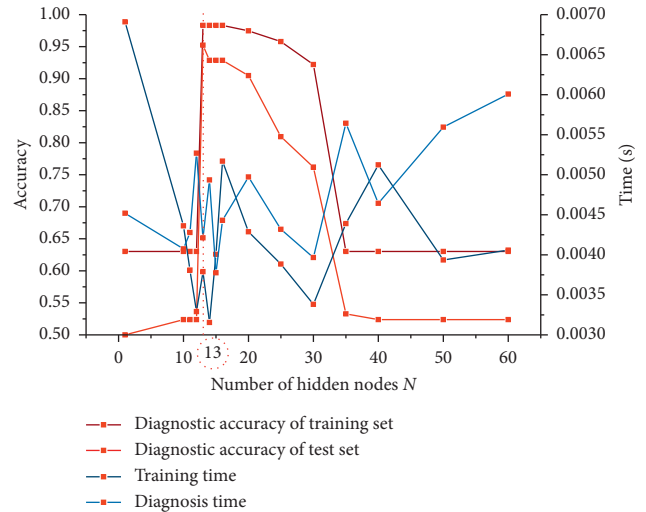
FIGURE 6: Selection of hidden layer node N in classifier RVFL1.

accuracy reaches a saturation peak for both the training and test sets, and the running time is slightly higher than the other hidden node numbers. When the hidden nodes $N \geq 20$, the network diagnostic capability gradually decreases. Considering that too many neurons in the hidden layer will complicate the network structure, prolong the convergence time, and even result in "overfitting," we choose the number of nodes $N = 13$ in the hidden layer as the best choice for the available sample data set.

When the characteristic variable is the standardized characteristic gas content, the value of the scaling factor S is adjusted to obtain the new simulation result, as shown in Figure 7.

It can be seen that when the scaling range increases, the accuracy of the RVFL neural network diagnostic training set changes greatly, while the data of the diagnostic accuracy test set changes less. When the scaling factor was greater than 2.1, it remained basically the same. After the diagnostic accuracy of the training set data reached the peak, the scale

factor value was greater than 4, which was difficult to adjust due to the small sample data set. In the experiment, if the value of S continues to increase, the diagnostic accuracy of the training set reaches a relatively high value floating around 0.9 after $S = 500$, but the diagnostic accuracy of the test set data does not improve. At the same time, considering the impact of running time, the scale factor $S = 2.1$ is selected as the best value of the existing data set, and its random range $[-S, S]$ is $[-2.1, 2.1]$.

When the characteristic variable is the standardized characteristic gas content, adjust the value of N, select the optimal number of hidden layer nodes, and sort out selected partial results, as shown in Figure 8.

When the input feature is the normalized gas content, the diagnostic accuracy of changing the number of hidden nodes is less than that of the feature vector being the characteristic gas content. The effect of this on the diagnostic accuracy of the test set is almost negligible. At the same time, it can be concluded that when the number of nodes in the hidden layer reaches 130, the RVFL network shows an obvious "overfitting" phenomenon. The diagnostic accuracy of trained data is high but the diagnosis of untrained data is often unsatisfactory. Therefore, considering the running time and diagnostic accuracy, the number of nodes in the hidden layer $N = 100$ is chosen as the optimal value for the existing data set.

The optimal parameters tested were $N = 13$, $S = 10$ for the characteristic gas content as a feature variable and $N = 100$, $S = 2.1$ for the normalized gas content as a feature variable. N is the number of nodes in the hidden layer and S is the scale factor. The specific process is not repeated here. Table 2 shows the diagnostic results for the characteristic gas content ratio and the results after normalizing the characteristic variable gas content. The results of the partial test of the RVFL1 classifier are shown in Table 2:

Fault type 1 is the normal state, and fault state 0 is the fault state. Yt_t and Y_t denote the network diagnosis results of the test set before and after the standardization process. The trY represents the true failure type of the test set. The diagnostic accuracy is 95.24% and 97.99%, respectively, when the characteristic gas content ratio and the standard gas content ratio are used as the characteristic variables. The test times are 0.004952 seconds and 0.004212 seconds.

4.2.2. RVFL2 Simulation and Test Results.
The classifier RVFL2 contains a total of 98 sets of characteristic gas data for electrical transformer fault operation states. Among them, there are 55 sets of data in electrical fault operation state, and the remaining 43 sets of data in thermal fault operation state. Seventy of the 98 sample data are selected as training data. The remaining 28 groups are used as test data.

The input data transformer characteristic gas content was reduced due to the sample data when the number of hidden layer nodes was $N = 13$ and the random range was still $[-10, 10]$. The diagnostic accuracy of the training set in RVFL2 with 70 sets of sample data is 82.86% and the convergence time is 0.008460 seconds. The diagnostic accuracy of the test set in the remaining 28 sets of sample data
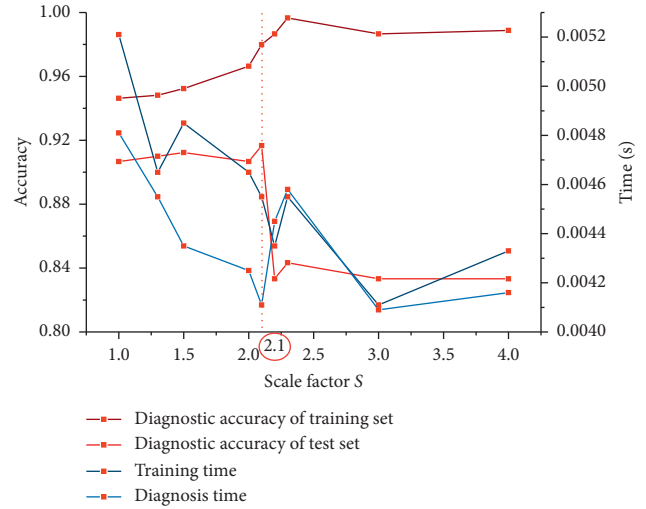
FIGURE 7: Selection of scale factor S of classifier RVFL1 after standardization.

FIGURE 8: Selection of hidden layer node N in classifier RVFL1 after standardization.

TABLE 2: Comparison of RVFL1 diagnosis results under standardized input characteristics.

	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	trY	Yt_t	Y_t
1	155	581	215	293	142	0	1	0
2	477	566	734	239	116	0	0	0
3	117	5.8	7	15.9	12.8	0	0	0
4	52.5	32.9	79.3	9.73	9.8	0	0	0
5	19.8	9.6	2.29	1.17	0	1	1	1
6	21.1	8.97	2.25	1.08	0	1	0	0
7	166.6	28.31	6.72	12.4	0.3	1	1	1

in RVFL2 is 73.57%, and the convergence time is 0.005415 seconds.

This shows that the simulation output is not satisfactory. From the results, it was found that the RVFL fault diagnosis model outputs all results as "electrical fault". The network fault diagnosis accuracy for both training data and test data

is not satisfactory. Therefore, it is necessary to adjust the relevant parameters in the model to ensure the effectiveness of the RVFL network training. Attempts to adjust the number of nodes N in the hidden layer were made, and it was found that adjusting the number of hidden nodes could not improve the accuracy of fault diagnosis and had very little effect on detection. Therefore, we adjusted the randomization range.

When the characteristic variable is the characteristic gas content, the value of the scaling factor S is adjusted to obtain the new simulation results, as shown in Figure 9.

According to the data in the above table, the randomization range is adjusted, and it is found that when the scale factor is within the range of $1 = S \leq 15$, the detection accuracy keeps increasing. In addition, after the range is exceeded, the model gradually appears as an overfitting phenomenon. Therefore, $S = 15$ is selected as the optimal scale factor value of classifier RVFL2. Adjust the random range $[-S, S]$ to $[-15, 15]$ and keep the number of hidden layer nodes N unchanged.

When the number of hidden layer nodes is $N = 13$ and the initial value $S = 10$ and the optimal value $S = 15$ are selected by the scale factor, the diagnosis results of transformer fault by RVFL network are compared. Where Yt1_temp and Y1_temp, respectively, represent the diagnostic output results of the network to the training set and the test set when the randomization range $[-S, S]$ is the initial value $[-10, 10]$. Yt2_temp and Y2_temp, respectively, represent the optimal threshold range $[-S, S]$ of debugging randomization $[-15, 15]$ and the diagnostic output results of the network to the training set and the test set. The diagnostic output results of the network to the training set and the test set are shown in Table 3.

The simulation test results with the number of hidden nodes $N = 13$ and the scale factor $S = 10$ are compared with the test results with the scale factor S adjusted to 15. When the classifier RVFL2 adjusted the parameters to $N = 13$, $S = 15$, the diagnostic accuracy of 70 sets of data in the training set of classifier RVFL2 was 94.29%, and the convergence time was 0.003247 seconds. The accuracy of 28 sets of test set data was 90.71% and the test time was 0.003544s.

The scaling range $[-15, 15]$ was set according to Table 3. The detection accuracy of the training data is improved by 11.43% and the detection accuracy of the test data is improved by 17.14% after parameter optimization. It can be seen that the parameter tuning of the RVFL2 classifier achieves the expected effect when the input feature is the characteristic gas content of the transformer.

The optimal parameters for the tested characteristic gas content as a feature variable are $N = 13$, $S = 15$, and the optimal parameters for the normalized gas content as a feature variable are still $N = 100$, $S = 2.1$. N is the number of hidden layer nodes and S is the scale factor. Table 4 shows the diagnostic results for the characteristic gas content ratio and the results after normalizing the gas content of the characteristic variables. Some test results of the RVFL2 classifier are shown in Table 4.

Failure type 1 is electrical failure, and 0 is thermal failure. The letters mean the same as above. The diagnostic accuracy
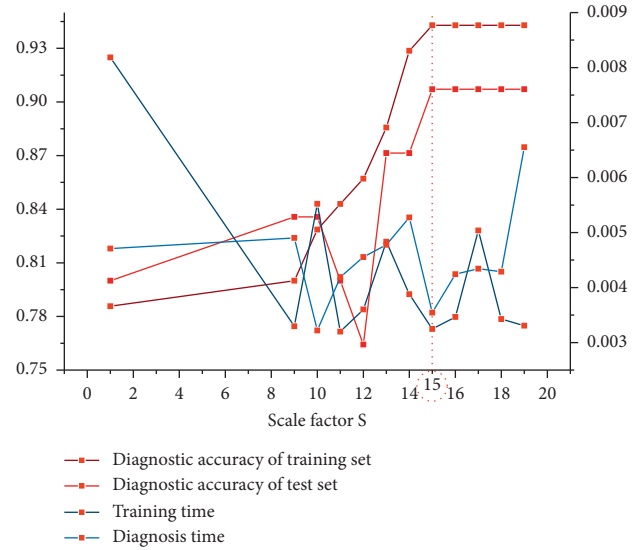
FIGURE 9: Selection of scale factor S of classifier RVFL2.

TABLE 3: Comparison of RVFL2 diagnosis results.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	1	1	1	1	1	0
2	1	1	0	1	0	0
3	1	1	1	1	1	1
4	1	0	1	1	1	1
5	1	1	1	1	1	1
6	1	1	1	1	1	1
7	1	1	0	1	1	0
8	1	1	1	1	1	1
9	1	0	0	1	1	1
10	1	0	1	1	1	1
11	1	1	0	1	1	1
12	0	0	0	1	1	0
13	1	0	0	1	0	0
14	1	1	1	1	1	1
15	1	1	0	1	1	0

TABLE 4: Comparison of RVFL4 diagnosis results under standardized input characteristics.

	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	trY	Yt_t	Y_t
1	117	5.8	7	15.9	12.8	1	1	0
2	52.5	32.9	79.3	9.73	9.8	1	0	1
3	421	135	351	27.7	374	1	1	1
4	46	98	26.3	41.3	0	0	0	0
5	25	85	178	9.6	1.4	0	0	0
6	56.4	42.7	1413	167	17.1	0	1	0
7	74.8	482	1409	297	7.5	0	1	0

was 90.71% and 93.10%, respectively. The test time was 0.002975 seconds and 0.004761 seconds, respectively.

4.2.3. RVFL3 Simulation and Test Results. The classifier RVFL3 contains a total of 55 groups of characteristic gas data of power transformers under the operation state of electrical faults, among which 22 groups are gas content data under

the operation state of low energy discharge, and the other 33 groups are gas content data under the operation state of high energy discharge. Take any 40 of the 55 sample data as training data. The remaining 15 groups are used as test data. The characteristic variable is the characteristic gas content, the hidden layer node $N = 13$, and the scale factor is $[10, 10]$. The diagnostic model of the test data and the resultant output of the training data are 1.

At this time, it can be concluded that the accuracy rate of 40 groups of sample data in the classifier RVFL3 diagnosis training set is 67.50%, and the training time is 0.004132 seconds. The accuracy of the 15 sample data in the classifier RVFL3 diagnostic test set is 64.57%, and the test time is 0.005092 seconds. After trying to adjust the number of nodes in the hidden layer N, it is found that under the sample data of classifier RVFL3, adjusting the number of hidden nodes cannot improve the fault diagnosis accuracy of RVFL3, and the impact on the detection time is extremely small. Therefore, the adjustment scale factor is selected. The test results of classifier RVFL3 under the change of scale factor are shown in Figure 10.

According to the above data, it is found that before the scale factor $S = 17$, with the expansion of the threshold range, the diagnosis time also increases rapidly. But accuracy does not change for either the training data or the test data. When the scale factor is within the range of $17 = S < 22$, the detection accuracy begins to rise slowly. After exceeding this range, the training effect of the model becomes significantly worse. From the above data, it is found that when the scale factor S increases, the diagnostic accuracy of the classifier RVFL3 to the training data is slightly improved. However, with the increase of the scale factor value, the diagnostic accuracy of the network for the test set data begins to decline. Therefore, the first consideration is that the diagnostic accuracy of the training set data is not greatly improved. Second, according to the figure above, the improvement of the diagnostic accuracy of training data does not improve the diagnostic performance of training data, but shows a small decline. Third, considering that the increase of scale factor S will prolong the training and detection time of the neural network, it is not applicable to the real power network. Therefore, under the condition of ensuring the diagnosis accuracy of RVFL network to the maximum extent, choosing the minimum random range will make the fault diagnosis network more practical, which is conducive to the RVFL fault detection model to judge the occurrence of faults in the first time, and reduce the fault loss to the minimum to ensure the economic efficiency of the power system.

Therefore, in the case that the number of hidden nodes $N = 13$, $S = 20$ are selected as the optimal scale factor values of classifier RVFL3.

As shown in Table 5, when the number of hidden layer nodes is constant, the input feature variable characteristic gas content scaling factor S is compared as its initial values 10 and 20 corresponding to the optimal value of classifier RVFL3 and the diagnosis results of transformer faults in RVFL network, respectively. Yt1_temp and Y1_temp, respectively, represent the diagnostic output results of the network to the training set and the test set when the
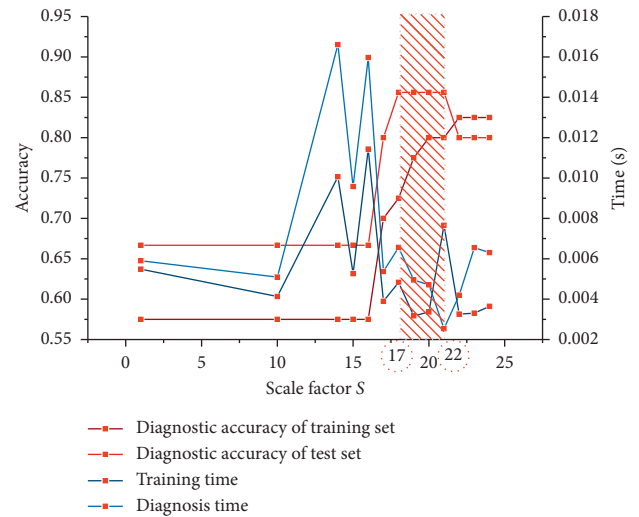
Figure 10: Selection of scale factor S of classifier RVFL3.

Table 5: Comparison of RVFL3 diagnosis results by scale factor optimization.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	0	0	0	1	1	1
2	1	1	1	1	1	1
3	0	0	0	0	0	0
4	0	0	0	0	1	1
5	0	0	0	0	1	0
6	1	1	1	0	0	0
7	0	0	0	0	0	0
8	1	1	1	0	0	0
9	0	0	0	0	1	1
10	1	1	0	1	1	1
11	0	1	1	0	0	0
12	0	0	0	1	1	1
13	0	1	1	0	1	1
14	0	0	0	0	0	0
15	0	0	0	0	0	1

randomization range $[-S, S]$ is the initial value $[-10, 10]$. Yt2_temp and Y2_temp, respectively, represent the diagnostic output results of the network to the training set and the test set when the optimal threshold range $[-S, S]$ is set $[-20, 20]$ after debugging.

Combined with the data in the table, the simulation test results in the case of the hidden node number $N = 13$ and scale factor $S = 10$ before adjusting the relevant parameters of classifier RVFL3 are compared with the test results in the case of tuning scale factor S to 20.

When the relevant parameters of the classifier RVFL3 are adjusted, that is, when the number of hidden nodes $N = 13$ and the scaling factor $S = 20$, the diagnostic accuracy of the 40 training sets of the classifier RVFL3 is 80.3% and the training time is 0.003361 seconds. The accuracy of the 15 sample data in the classifier RVFL3 diagnostic test set is 85.6%, and the test time is 0.004706 seconds. It is clear that after parameter optimization, the detection accuracy of the network model improved by 12.8% for the training data and 21.03% for the test data. It can be seen that the parameter

tuning of the RVFL3 classifier is very successful when the input feature is the characteristic gas content of the transformer.

The input data are the standardized transformer characteristic gas content, the number of hidden layer nodes $N = 100$, and the randomization range is $[-2.1, 2.1]$. At this point, the diagnostic results of the RVFL network model are acceptable for the training data set but not optimal for the test data set. The effectiveness of the network training process is poor.

At this time, the accuracy of 40 sets of sample data in the classifier RVFL3 diagnostic training set was 67.28%, and the training time was 0.004133 seconds. The accuracy of 15 sets of sample data in the RVFL3 diagnostic test set classifier was 74.67%, and the test time was 0.004709 seconds. By debugging the relevant parameters, it was found that adjusting the scaling factor did not significantly improve the test results of the classifier RVFL3 for the corresponding sample data under the input type. Therefore, we try to improve the training effect by changing the number of hidden layer nodes N.

When the scaling range remains unchanged as $[-2.1, 2.1]$, change the number of nodes in the hidden layer N to observe the changes in the simulation results of classifier RVFL3, as shown in Figure 11.

According to the data in the figure, when the scale factor S is still 2.1 and the number of hidden layer nodes is selected, $N = 20$, the fault diagnosis accuracy of classifier RVFL3 reaches the highest value. When the number of hidden nodes is $N > 20$, the diagnostic accuracy of the RVFL network model to the training data still increases, but the diagnostic accuracy of the test data does drop one level. Although with the increase of the number of hidden nodes, the diagnostic accuracy of the network for the training set data is improved to some extent, the first consideration is that the diagnostic accuracy of the training set data is not greatly improved. Second, the improvement of the diagnostic accuracy of training data does not improve the diagnostic performance of test data but shows a small decline. Third, considering that the increase in the number of hidden layer nodes will prolong the training and detection time of the neural network, it can be seen that it is not applicable to the real power network.

The results show that on the premise of ensuring the transformer fault diagnosis of the RVFL neural network, choosing the number of hidden nodes N can effectively reduce the complexity of the network model. This method is helpful to determine the RVFL fault detection model to reduce the loss degree for the first time and ensure the safety of the power system. By analyzing the relationship between the number of hidden nodes in Figure 11 and the test accuracy and the length of test time, the number of hidden layer nodes $N = 20$ is finally selected as the optimal value of this classifier.

As shown in Table 6, the scale factor is $[-2.1, 2.1]$, and the number of nodes in the hidden layer is used as the initial optimum when the normalized characteristic gas content is the input feature compared with the optimum value of the classifier RVFL3. Yt1_temp and Y1_temp denote the
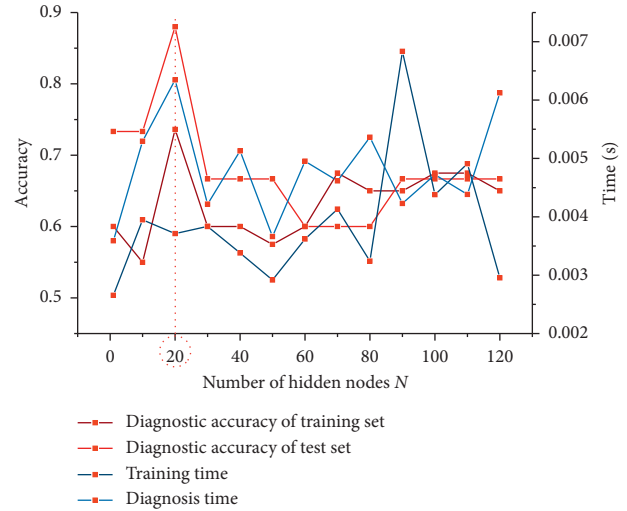
FIGURE 11: Selection of hidden layer node N in classifier RVFL3 after standardization.

TABLE 6: Hidden node optimization RVFL3 diagnosis comparison.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	1	0	0	1	0	1
2	0	0	0	0	0	0
3	0	0	1	1	1	1
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	1	1	1	1	0	1
7	1	0	0	0	0	0
8	1	1	1	0	0	1
9	0	0	0	1	1	0
10	0	0	1	0	0	0
11	0	0	0	1	1	1
12	0	1	0	0	0	0
13	1	1	1	0	0	1
14	0	0	0	0	0	0
15	1	0	0	1	0	1

diagnostic outputs of the network for the training and test sets, respectively, when the number of hidden nodes N is an initial value of 100. Yt2_temp and Y2_temp represent the diagnostic output of the network for both the training and test sets, respectively. The optimal value of 20 is obtained when the number of hidden nodes is optimized.

When the relevant parameters of the classifier RVFL3 are adjusted, that is, when the number of hidden nodes is $N = 20$ and the scale factor is $S = 2.1$, the accuracy rate of the 40 groups of sample data in the diagnostic training set of the classifier RVFL3 is 73.62% and the training time is 0.003713 seconds. The accuracy of the 15 sample data in the classifier RVFL3 diagnostic test set is 88.00% and the test time is 0.006346 seconds.

Obviously, after parameter optimization, the detection accuracy of network model training data increases by 6.34%, and the detection accuracy of test data increases by 13.33%. It can be seen that the parameter adjustment of classifier RVFL3 is very successful when the input feature is the normalized characteristic gas content of transformer.

Through the test, the characteristic gas content is obtained as the best parameter of the characteristic variable, $N = 13$, $S = 20$. After standardized treatment on characteristic variables, the optimal gas content parameters were $N = 20$ and $S = 2.1$. Where N is the number of hidden layer nodes, and S is the scale factor. The diagnostic results shown in Table 7 are the characteristic gas content ratio and the standardized treatment of gas content as the characteristic variables, respectively. Some test results for the RVFL3 classifier are shown in Table 7.

Fault 1 is low energy discharge, and fault 0 is high energy discharge. The letters mean the same as above. The diagnostic accuracy is 85.6% and 88%, respectively. The test times are 0.003132 seconds and 0.006254 seconds, respectively.

4.2.4. RVFL4 Simulation and Test Results. The classifier RVFL4 contains a total of 43 groups of characteristic gas data of power transformers under thermal failure operation state, among which 14 groups are gas content data under medium and low temperature heating operation state, and the remaining 27 groups are gas content data under high temperature heating operation state. Take 35 groups of gas characteristic variables as training data. The remaining 8 groups of characteristic variables are test data.

When the input data are the characteristic gas content of the transformer, that is, when the number of hidden layer nodes is $N = 13$ and the randomization range is still $[-10, 10]$, the simulation results of the classifier RVFL4 are good. When the input data are the standardized characteristic gas content and the parameters are not optimized, that is, the number of hidden layer nodes $N = 100$ and the randomization range is $[-2.1, 2.1]$, the detection accuracy of classifier RVFL4 for 35 groups of sample data in the training set is 72.14%, and the training time is 0.002364 seconds. However, the diagnostic accuracy of the classifier RVFL4 for 8 groups of sample data in the test set is 62.50% and the test time is 0.001845 seconds.

It can be seen that the RVFL network model has an ideal simulation result for the training set but has low simulation accuracy for the test set. According to the model test results, it can be found that the diagnostic accuracy of the training set data is much smaller than that of the test set by 9.64%. Therefore, the training process of the network model is basically ineffective at the current parameter settings.

According to the above assumptions, it is found through simulation and debugging related parameters that under the input type, for the corresponding sample data, the adjustment scale factor S does not significantly improve the test results of the classifier RVFL4. Therefore, by changing the value of the hidden layer node N, a certain rule can be observed. The relationship between the simulation results of classifier RVFL4 and the number of hidden layer nodes is shown in Figure 12:

As shown in the figure, it can be seen that when the number of hidden layer nodes $N = 1$, the test results for both the training data and the test data set are the best, and the fault diagnosis accuracy is also the highest. When the

Table 7: Comparison of RVFL3 diagnosis results under standardized input characteristics.

	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	trY	Yt_t	Y_t
1	26.5	29.8	69.3	0	8.6	1	1	1
2	34.4	21.9	45	3.2	19.6	1	0	1
3	316	36.1	109	37.3	440	1	1	1
4	29	7.52	14.8	5.5	5.4	0	0	0
5	6.3	7.87	3.03	6.97	11.1	0	0	1
6	413	6.8	4.6	40.4	8.15	0	1	0
7	23.4	14.3	9.71	7.12	3.44	0	0	0
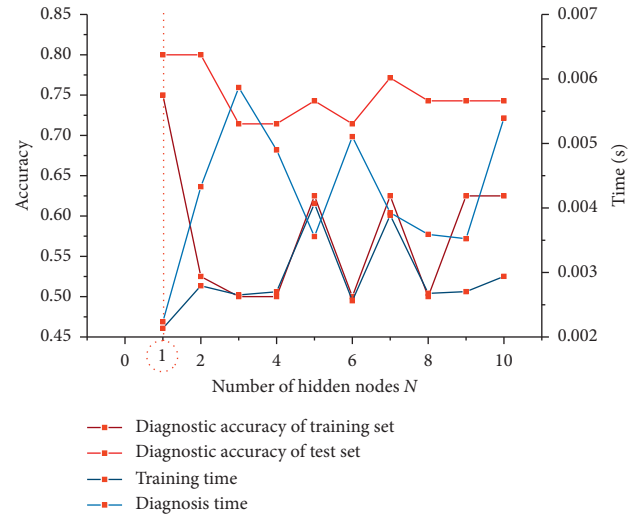
Figure 12: Selection of the hidden layer node N in classifier RVFL4 after standardization.

number of nodes in the hidden layer is $N > 2$, the diagnostic accuracy of the network for the training data set is basically stable, while the diagnostic accuracy of the data for the test set is also gradually decreasing, and the network gradually shows the phenomenon of overfitting. Because the increase of the number of hidden layer nodes prolong the training and detection time of the neural network, it is not applicable to the real power network. Therefore, in order to ensure the diagnostic accuracy of the RVFL network to the maximum extent, the least hidden layer neurons are selected. The fault type can be diagnosed and processed in the shortest time with the lowest model complexity, and the loss can be reduced to the lowest degree. Therefore, the number of hidden layer nodes $N = 1$ is selected as the optimal value of classifier RVFL4.

As shown in Table 8, the diagnosis results of transformer faults by the RVFL network are compared under two different conditions, in which the initial value of 100 is selected for the number of hidden layer nodes and the optimal value 1 corresponding to classifier RVFL4 is selected. Yt1_temp and Y1_temp represent the diagnostic output of the network for the training and test sets, respectively, of classifier RVFL4 when the number of hidden nodes N is the initial value of 100. Yt2_temp and Y2_temp represent the diagnostic output of the network for the training and test sets, respectively, when the number of hidden nodes is adjusted and the optimal value of 1 is chosen with the same scaling range.

TABLE 8: Hidden node optimization RVFL4 diagnosis comparison.

	Yt1_temp	Yt2_temp	testY	Y1_temp	Y2_temp	trainY
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	1	1	1
4	0	1	1	0	0	1
5	0	0	0	0	1	1
6	1	0	1	0	0	0
7	0	1	1	0	0	0
8	0	0	1	0	0	1

TABLE 9: Comparison of RVFL4 diagnosis results under standardized input characteristics.

	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	trY	Yt_t	Y_t
1	18	20	12	6	2	1	1	1
2	46	98	26.3	41.3	0	1	0	1
3	25	85	178	9.6	1.4	1	1	1
4	155	581	215	293	142	0	0	0
5	477	566	734	239	116	0	1	1
6	56.4	42.7	1413	167	17.1	0	1	0
7	74.8	482	1409	297	7.5	0	0	0

The comparison of the data in Table 8 illustrates the simulation test results of the classifier RVFL4 for both the test data and the training data. The simulation result data when the number of hidden nodes N is set to 1 and the scale factor S remains 2.1 are compared with the simulation data when the number of hidden nodes $N = 100$ and the scale factor $S = 2.1$. The accuracy of the diagnostic training set classifier RVFL4 for 35 sets of sample data in RVFL4 with the optimized number of hidden nodes $N = 1$ and scale factor $S = 2.1$ is 75.00%, and the training time is 0.002135 seconds. The accuracy of classifier RVFL4 for diagnostic test set of 8 sample data is 80.00%, and the test time is 0.002235 seconds.

It can be found that by optimizing the number of hidden nodes in the classifier RVFL4, the diagnostic accuracy of the RVFL neural network for training data is improved by 2.86%, and the diagnostic accuracy of test data is increased by 17.50%. It can be seen that the parameter adjustment of classifier RVFL4 is successful when the input feature is the transformer gas content ratio.

The simulation experiment shows that the optimal parameter of characteristic gas content as the characteristic variable is $N = 1$, $S = 2.1$, and the optimal parameter of characteristic variable after standardized treatment is $N = 13$, $S = 10$. Where N is the number of hidden layer nodes and S is the scale factor. The diagnostic results shown in Table 9 are the characteristic gas content ratio and the standardized treatment of gas content as the characteristic variable, respectively. Some test results shown in Table 9.

Fault type 1 is a medium to low temperature heating fault, and fault type 0 is a high temperature heating fault. The meaning of these letters is the same as above. The correct diagnosis rate is 75.7% and 80.00%, respectively. The convergence times are 0.003525 and 0.002235 seconds, respectively.

4.3. Comparison with the Five Classification Methods

4.3.1. RVFL Simulation and Test Results. A classifier is constructed to classify the five operating conditions of transformers using the available 162 sets of transformer gas characteristic data. There are 64 sets of transformer normal working condition operation data, 22 sets of low energy discharge fault data, 33 sets of high energy discharge data, 14 sets of medium and low temperature overheating fault data, and 27 sets of high temperature overheating fault data. The randomly selected 120 sets of data are used as the training set, and the remaining 42 sets are used as the test set. Considering

the influence of input characteristics and related parameter values on the simulation test results, the unification of parameter values with input characteristics of order of magnitude will improve the network diagnosis function.

When the input feature is the gas content extracted from transformer oil, and when the scaling range $[-S, S]$ is still $[-10, 10]$, the influence of the numerical selection of the number of hidden layer nodes N on the five classifiers is first considered, as shown in Figure 13:

With the gradual increase in the number of nodes in the hidden layer, the fault diagnosis accuracy of the network steadily improves for both training and test data. The convergence speed increases slightly when the network structure tends to be complex. The accuracy of RVFL neural network for training set and test set data diagnosis is the highest when the number of hidden nodes $N = 290$. When $N = 290$, the diagnostic accuracy of the training set decreases slightly and then continues to increase. Meanwhile, the diagnostic accuracy of the neural network for the test set data will remain at the peak after a sharp drop until the neural network overfitting occurs. Therefore, in order to minimize the complexity of network construction, shorten the convergence time and improve the learning effect of the model; the number of hidden layer nodes $N = 290$ is selected as the characteristic input for five characteristic gas contents of the transformer, and the optimal value of hidden nodes in the overall classification is obtained.

The parameter scale factor S has a significant effect on the gas content of the relative dispersion of the input features. The effect of different values of the scale factor S on the overall classification simulation test results is shown in Figure 14.

The scale factor range increases with the classification accuracy of the training data and test data of the RVFL network model, and the diagnostic accuracy of the training data network slowly steps up. The accuracy of both training data and test data showed an overall increasing trend. When the value of the scale factor is in the range of $23 = S \le 39$, the detection accuracy starts to rise slowly. Beyond this range, the training effect becomes significantly worse.

Therefore, choosing the smallest scale factor value $S = 35$ under the condition of ensuring the diagnostic accuracy of RVFL will make the fault diagnosis network more practical. This facilitates the RVFL fault detection model to determine the occurrence of faults in the first time and reduce the fault loss to a minimum. $S = 35$ was chosen as the optimal scale factor value for the overall classification of the model, and the random range $[-S, S]$ was adjusted to $[-35, 35]$.
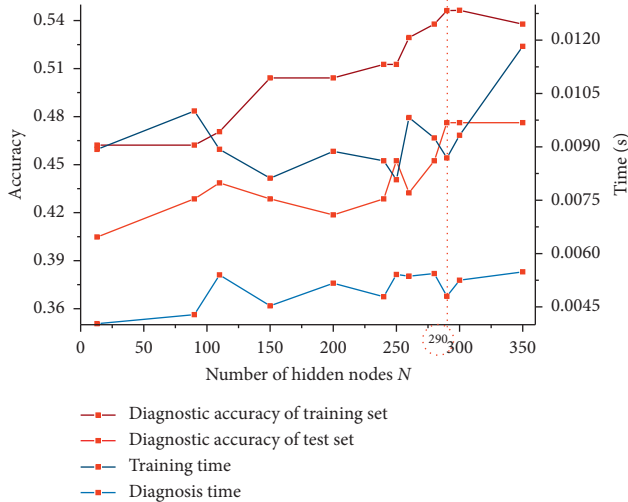
FIGURE 13: Selection of hidden layer node N in classifier RVFL.
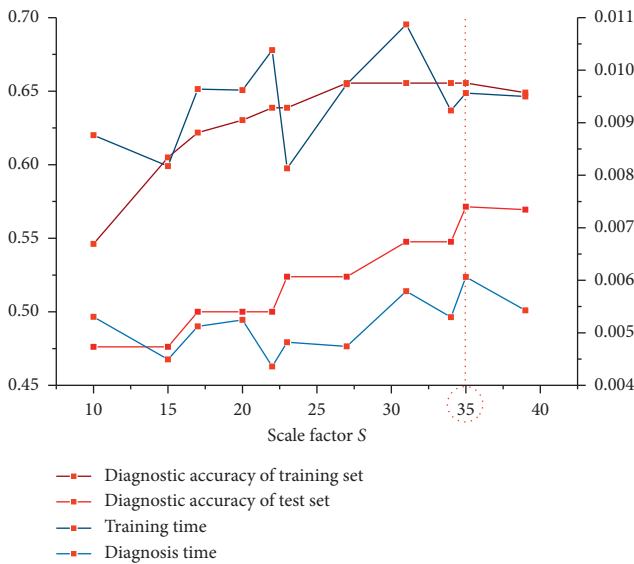
FIGURE 14: Selection of scale factor S of classifier RVFL.

FIGURE 15: Selection of scale factor S of classifier RVFL after standardization.

When the input feature is the characteristic gas content of the transformer after standardization, and when the number of hidden layer neurons N is still 100, the influence of the value selection of scale factor S on the five classifiers is first considered, as shown in Figure 15.

From Figure 15, it can be seen that the fault diagnosis accuracy of both training and test data has increased but the change in model convergence time is small. When the scale factor $[S, S]$ is $[2.4, 2.4]$, the diagnostic accuracy set saturation of the network model for the test data is fixed at 69.05% without rising. If we continue to debug the scale factor, we can find that the diagnostic accuracy of both network data and test data decreases significantly when $S > 6$. Therefore, $S = 2.4$ is selected as the optimal scale factor value of the five classifiers. The random range $[-S, S]$ is adjusted to $[-2.4, 2.4]$.

Then, under the condition that the scaling range $[-S, S]$ is $[-2.4, 2.4]$, the number of hidden layer nodes N, which has

a great influence on the input characteristics, is studied. As shown in Figure 16, it is the influence of different values of hidden node N on the simulation test results of five categories.

From the data in the figure, it can be seen that the highest fault diagnosis accuracy is achieved for the five classifiers with the selected number of hidden layer nodes $N = 200$ when the scale factor S is 2.4. When the number of hidden nodes is in the range of $200 = N < 250$, the network saturates the diagnostic accuracy of the test data set. With the increasing number of hidden nodes N, the diagnostic accuracy of the network on the training set data is improved to some extent. However, the diagnostic accuracy of the training set data is not greatly improved. Secondly, the improvement of the diagnostic accuracy of the training data did not improve the diagnostic performance of the test data but slightly decreased. In addition, the increase in the number of nodes in the hidden layer prolongs the training and detection time of the neural network. This indicates that it is not applicable in the real grid.

While ensuring the diagnostic accuracy of the RVFL neural network, improving the number of hidden nodes can effectively reduce the complexity of the network model. This method can reduce the loss of RVFL fault diagnosis model and ensure the safety of power system. By analyzing the relationship between the number of hidden nodes in Figure 16 and the test accuracy and the length of test time, the number of hidden layer nodes $N = 200$ is finally selected as the optimal value of this classifier.

The optimal parameters tested for the characteristic gas content as a characteristic variable are $N = 13$, $S = 10$, and the optimal parameters tested for the standardized gas content as a characteristic variable are $N = 100$, $S = 2.1$. N is the number of hidden layer nodes, and S is the scale factor. Some of the diagnostic results are shown in Table 10 for the characteristic gas content ratio and the results after normalizing the gas content of the characteristic variables.
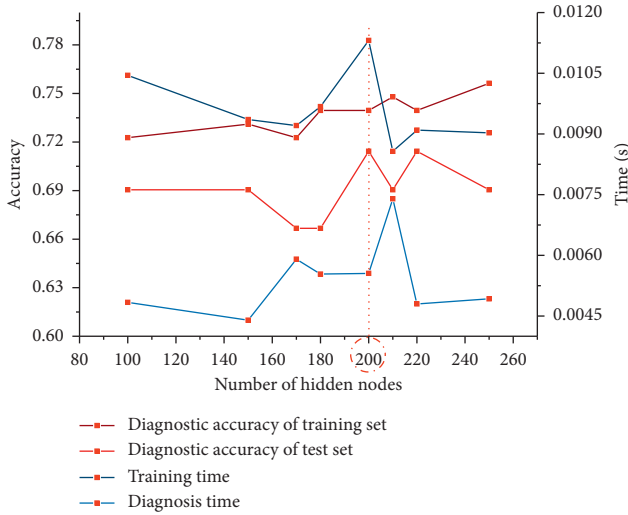
Figure 16: Selection of hidden layer node N in classifier RVFL after standardization.

Table 10: Comparison of RVFL diagnosis results under standardized input characteristics.

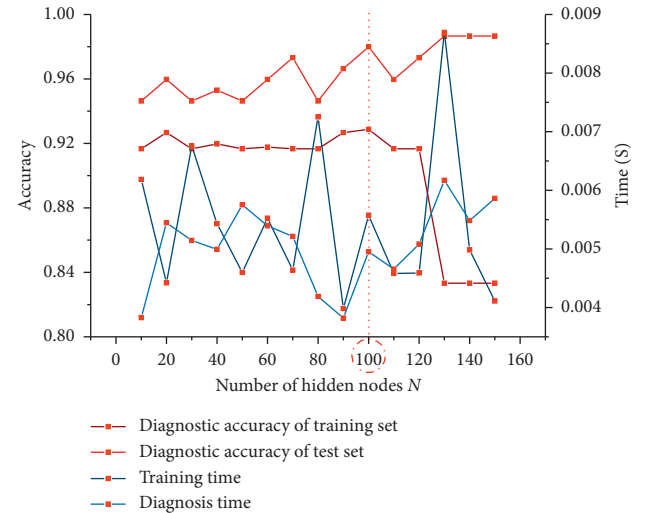|	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2	trY	Yt_t	Y_t
1	20.9	8.85	2.47	1.04	0	1	1	1
2	22.5	8.5	7.3	1.2	8.4	3	3	4
3	19.8	8.18	2.47	1.11	0	1	3	1
4	155	581	215	293	142	4	4	2
5	20.63	9.12	2.37	1.07	0	1	2	1
6	2054	370	535	26	805	2	2	2
7	20.8	8.97	2.39	1.13	0	1	3	1
8	4430	941	2475	76	55585	2	3	2
9	11	31	52	26	0	5	5	4
10	18.2	9.56	2.27	1.07	0	1	1	4
11	179	413	544	27	473	2	4	2
12	143	305	721	78	2	4	4	1
13	18.8	9.32	2.49	1.12	0	1	1	1
14	5.5	9.6	50	3.9	1.5	5	2	1
15	17	21	156	11	4.1	4	4	2
16	18.9	8.86	2.48	0.99	0	1	2	1
17	246	432	674	141	7.2	4	4	4
18	20	9.76	2.42	1.06	0	1	5	1
19	20.3	8.38	2.38	1.15	0	1	1	1
20	21.5	9.45	2.5	1.18	0	1	1	2

The data in the table show the diagnosis results of test set and training set, respectively, by RVFL network. Among them, fault type 1 represents normal operation state, 2 represents high energy discharge fault, 3 represents low energy discharge fault, 4 represents high temperature heating fault, and 5 represents medium and low temperature heating fault. Yt_t and Y_t denote the fault diagnosis results of the test set before and after normalization, respectively. TrY denotes the true failure type of the tester. The correct diagnosis rate is 51.14% and 71.43%, respectively. The convergence time is 0.005356 seconds and 0.006428 seconds, respectively.

In summary, the fault diagnosis accuracy of the RVFL neural network for overall classification is much lower than that of the fault detection model consisting of four binary classifiers. The convergence rate of the overall classification, which is nearly twice faster than that of the step classification, is not favorable for transformer fault diagnosis in real power grids. Meanwhile, the weight scaling effect of the classification model network becomes worse for the input feature of gas content. The diagnostic accuracy of the training set with standardised gas content as input feature is 8.4% higher than that with gas content as input. When the input feature is normalized gas content, the diagnostic accuracy of the test data is 14.29% higher than that when the input feature is gas content ratio.

4.3.2. Analysis of Test Results. In part B, a three-layer, four-classifier model is constructed for five different operating states of the power transformer. Under two different types of input characteristics, four different sets of sample data are obtained by experimenting on the test set and training set data through the classifier. Two different types of feature variables, pre-standardized and post-standardized gas content, are taken as input data, and the feature scaling of different input sample data is performed according to the optimal threshold range, so that the feature data of different dimensions can be compared. At the same time, it speeds up the search for the optimal solution and greatly improves the accuracy of fault diagnosis.

In the five-classification model, the number of data types for the five samples varies considerably, resulting in an inadequate theoretical basis for the weighting values of a particular fault type when the training set data is small. The network needs to acquire more fault information and data volumes to learn, so that a significant increase in convergence time can be observed. Not surprisingly, the fault diagnostic accuracy of the RVFL neural network is lower than that of the fault detection model consisting of four dichotomous classifiers. In addition, the convergence of the five-classifier algorithm is slower than the stepwise classification, which is not conducive to the diagnosis of transformer faults in real power grids. Meanwhile, the weight scaling of the network is poor for the classification model with gas content as the input feature.

Thus, the three-layer, four-type classification model constructed in this paper for five different types of power transformer faults is superior to the five-class classification processing of the sample data.

5. Conclusion

In this paper, the relationship between RVFL neural network-based transformer fault diagnosis and the values of the associated parameters is investigated. The method yields stable ranges and diagnostic accuracies of the relevant parameters of the network model under two different input features and data sets, and the optimal range of values is discussed. To improve the diagnostic accuracy of the model, the traditional RVFL neural network model is modified and the input data are standardized. The method also reduces the complexity of the network model and simplifies the network model. The experiments validate the validity of the

parameters associated with the RVFL neural network. After adjusting the parameters, it was found that RVFL2 improved the diagnostic accuracy by 11.43% for the training set and 17.14% for the test set. The diagnostic accuracy of RVFL3 standardized on the training set improved by 6.34% and the diagnostic progress of the test set improved by 13.33%. Diagnostic accuracy improved by 2.86% for the RVFL4 training set and by 17.5% for the test set. The diagnostic accuracy of each classifier is substantially improved. Adjusting the number of hidden nodes when the input sample data are uniform can improve the accuracy of network diagnostics. When the input sample data are scattered, adjusting the scale range can improve the accuracy of network testing. For different sample data, the network testing accuracy is more sensitive to the change of scale factor than the change of the number of hidden nodes under the premise of ensuring the validity of model training.

This paper describes the theoretical basis for the number of implied nodes and scale factor assignment in the RVFL network model. It effectively improves the diagnostic accuracy of transformer faults. This study is of reference value for the analysis of other neural network engineering applications.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

References

[1] K. Wu, J. Kang, and K. Chi, "Power transformer fault diagnosis method based on improved multi-classification algorithm and related vector machine," *High Voltage Technology*, vol. 42, no. 9, pp. 3011–3017, 2016.

[2] A. Abu-Siada and S. Hmood, "A new fuzzy logic approach to identify power transformer criticality using dissolved gas-in-oil analysis," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 401–408, 2015.

[3] H. Wu, X. Li, and D. Wu, "RMP neural network based dissolved gas analyzer for fault diagnostic of oil-filled electrical equipment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 2, pp. 495–498, 2011.

[4] J. Fan, Z. Liu, A. Meng et al., "Characteristics of tin oxide chromatographic detector for dissolved gases analysis of transformer oil," *IEEE Access*, vol. 7, pp. 94012–94020, 2019.

[5] C. Xiang, Z. Huang, J. Li, Q. Zhou, and W. Yao, "Graphic approaches for faults diagnosis for Camellia insulating liquid filled transformers based on dissolved gas analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 5, pp. 1897–1903, 2018.

[6] J. Jiang, M. Zhao, G.-M. Ma et al., "TDLAS-based detection of dissolved methane in power transformer oil and field application," *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2318–2325, 2018.

[7] Standard IEC 60599, *Guide for the Interpretation of Dissolved Gas Analysis and Gas Free*, International Electrotechnical Commission, Geneva, Switzerland, 2007.

[8] J. Faiz and M. Soleimani, "Dissolved gas analysis evaluation in electric power transformers using conventional methods a review," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 2, pp. 1239–1248, 2017.

[9] H.-C. Sun, Y.-C. Huang, and C.-M. Huang, "Fault diagnosis of power transformers using computational intelligence: a review," *Energy Procedia*, vol. 14, no. 2, pp. 1226–1231, 2012.

[10] B. Zhao, X. Zhang, H. Li, and Z. Yang, "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions," *Knowledge-Based Systems*, vol. 199, Article ID 105971, 2020.

[11] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mechanical Systems and Signal Processing*, vol. 110, pp. 349–367, 2018.

[12] Z. Meng, X. Guo, Z. Pan, D. Sun, and S. Liu, "Data segmentation and augmentation methods based on raw data using deep neural networks approach for rotating machinery fault diagnosis," *IEEE Access*, vol. 7, pp. 79510–79522, 2019.

[13] J. Liang, T. Jing, H. Niu, and J. Wang, "Two-terminal fault location method of distribution network based on adaptive convolution neural network," *IEEE Access*, vol. 8, pp. 54035–54043, 2020.

[14] S. Chen, H. Ge, J. Li, and M. Pecht, "Progressive improved convolutional neural network for avionics fault diagnosis," *IEEE Access*, vol. 7, pp. 177362–177375, 2019.

[15] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 100, Article ID 106724, 2020.

[16] H. Zhao, S. Zuo, M. Hou et al., "A novel adaptive signal processing method based on enhanced empirical wavelet transform technology," *Sensors*, vol. 18, no. 10, Article ID 3323, 2018.

[17] J. Yu, H. Huang, H. Chen et al., "Application of neural network in transformer fault diagnosis research," *Control Engineering*, vol. 25, no. 10, pp. 1898–1904, 2008.

[18] Y. Liu, T. Yan, L. Jiang et al., "An improved neural network algorithm optimized by genetic algorithm and its application in power transformer fault diagnosis," *Electronic Technology*, vol. 46, no. 10, pp. 54–56, 2017.

[19] S. Wang, S. Zhang, T. Wu et al., "A first-order Markov dynamic Bayesian network classifier with continuous

attributes," *Knowledge-Based Systems*, vol. 195, Article ID 105638, 2020.

[20] G. Chen, M. Liu, and J. Chen, "Frequency-temporal-logic-based bearing fault diagnosis and fault interpretation using Bayesian optimization with Bayesian neural networks," *Mechanical Systems and Signal Processing*, vol. 145, Article ID 106951, 2020.

[21] K. Jia, T. Feng, Q. Zhao, C. Wang, and T. Bi, "High frequency transient sparse measurement-based fault location for complex DC distribution networks," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 312–322, 2020.

[22] C. Zhang and Z. Lin, "Transformer fault diagnosis based on normal cloud model and improved Bayesian classifier," *Electrical Measurement & Instrumentation*, vol. 54, no. 4, pp. 50–56, 2017.

[23] Y. Li and X. Gu, "Application of online SVR in very short-term load forecasting," *International Review of Electrical Engineering*, vol. 8, no. 1, pp. 277–282, 2013.

[24] X. W. Liang, A. P. Jiang, T. Li, Y. Y. Xue, and G. T. Wang, "LR-SMOTE—an improved unbalanced data set oversampling based on K-means and SVM," *Knowledge-Based Systems*, vol. 196, Article ID 105845, 2020.

[25] R. Gao, L. Zhang, and X. Liu, "Short-term load forecasting based on least square support vector machine combined with fuzzy control," in *Proceedings of the 10th World Congress on Intelligent Control and Automation*, pp. 1048–1051, Beijing, China, July 2012.

[26] L. Limei and H. Xuan, "Study of electricity load forecasting based on multiple kernels learning and weighted support vector regression machine," in *Proceedings of the 2017 29th Chinese Control and Decision Conference (CCDC)*, pp. 1421–1424, Chongqing, China, May 2017.

[27] M. Ali, Z. A. Khan, S. Mujeeb, S. Abbas, and N. Javaid, "Short-term electricity price and load forecasting using enhanced support vector machine and K-nearest neighbor," in *Proceedings of the 2019 Sixth HCT Information Technology Trends (ITT)*, pp. 79–83, Ras Al Khaimah, UAE, November 2019.

[28] G. Mitchell, S. Bahadoorsingh, N. Ramsamooj, and C. Sharma, "A comparison of artificial neural networks and support vector machines for short-term load forecasting using various load types," in *Proceedings of the 2017 IEEE Manchester PowerTech*, pp. 1–4, Manchester, UK, June 2017.

[29] Y. Li, "Short-term load forecasting based on LS-SVM optimized by BCC algorithm," in *Proceedings of the 2009 15th International Conference on Intelligent System Applications to Power Systems*, pp. 1–5, Curitiba, Brazil, November 2009.

[30] L. Shi and P. Yu, "Application of fuzzy support vector machine in transformer fault diagnosis," *Electrical Measurement & Instrumentation*, vol. 52, no. 8, pp. 115–119, 2015.

[31] L. Zhang, H. Ma, N. Jiang, T. Gao, and H. Xu, "Recognition and location of transformer winding looseness based on singular value spectrum entropy and SVM," *Power System Protection and Control*, vol. 45, no. 18, pp. 69–75, 2017.

[32] X. Xue, W. Sun, J. Wang et al., "RVFL-LQP: RVFL-based link quality prediction of wireless sensor networks in smart grid," *IEEE Access*, vol. 8, pp. 7829–7841, 2020.

[33] Y. Zhang, B. Li, Y. Wang, and Q. Tian, "The forecast of the temperature in subway station based on RVFL neural network," in *Proceedings of the 2018 11th International Congress on Image and Signal Processing, Bio Medical Engineering and Informatics (CISP-BMEI)*, pp. 1–5, Beijing, China, October 2018.

[34] T. Han, C. Liu, W. Yang, and D. Jiang, "Deep transfer network with joint distribution adaptation: a new intelligent fault diagnosis framework for industry application," *ISA Transactions*, vol. 97, pp. 269–281, 2020.

[35] T. Han, C. Liu, W. Yang, and D. Jiang, "Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions," *ISA Transactions*, vol. 93, pp. 341–353, 2019.

[36] K. Patan and M. Patan, "Neural-network-based iterative learning control of nonlinear systems," *ISA Transactions*, vol. 98, pp. 445–453, 2020.

[37] Y. Wu, Y. Fang, S. Shang, J. Jin, L. Wei, and H. Wang, "A novel framework for detecting social bots with deep neural networks and active learning," *Knowledge-Based Systems*, vol. 211, Article ID 106525, 2021.

[38] C. Cui and D. Wang, "High dimensional data regression using lasso model and neural networks with random weights," *Information Sciences*, vol. 372, pp. 505–517, 2016.

[39] S. Scardapane, D. Wang, M. Panella, and A. Uncini, "Distributed learning for random vector functional-link networks," *Information Sciences*, vol. 301, pp. 271–284, 2015.

[40] D. Wang, "Editorial: randomized algorithms for training neural networks," *Information Sciences*, vol. 364-365, pp. 126–128, 2016.

[41] W. F. Schmidt, M. A. Kraaijveld, and R. P. Duin, "Feedforward neural networks with random weights," in *Proceedings of the 11th IAPR International Conference on Pattern Recognition. Vol. II: Conference B: Pattern Recognition Methodology and Systems*, pp. 1–4, The Hague, Netherlands, September 1992.

[42] A. Elisseeff, T. Evgeniou, and M. Pontil, "Stability of randomized learning algorithms," *Journal of Machine Learning Research*, vol. 6, no. 1, pp. 55–79, 2005.

[43] D. Husmeier, "Random vector functional link (RVFL) networks," *Neural Networks for Conditional Probability Estimation: Forecasting Beyond Pint Predictions*, Springer, vol. 6, London, UK, 1999.

[44] B. Igelnik and Y.-H. Pao, "Stochastic choice of basis functions in adaptive function approximation and the functional-link net," *IEEE Transactions on Neural Networks*, vol. 6, no. 6, pp. 1320–1329, 1995.

[45] K. Luo, Z. Y. Jia, and W. Gao, "Strong and weak stability of randomized learning algorithms," in *Proceedings of the 2012 IEEE 14th International Conference on Communication Technology*, pp. 9–11, Chengdu, China, November 2012.

[46] Y. LeCun, L. Bottou, G. B. Orr, and K. R. Müller, "Efficient backprop," in *Lecture Notes in Computer Science*, vol. 1524, pp. 9–50, Springer, Berlin, Germany, 1998.

[47] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[48] Y.-H. Pao, S. M. Phillips, and D. J. Sobajic, "Neural-net computing and the intelligent control of systems," *International Journal of Control*, vol. 56, no. 2, pp. 263–289, 1992.

[49] E. Klinefelter and J. A. Nanzer, "Interferometric microwave radar with a feedforward neural network for vehicle speed-over-ground estimation," *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 3, pp. 304–307, 2020.

[50] P. Musikawan, K. Sunat, Y. Kongsorot, P. Horata, and S. Chiewchanwattana, "Parallelized metaheuristic-ensemble of heterogeneous feedforward neural networks for regression problems," *IEEE Access*, vol. 7, pp. 26909–26932, 2019.

[51] G. Han, Q. Cheng, X. Sun, L. Li, and W. Di, "A biological mechanism based structure self-adaptive algorithm for feedforward neural network and its engineering applications," *IEEE Access*, vol. 7, pp. 25111–25122, 2019.

[52] L. Zhang and P. N. Suganthan, "A comprehensive evaluation of random vector functional link networks," *Information Sciences*, vol. 367-368, pp. 1094–1105, 2016.

[53] L. Cui, H. Zhai, B. Wang et al., "Improved extreme learning machine based on the sensitivity analysis," *IOP Conference Series Materials Science and Engineering*, vol. 320, no. 1, Article ID 012015, 2018.

[54] H. Han, L. Gan, and L. He, "Improved variations for extreme learning machine: space embedded ELM and optimal distribution ELM," in *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, pp. 1–8, Xi'an, China, July 2017.

[55] L. Yang, Q. Zhao, and Y. Jing, "Channel equalization and detection with ELM-based regressors for OFDM systems," *IEEE Communications Letters*, vol. 24, no. 1, pp. 86–89, 2020.

[56] J. B. Leite and J. R. S. Mantovani, "Distribution system state estimation using the Hamiltonian cycle theory," in *Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM)*, p. 1, Boston, MA, USA, July 2016.

Research Article

Alighting Stop Determination of Unlinked Trips Based on a Two-Layer Stacking Framework

Ziwei Cui [ID],[1,2] **Cheng Wang** [ID],[1] **Yueer Gao** [ID],[3] **Dingkang Yang** [ID],[1] **Wei Wei** [ID],[4]
Jianwei Chen,[5] **and Ting He** [ID][1]

[1]*College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China*
[2]*School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou 510006, China*
[3]*School of Architecture, Huaqiao University, Xiamen 361021, China*
[4]*School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China*
[5]*Department of Mathematics and Statistics, San Diego State University, San Diego 92182, CA, USA*

Correspondence should be addressed to Cheng Wang; wangcheng@hqu.edu.cn

Smart card data of conventional bus passengers are important basic data for many studies such as bus network optimization. As only boarding information is recorded in most cities, alighting stops need to be identified. The classical trip chain method can only detect destinations of passengers who have trip cycles. However, the rest of unlinked trips without destinations are hard to analyze. To improve the accuracy of existing methods for determining alighting stops of unlinked trips, a two-layer stacking-framework-based method is proposed in this work. In the first layer, five methods are used, i.e., high-frequency stop method, stop attraction method, transfer convenience method, land-use type attraction method, and improved group historical set method (I-GHSM). Among them, the last one is presented here to cluster records with similar behavior patterns into a group more accurately. In the second layer, the logistic regression model is selected to get the appropriate weight of each method in the former layer for different datasets, which brings the generalization ability. Taking data from Xiamen BRT Line Kuai 1 as an example, I-GHSM given in the first layer has proved to be necessary and effective. Besides, the two-layer stacking-framework-based method can detect all destinations of unlinked trips with an accuracy of 51.88%, and this accuracy is higher than that of comparison methods, i.e., the two-step algorithms with KNN (k-nearest neighbor), Decision Tree or Random Forest, and a step-by-step method. Results indicate that the framework-based method presented has high accuracy in identifying all alighting stops of unlinked trips.

1. Introduction

In the smart card system, the smart card allows the continuous collection of individualized transactional data about the use of public transport networks, and each card is always corresponding to one user. Therefore, smart card data have become a valuable source of information due to their larger scale than surveys permit and over long periods [1–3].

Smart card data can be used for data mining for identification of trip purpose [4], development of origin-destination matrices [5], estimation of vehicle load profile, and other network performance measures [6, 7]. All of these studies require known trip destinations [4–7]. However, in most cities'

smart card system, only passengers' boarding transactions are recorded because users do not validate when they leave the buses. Thus, the alighting stops must be estimated [8, 9].

The trip chain method, described in the next section, is normally used to infer alighting stops [5]. Relying on the chaining of trips during the day, it is assumed that users will alight at the stop near the next boarding stop, and for the last trip of the day, the first boarding point of this day or the next day can be regarded as the next boarding stop to form the trip chain. But destinations that are not linked to the sequence cannot be estimated with this approach. These unlinked trips do not satisfy the approach criteria or they are alone within the day, which are the focus of this paper.

To identify the alighting stops of unlinked trips, most existing researchers use only one method or machine learning algorithm, and some other researchers use several fixed methods in sequence. However, which method or combination is the best for a different dataset is difficult to determine. This work proposes a two-layer stacking framework, which can get the appropriate contributions of several methods in a different dataset simultaneously, so this method has strong generalization ability and can detect all destinations with high accuracy.

Among existing methods of alighting stop determination of unlinked trips, the individual historical set method is widely used. In this method, an individual's records with alighting stops detected by the trip chain method are picked as the individual historical set, and then the stop with the highest alighting frequency in the historical dataset is the destination. But the individual historical set is small when passengers have fewer trips or more unlinked trips, which leads to limited use and a low identification rate [10, 11]. To expand the historical dataset, our previous study reported a method for alighting stop determination of transit passengers based on expanded history trip records [12], which is a step-by-step method. Firstly, the trip chain method with multisource data is presented, and data with destinations are used as the historical dataset. Then, the individual historical set method is used. Finally, for the rest data without alighting points, the group historical set method (GHSM) is used. Records of other passengers boarding at the same stop on the same line are selected as the group history dataset, and alighting stops are determined based on similar trips in the group history dataset. However, in the GHSM, the records' group can be clustered more carefully based on more features, so the improved group historical set method (I-GHSM) is proposed in this paper. Besides, different from researches dividing similar passengers into groups and making records of them in the same group [8, 13–17], this study takes each record as the smallest research unit and clusters them to make records with similar behavior patterns in the same group more directly and accurately.

The rest of the paper is organized as follows. Section 2 illustrates problem descriptions, data used, and related works. Methods to determine alighting stops of unlinked trips are introduced in Section 3. Section 4 manifests a case based on the presented method, and Section 5 holds the conclusions of this research.

2. Background

2.1. Problem Description.
There must be an alighting stop for a user on the bus, but the alighting transaction is not recorded in most cities, so which stop is the alighting point is an important problem. The trip chain method is normally used to infer destinations [5], and it is introduced first as follows. However, some records' destinations cannot be identified by this method, and their alighting points are more difficult to determine. These records of unlinked trips are research objectives and their alighting point determinations are the study problems in this work.

The trip chain method is a classical model to estimate alighting points when a card with multiple records one day, and there are two hypotheses of this method [10, 11]:

(1) For a trip, users alight at the stop where the distance between the current alighting destination and the next boarding point is minimal, and the distance must be lower than a specified maximum value

(2) At the last trip of the day, users return to the first boarding station of this day or the next day, which is assumed to be the stop closest to home

Many algorithms are improved and generated from the trip chain method. For instance, Kumar et al. described a method to relax assumptions on various parameters, such as transferring walking distance threshold, buffering distance for selecting the boarding location, and developing a time window for selecting the vehicle trips [18]. Nassir et al. detected short activity locations among the stops visited on a trip chain and proposed a new heuristic to estimate the stop-level origins and destinations based on the traveler activities determination in the observed transactions [19]. Nunes et al. presented some new spatial validation features to increase destination inference results [20]. Compared with earlier methods on a heuristic basis, Sánchez-Martínez proposed a dynamic programming model to infer destinations with a generalized disutility minimization objective, which took the disutility of waiting, transferring, riding, and walking into consideration [21]. In terms of the use of data, the common trip chain method found passengers' trip chains based on the data of conventional bus IC cards, bus GPS, and static line station information. Cui et al. not only used the IC data of the conventional bus but also considered the IC card data of other public transport modes (such as the rail transit and the Bus Rapid Transit System (BRT)), to make the public transport trip chains more complete. The study case in Cui's research showed that using multisource IC card data can estimate more alighting stops with higher accuracy than single-source IC card data [12].

However, some passengers don't take public transportations frequently, and they have only one trip in a day or broken trips, which do not satisfy the approach criteria. Therefore, the destinations of unlinked trips cannot be estimated through the methods mentioned above, and they are the focus of this paper. The relationship of smart card data in the method based on a stacking framework can be found in Figure 1.

There are some examples of alighting stop determination in Figure 2:

(i) For the first trip, when a passenger boards at the k_1 stop of line A in the up direction, we can get the potential alighting stop $\{k_2, k_3, k_4, k_5, k_6\}$, and the k_5 stop will be the destination based on the 2nd boarding stop. Because the distance and time between two stops match rules in the trip chain method, these two trips build a trip chain.

(ii) For the second trip, we cannot confirm which potential alighting stop is the destination. Because
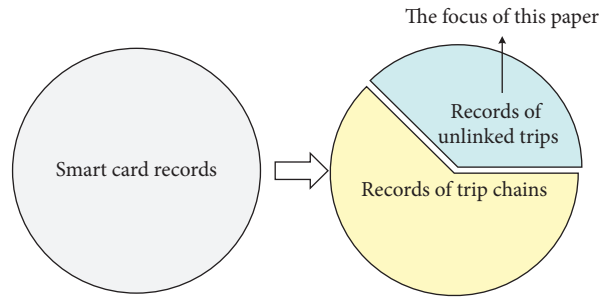
FIGURE 1: The relationship of smart card records in this paper.
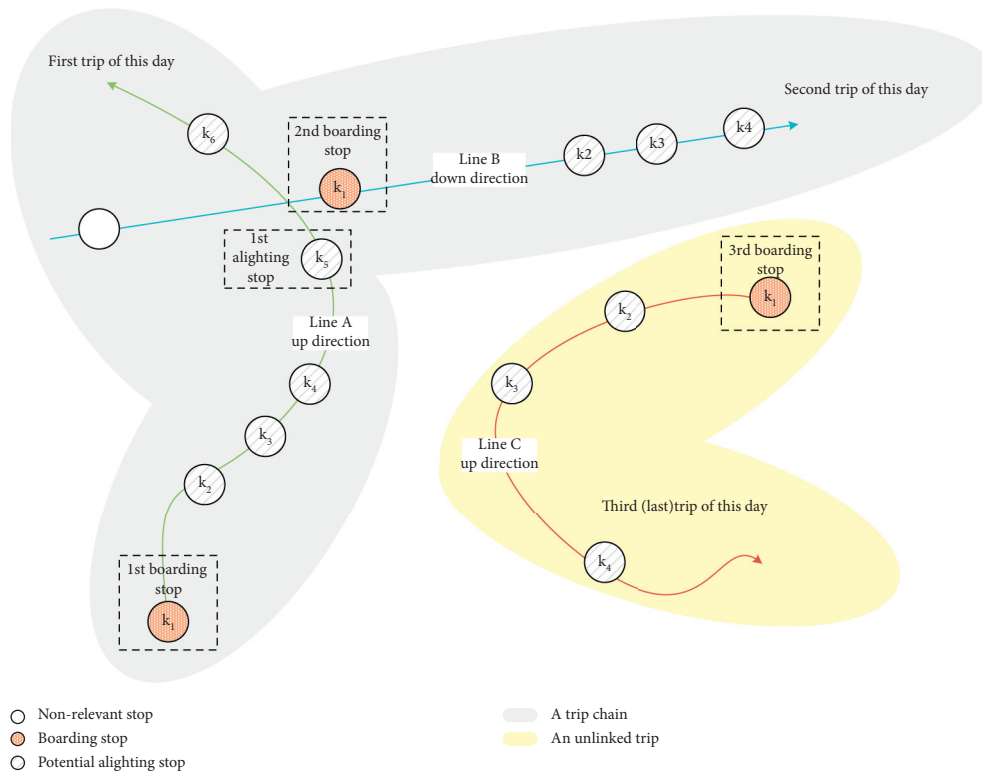
FIGURE 2: Some examples of alighting stop determination.

the next trip's line C is far away from line B, and the distances between the 3rd boarding stop and potential alighting stops in the second trip are not within the range, the second trip does not satisfy the approach criteria, and it is unlinked.

(iii) For the third (last) trip, similarly, we cannot detect the destination because the first boarding stop of the day is far away from potential alighting stops, and it is also an unlinked trip.

In this paper, we make contributions to identify alighting points of the second and third trips.

2.2. Data Description

2.2.1. Static Bus Stop Information.
Static bus stop information contains the line number, the line's direction (up or

down), and the index, name, longitude, and latitude of each stop along this direction of this line in Table 1. It is needed to count the number of lines passing through each stop, which is useful in the two-layer stacking framework method.

2.2.2. Bus GPS Data.
As can be seen from Table 2, GPS data embedded on all buses have ten fields, line number, direction, stop index, stop name, bus license plate number, longitude, latitude, date, vehicle operation shift, and timestamp. For a smart card record, when the interval of the record's transaction time and the GPS timestamp is minimal, the corresponding stop index, name, longitude, and latitude will be regarded as the boarding stop information.

2.2.3. Land-Use Type Attraction Coefficient.
The urban construction land with the bus stop as the center and within

TABLE 1: Static bus stop information.

Line no.	Direction	Stop index	Stop name	Longitude	Latitude
9 * 5	Up	1	A	118.056	24.6186
9 * 5	Up	2	B	118.058	24.6145
9 * 5	Up	3	C	118.060	24.6129

TABLE 2: Bus GPS data.

Line no.	Direction	Stop index	Stop name	Bus license plate no.	Longitude	Latitude	Date	Vehicle operation shift	Timestamp
9 * 5	Up	1	A	D0 * 1 *	118.056	24.6186	Aug. 8, 2018	12	8 : 00 : 15
9 * 5	Up	2	B	D0 * 1 *	118.058	24.6145	Aug. 8, 2018	12	8 : 06 : 45
9 * 5	Up	3	C	D0 * 1 *	118.060	24.6129	Aug. 8, 2018	12	8 : 08 : 55

the radius of $\mathrm{Dis}^{\mathrm{land-use}}$ meters is taken as the research scope in this paper to get the land-use type attraction coefficient around the stop. There are eight land-use types around stops within the study radius [22]: Residential; Commercial and Business Facilities; Administration and Public Services; Industrial, Manufacturing; Logistics and Warehouse; Municipal Utilities; Green Space and Square; Road, Street, and Transportation.

As there are not only urban construction land but also unused land and other types of land in the study area around bus stops, the total area of urban construction land around each station is not necessarily equal. However, the nonurban construction land has less attraction to the daily life of urban residents, so they are not considered here. What is more, the attraction coefficient is determined according to the scale of the city, and there is no significant difference in its values of cities with similar scales.

2.2.4. Smart Card Data. Raw smart card data in most cities in China are illustrated in Table 3, which contains ID, card type, boarding date, transaction time, line number, direction, and bus license plate number. Users tap smart cards only when boarding, but the boarding location and alighting information are not recorded for efficiency [8].

To get the boarding location, bus GPS data help to detect the minimal time interval of the transaction time and the GPS timestamp, and raw smart card data can be identified with the corresponding stop index, name, longitude, and latitude, as shown in Table 4.

For the identification of records' alighting information, it is necessary to identify destinations by the trip chain method firstly, and the rest of the data of unlinked trips are the focus of this paper. In the case of He and Trépanier, for the last trip of the day, the identification rate when destinations are found in the first trip of the next day is 4.17%, which is lower than that when alighting points are found in the first boarding station of the current day [23]. Besides, Cui's research showed that using multisource IC card data can estimate more alighting stops with higher accuracy than single-source IC card data [12]. Thus, the trip train method used in this paper is based on multisource data and two hypotheses mentioned above without destination found in the first trip of the next day. Records without alighting stop

determination by the trip chain method are the study objects in this paper.

2.3. Related Works. Methods used to determine alighting stops of unlinked trips can be divided into aggregate and disaggregate models [24]. The aggregate model's research object is the group, and models can determine the alighting passenger flow at each stop but cannot infer the destination of every smart card record accurately. The method based on the allure of bus stop is popular from aggregate models, and every potential alighting stop attraction can be obtained with factors such as trip distance, nature of land use around the stop, and the transferability of each bus site [25, 26]. In contrast, disaggregate models can identify every record's alighting stop, and they are widely used in more researches, such as the estimation of vehicle load profile and other network performance measures [6, 7]. Therefore, in this work, the framework-based method proposed is disaggregate. More methods of disaggregate models are introduced as below.

2.3.1. Probability Methods. Many scholars calculated the alighting probability at each potential stop in many ways, and the one with the highest probability was the estimated destination. Most researchers used records identified with destinations by the trip chain method as historical datasets to get the alighting probability. He and Trépanier constructed spatial probability and temporal probability in a kernel estimation via the probability of discrete variables with continuous variables, which were multiplied to predict the alighting probability at each potential stop, and the one with the highest probability was the estimated destination [23]. Also, they found a suitable threshold for the distance between estimated and observed alighting stops with the changes in accuracy [27].

However, these approaches failed to cover all destinations of records in unlinked trips with a low identification rate [23, 27].

2.3.2. Step-by-Step Method. Some scholars used several fixed methods to infer alighting stops of unlinked trips, and most of these methods were selected from probability methods.

TABLE 3: Raw smart card records.

ID	Card type	Boarding date	Transaction time	Line no.	Direction	Bus license plate no.
$8*\!*350$	Common	Aug. 8, 2018	$8:06:48$	$9*5$	Up	$D0*1*$
$8*\!*170$	Common	Aug. 8, 2018	$8:06:51$	$9*5$	Up	$D0*1*$
$8*\!*630$	Common	Aug. 8, 2018	$8:06:54$	$9*5$	Up	$D0*1*$

TABLE 4: Smart card records with boarding stops determination.

ID	Card type	Boarding date	Transaction time	Line no.	Direction	Bus license plate no.	Boarding stop			
							Index	Name	Longitude	Latitude
$8*\!*350$	Common	Aug. 8, 2018	$8:06:48$	$9*5$	Up	$D0*1*$	2	B	118.058	24.6145
$8*\!*170$	Common	Aug. 8, 2018	$8:06:51$	$9*5$	Up	$D0*1*$	2	B	118.058	24.6145
$8*\!*630$	Common	Aug. 8, 2018	$8:06:54$	$9*5$	Up	$D0*1*$	2	B	118.058	24.6145

Hu et al. used individual characteristics for alighting attraction weighting and identified all records' destinations in their case. Firstly, they used the high-frequency stop method, which picked the potential alighting stop with the highest individual boarding frequency as a destination, and the rest of the data were turned to the next step. Secondly, the ratio of boarding passenger flow in each potential stop to all potential stops was calculated, which helped allocate potential destinations to smart card records randomly. However, the results only showed high reliability in a cluster analysis, and they didn't validate the accuracy of alighting stop determination [24]. Li et al. did a similar study, but at the second step, they chose the stop attraction method, which used the stop with the largest boarding passenger flow in the same shift as the destination. But if no one boarded the bus at the last few stops, the attraction coefficients of these stops were all zero, and selecting the stop with the largest probability was impossible. Therefore, few records may not get alighting stops and the recognition rate was difficult to reach 100%. It was worth mentioning that Li's research got the identification accuracy based on the distance and weight between the true and the estimated destinations [28].

Step-by-step method always used several fixed methods for different datasets, which led to poor generalization ability and low accuracy, and some methods were not enough to identify all alighting stops [24, 28].

2.3.3. Machine Learning Algorithms. Machine learning models could find hidden insights and produce reliable decisions based on the learning from historical data, which were applied on bus alighting stop determination in recent research. Yan et al. developed two-step algorithms with KNN, Decision Tree, Random Forest, or other machine learning algorithms to cover all records of unlinked trips [8]. Destinations were detected by each machine learning algorithm with several features such as origin location, boarding time, bus line number, and two features (number of POIs, distribution of POI points) from POI data instead of land-use information. Besides, this research divided passengers into different groups by using K-means clustering to accurately estimate the alighting stop.

However, machine learning algorithms were more complex and difficult to explain, and parameters needed to

be determined in advance and their values had a great influence on the results [8]. In the two-step algorithm with KNN, the value of the nearest neighbor sample was usually subjective and lacked an objective basis. In the two-step algorithm with Decision Tree, if there was no reasonable restriction and pruning for the tree's growth, the small probability events in the training dataset would be completely included, which was prone to overfitting, leading to low prediction accuracy. The two-step algorithm with Random Forest contained multiple decision trees and could reduce overfitting to a great extent. The alighting stop was determined by the output results from multiple decision trees, and the number of trees established in this method was needed to be set manually. In large datasets, too many trees would cost too much time and space, while too few trees reduced the accuracy, so it was difficult to get a suitable value for the number of trees in practical application.

From the literature survey, it can be observed that most of the researchers use only one method (or machine learning algorithm) or several fixed methods in sequence to identify the alighting stops of unlinked trips, and some of them cannot determine all destinations. However, which method or combination is the best for a different dataset is hard to determine, and all alighting stops are expected to be identified as possible. In this work, a two-layer stacking framework is proposed to get the appropriate contributions of several methods in a different dataset simultaneously, which has strong generalization ability and can detect all destinations with high accuracy. Moreover, the step-by-step method of Li [28] and the two-step algorithms with KNN, Decision Tree, or Random Forest of Yan's [8] are chosen as comparison methods.

3. Method

A two-layer stacking framework based on alighting stop determination of unlinked trips can be seen in Figure 3. First, five methods in the first layer are described, and they are high-frequency stop method [28], stop attraction method [28], transfer convenience method [25], land-use type attraction method [26], and I-GHSM presented here. Then, the logistic regression model in the second layer is presented. All results from five methods in the first layer and
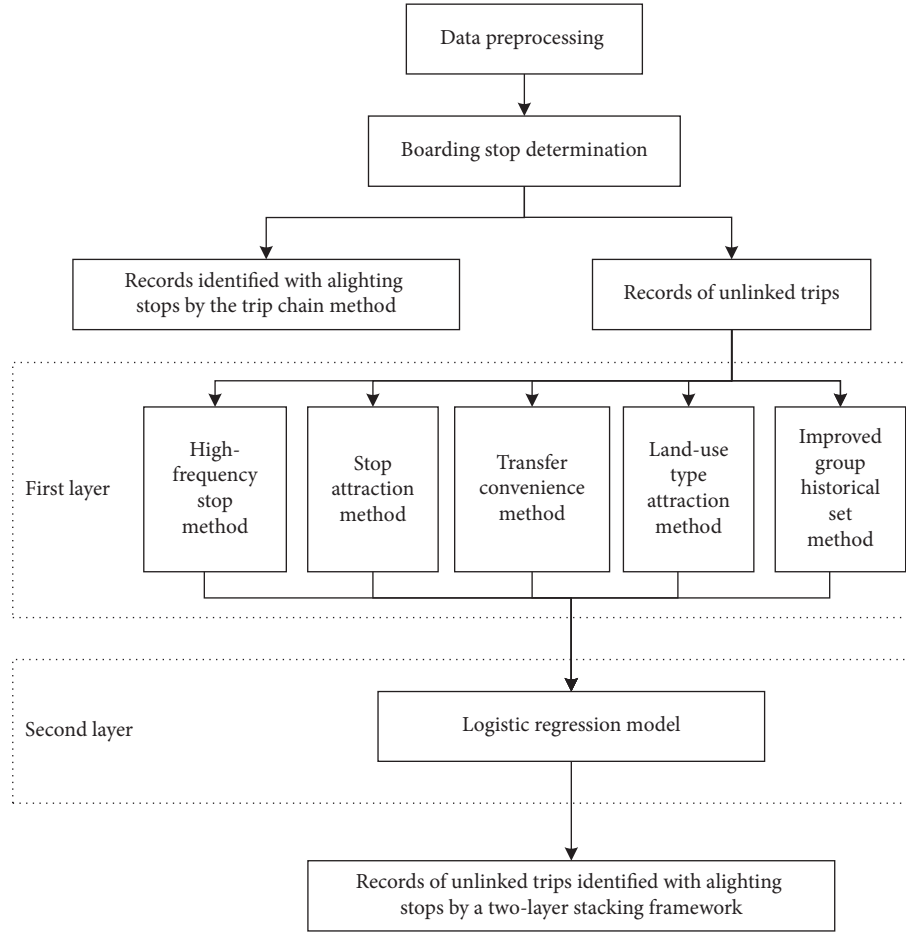
FIGURE 3: Alighting stop determination of unlinked trips based on a two-layer stacking framework.

the contribution weight of each method can be obtained via the regression model in the second layer [29]. The two-layer stacking framework uses outputs from the first layer as learning inputs of the second layer, which can correct the systematic deviation in the learning algorithm to improve the accuracy. At last, the destination of an unlinked trip is detected when the probability is maximal.

3.1. Methods in the First Layer.

During the research period D, when the passenger whose ID is m boards at the stop $s_{m,d,b}^{in}$ of trip b on day $d (\in D)$, the stop $s_{m,d,b}^{in}$ is the k_1th stop on direction f of line l, and there are total $K_{l,f}$ stops on the same direction of line l. For the potential alighting stops numbered k_2 ($k_1 < k_2 \le K_{l,f}$), alighting probabilities can be obtained by high-frequency stop method [28], stop attraction method [28], transfer convenience method [25], land-use type attraction method [26], and improved group historical set method (I-GHSM), respectively, as shown in Table 5.

3.1.1. High-Frequency Stop Method.

There are round-trip characteristics in residents' public transportation [9], which leads to bus stops with higher boarding frequency that have more attraction to alight. For the passenger with ID m, the

boarding frequency at every potential stop numbered k_2 is Nup_{k_2} during D days, and the alighting probability of the k_2th stop of trip b can be seen in [28]

$$
P_{m,d,b}^1 (k_1, k_2) = \begin{cases} \dfrac{\mathrm{Nup}_{k_2}}{\sum_{k_2=k_1+1}^{K_{l,f}} \mathrm{Nup}_{k_2}}, & \exists \mathrm{Nup}_{k_2} > 0, \\[12pt] 0, & \forall \mathrm{Nup}_{k_2} = 0. \end{cases} \tag{1}
$$

3.1.2. Stop Attraction Method.

Stop attraction means passengers prefer to alight at popular stops with more people getting on the bus. For the trip b of the passenger with ID m, passengers in the same bus shift can be found, and their boarding frequency at stop numbered k_2 is Num_{k_2}. Therefore, the alighting probability of the k_2th stop by the stop attraction method is given as [28]

$$
P_{m,d,b}^2(k_1, k_2) = \begin{cases} \dfrac{\mathrm{Num}_{k_2}}{\sum_{k_2=k_1+1}^{K_{l,f}} \mathrm{Num}_{k_2}}, & \exists \mathrm{Num}_{k_2} > 0, \\[12pt] 0, & \forall \mathrm{Num}_{k_2} = 0. \end{cases} \tag{2}
$$

TABLE 5: Five methods in the first layer.

Method	Data source	Alighting probability of the k_2th stop
High-frequency stop method [28]	Smart card data with boarding stops	$P^1_{m,d,b}(k_1, k_2)$
Stop attraction method [28]	Smart card data with boarding stops	$P^2_{m,d,b}(k_1, k_2)$
Transfer convenience method [25]	Static bus stop information	$P^3_{m,d,b}(k_1, k_2)$
Land-use type attraction method [26]	Land-use type attraction coefficient	$P^4_{m,d,b}(k_1, k_2)$
I-GHSM (improved group historical set method)	Smart card data with boarding stops, smart card data determined with destinations by the trip chain method	$P^5_{m,d,b}(k_1, k_2)$

3.1.3. Transfer Convenience Method. If more bus lines are passing through a stop, it is convenient to transfer at this stop, and its attraction is greater than others. There are L_{k_2} bus lines passing through the k_2th stop based on static bus stop information, so the alighting probability of each potential stop by the transfer convenience method can be obtained via equation (3) [25].

$$P^3_{m,d,b}(k_1, k_2) = \frac{L_{k_2}}{\sum_{k_2=k_1+1}^{K_{l,f}} L_{k_2}}. \quad (3)$$

3.1.4. Land-Use Type Attraction Method. Land-use type is one of the determinant factors for bus passengers' destinations. If there are shopping malls near the bus stop, the attraction is greater for people to get off the bus. There are eight types of urban construction land use around each stop, and C_h is the attractive coefficient of the $h \in \{1, 2, \ldots, 8\}$th land-use type. Within the radius of $\text{Dis}^{\text{land-use}}$ meters around the k_2th stop, the land occupation ratio of the hth land-use type is $C_{k_2, h}$. The alighting probability of k_2th stop can be illustrated in equation (4) with the land-use type attraction method [26].

$$P^4_{m,d,b}(k_1, k_2) = \frac{\sum_{h=1}^{H}(C_{k_2,h} \cdot C_h)}{\sum_{k_2=k_1+1}^{K_{l,f}} \sum_{h=1}^{H}(C_{k_2,h} \cdot C_h)}. \quad (4)$$

3.1.5. Improved Group Historical Set Method. Bus stops have similar attractions for records with similar behavior patterns. To cluster these records into a group more directly and accurately, based on the GHSM in our previous study, the I-GHSM is proposed here with more indicators considered and using each datum as the smallest research unit. Specific steps of the I-GHSM are described as follows:

(i) Construction of clustering indicators. Several indicators are used to classify records, and there are two types of indicators in this paper [17], as shown in Table 6. In the first type, some fields are picked as base indicators to ensure that the record is the smallest unit. In the second type, we set up some indicators to mine more information. For example, the type of boarding stop is the grade of the passenger flow center to which the station belongs. Records with the same type of boarding stop have

similar travel demands. There are three grades in this paper, and they are clustered by the k-means algorithm based on the boarding stop index and its passenger flow. The loyalty of passengers makes records with similar behavior rules closer via clustering the boarding times and the amounts of smart cards into three categories by the k-means algorithm. Besides, the k-means algorithm mentioned above is described specifically in [30] and [31].

(ii) Normalization of indicators. There are different ranges among eleven indicators in Table 6, so min-max standardization is used to normalize indicators, and the value of each indicator is made to be scaled to unit size.

(iii) Clustering is based on the K-means algorithm. The most widely used K-means clustering is picked to cluster records, and we can get the optimal number of clustering categories C with groups $\{R_1, \ldots, R_c, \ldots, R_C\}$ via the elbow rule. Data in each group have a similar behavior pattern based on the indicators mentioned above.

(iv) Determining the alighting probability at each possible stop is the last step. All records in the same group R_c with destinations by the trip chain method are used as the group historical dataset. The alighting frequency at every potential stop numbered k_2 is $M^{\text{down}}_{k_2}$ according to the historical dataset, and the alighting probability of the k_2th stop is given as

$$P^5_{m,d,b}(k_1, k_2) = \begin{cases} \dfrac{M^{\text{down}}_{k_2}}{\sum_{k_2=k_1+1}^{K_{l,f}} M^{\text{down}}_{k_2}}, & \exists M^{\text{down}}_{k_2} > 0, \\ \\ 0, & \forall M^{\text{down}}_{k_2} = 0. \end{cases} \quad (5)$$

3.2. Logistic Regression Model in the Second Layer. The logistic regression model with strong interpretability is used in the second layer, which can get each method's contribution weight directly.

3.2.1. Model Constructing. The alighting probabilities at every potential destination are obtained by each method in the first layer, and they are inputs of the logistic regression model. The range of values in the input is [0, 1], so no

TABLE 6: Indicators of smart card data for clustering.

Type	Index	Indicators	Quantification
Base indicators	1	ID	——
	2	Type	——
	3	Date	Making data into numbers, such as using 20181101 rather than November 1, 2018
	4	Transaction time	Changing transaction time to numbers in seconds with $00:00:00$ a day as a reference, such as using 28800 rather than $8:00:00$
	5	Boarding stop index	——
Constructed indicators	6	Weekday	$=0$: The boarding day is not a weekday $=1$: The boarding day is a weekday
	7	Weather	$=0$: It rains on the boarding day $=1$: It doesn't rain on the boarding day
	8	Peak hours	$=0$: Transaction time is not in peak hours $=1$: Transaction time is in peak hours
	9	Type of boarding stop	$=1$: The boarding stop is the first-class passenger flow center $=2$: The boarding stop is the second-class passenger flow center $=3$: The boarding stop is the third-class passenger flow center
	10	Administrative region of boarding stop	$=1$: Boarding stop is in administrative region 1 $=2$: Boarding stop is in administrative region 2 The rest can be done in the same manner
	11	Loyalty of passenger	$=1$: Bus is an occasional selection $=2$: Bus is an alternative selection $=3$: Bus is a loyal selection

normalization operation is needed. Each record with a boarding stop is combined with a potential alighting stop to be a pair, and every pair has a label. The label is 1 when the potential destination is the true destination, and 0 stands for other situations.

When the passenger whose ID is m boards at the stop $s_{m,d,b}^{\text{in}}$ on day d of unlinked trip b, the second layer's output is the probability when the label is 1 at the potential k_2th stop in equation (6). Similarly, the probability when the label is 0 can be calculated.

$$P\left(Y_{k_2} = 1 | \overrightarrow{x}_{k_2}\right) = \frac{\exp\left(\overrightarrow{w}^T \cdot \overrightarrow{x}_{k_2}\right)}{1 + \exp\left(\overrightarrow{w}^T \cdot \overrightarrow{x}_{k_2}\right)}, \quad (6)$$

where $\overrightarrow{x}_{k_2} = [P_{m,d,b}^1(k_1,k_2), P_{m,d,b}^2(k_1,k_2), P_{m,d,b}^3(k_1,k_2), P_{m,d,b}^4(k_1,k_2), P_{m,d,b}^5(k_1,k_2), 1]^T$ is the input vector; $\overrightarrow{w} = [w_1, w_2, w_3, w_4, w_5, b]^T$ is the weight vector to be learned, where b is bias.

3.2.2. Model Learning.
When learning the model, records identified with alighting stops by the trip chain method based on multisource data are used as the training and testing datasets. Records of unlinked trips can be determined with destinations via the learned model.

Every boarding record only has one alighting stop, and records with the label 0 are more than that with label 1. To solve the problem that the smart card data is unbalanced, the large sample data labeled 0 is sampled randomly, and the small sample labeled 1 is sampled based on the SMOTE algorithm. Then the two parts of data merge as a new dataset, and 90% of the set is selected randomly as the training set, while the remaining 10% is the testing set to verify the validity of the model with the evaluation index of F1-Score.

The maximum likelihood estimation method is used to estimate the parameters, and the Limited-memory BFGS (L-BFGS) algorithm for large-scale data computing is chosen to solve the optimization problem aiming to maximize the likelihood function. When the maximum likelihood estimation of $\overrightarrow{w}$ is $\widehat{\overrightarrow{w}}$, we can get the probability of k_2th stop when the label is 1 as in equation (7), and the probability when the label is 0 can be calculated similarly.

$$P\left(Y_{k_2} = 1 | \overrightarrow{x}_{k_2}\right) = \frac{\exp\left(\widehat{\overrightarrow{w}}^T \cdot \overrightarrow{x}_{k_2}\right)}{1 + \exp\left(\widehat{\overrightarrow{w}}^T \cdot \overrightarrow{x}_{k_2}\right)}. \quad (7)$$

3.3. Alighting Stop Determination of Unlinked Trips.
To ensure that every boarding record has one alighting stop, we choose the potential alighting stop with the maximum probability when the label is 1 as the destination.

When the passenger whose ID is m boards at the stop $s_{m,d,b}^{\text{in}}$ on day d of unlinked trip b, the alighting stop is the k_2th stop as shown in equation (8). Furthermore, alighting stop name, longitude, and latitude can be obtained by combining the stop index with static bus stop information.

$$k_2 = \arg\max_{k_2}\left\{P\left(Y_{k_1+1} = 1|\overrightarrow{x}_{k_2}\right), \ldots, P\left(Y_{k_2} = 1|\overrightarrow{x}_{k_2}\right), \ldots, P\left(Y_{K_{l,f}} = 1|\overrightarrow{x}_{k_2}\right)\right\}. \tag{8}$$

4. Case Analysis

4.1. Dataset. Xiamen BRT (Bus Rapid Transit System) is the first BRT system adopting the viaduct mode in China. The smart card data of BRT contains complete information of boarding and alighting stops, and the physical isolation of lanes makes several BRT lines form a small bus network. Smart card records with boarding and alighting stops along BRT Line Kuai 1 in Xiamen are research subjects in this paper. Besides, data of other BRT lines in Xiamen are used to complete passengers' trip trains (see Figure 4).

Selecting data of IC card owners who are on BRT Line Kuai 1 in November 2018, 3673184 records with boarding stops are obtained. Among them, 2425101 records can be identified with alighting stops by the trip chain method with multisource data, and the accuracy is 80.96%, as shown in Figure 5 [12]. Based on all records with boarding stops, more than 80% of passengers traveled only 8 times or less this month, so the individual historical set (the card history) is small, and the group historical set method proposed is needed. Also, there are four types of IC cards including common cards, student cards, elderly cards, and special cards. The morning peak is $7:00:00–9:00:00$ and the evening peak is from $17:00:00–19:00:00$ in November 2018, and the weather in each day can be obtained from the China Meteorological Administration.

BRT Line Kuai 1 has 27 stops, and its up and down directions run through the same stops. These stops are distributed in three administrative areas: Siming District, Huli District, and Jimei District. The land-use type distributions around 27 stops within 800 meters are illustrated in Figure 6. The attraction coefficient varies because of the different land-use nature, but it has good portability in similar cities. This paper sets the attraction coefficient of each land-use type in Table 7, according to experiences and relevant studies [32]. What's more, the number of lines passing through each stop along BRT Line Kuai 1 can be counted with static bus stop information and is shown in Figure 7.

4.2. Evaluation Methods. The identification rate Iden and the accuracy Acc are used as evaluation indexes to measure the method's performance, which is shown in equations (9) and (10), respectively.

$$\text{Iden} = \frac{N_{\text{iden}}^{\text{un}}}{N^{\text{un}}}, \tag{9}$$

$$\text{Acc} = \frac{N_{\text{iden}-r}^{\text{un}}}{N_{\text{iden}}^{\text{un}}}, \tag{10}$$

, where N^{un} is the total number of records that need to be identified for alighting stops in unlinked trips, $N_{\text{iden}}^{\text{un}}$ is the number of records that can be determined with destinations, and $N_{\text{iden}-r}^{\text{un}}$ is the number of records with destinations

predicted correctly. The higher the values of Iden and Acc, the higher the prediction accuracy of the model.

4.3. Parameters Setting. Parameters involved in this case are determined as follows:

4.3.1. Distance Threshold Setting Based on the Trip Chain Method. According to the distance between BRT stops in Xiamen, this paper sets the radius of $\text{Dis}^{\text{land-use}}$ = 800 meters and the distance threshold in the trip chain method as 2000 meters.

4.3.2. Determination of Penalty Coefficient in the Logistic Regression Model. In the second layer of our method, after many experiments, the best value of penalty coefficient in the logistic regression model is 100.

4.3.3. Parameters Setting of Comparison Methods. Two-step algorithms with KNN, Decision Tree, Random Forest [8], and the step-by-step method of Li [28] are selected for comparison. The theoretical analysis and comparison between these methods and the method proposed have been introduced in Section 3.1.3.

In two-step algorithms with KNN (K-Nearest Neighbor), Decision Tree, and Random Forest, existing studies use POI data to replace the land-use type around stops. However, the land-use type is known in this case, so there is no need to use POI data instead. After many experiments, it is determined that the nearest neighbor sample value is 1000 in the KNN, the number of trees established is 2000 in the two-step algorithms with Random Forest, and the Gini coefficient is selected as the standard in the two-step algorithm with Decision Tree.

The step-by-step method of Li [28] and our approach all need to use the passenger flow at each stop of the same bus shift. However, BRT's smart card records only have boarding and alighting stops, and we cannot sort out the shift passengers took. Therefore, each station's passenger flow within the hour of the passenger's timestamp is taken as the passenger flow needed in this case.

4.4. Results

4.4.1. Methods in the First Layer. In the first layer, I-GHSM is proposed, whose results of clusters are introduced as given below. Besides, the comparisons of I-GHSM and GHSM in the method based on expanded history trip records are shown. At last, results of unlinked trips' destinations determined by each method in the first layer are calculated.

In I-GHSM, when constructing clustering indicators, the passenger loyalty index should be calculated firstly, which is

```
┌─────────────────────────────────┐     ┌─────────────────────────────┐
│ IC card data with boarding and  │     │ IC card records of other    │
│ alighting stops along BRT line  │     │ BRT stops (dataset B)       │
│ Kuai 1 (dataset A)              │     │                             │
└─────────────────────────────────┘     └─────────────────────────────┘
                    │                                    │
                    ▼                                    │
┌───────────────────────────────────────────────────────────────────────┐
│         Alighting stop determination by the trip chainmethod            │
└───────────────────────────────────────────────────────────────────────┘
                    │                                    │
        ┌───────────┘                                    └────────┐
        ▼                                                         ▼
┌─────────────────────────────────┐     ┌─────────────────────────────┐
│ Records identified with         │     │ Records of unlinked trips   │
│ alighting stops by the trip     │     │ in dataset a                │
│ chain method in dataset A       │     │                             │
└─────────────────────────────────┘     └─────────────────────────────┘
                                                          │
                                                          ▼
                                         ┌─────────────────────────────┐
                                         │ Records of unlinked trips   │
                                         │ identified with alighting   │
                                         │ stops by A two-layer        │
                                         │ stacking framework in       │
                                         │ dataset A                   │
                                         └─────────────────────────────┘
```

FIGURE 4: Schematic diagram of data relations used in this case.

- Records of Unlinked Trips
- Records Identified with Alighting Stops by the Trip Chain Method (Accuracy is 80.96%)
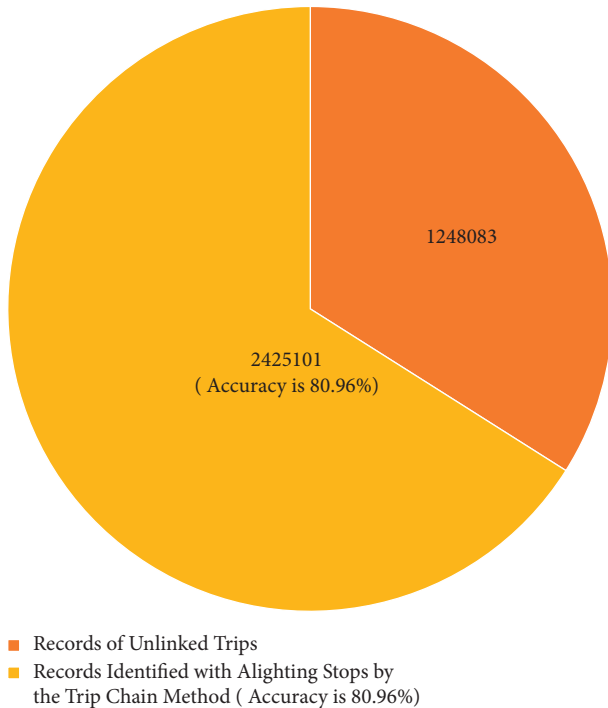
FIGURE 5: The relationship of smart card data in this case.

based on clustering the boarding times and the number of smart cards into three categories with the k-means algorithm. In this case, when the boarding frequency is 1 or 2 times per month, the passenger occasionally takes BRT with the value of passenger loyalty being 1; when the boarding frequency is 3 to 8 times per month, the passenger occasionally takes BRT as an alternative trip mode with the value being 2; when the boarding frequency is more than 8 times per month, the passenger is a loyal user of BRT and the value is 3. All eleven indicators in the I-GHSM can be determined and then processed by min-max standardization. After that, K-means clustering is done to cluster records, and the variation of average distortion degree with the number of

clusters is obtained, as shown in in Figure 8. As a result, the best number of clusters is 2 due to the elbow rule, and 3673184 records are divided into clusters R_1 and R_2, as shown in Table 8.

In the method based on expanded history trip records, the trip chain method with multisource data is used to estimate alighting stops firstly. For records in the unlinked trips, the individual historical set method and the GHSM are used in sequence for estimating alighting stops. However, using the I-GHSM as the third method can get better results, as shown in Table 9.

There are five methods in the first layer: the high-frequency stop method [28], the stop attraction method [28], the transfer convenience method [25], the land-use type attraction method [26], and the I-GHSM. If the potential alighting stop with the maximum probability is the destination, each method's results of alighting stop determination of unlinked trips are shown in Table 10.

4.4.2. Logistic Regression Model in the Second Layer. In this part, 2425101 records detected with alighting stops by the trip train method and multisource data help to learn and test the logistic regression. In the model training, each datum's boarding stop should be combined with each potential alighting stop to be a new pair, and there are a total of 40113752 new pairs in this case. Among them, 2425101 pairs' labels are 1 with true destinations, and the rest 37688651 pairs' labels are 0. Each pair can be considered as a new complete record. Considering the imbalance of data with different labels, records are resampled, as shown in Table 11.

To prevent the model from underfitting caused by the training dataset being too small, the data with label 0 after sampling is still more than expected, as seen in Table 10. However, the data ratio (Label is 0: Label is 1) has been reduced from 15.54 : 1 to 1.82 : 1 after sampling, which is a good improvement.
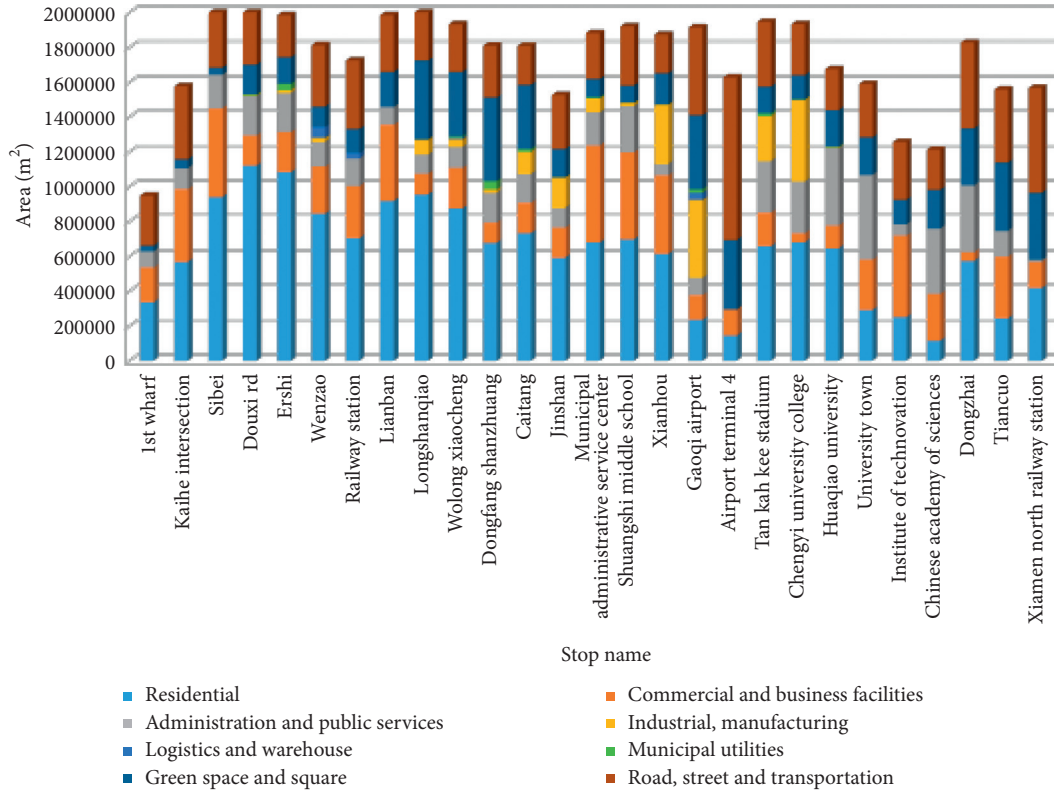
FIGURE 6: Land-use types and area coverage within 800 meters around stops along Xiamen BRT Line Kuai 1.

TABLE 7: Attractive coefficient of various land-use types.

Land-use type	Attractive coefficient
Residential	1
Commercial and business facilities	1.2
Administration and public services	1.1
Industrial, manufacturing	1
Logistics and warehouse	0.6
Municipal utilities	0.8
Green space and square	0.7
Road, street, and transportation	1.3

For 75178131 records after sampling, 90% of them are randomly selected as the training set of the logistic regression model and the remaining 10% of them as the test set. When the penalty coefficient is 100, the F_1-Score is 0.67, and the parameters of the trained logistic regression model are shown as equation (11).

$$\vec{w} = [5.7141, 0.0769, 0.1527, -0.1713, 0.4785, 2.4893]^T. \quad (11)$$

4.4.3. Alighting Stop Determination of Unlinked Trips. For 1248083 records of unlinked trips, identification rate and accuracy of existing methods [8, 28] and the two-layer stacking framework method proposed are in Table 12.

The accuracy of destination recognition based on a two-layer stacking framework changes in different conditions, and it can be calculated as follows:

(1) According to different days of one week in our study period, passenger transaction time can be divided into hours, and the accuracy of each period is shown in Figure 9

(2) According to different card types, passenger transaction time can be divided into hours, and the accuracy of each period is shown in Figure 10

(3) According to different card types and loyalty, each division's accuracy rate is counted and shown in Figure 11

4.5. Results Analysis.

(1) According to Table 9, the accuracy of unlinked trips' alighting stop determination with the individual historical set method and I-GHSM is 33.20%, which is 7.83% higher than that of the previous method. Therefore, I-GHSM achieves the goal that records with similar behavior patterns are clustered into a group more accurately than GHSM.

(2) As can be seen from Tables 10 and 12, the identification rate or the accuracy of each method in the first layer is lower than that of the two-layer stacking framework method. These verify that the contribution weight of each method obtained by the logistic regression model in the second layer is suitable enough to get better results, which brings the generalization ability for this dataset.
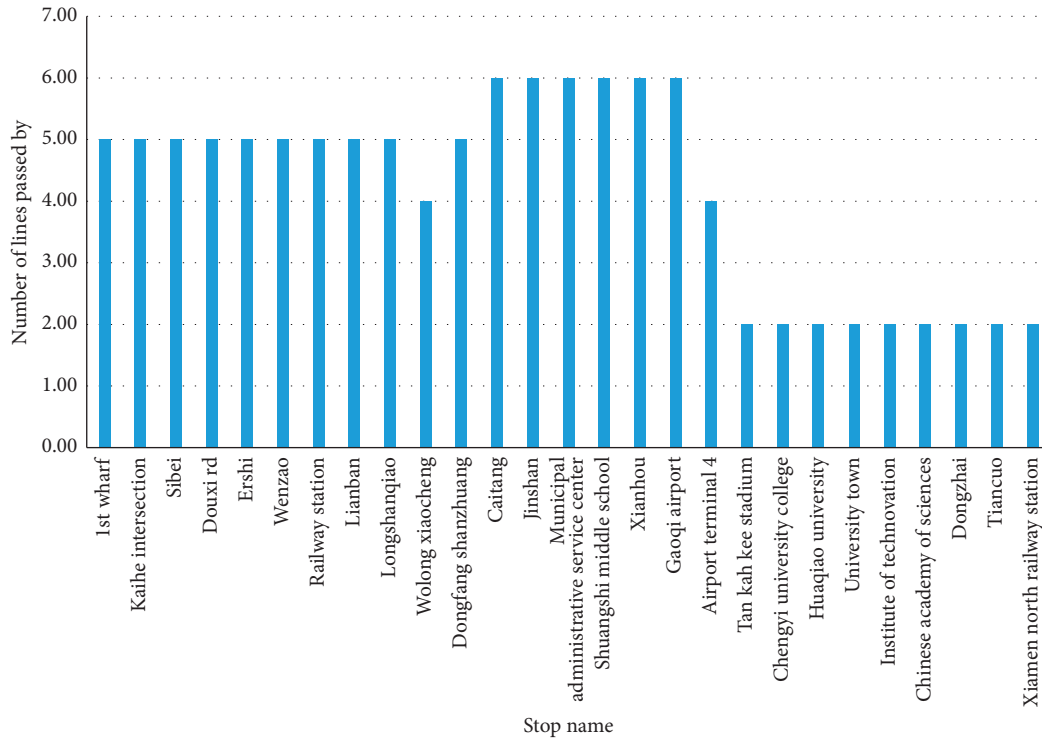
FIGURE 7: The stop name and the number of lines passed by.
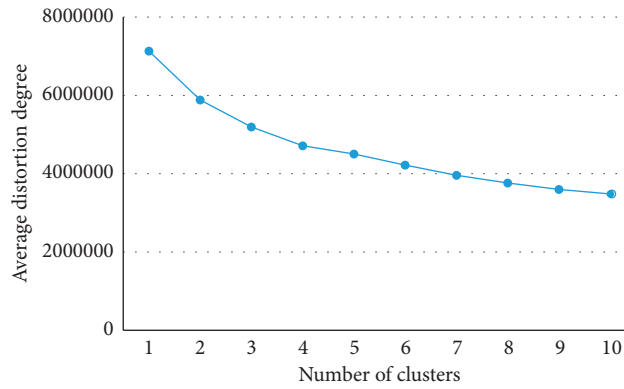
FIGURE 8: The average distortion degree varies with the number of clusters.

TABLE 8: Number of records in different clusters.

Cluster	Records with boarding stops	Records determined with alighting stops by the trip chain method	Records of unlinked trips
R_1	2129790	1351757	778034
R_2	1543394	1073344	470049
Total	3673184	2425101	1248083

(3) In Table 12, the accuracy of the two-layer stacking framework method is the highest than comparison methods, which indicates that using several models simultaneously is more effective than choosing only one or several fixed methods in sequence. Besides, the two-layer stacking framework-based method

detects all alighting points of unlinked trips, which is better than the step-by-step method of Li [28].

(4) As shown in Figure 9, the accuracy obtained during the morning peak (7 : 00 : 00–9 : 00 : 00) on weekdays was higher than that on weekends. In Figure 10, records from common and student cards account for

TABLE 9: Results of unlinked trips' alighting stop determination with the GHSM or the I-GHSM.

Method	Records identified with alighting stops	Identification rate	Records identified with alighting stops correctly	Accuracy (%)
Individual historical set method	446915	35.81	273628	61.23
GHSM/**I-GHSM**	801168	64.19	43042/**140721**	5.37/**17.56**
Total	1248083	100.00	316670/**414349**	25.37/**33.20**

TABLE 10: Results of unlinked trips' alighting stop determination by each method in the first layer.

Method	Records identified with alighting stops	Identification rate	Records identified with alighting stops correctly	Accuracy
High-frequency stop method [28]	926202	74.21	582581	62.90
Stop attraction method [28]	1248083	100.00	146400	11.73
Transfer convenience method [25]	1248083	100.00	118818	9.52
Land-use type attraction method [26]	1248083	100.00	141907	11.37
I-GHSM	1248083	100.00	225653	18.08

TABLE 11: The number of records before and after data sampling.

Type		Label is 0	Label is 1	Total
Before sampling		37688651	2425101	40113752
After sampling	Random under-sampling	24251010	2425101	75178131
	Oversampling based on SMOTE algorithm	24251010	24251010	

TABLE 12: Results of alighting stop determination of unlinked trips in different methods.

Method	Records identified with alighting stops	Identification rate (%)	Records identified with alighting stops correctly	Accuracy (%)
Two-step algorithms with KNN [8]	1248083	100.00	124808	10.00
Two-step algorithms with decision tree [8]	1248083	100.00	160379	12.85
Two-step algorithms with random forest [8]	1248083	100.00	202689	16.24
Step-by-step method of Li [28]	1247536	99.96	623876	50.01
Two-layer stacking framework method	**1248083**	**100.00**	**647538**	**51.88**
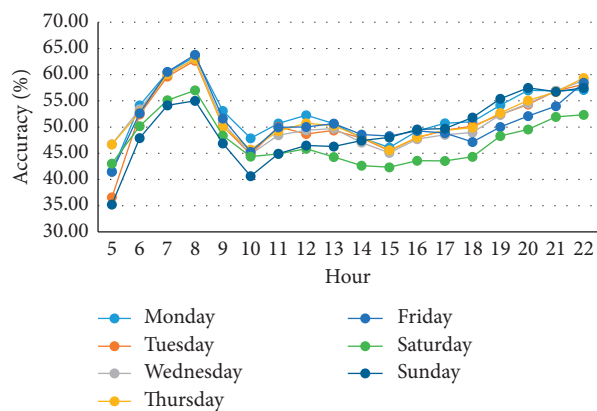
FIGURE 9: The accuracy based on the method proposed in different hours and days of the week.
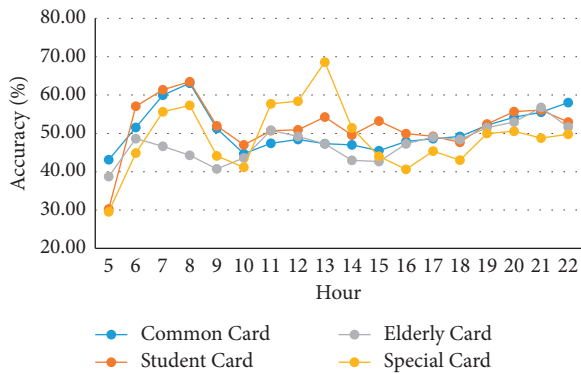
FIGURE 10: The accuracy based on the method proposed in different hours and card types.
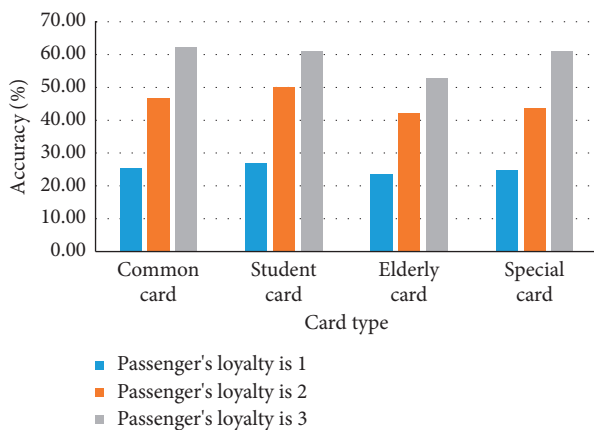
FIGURE 11: The accuracy based on the method proposed in different card types and loyalty.

more than 93.5% of the total smart card data, and their accuracy in the morning peak is higher than that in other periods. Therefore, the morning peak on weekdays is the most regular time for passengers to travel by public transport, which is consistent with the fact that there are many commuters (/students) from their fixed residence to their fixed workplace (/school) during this period. From the analysis of Figure 11, all card types meet the condition that higher boarding frequency leads to higher accuracy of alighting stops determination, which shows that high loyalty passengers' travel behavior is more regular than that of other passengers with lower loyalty.

5. Conclusions

In this paper, a method based on a two-layer stacking framework is proposed to get better accuracy in unlinked trips' alighting stop determination, and the improved group historical set method is presented in the first layer.

Xiamen's case shows that the I-GHSM can cluster records with similar behavior patterns into a group more accurately than the GHSM. Together with the individual historical set method, the I-GHSM improves accuracy by 7.83% in unlinked

trips' destination determination than our previous study. Besides, the method based on the two-layer stacking framework can detect all alighting points with a higher accuracy of 51.88%, which is better than the step-by-step method of Li [28], and two-step algorithms with KNN, Decision Tree, or Random Forest of Yan [8] in this case.

When comparing the identification results of each method in the first layer and the two-layer stacking framework method, the logistic regression model in the second layer is verified to bring the appropriate contribution weight of each method and the generalization ability. And, because the accuracy of the two-layer stacking framework method is the highest than comparison methods, using several models simultaneously is more effective than choosing only one or several fixed methods in sequence. After analyzing changes in identification accuracy based on the framework-based method in different conditions, the morning peak on weekdays is the most regular time for passengers to travel. And, the higher the boarding frequency is, the more regular their travel behavior will be.

However, in the second layer of the two-layer stacking-framework-based method, smart card data are still unbalanced after sampling when determining alighting stops of unlinked trips. Therefore, the sampling method needs to be further studied. What's more, our method can be applied to different datasets of other cities in the future to further verify its effectiveness.

Data Availability

The data used to support the findings of this study are available in Cui, Z., Wang, C., Chen, D., Lei, L. "Alighting stop determination of transit passengers based on expanded history trip records". Journal of Nanjing University (Natural Science), 2020, 56(2), 227–235. https://doi.org/10.13232/j.cnki.jnju.2020.02.008.

Conflicts of Interest

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

Acknowledgments

References

[1] M. Bagchi and P. R. White, "The potential of public transport smart card data," *Transport Policy*, vol. 12, no. 5, pp. 464–474, 2005.

[2] M. Utsunomiya, J. Attanucci, and N. Wilson, "Potential uses of transit smart card registration and transaction data to improve transit planning"" *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1971, no. 1, pp. 118–126, 2006.

[3] C. Morency, M. Trépanier, and B. Agard, "Measuring transit use variability with smart-card data," *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.

[4] F. Devillaine, M. Munizaga, and M. Trépanier, "Detection of activities of public transport users by analyzing smart card data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2276, no. 1, pp. 48–55, 2012.

[5] N. Nassir, A. Khani, S. G. Lee, H. Noh, and M. Hickman, "Transit stop-level origin-destination estimation through use of transit schedule and automated data collection system," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2263, no. 1, pp. 140–150, 2011.

[6] M. Trépanier and F. Vassiviere, "Democratized smartcard data for transit operator," in *Proceedings of the 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual MeetingITS AmericaERTICOITS JapanTransCore*, New York, NY, USA, 2008.

[7] M. Trépanier, C. Morency, and B. Agard, "Calculation of transit performance measures using smartcard data," *Journal of Public Transportation*, vol. 12, no. 1, pp. 79–96, 2009.

[8] F. Yan, C. Yang, and S. V. Ukkusuri, "Alighting stop determination using two-step algorithms in bus transit systems," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1522–1542, 2019.

[9] W. Liu, Q. Tan, and L. Liu, "Destination estimation for bus passengers based on data fusion," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8305475, 10 pages, 2020.

[10] M. A. Munizaga and C. Palma, "Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile," *Transportation Research Part C: Emerging Technologies*, vol. 24, pp. 9–18, 2012.

[11] Q. Zou, X. Yao, P. Zhao, H. Wei, and H. Ren, "Detecting home location and trip purposes for cardholders by mining smart card transaction data in Beijing subway," *Transportation*, vol. 45, no. 3, pp. 919–944, 2018.

[12] Z. Cui, C. Wang, D. Chen, and L. Lei, "Alighting stop determination of transit passengers based on expanded history trip records," *Journal of Nanjing University (Natural Science)*, vol. 56, no. 2, pp. 227–235, 2020.

[13] J. L. Machado, R. De Oña, F. Diez-Mesa, and J. De Oña, "Finding service quality improvement opportunities across different typologies of public transit customers," *Transportmetrica A: Transport Science*, vol. 14, no. 9, pp. 761–783, 2018.

[14] A. Bhaskar and E. Chung, "Passenger segmentation using smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1537–1548, 2014.

[15] C. Yang, F. Yan, and S. V. Ukkusuri, "Unraveling traveler mobility patterns and predicting user behavior in the Shenzhen metro system," *Transportmetrica A: Transport Science*, vol. 14, no. 7, pp. 576–597, 2018.

[16] X. Ma, C. Liu, H. Wen, Y. Wang, and Y.-J. Wu, "Understanding commuting patterns using transit smart card data," *Journal of Transport Geography*, vol. 58, pp. 135–145, 2017.

[17] K. Mohamed, E. Côme, L. Oukhellou, and M. Verleysen, "Clustering smart card data for urban mobility analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 3, pp. 712–728, 2016.

[18] P. Kumar, A. Khani, and Q. He, "A robust method for estimating transit passenger trajectories using automated data," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 731–747, 2018.

[19] N. Nassir, M. Hickman, and Z.-L. Ma, "Activity detection and transfer identification for public transit fare card data," *Transportation*, vol. 42, no. 4, pp. 683–705, 2015.

[20] A. A. Nunes, T. G. Dias, and J. F. Cunha, "Passenger journey destination estimation from automated fare collection system data using spatial validation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 1, pp. 133–142, 2015.

[21] G. E. Sánchez-Martínez, "Inference of public transportation trip destinations by using fare transaction and vehicle location data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2652, no. 1, pp. 1–7, 2017.

[22] Chinese Standard Net, *Code for Classification of Urban Land Use and Planning Standards of Development Land: GB 50137-2011*, Chinese Standard Net, Beijing, China, 2010.

[23] L. He and M. Trépanier, "Estimating the destination of unlinked trips in transit smart card fare data," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2535, no. 1, pp. 97–104, 2015.

[24] J. Hu, J. Deng, and Z. Huang, "Trip-chain based probability model for identifying alighting stations of smart card passengers," *Journal of Transportation Systems Engineering and Information Technology*, vol. 14, no. 2, pp. 62–67, 2014.

[25] M. Zhang, Y. Guo, and Y. Ma, "A probability model of transit OD distribution based on the allure of bus station," *Journal of Transport Information and Safety*, vol. 3, pp. 57–61, 2014.

[26] W. Xu, C. Deng, and B. Liu, "Approach on public traffic passenger flow statistics based on IC data," *China Journal of Highway and Transport*, vol. 26, no. 5, pp. 158–163, 2013.

[27] L. He, N. Nassir, M. Trépanier, and M. Hickman, *Validating and Calibrating a Destination Estimation Algorithm for Public Transpot Smart Card Fare Collection Systems*, CIRRELT), Montreal, Canada, 2015.

[28] J. Li, J. Zhang, J. Zhang, Q. Wang, and D. Peng, "An algorithm to identify passengers' alighting stations and the effectiveness evaluation," *Geomatics and Information Science of Wuhan University*, vol. 43, no. 8, pp. 1172–1177, 2018.

[29] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[30] X. S. Lu, M. Zhou, L. Qi, and H. Liu, "Clustering-algorithm-based rare-event evolution analysis via social media data," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 2, pp. 301–310, 2019.

[31] X. Xu, J. Li, M. Zhou, J. Xu, and J. Cao, "Accelerated two-stage particle swarm optimization for clustering not-well-separated data," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4212–4223, 2020.

[32] Q. Xiao and W. Xu, *Urban Traffic Planning*, pp. 212–282, China Communications Press Co., Ltd, Beijing, China, 1990.

Research Article

An Adaptability Analysis of the Space-Vehicle Traffic State Estimation Model for Sparsely Distributed Observation Environment

Han Yang [ID] [1] **and Qing Yu** [ID] [2]

[1]*Research Institute of Highway Ministry of Transport, Xitucheng Road, Beijing 100088, China*
[2]*The Key Laboratory of Road and Traffic Engineering of the Ministry of Education School of Transportation Engineering, Tongji University, Shanghai, China*

Correspondence should be addressed to Qing Yu; 1810514@tongji.edu.cn

The autonomous driving has shown its enormous potential to become the new generation of transportation in the last decade. Based on the automated technology, vehicles can drive in a new form, vehicle platoon, which can significantly increase the efficiency of the road system and save road resources. The space-vehicle traffic state estimation model has shown its benefits in modeling autonomous vehicle platoon in nonpipeline corridors with on- and off-ramps in ideal observation environment. However, in the current initial stage of automated connected vehicles' application, the observation environment is quite imperfect. Limited by financial and investment, traffic flow observation equipment is sparsely distributed on the road. How to adapt to the sparse observer layout is a critical issue in the current application of the space-time traffic state estimation, which is originally designed for the autonomous transportation. Therefore, this manuscript overviews the observation environment in practice and summarizes three key observation problems. This article designs 22 numerical experiments focusing on the three key issues and implements the space-time estimation model in different observation scenarios. Finally, the observation environment adaptability is analyzed in detail based on the experiment results. It is found that the accuracy of the estimation results can be improved with the highest efficiency under the premise of limited equipment input by reducing the observation interval to 1000 m and increasing the density of the observer to 1/km. For the road sections with relatively homogeneous traffic conditions, the layout of observation equipment can be relatively reduced to save the investment input. Also, the maintenance of observation equipment for the ramp with larger flow can be slowed down appropriately in limited equipment investment. This manuscript is of great practical significance to the popularization and application of connected automatic transportation.

1. Introduction

Autonomous driving has obtained abundant attractions in the last decade. Road automated driving as a new generation of information technology and the integration of the transport industry's development has become a new round of global scientific and technological innovation and industrial transformation. This technology will promote the continuous upgrading of the field of road traffic. At present, the government, enterprises, and investors all take this as the goal and direction to accelerate automated driving around the world [1–3].

In the ideal application scenario of autonomous driving, the next generation of sustainable urban development requires a new wave of infrastructure construction. However, in the early development stage of autonomous driving, due to financial and investment constraints, existing infrastructure needs to adapt to the application and promotion of autonomous driving. Therefore, we need to propose a traffic model that is not only applicable to the operation

mechanism of automatic driving but also can adapt to the existing observation infrastructure.

In the field of traffic flow modeling for connected automatic vehicles, the space-vehicle traffic flow model (referred to as "N-X traffic flow model") with vehicles as the research object has been proposed since 2007 [4, 5]. Besides the traditional space-time model (referred to as "X-T traffic flow model") based on space (referred to as "X") and time (referred to as "T") dimension [6–9], the traffic flow can also be described in vehicle (referred to as "N") dimension. In the X-T model, traffic flow can be described as the traffic character variables such as flow rate and speed varying with time index. Similarly, In the N-T model, traffic flow can be described in a different form: traffic character variables such as headway and pace which are varying with the vehicle index. In the X-T model, the time index can be integrated into 10 seconds or 5 minutes. Similarly, the vehicle index in the N-X model can also be integrated into a vehicle packet with 5 vehicles or 10 vehicles as well. The N-X model has attracted many researchers' interest and has shown its advantages in modeling vehicle platoon, which is the essential form of connected automatic traffic flow. In other words, the N-X model has shown its enormous potentials in the connected automatic transportation system. In 2014, Jin proposed the traffic state estimation framework based on the N-X traffic flow model (referred to as the N-X traffic state estimation model) for a pipeline corridor without inflow and outflow [10]. In 2019, Yang proposed an N-X traffic state estimation model adding the node model [11]. Then, this model can be applied into nonpipeline corridors with on- and off-ramps. Then, in 2019, Yang proposed a revised nonpipeline N-X traffic state estimation model by reducing the node assumption [12]. The proposed N-X traffic state estimation model has shown its adaptability in fixed-point vehicle matching detection environment. It is worth noting that the adaptability is mainly verified in the ideal observation environment. In the proposed model, the ideal observation environment refers to the relatively dense layout of traffic flow observers.

In recent years, with the reduced cost and wide use of data acquisition equipment, high quality, reliable data, and robust models are essential to the development of the intelligent transportation system, and transportation data mining based on big geospatial data has become a research hotspot and earns a lot of attention [13–15]. These emerged big data-based technologies have been proved as central to the development of transportation operation and management [16, 17].

With the rapid development in automobile manufacturing, industrial investment, and financing in the last two years, connected automatic vehicles are gradually finishing the theoretical verification and starting to enter the practical road environment. However, the development of the data environment of automatic driving is still in its infancy, and the infrastructure cannot generate high quality and reliable data to support the application of new big data mining technology. In other words, compared with the dense detector layout environment, the sparse layout detection environment will be widely applied in practical applications.

Therefore, for the initial application and promotion of connected automatic vehicles, it is very important to study the adaptability of the N-X traffic state estimation model in the practical imperfect detection environment with sparse detector layout, which has high potential to support the development of the next-generation intelligent transportation system.

Therefore, in this paper, the adaptability of the N-X traffic state estimation model is analyzed in detail in view of the practical imperfect detector layout environment, which is of great practical significance to the popularization and application of connected automatic transportation. This manuscript first analyzes the existing problems in the practical observation environment. Secondly, the observation infrastructure environment is designed and simulated based on the urban channel. The ability to adapt to the actual observation infrastructure environment is discussed from three aspects: trunk road observation equipment damage, trunk road observation interval distance, and ramp observation equipment missing. Finally, this article puts forward the design and planning suggestion of the observation infrastructure's layout in connected automated vehicle scenario.

2. Overview of Imperfect Observation Environment

Automatic vehicle license plate recognition (ALPR) is one of the important components in modern intelligent transportation systems [18]. ALPR technology can detect vehicles on the monitored road surface through roadside fixed observation equipment and automatically extract vehicle license plate information (including Chinese characters, English letters, Arabic numerals, and plate colors) for processing. At present, this technology has been widely used in highway vehicle management [19]. In practice, affected by natural conditions, financial investment, operation, and maintenance, the actual observation environment of ALPR is often "imperfect." For example, the damage of the equipment on main roads and ramps often occurs because of the maintenance. Also, the limit of financial investment and natural conditions often lead to the too wide layout intervals of the equipment on main roads.

Problem 1. Damaged main road observer

Affected by factors such as equipment operation and maintenance, the vehicle license plate recognition observer deployed on arterial roads is often damaged. The damaged observation equipment cannot transmit effective arterial road observation data, which will result in a lack of observation space for arterial road. Figure 1 shows the distribution of vehicle license plate recognition observers in Shenzhen, among which about 10% of vehicle license plate observers are damaged.

Problem 2. Different observation interval on arterial road

The arterial road observation interval size refers to the spatial distance between adjacent observers on the arterial road. The interval of the main road observation interval will be affected by factors such as financial investment and equipment layout environment. In addition, the granularity of the observation interval on arterial roads is often various

⊕ ALPR observer (work)
● ALPR observer (broken)

Figure 1: Layout and distribution of vehicle license plate recognition observation equipment in Shenzhen

for different observation scenarios such as expressways in central areas, expressways in suburbs, and expressways outside cities. Taking Shanghai as an example, the VLPR observation interval within the inner ring expressway is about 300m–500 m, on the middle ring expressway is about 1 km, and on the outer city expressway can reach 2 km or more.

Problem 3. Ramp observer missing

The ALPR observation equipment is also deployed on the up-ramp and the down-ramp to effectively detect the information of vehicles entering and exiting the arterial road. However, because the license plate recognition system is more inclined to recognize vehicles entering the main road, the observation equipment located on the off-ramp is often missing. In addition, with the deepening of urbanization, urban road networks are highly dense, which leads to the numerous ramps. Then, in practice, it is quite normal to cut down the ramp observation equipment to save investment. For example, in Shanghai's ALPR system, there is no observation equipment on off-ramps, and about 70% of the on-ramps do not install observation equipment.

3. Methodology

3.1. N-X State Process Model. The classic traffic state estimation framework includes two key components: a state process model and a measurement model [20–26]. Based on the proposed travel time transmission model, the state process model for the nonpipeline N-X Kalman filter framework is formulated as

$$T_i(n+1) = T_i(n) + \Delta n_i(p_{i+1}(n) - p_i(n))$$

$$+ \Delta n_i \left(\frac{p_{i+1}(n)^2 \left((1/p_i^r(n)) - (1/p_i^s(n)) \right)}{1 - p_{i+1}(n) \left((1/p_i^r(n)) - (1/p_i^s(n)) \right)} \right)$$

$$+ \zeta_i(n), \tag{1}$$

$$p_i(n) = \max \left\{ p^*, \frac{1}{C_i}, \frac{h_{\mathrm{jam},i}\left(T_i(n) - T_{\mathrm{free},i}\right)}{\Delta x_i} + \frac{1}{C_i} \right\}, \tag{2}$$

where $\zeta_i(n)$ denotes the state noise in the estimation. Travel time $T_i(n)$ is the state variable; $p_i(n)$ is a function of $T_i(n)$ based on equation (2) from the fundamental diagram, and $p_i^r(n)$ and $p_i^s(n)$ are system external input variables. System parameters include the capacity C_i, jam distance headway $h_{\mathrm{jam},i}$, and free-flow travel time $T_{\mathrm{free},x}$ for each segment i. System boundary variables include $p_1(n)$ and $p_{M+1}(n)$. In this paper, for the free-flow regime in the N-X fundamental diagram, p^* is defined as follows:

$$p^* = \max \left\{ p_i(n-1), \frac{1}{C_i} \right\}. \tag{3}$$

3.2. N-X Measurement Model. In practice, travel time cannot be fully observed because of the limited penetration of probe vehicle. Thus, we formulate the N-X measurement model as

$$\widehat{T}_i(n) = T_i(n) + \xi_i(n) = \Delta \mathrm{x} \cdot \tau_{\mathrm{x}}(\mathrm{n}) + v_x^\tau(\mathrm{n}), \tag{4}$$

where $\widehat{T}_i(n)$ denotes the estimated measurement of travel time $T_i(n)$ and $\xi_i(n)$ is the measurement noise.

3.3. State-Space Model. The state-space representation for the N-X traffic state estimation framework is given as follows:

$$T(n) = F(T(n-1), b(n-1), u(n-1), \alpha(n-1), \zeta(n-1)),$$

$$\widehat{T}(n) = H(T(n), \xi(n)), \tag{5}$$

where

$$\begin{bmatrix} T_1(n) \\ T_2(n) \\ \cdots \\ T_M(n) \end{bmatrix} = \begin{bmatrix} f(T_1(n-1), b_1(n-1), u_1(n-1), \alpha_1(n-1),) \\ f(T_2(n-1), b_2(n-1), u_2(n-1), \alpha_2(n-1),) \\ \cdots \\ f(T_M(n-1), b_M(n-1,X), u_M(n-1,X), \alpha_M(n-1)) \end{bmatrix} + \begin{bmatrix} \zeta_1(n-1) \\ \zeta_2(n-1) \\ \cdots \\ \zeta_M(n-1) \end{bmatrix},$$

$$\widehat{T}(n) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \cdot T(n) + \xi(n). \tag{6}$$

F and H are the vector functions based on equation (1)–(4). Vectors T, b, u, α, ζ, $\widehat{T}$, and ξ are system state variable, boundary variable, control input variable, parameter, process noise, system observation variable, and measurement noise, respectively. These vectors are defined as follows:

$$T(n) = \begin{bmatrix} T_1(n) & T_2(n) & \ldots & T_M(n) \end{bmatrix}^T,$$

$$b(n) = \begin{bmatrix} p_1(n) & p_{M+1}(n) \end{bmatrix}^T,$$

$$u(n) = \begin{bmatrix} p_1^r(n) & p_1^s(n) & \ldots & p_M^r(n) & p_M^s(n) \end{bmatrix},$$

$$\alpha(n) = \begin{bmatrix} h_{\mathrm{jam},1}(n) & \tau_{\mathrm{free},1}(n) & C_1(n) \\ \vdots & \vdots & \vdots \\ h_{\mathrm{jam},M}(n) & \tau_{\mathrm{free},M}(n) & C_M(n) \end{bmatrix}, \quad (7)$$

$$\zeta(n) = \begin{bmatrix} \zeta_1(n) & \zeta_2(n) & \ldots & \zeta_M(n) \end{bmatrix}^T,$$

$$\widehat{T}(n) = \begin{bmatrix} \widehat{T}_1(n) & \widehat{T}_2(n) & \ldots & \widehat{T}_M(n) \end{bmatrix}^T,$$

$$\xi(n) = \begin{bmatrix} \xi_1(n) & \xi_2(n) & \ldots & \xi_M(n) \end{bmatrix}^T.$$

3.4. Unscented Kalman Filter Configuration.

Because state process model equation (1) is nonlinear, unscented Kalman filter (UKF) is applied to estimate the traffic state. The UKF is an effective filter method for nonlinear systems, avoiding the linear approximation error and Jacobian matrix calculation. The UKF relies mainly on unscented transform (UT) to process the mean and covariance transition in a nonlinear system [27, 28].

3.5. Performance Evaluation.

To illustrate the estimation result, the traffic state travel time $T_i(n)$ in the N-X model is converted to travel speed $u_i(n)$. Furthermore, the N-X result recorded in the vehicle dimension is converted into the time dimension $u_i(t)$ via the N-X interpolation method [21]. The method tracks the link entry time of each flow packet and conducts linear interpolation of the traffic states.

The performances of the results are evaluated by the following performance indexes: the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). They are defined as follows:

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{\Theta} \sum_{t=1}^{t=\Theta} \left(u_i^{GT}(t) - \widetilde{u}_i(t) \right)^2},$$

$$\mathrm{MAE}_i = \frac{1}{\Theta} \sum_{t=1}^{t=\Theta} u_i^{GT}(t) - \widetilde{u}_i(t), \quad (8)$$

$$\mathrm{MAPE}_i = \frac{1}{\Theta} \sum_{t=1}^{t=\Theta} \left| \frac{u_i^{GT}(t) - \widetilde{u}_i(t)}{u_i(t)} \right| \times 100\%,$$

where $i = 1, \ldots, M$. Θ is the total number of time index during the estimation period:

$$\Theta = 3.5 \times \frac{3600}{\Delta t}. \quad (9)$$

M is the total number of segments in the corridor. $u_i^{GT}(t)$ and $\widetilde{u}_i(t)$ are the ground truth state and estimated state of segment i at timestamp t, respectively.

4. Design of Numerical Experiment

4.1. Simulation Model and Parameters.

The proposed N-X estimation model is evaluated through the VISSIM simulation model, which is calibrated by field loop data. VISSIM simulation allows for flexibility in generating vehicle-based paired detector data, along with high-resolution ground truth data for evaluating the congestion patterns. The VISSIM simulation model is established and calibrated by field loop detector data from IH-894 in Milwaukee, Wisconsin, during the morning peak hours from 5:30 am to 9:30 am [29]. The simulated area is a 4-mile segment of a nonpipeline freeway corridor with five on-ramps and three off-ramps, as shown in Figure 2. The nonpipeline freeway includes a significant bottleneck, which can be used to further analyze the accuracy of the model in the congestion pattern.

The entire corridor is divided into 12 segments with the same length of 500 m. Propagation packet size Δn is set as 50 vehicles for all the segments with (or without) freeway ramps. Uniform traffic parameters are assumed across the entire corridor to simplify the calculation. Then, C, k_{jam}, and u_{free} are 7,200 veh/h, 500 veh/km, and 100 km/h, respectively. In the estimation process, the system and observation noise are assumed to be 1% of the average traffic state value and observation value, respectively. The assumption for traffic system state and observation noise is as the same as the noise assumption in [21]. The uncertainty terms, R and Q, are defined as $0.5 * I(M)$, where $I(M)$ is a unit matrix with M dimensions.

4.2. Observation Environment.

This paper designs the numerical observation environment according to the actual observation situation of ALPR. The designed scene is a partial passage of Highway 894 (I-894) in Milwaukee, Wisconsin, USA, with a total length of 6.5 km. The designed scene channel contains 13 sequentially numbered cells, and Cell 1 and Cell 13 are the boundary cells of the channel. License plate observers are set up between each adjacent channel, as the blue rectangle identified in Figure 3, numbered A0, A1, A2, ..., A11, A12, and A13. Through the vehicle license plate observers placed at the boundary position at both ends of the cell, the time information of the vehicle passing the cell boundary can be recorded, and the travel time of the vehicle through the cell can be obtained. Observers A0, A1, A12, and A13 are set as the boundary observers of the channel to obtain the travel time of the channel boundary cell (Cell 1 and Cell 12). These observers' layout conditions remain unchanged and the passing time information of all passing vehicles can be obtained. Observers A2, A3, ..., A10, and A11 are experimentally designed ALPR observation equipment. In addition, the green line segments identified in the figure are the license plate recognition
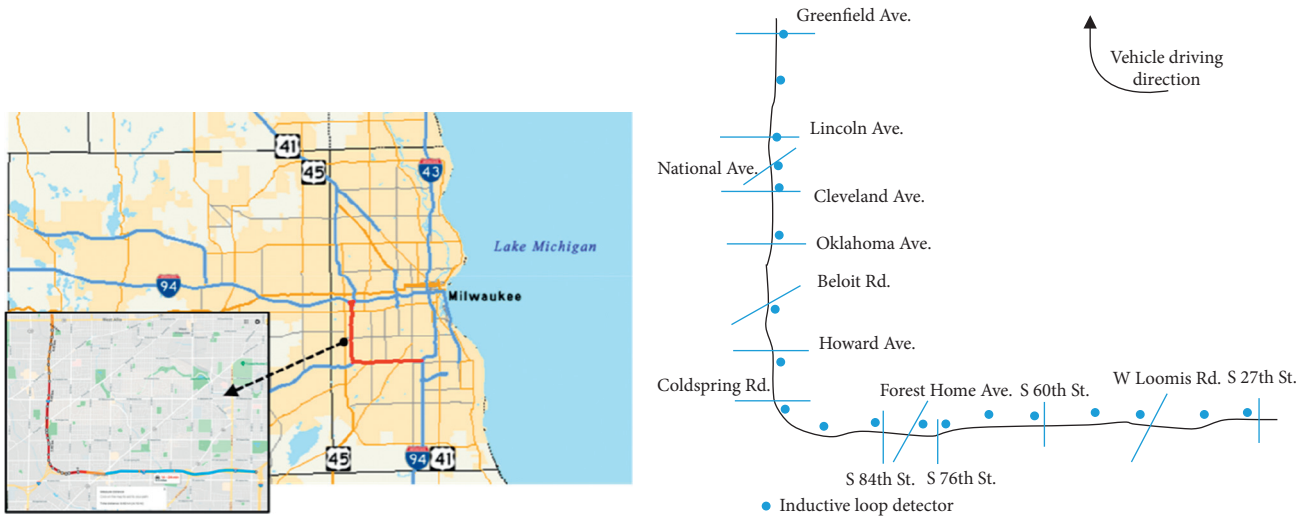
FIGURE 2: Detail of IH-894 in Milwaukee.
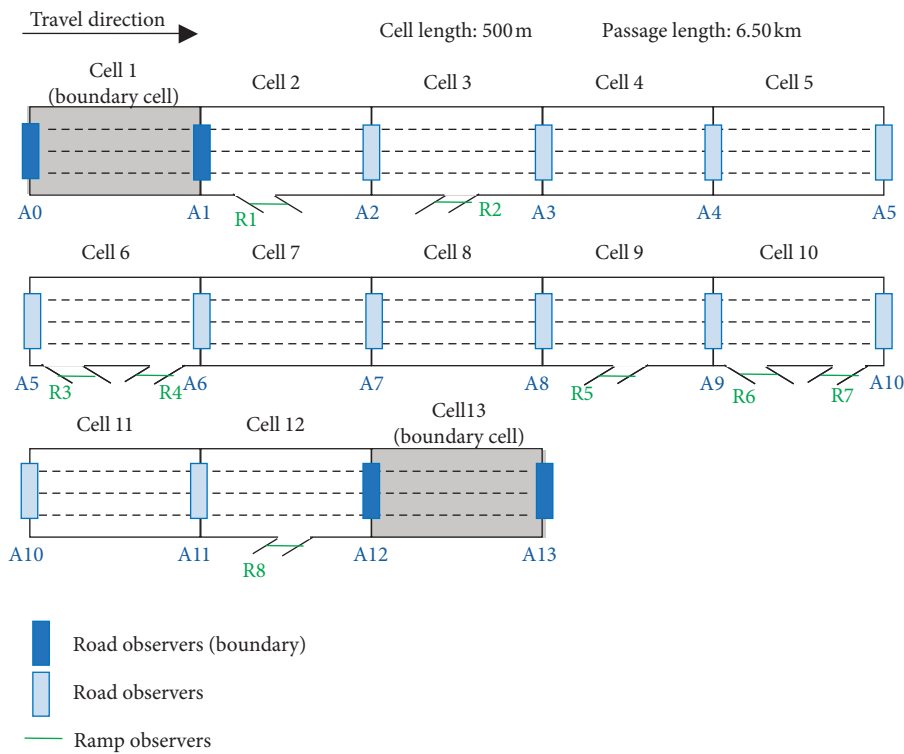
FIGURE 3: Simulation of observation environment.

observers placed on the up- and down-ramps, which are numbered as R1, R2, ..., R7, and R8 in sequence. This paper will design different observation equipment layout based on different experimental observation scenarios.

4.3. Observation Scenarios

Scenario 1. Damaged arterial road observer

The damage rate of the arterial road observers in the actual observation environment is about 10%. Therefore, for the 10 trunk road observers designed in the experimental scene, we assume that there is a damaged observer in the channel. According to the 10 trunk road observers laid out in the passage, 10 groups of cases are designed in this section, corresponding to damaged observers A2, A3, ..., A10, and A11, respectively, as shown in Table 1. For example, in CASE A-2, observer A3 is damaged, as shown in Figure 4. Then, red observer A3 is the damage observer, and the travel time of the vehicle in Cell 3 and Cell 4 cannot be obtained directly. In this case, only the travel time of the vehicle passing through Cell 3 and Cell 4 can be obtained. In addition, because Cell 3 contains an up-ramp, the number of vehicles passing through Cell 4 is higher than that passing through

TABLE 1: Scenario design of arterial road observer damaged.

Scenario	Damaged observer	Upstream cell	Downstream cell
CASE A-1	A2	Cell 2	Cell 3
CASE A-2	A3	Cell 3	Cell 4
CASE A-3	A4	Cell 4	Cell 5
CASE A-4	A5	Cell 5	Cell 6
CASE A-5	A6	Cell 6	Cell 7
CASE A-6	A7	Cell 7	Cell 8
CASE A-7	A8	Cell 8	Cell 9
CASE A-8	A9	Cell 9	Cell 10
CASE A-9	A10	Cell 10	Cell 11
CASE A-10	A11	Cell 11	Cell 12
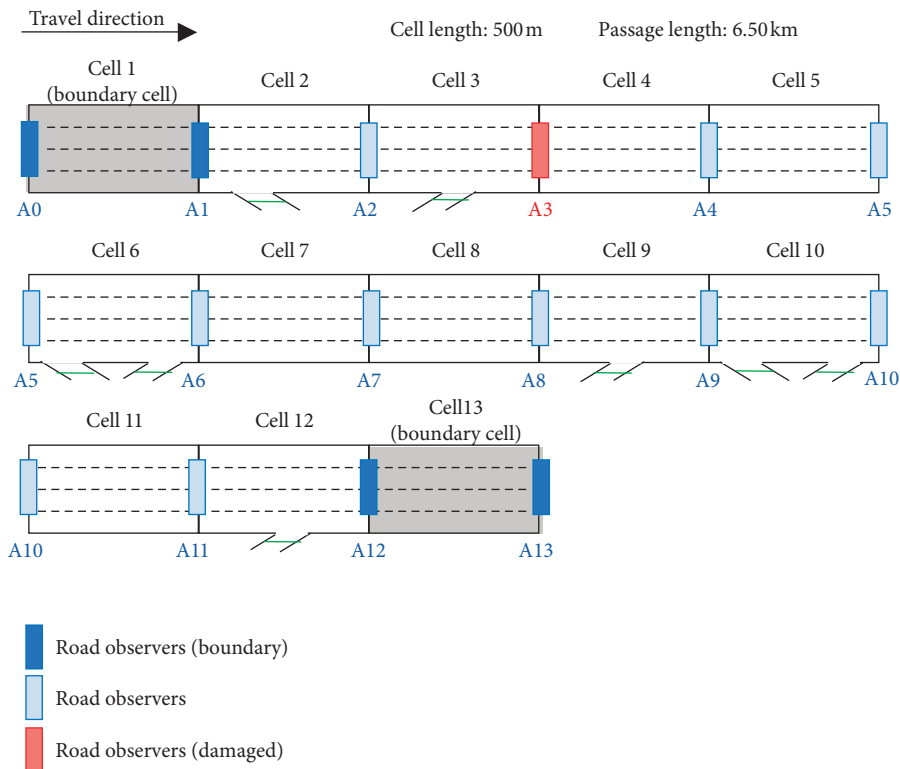
FIGURE 4: Example of scenario 1 (CASE A-2).

Cell 3. In other words, several vehicles' channel entering time cannot be found in the ALPR system. Then, according to the travel time calculation principle of license plate recognition and matching, only a part of vehicles' travel time information in Cell 4 can be obtained.

Scenario 2. Different observation intervals on arterial road

The granularity of arterial road observation interval refers to the spatial distance between adjacent observers on the arterial road. Because of the limitation of funds and technology, the distance between adjacent trunk road observers in the actual observation environment is between 500 m and 2 km. For the experimental channel, four arterial road observers' interval granularities are designed: 500 m, 1000 m, 1500 m, and 2000 m, as shown in Table 2. For example, in CASE B-2, the observation interval distance is set

as 1000 m as, as shown in Figure 5. Limited by the designed observation interval granularity, the travel time of the vehicle in Cell 2 and Cell 3 cannot be obtained directly, and only the travel time of the vehicle passing through Cell 2 and Cell 3 can be obtained. In addition, since Cell 2 contains a down-ramp and Cell 3 contains an up-ramp, only a part of vehicles' travel time can be recorded. According to the travel time calculation principle, the ALPR system can obtain the travel time information of vehicles passing through observers A1 and A3 successively and cannot record the information of vehicles leaving the Cell 2 through off-ramp or entering the Cell 3 through on-ramp.

Scenario 3. Ramp observer missing

Due to the limitations of funds, technology framework, and functional requirements of the ALPR system, license

TABLE 2: Scenario design of arterial road observation interval granularity.

Scenario	Observation interval (m)	Division of observation section
CASE B-1	500	[Cell 2]; [Cell 3];...; [Cell 11]; [Cell 12]
CASE B-2	1000	[Cell 2-Cell 3]; [Cell 4-Cell 5]; [Cell 6-Cell 7]; [Cell 8-Cell 9]; [Cell 10-Cell 11]; [Cell 12]
CASE B-3	1500	[Cell 2-Cell 3-Cell 4]; [Cell 5-Cell 6-Cell 7]; [Cell 8-Cell 9-Cell 10]; [Cell 11-Cell 12]
CASE B-4	2000	[Cell 2-Cell 3-Cell 4-Cell 5]; [Cell 6-Cell 7-Cell 8-Cell 9]; [Cell 10-Cell 11-Cell 12]
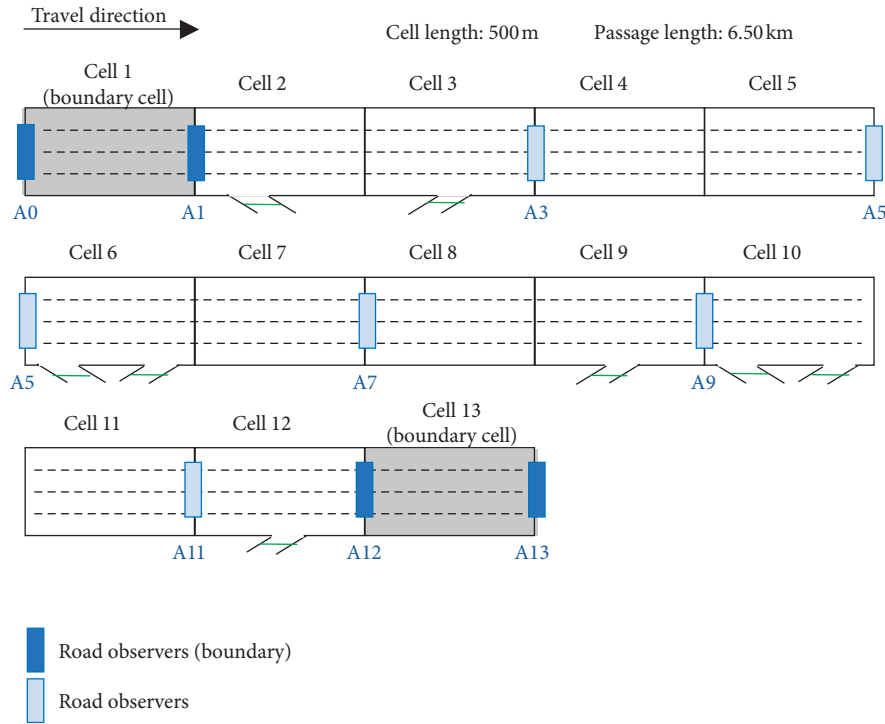
FIGURE 5: Example of scenario 2 (CASE B-2).

plate identification observer on-ramps are often missing in the actual observation environment. For example, the ALPR system in Shanghai does not install the corresponding observation equipment in the off-ramp part. However, when other observation data, such as inductive loop data, are used to supplement the observation function, there are often data mismatching problems in heterogeneous granularity, inconsistent mean variance, and inconsonant spatial and temporal range. Therefore, in the experimental channel, we design eight scenarios of missing ramp observation equipment, as shown in Table 3 and Figure 6. The detailed cell number, ramp observer number, ramp type, and average ramp flow were recorded, respectively, in the table.

5. Results and Analysis

The estimation results of N-X traffic state estimation framework in scenarios of different experimental observation scenes are shown in Figures 7–9. The estimation results are transformed from travel time into travel speed to make results easy to be understood. The result figures have two forms: estimated travel speed figures and estimation error figures. In the estimated travel speed figures, the dark blue areas represent free-flow situation of 110 km/h, and the dark

red areas represent blocked flow situation of 10 km/h. Black outline draws the contour of space-time congestion area of 30 km/h, which is a congestion time and space area surrounded by 30 km/h boundary. The figure can show the spatial and temporal development process of the generation, spread, and dissipation of congestion in the case channel during the morning rush hour. In the estimation error figures, the blue area represents the estimated result which is higher than the actual situation. In this case, the travel speed is overestimated, and the congestion is underestimated. The red area represents the estimated result which is lower than the actual situation. In this case, travel speed is underestimated, and congestion is overestimated. The green area represents the estimation error which is relatively close to 0. In this case, the estimated result is very close to the actual situation.

5.1. Adaptability Analysis of Damaged Arterial Road Observer. The estimation results of N-X traffic state estimation framework in scenarios of different damaged arterial road observers are shown in Figure 10 and Figure 8. It is found that the damage of the trunk road observer has a small impact on the estimation results of congestion status, but an

TABLE 3: Scenario design of ramp observer missing.

Scenario	Cell without ramp observer	Ramp observer missing	Ramp type	Traffic flow on-ramp (veh/h)
CASE C-1	Cell 2	R1	Off-ramp	972.14
CASE C-2	Cell 3	R2	On-ramp	1566.55
CASE C-3	Cell 6	R3	Off-ramp	218.75
CASE C-4	Cell 6	R4	On-ramp	332.09
CASE C-5	Cell 9	R5	On-ramp	481.12
CASE C-6	Cell 10	R6	Off-ramp	261.91
CASE C-7	Cell 10	R7	On-ramp	337.77
CASE C-8	Cell 12	R8	On-ramp	458.70
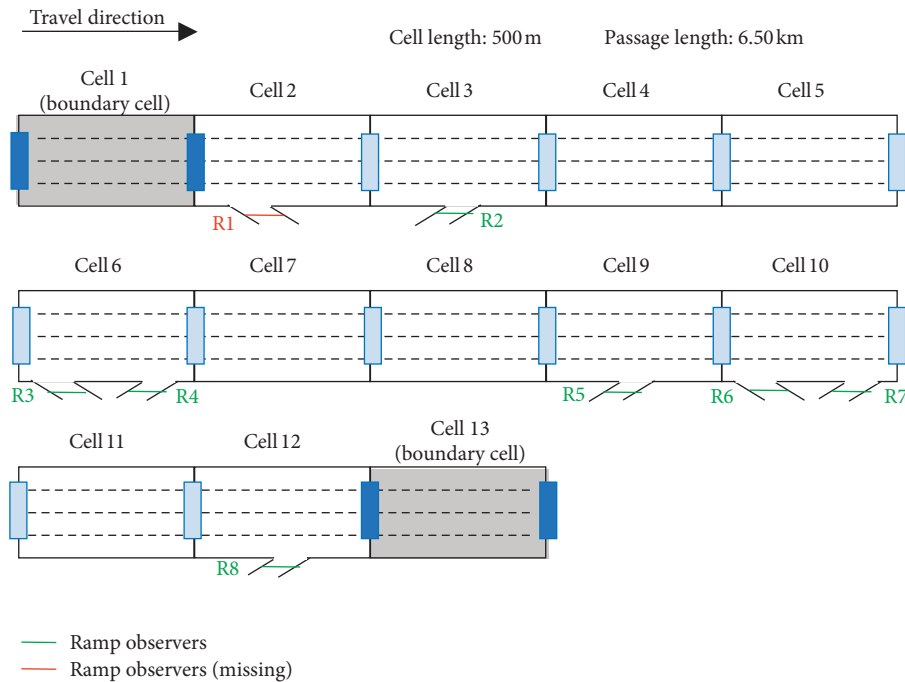
FIGURE 6: Example of scenario 3 (CASE C-1).

obvious impact on the estimation results of free-flow status. In addition, the damage of the trunk road observer has a significant impact on the travel speed estimation results of adjacent cells on both sides of the damage observer. For example, in CASE A-4, trunk road observer A5 was damaged, and the adjacent Cells 5 and 6 on both sides of observer A5 could not be directly observed. In other words, the space position cannot be directly observed within the range of 2.0 km to 3.0 km. Then, it is found that the estimation error near the broken observation increases, while the estimation error in other functioning observation areas does not increase significantly.

5.2. Adaptability Analysis of Different Observation Intervals. The estimation results of different trunk road observations' distance interval are as shown in Table 4 and Figure 8. The trunk road observation interval size refers to the space distance between the adjacent road observers. The number of the trunk observers refers to the number of installed observations on the channel. The trunk road observer layout

density refers to the arterial observer number on unit space length, which is equal to the number of arterial road observer divided by channel distance. It is found that the estimation error of the channel increases gradually with the increasing distance between the adjacent observers of the trunk road. In addition, the estimation error of free-flow condition increases significantly, while the estimation error of congestion condition changes relatively little. Compared with the scenario of the damaged trunk road observers, the effect of changing the granularity of the trunk road observation interval on the estimated results is significantly higher.

The estimation results vary with trunk road observers' distribution density and distance interval, which is shown in Figure 11. The blue curve is the change of RMSE, and the green curve is the change of MAPE. From the figure, with the increase of the trunk road observations' distance interval, the density of trunk road observations gradually decreases, RMSE and MAPE both gradually increase, and the accuracy of estimation model decreases gradually. There is an inflection point of RMSE and MAPE indices between 1000 m and 1500 m in the observation range and between 0.91/km

(a)

(b)

(c)

(d)

(e)

(f)
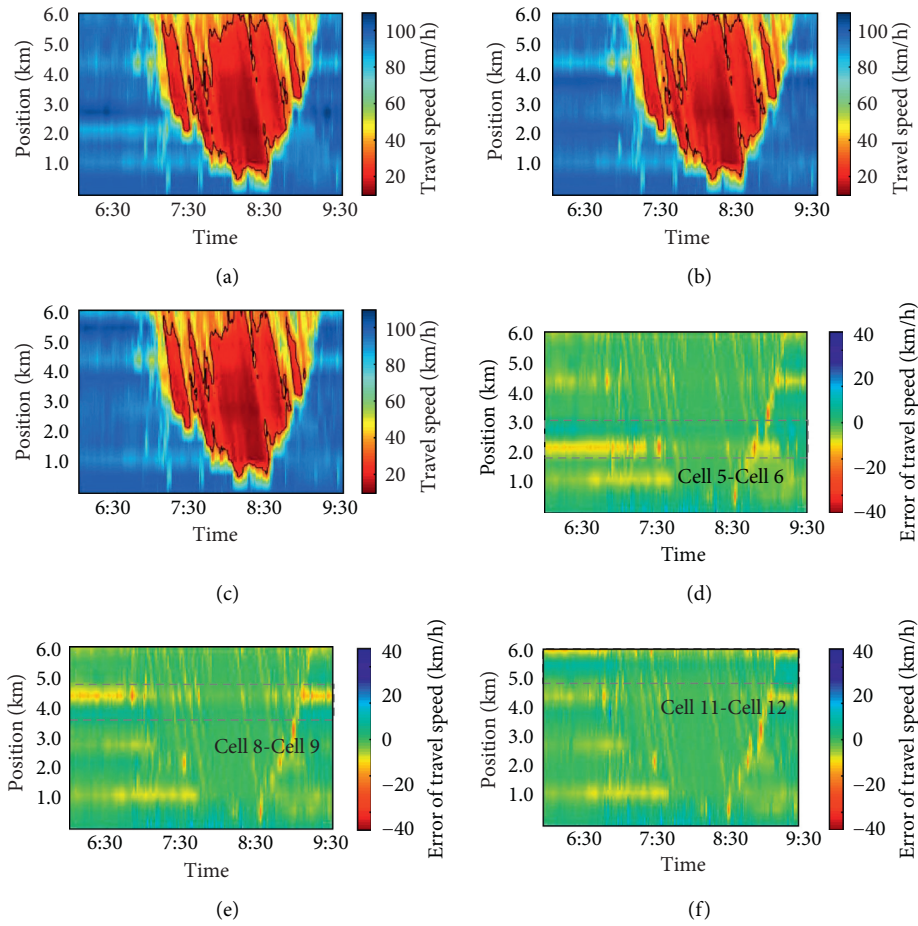
FIGURE 7: Result and error of scenario 1: (a) result of Case A-4, (b) result of Case A-7, (c) result of Case A-10, (d) error of Case A-4, (e) error of Case A-7, and (f) error of Case A-10

(a)

(b)

(c)

(d)
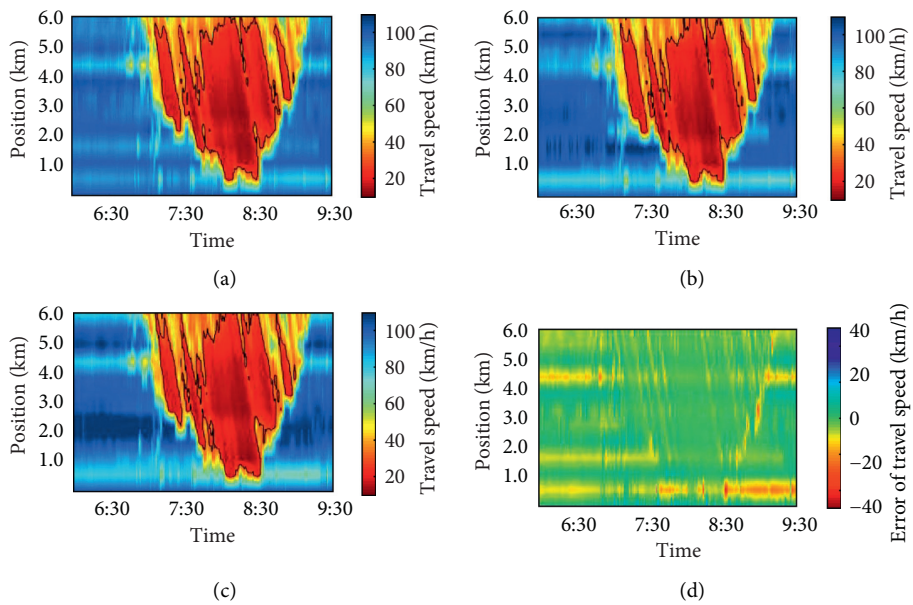
FIGURE 8: Continued.

(e)

(f)
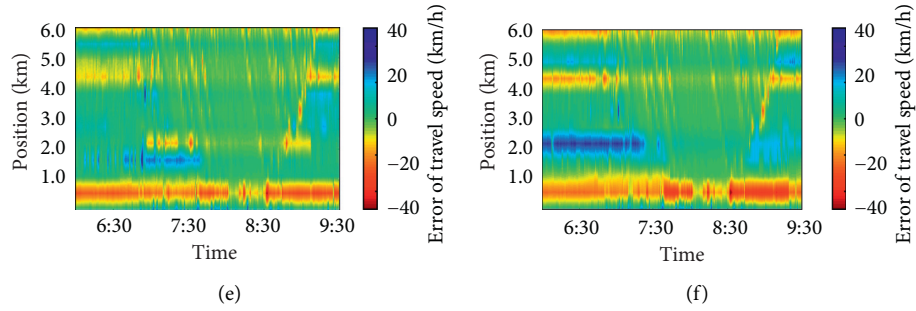
FIGURE 8: Result and error of scenario 2: (a) result of Case B-2, (b) result of Case B-3, (c) result of Case B-4, (d) error of Case B-2, (e) error of Case B-3, and (f) error of Case B-4.

(a)

(b)

(c)

(d)

(e)

(f)
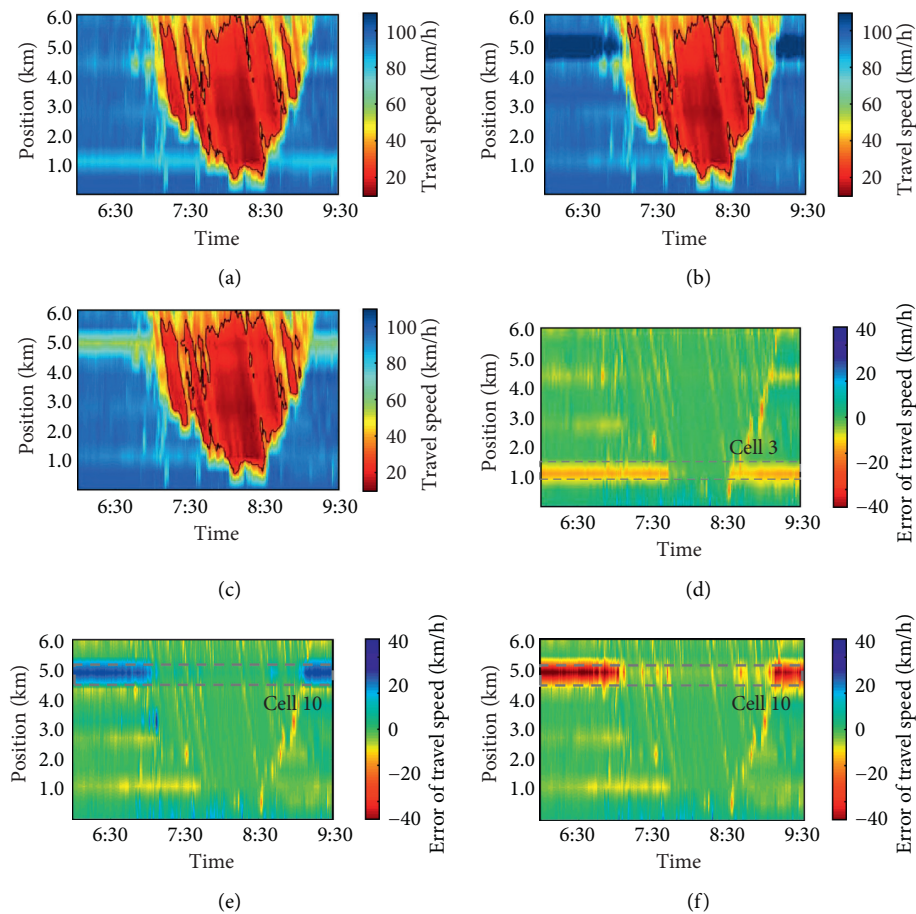
FIGURE 9: Result and error of scenario 3: (a) result of Case C-2, (b) result of Case C-6, (c) result of Case C-7, (d) error of Case C-2, (e) error of Case C-6, and (f) error of Case C-7.

and 1.27/km in the distribution density. Therefore, the accuracy of the estimation results can be improved with the highest efficiency by reducing the particle size of the observation interval to 1000 m and increasing the observer density to 1/km.

5.3. Adaptability Analysis of Missing Ramp Observers. The estimated results for different missing ramp observers are

shown in Table 5 and Figure 10. It is found that the absence of ramp observation environment has a higher impact on the free-flow conditions but a smaller impact on the congestion conditions. In addition, for high inflow and outflow ramps (take CASE C-1 and CASE C-2 as examples), the absence of ramp observation equipment has a relatively small impact on the estimation results, while for low flow ramps, the absence of ramp observation equipment will significantly increase the estimation error and reduce the estimation accuracy.
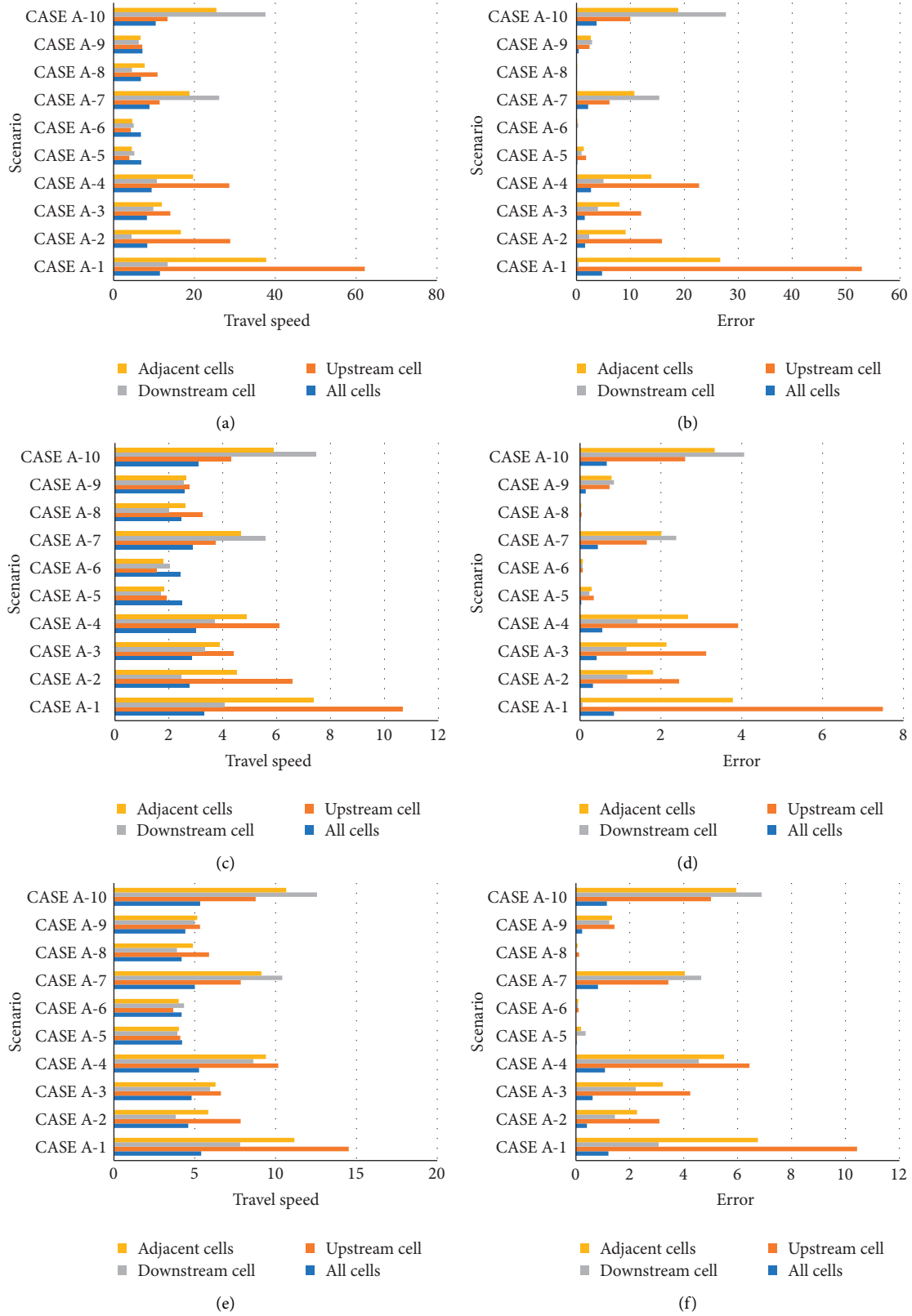
Figure 10: Result and error of scenario 1 under different evaluation methods: (a) result (RMSE: km/h), (b) result (RMSE: km/h), (c) result (MAE: km/h), (d) error (MAE: km/h), (e) result (MAPE: %), and (f) error (MAPE: %).

TABLE 4: Result of scenarios with different observation distance intervals.

Scenario	Observation interval (m)	Number of observers	Density of observers per km	RMSE (km/h)	MAE (km/h)	MAPE (%)
CASE B-1	500	12	2.18	6.73	2.46	4.19
CASE B-2	1000	7	1.27	15.96	3.99	6.80
CASE B-3	1500	5	0.91	35.89	5.80	9.57
CASE B-4	2000	4	0.73	49.21	6.77	10.85

(a)

(b)
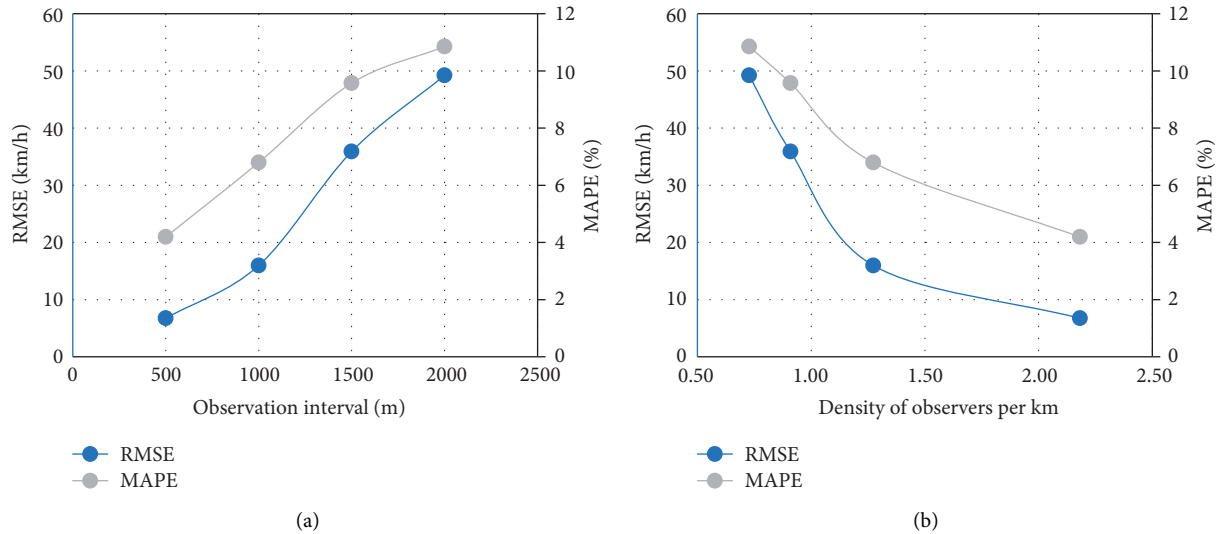
FIGURE 11: Adaptability of different observation intervals: (a) result under different observation intervals; (b) result under different density of observers.

TABLE 5: Result of scenarios with missing ramp observers.

Scenario	All cells in the channel						Cell without ramp observers					
	Result			Error			Result			Error		
	RMSE (km/h)	MAE (km/h)	MAPE (%)	RMSE (km/h)	MAE (km/h)	MAPE (%)	RMSE (km/h)	MAE (km/h)	MAPE (%)	RMSE (km/h)	MAE (km/h)	MAPE (%)
CASE C-1	7.59	2.53	4.34	0.86	0.08	0.15	19.54	4.14	5.81	10.34	0.96	1.71
CASE C-2	11.74	3.01	4.81	5.01	0.55	0.62	69.06	10.34	11.62	56.05	6.20	6.87
CASE C-3	51.61	4.17	6.12	44.88	1.72	1.93	501.15	21.44	25.66	495.45	19.17	21.59
CASE C-4	48.50	4.31	6.55	41.77	1.86	2.36	461.33	22.13	29.29	455.62	19.86	25.22
CASE C-5	28.63	3.64	6.22	21.90	1.19	2.03	253.58	16.87	28.71	242.80	13.65	22.95
CASE C-6	28.40	3.66	5.78	21.67	1.21	1.59	228.78	14.12	19.67	224.11	12.10	15.77
CASE C-7	27.79	3.77	6.18	21.06	1.31	1.99	231.52	15.82	24.76	226.84	13.79	20.86
CASE C-8	31.51	3.95	6.56	24.78	1.49	2.37	283.22	19.95	31.91	273.36	16.53	26.24

Moreover, for the same cell (CASE C-6 and CASE C-7, for example), when the ramps' inflow and outflow are relatively similar, the error caused by the absence of the off-ramp's observation equipment is smaller than that caused by the absence of the on-ramp observation equipment. In other words, the impact of the absence of the off-ramp's observation equipment is relatively small. Also, the impact of ramp observer loss on the estimated results is lower than that with 1000m or longer distance interval and is close to that of the road observer damage environment.

6. Conclusion

Under the global development of autonomous driving, the new generation of urban infrastructure is required. However, limited by the financial and investment input, it is quite necessary to model a new traffic flow model which is not only applicable to autonomous driving vehicles but also can adapt to the existing urban observation infrastructure. Therefore, this paper discusses the applicability of the N-X traffic state estimation model in the actual observation environment by

taking the ALPR technology as an example. This manuscript first analyzes the existing problems in the ALPR system's observation environment in practice. Secondly, based on the observation mechanism of vehicle license plate recognition technology, the ALPR observation environment is simulated based on the simulated channel modeling. Finally, through the observation scene design, the ability to adapt to the actual measurement environment is discussed from three aspects: trunk road observation equipment damage, trunk road observation interval distance, and ramp observation equipment missing. The experimental results reveal several key features of the adaptability of the N-X traffic state estimation model in actual observation environment:

(1) The trunk road observation interval has the most significant influence on the estimation results of the N-X traffic state estimation model. The accuracy of the traffic state estimation model can be significantly improved by shortening the distance between adjacent observation devices. In addition, it is found that the accuracy of the estimation results can be improved with the highest efficiency under the premise of limited equipment input by reducing the observation intervals to 1000 m and increasing the density of observer layout to 1 device per kilometer.

(2) The influence of the damage of trunk road observation equipment on the results of n-X traffic state estimation model is less than the influence of excessive interval. When the traffic conditions on both sides of the observation equipment are relatively similar, the damage of the observation equipment has little influence on the traffic state estimation results. However, when the traffic conditions on both sides of the observation equipment are significantly different, the damage of the observation equipment has a great impact on the traffic state estimation results. Therefore, for the section with relatively homogeneous traffic conditions, the layout of observation equipment can be relatively reduced. However, for the road sections with different traffic conditions, the normal operation of observation equipment should be ensured in priority, and the arrangement of observation equipment can be increased relatively.

(3) The impact of ramp metering on the estimated results is smaller than that of the oversize (1.5 km or more) of the observation interval of the trunk road and is close to the damage of the observation equipment of the trunk road. In addition, for ramps with high flow rate, the absence of ramp observation equipment has a relatively small impact on the estimated results. For ramps with low flow rate, the absence of ramp observation equipment will significantly increase the estimation error and reduce the estimation accuracy. In addition, compared with the absence of on-ramp observation equipment, the absence of off-ramp observation equipment has relatively little impact on traffic status estimation

results. Therefore, in the case of limited equipment investment, observation equipment should be placed on the on-ramp in priority, and maintenance of observation equipment on the ramp with large flow can be slowed down appropriately.

Future work following this study may include the N-X estimation framework to more complicated automated connected applications in the intelligent transportation system. Furthermore, field probe vehicle data-based estimation analysis is desired to reveal the actual potentials of the proposed model in practice.

Notations

n: The order of the vehicle packet and vehicle coordinate
x: Location, distance on a road, and space coordinate
t: Time, duration, and time coordinate
k: Density, $k = -(\partial n / \partial x)$
q: Flow, $q = (\partial n / \partial t)$
u: Speed, $u = (\partial x / \partial t)$
p: Time headway between two vehicles, $p = (\partial t / \partial n)$ and $p = (1/q)$
τ: Travel time density and travel time over a unit distance, $\tau = (\partial t / \partial x)$ and $\tau = (1/u)$
h: Distance headway between two vehicles, $h = -(\partial x / \partial n)$ and $h = (1/k)$
Δx: Segment length
Δn: Number of vehicles in a flow packet n and packet size
T: Travel time, $T = \tau \times \Delta x = p \times \Delta n$.

Data Availability

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

Conflicts of Interest

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

Authors' Contributions

Han Yang was responsible for supervision and conceptualization, and Qing Yu carried out writing: review and editing.

Acknowledgments

References

[1] J. Levinson, J. Askeland, J. Becker et al., "Towards fully autonomous driving: systems and algorithms," in *Proceedings of the 2011 IEEE Intelligent Vehicles Symposium*, pp. 163–168, Dearborn, MI, USA, 2011.

[2] K. Jo, J. Kim, D. Kim, C. Jang, and M. Sunwoo, "Development of autonomous car-Part II: a case study on the implementation of an autonomous driving system based on distributed architecture," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 8, pp. 5119–5132, 2015.

[3] J. Capus, "POWDERMET2017: the future of electrification and autonomous driving: coming sooner than you think," *Metal Powder Report*, vol. 72, no. 5, pp. 314–316, 2017.

[4] F. van Wageningen-Kessels, H. Van Lint, S. P. Hoogendoorn, and K. Vuik, "Lagrangian formulation of multiclass kinematic wave model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2188, no. 1, pp. 29–36, 2010.

[5] L. Leclercq, J. Laval, and E. Chevallier, "The lagrangian coordinate system and what it means for first order traffic flow model," in *Proceedings of the 17th International Symposium on Transportation and Traffic Theory*, New York, USA, 2007.

[6] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.

[7] M. J. Lighthil, "A theory of traffic flow on long crowded roads," *Proceedings of The Royal Society*, vol. 2291178 pages, 1955.

[8] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.

[9] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[10] P. J. Jin, K. Han, and B. Ran, "Some theoretical and practical perspectives of the travel time kinematic wave model: generalized solution, applications, and limitations," in *Proceedings of the Transportation Research Board 93rd Annual Meeting*, Washington, DC, USA, 2014.

[11] H. Yang, P. J. Jin, B. Ran, D. Yang, Z. Duan, and L. He, "Freeway traffic state estimation: a Lagrangian-space Kalman filter approach," *Journal of Intelligent Transportation Systems*, vol. 23, no. 6, pp. 525–540, 2019.

[12] H. Yang, J. P. Jin, Z. Duan, B. Ran, D. Yang, and L. He, "Vehicle-space traffic-state estimation of a motorway corridor with slip roads," *Proceedings of the Institution of Civil Engineers - Transport*, vol. 172, no. 1, pp. 47–56, 2019.

[13] Q. Yu, H. Zhang, W. Li et al., "Mobile phone data in urban bicycle-sharing: market-oriented sub-area division and spatial analysis on emission reduction potentials," *Journal of Cleaner Production*, vol. 254, Article ID 119974, 2020.

[14] Q. Yu, H. Zhang, W. Li, X. Song, D. Yang, and R. Shibasaki, "Mobile phone GPS data in urban customized bus: dynamic line design and emission reduction potentials analysis," *Journal of Cleaner Production*, vol. 272, Article ID 122471, 2020.

[15] W. Jiang, H. Zhang, Y. Long et al., "GPS data in urban online ride-hailing: the technical potential analysis of demand prediction model," *Journal of Cleaner Production*, vol. 279, Article ID 123706, 2021.

[16] G. Harrison, S. M. Grant-Muller, and F. C. Hodgson, "New and emerging data forms in transportation planning and policy: opportunities and challenges for "Track and Trace" data," *Transportation Research Part C: Emerging Technologies*, vol. 117, Article ID 102672, 2020.

[17] Q. Yu, W. Li, D. Yang, and Y. Xie, "Policy zoning for efficient land utilization based on spatio-temporal integration between the bicycle-sharing service and the metro transit," *Sustainability*, vol. 13, no. 1, p. 141, 2021.

[18] S. Du, M. Ibrahim, M. Shehata, and W. Badawy, "Automatic license plate recognition (ALPR): a state-of-the-art review," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 311–325, 2012.

[19] Z.-X. Chen, C.-Y. Liu, F.-L. Chang, and G.-Y. Wang, "Automatic license-plate location and recognition based on feature salience," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 7, pp. 3781–3785, 2009.

[20] Y. Wang, P. Coppola, A. Tzimitsi, A. Messmer, M. Papageorgiou, and A. Nuzzolo, "Real-time freeway network traffic surveillance: large-scale field-testing results in southern Italy," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 548–562, 2011.

[21] Y. Wang, M. Papageorgiou, A. Messmer, P. Coppola, A. Tzimitsi, and A. Nuzzolo, "An adaptive freeway traffic state estimator," *Automatica*, vol. 45, no. 1, pp. 10–24, 2009.

[22] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: adaptive capabilities and real data testing," *Transportation Research Part A: Policy and Practice*, vol. 42, no. 10, pp. 1340–1358, 2008.

[23] Y. Wang, M. Papageorgiou, and A. Messmer, "Investigation of the adaptive features of a real-time nonlinear freeway traffic state estimator," *Nonlinear Dynamics*, vol. 49, no. 4, pp. 511–524, 2007.

[24] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: a case study," *Transportation Science*, vol. 41, no. 2, pp. 167–181, 2007.

[25] Y. Wang, M. Papageorgiou, and A. Messmer, "A real-time freeway network traffic surveillance tool," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 1, pp. 18–32, 2005.

[26] Y. Wang, M. Papageorgiou, and A. Messmer, "Renaissance - a unified macroscopic model-based approach to real-time freeway network traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 3, pp. 190–212, 2006.

[27] M. Treiber, A. Hennecke, and D. Helbing, "Derivation, properties, and simulation of a gas-kinetic-based, nonlocal traffic model," *Physical Review E*, vol. 59, no. 1, p. 239, 1999.

[28] L. Mihaylova, A. Hegyi, A. Gning, and R. K. Boel, "Parallelized particle and Gaussian sum particle filters for large-scale freeway traffic systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 36–48, 2012.

[29] P. J. Jin and S. D. Boyles, "The Travel Time Transmission model for dynamic network loading," in *Proceedings of the Transportation Research Board 93rd Annual Meeting*, Washington, DC, USA, 2014.

Research Article

Has the Reform of the Administrative Examination and Approval System Increased the Efficiency of Resource Allocation: Evidence from China

Yangyang Zhong,[1] Yilin Zhong,[2] and Longpeng Zhang (ID)[3]

[1]*School of Marxism, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2]*School of Economic and Finance, South China University of Technology, Guangzhou 510405, China*
[3]*School of Public Affairs and Administration, University of Electronic Science and Technology of China, Chengdu 611731, China*

Correspondence should be addressed to Longpeng Zhang; zlp1988@uestc.edu.cn

Based on data from the China Industry Business Performance Database from 1998 to 2007 and the time of establishment of the administrative examination and approval center in each administrative division in China, this study attempts to empirically determine the effects of the reform of the administrative examination and approval system on the efficiency of resource allocation from the perspectives of the degree of enterprise-level productivity dispersion. The empirical results showed that the reform of the administrative examination and approval system significantly reduced the degree of productivity dispersion among enterprises in an industry, in addition to enhancing the efficiency of resource allocation in the industry. A further analysis of heterogeneity revealed that the reform of the administrative examination and approval system yielded greater enhancements of the efficiency of resource allocation in industries with lower entry and exit rates.

1. Introduction

The efficiency of resource allocation is not only an important determinant of the existing gap between the productivity and income levels of different economies [1–3] it is also a key factor that influences a country's economic growth [4]. As a result, there has been an increase in the number of studies on the factors influencing the efficiency of resource allocation, with the goal of achieving an optimal allocation of resources and, thereby, increasing the total factor productivity of an economy. In a perfect market, an inefficient enterprise will exit the market due to natural selection, and its former resources will be reallocated to efficient enterprises, thereby increasing the efficiency of resource allocation [5]. However, if market mechanisms become distorted, an inefficient enterprise will remain in the market and continue to occupy a portion of resources, thereby distorting the process of resource allocation and causing resource misallocation. There

are a multitude of institutional factors impeding the development of market mechanisms in countries with different types of economic systems. Consequently, the efficiency of resource allocation is reduced. Government regulation or government intervention is one of the key factors that reduce the efficiency of resource allocation [6].

Researchers agree that a relatively serious resource misallocation effect exists in China, as the factors of production are not effectively allocated among enterprises [2, 7, 8]. Government regulation or intervention factors such as the hukou (household registration) system, localism, government subsidies, and credit policies have systematically reduced the efficiency of the allocation of China's resources. Hsieh and Klenow [2] pointed out that China's reform of factors hindering the efficient allocation of resources in its economic system had significantly enhanced its economic output. On this basis, the Chinese government has implemented a series of institutional reforms with the goals

of increasing the efficiency of resource allocation, unleashing China's economic development potential, reducing the impacts of government interventions on the economy, and stimulating the resource allocation function of market mechanisms.

China's reform of its administrative examination and approval system is one of the most crucial measures in this series of institutional reforms. The administrative examination and approval system reform was intended to clarify the boundaries between the government and the market by easing regulations and reducing government intervention, thus realizing the market's decisive role in economic development. According to the hypotheses proposed by the two competing theories of public interest theory and public choice theory, government regulations have positive and negative effects on an economy. In public interest theory, government regulations can remedy market failures and produce optimal social outcomes [9]. Strict regulations play the role of screening new enterprises so as to select enterprises that are able to provide high quality goods or services. On the other hand, public choice theory argues that strict government regulations are a hindrance to market entry, thus reducing market competitiveness and enhancing the interests of those in power [10–12]. Moreover, government regulations are regarded as rent-seeking tools used by politicians and officials against market entrants [13]. Therefore, this study sought to resolve the issue of whether the efficiency of resource allocation can be enhanced by the reform of the administrative examination and approval system, which centers on easing regulations and reducing government intervention.

Section 2 provides a review of the literature; Section 3 provides an analysis of institutional backgrounds and theories; Section 4 introduces the data sources, econometric models, and measurement of variables; Section 5 offers empirical validation methods which consist of regression analyses and robustness tests; Section 6 further analyses the heterogeneity of the effects of the reform on the efficiency of resource allocation; and Section 7 provides the conclusions of this research.

2. Literature Review

This research is closely related to two fields of study. One field consists of studies on the economic effects of government regulatory reforms, while another consists of studies on the factors governing the efficiency of resource allocation. The authors have organized and summarized the two types of studies so as to clarify the academic contributions of the current study.

2.1. Review of Literature on the Economic Effects of Government Regulatory Reforms. Many scholars have performed comprehensive and in-depth assessments of the economic effects of the government regulatory reforms. In particular, the study by Djankov et al. [14] has great importance. They investigated the time and cost spent by start-up companies in 85 countries in 1999 to meet all governmental requirements. This served as the basis for measuring a country's regulation of entry. The study indicated that a stricter regulation of entry will generate greater corruption and larger informal economies in addition to not providing goods or services with better qualities. Since then, many studies have examined the economic effects of regulation of entry reforms. A multitude of scholars, such as Branstetter et al. [15], Bruhn [16], Kaplan et al. [17], and Rostam-Afschar [18], have studied the effects of regulatory reforms on market entry. The results indicated that easing regulations facilitates market entry and encourages economic and employment growth. Schivardi and Viviano [19] studied the sectoral performance of the regulations of entry reform of the Italian retail trade industry in 1998. Furthermore, the economic effects of other regulatory reforms have received much attention. The studies by Black and Strahan [20], Lee et al. [21], and Nystrom [22] revealed that deregulatory measures in the banking industry, job market, and enterprise bankruptcy had effectively enhanced the viability of market entry. Casu et al. [23] studied the effects of postfinancial crisis regulatory reforms in Asian countries on the performance of the banking industry. Their empirical results supported the positive effects of the easing of regulations on the banking industry. Guerrini et al. [24] evaluated the impacts of regulatory reforms in the Italian water sector on the performance of water utilities. The results showed that the positive effects of the reforms in the water utilities sector were limited.

In general, the studies discussed above suggest that government regulatory reforms can generate positive economic effects to a certain extent. Strict regulations are not socially optimal, an outcome that supports public choice theory. However, these studies merely evaluated the economic effects of government regulations in a certain sector and did not evaluate the economic impacts generated by overall government regulations. As a result, the assessment results are restricted to a certain extent, whereas China's reform of the administrative examination and approval system was centered on a comprehensive reform and easing of government regulations. Therefore, China's reform of the administrative examination and approval system can be used to conduct a comprehensive assessment of the economic effects of government regulation reforms.

2.2. Review of Literature on Factors Governing the Efficiency of Resource Allocation. On the basis of measuring the efficiency of resource allocation, scholars have explored the key factors affecting the efficiency of resource allocation from different perspectives. Balasubramanian and Sivadasan [25] assessed the impacts of sunk costs on the efficiency of resource allocation. Their results revealed that reducing sunk costs can reduce industry concentration while significantly increasing the allocative efficiency of enterprises in an industry. Guner et al. [26] observed the resource misallocation effects generated by policies dependent on enterprise size. The results suggested that tax policies and employment protection are vitally dependent on enterprise size, as effective tax rates and effective protection rates can reflect the

characteristics of enterprise size dependence. Ding et al. [27] studied the effects of free trade on the efficiency of resource allocation. They reported that import penetration achieves a dynamic adjustment of intraindustry resources through competition and promotes inefficient enterprises to exit the market, thereby elevating the efficiency of resource allocation. In recent years, a batch of scholars has turned their attention to the spatial allocation of resources. Based on the perspectives of real estate supply restrictions and tax competition, these scholars sought to determine the factors of spatial misallocation of resources. Desmet and Rossi-Hansberg [28] divided market size into three components: efficiency, amenities, and friction. On this basis, they verified that reducing friction within cities can lead to a spatial reallocation of population resources, thereby increasing welfare levels. Brandt et al. [7] utilized China's database to explore the intraprovincial and interprovincial distortion of labor resources and output losses. Their results delineate a significant declining trend in the misallocation of labor resources in the intraprovincial level. However, the misallocation of resources in the interprovincial level did not improve. Hsieh and Moretti's [29] study revealed that a large wage gap exists between American cities, as real estate supply restrictions had limited the expansion of efficient cities, which ultimately results in the spatial misallocation of American resources and production losses. Based on the perspectives of tax rates in different American states, Fajgelbaum et al. [30] identified the spatial misallocation of resources and corresponding production losses generated by the regional differences in tax rates.

Based on the existing literature, even though institutional distortion is an important factor of resource misallocation, the effects of institutional reforms on the efficiency of resource allocation are worth discussing. Regretfully, studies on the factors governing the efficiency of resource allocation or the performance evaluation of regulatory reforms have not investigated the impacts of the administrative examination and approval system reform on the efficiency of resource allocation. In view of this gap in the existing literature, this study performed an empirical evaluation of the impacts of the administrative examination and approval system reform on the efficiency of resource allocation.

3. Institutional Background and Theoretical Analysis

3.1. Institutional Background. In a planned economy, the government is the sole entity controlling the activities in the economy or society. Hence, the administrative examination and approval system is China's key approach to maintaining the stability of the social order, guaranteeing the successful implementation of mandatory plans, and ensuring the rational allocation of limited resources. Since 1992, China's economic system has transformed from a socialist planned economy to a market economy. A large shift in an economic system must require government institutional reforms to comply with it. The administrative examination and approval system under a planned economy is no longer able to

adapt to the development needs of a socialist market economy. Therefore, the reform of the administrative examination and approval system is a crucial approach for standardizing and optimizing the relationship between the government and the market. The reform of China's administrative examination and approval system first began in the open coastal cities of Shenzhen and Jiangmen. The reform came into effect on a nationwide level after China had joined the World Trade Organization (WTO) in 2001, and the General Office of the State Council had issued the Notice on the Establishment of the State Council Leading Group for the Reform of the Administrative Examination and Approval System. More specifically, the reform of China's administrative examination and approval system has consisted of three phases from 2001 to the present.

3.1.1. Phase 1 (2001–2004). This phase consisted primarily of institutional developments implemented to stimulate market vitality. Reforms in this phase aimed to reduce administrative restrictions, open up the market, and promote market vitality. Many invalid and inefficient administrative examination and approval items were either omitted or adjusted, and the Advice on Implementing the Reform of the Administrative Examination and Approval System was introduced.

3.1.2. Phase II (2004–2013). This phase involved the implementation and standardization of government-enterprise relations. Following the promulgation of the Administrative Law of the People's Republic of China in 2004, the Notice on the Further Cancellation and Adjustment of the Administrative Examination and Approval Items in 2007, and the Notice on Deepening the Reform of the Administrative Examination and Approval System in 2008, the administrative examination and approval system became legalized and standardized, thereby effectively promoting the realization of market mechanisms and reducing rent-seeking opportunities.

3.1.3. Phase III (2013-Present). Based on the core principle of "streamline administration, delegate power, strengthen regulation, and improve service," this phase aims to deepen and innovate the reform. This stage places an emphasis on clarifying government-market relations and realizing the determinant function of the market in resource allocation. Major restrictions were imposed on government intervention on enterprise operation, and the scope of administrative examination and approval was largely reduced. In 2016, the State Council promulgated the Guiding Opinions on Accelerating the Promotion of the "Internet + Government Services" Work, which stressed the integration of online and offline reforms. This signified an important transformation in the reform of China's administrative examination and approval system. Between 2013 to 2018, 44% of the 1700-plus administrative examination and approval items established from the previous regime for State Council departments were omitted, leaving 900-plus items for administrative

approval and bringing an end to nonadministrative licensing examination and approval.

Prior to the reform, the authority for administrative examination and approval was dispersed among various functional departments. This meant that many items had to be examined and approved repeatedly and many forms had to be filled out repeatedly. As a result, the general public had to go through repeated procedures for a single administrative examination and approval application, and this greatly reduced their satisfaction. In order to comply with the requirements of the reform, decentralize the authority for administration examination and approval, and further integrate and optimize the process of the reform, administrative examination and approval centers were established in various locations. These centers enabled the dispersed functional departments to complete their examination and approval duties in a centralized "one-stop" location that offers joint and integrated services. This approach not only saved time and reduced costs for administrative examination and approval but also enhanced the quality of services provided to the public. In addition, the transparency of administrative examination and approval was increased, which prevents rent-seeking operations. In other words, the establishment of administrative examination and approval centers by regional governments is an important step in the implementation of China's administrative examination and approval system reform.

In 1995, Shenzhen became the first city to set up an administrative examination and approval pilot center. From 2001 onwards, the nationwide promotion of the reform caused establishment administrative examination and approval centers to be set up in multiple locations. Figure 1 shows the distribution of the time of establishment of administrative examination and approval centers in Chinese cities. From the figure, it can be seen that the establishment of the centers peaked during 2001 to 2002. After 2002, the number of newly established prefecture-level administrative examination and approval centers declined annually. As of 2015, 316 Chinese cities have established administrative examination and approval centers, accounting for 94.9% of the total number of cities in China. Hence, the administrative examination and approval center can be regarded as an important agent for China's reform of its administrative examination and approval system.

3.2. Theoretical Analysis.

Resource allocation refers to the distribution of factors of production, such as capital and labor, among enterprises. When resources flow from an enterprise with low efficiency to one with high efficiency, the optimal allocation of resources is achieved, thereby promoting a growth in total factor productivity. An important premise for optimal resource allocation and enhanced allocative efficiency is the unrestricted flow of resources. If the flow of resources becomes obstructed, the factors of production are unable to flow from inefficient enterprises to efficient enterprises. Hence, market entry and exit mechanisms must be ideal to allow the unrestricted flow of resources. If an efficient enterprise is unable to enter a market

while an inefficient enterprise is unable to exit a market, the latter will continue to occupy a share of production resources and the former will not be able to access these resources. Therefore, market entry and exit are crucial for the allocation of resources.

Government regulation or intervention happens to be an important factor behind a decline in the efficiency of resource allocation [6]. The efficiency of resource allocation in an economy can be improved effectively by easing up government regulations and standardizing government actions. Hence, the administrative examination and approval reform is an important measure for changing governmental functions since it clarifies government-market relations and realizes the function of the market as a determinant for resource allocation. Thus, the reform could generate important effects on the efficiency of resource allocation. From the perspectives of market entry and exit, this study sought to delineate the mechanisms of the administrative examination and approval system reform affecting the efficiency of resource allocation.

Based on the views of market entry, the reform of the administrative examination and approval system aims to reduce enterprise start-up time and cost through an integration and optimization of administrative procedures and removal of minimum paid-in capital. As a result, the business start-up requirements are lowered, which, in turn, boosted start-up plans among Chinese citizens and increased the number of new establishments [17, 18]. The market entry of an enterprise achieves optimal resource allocation through two methods: market selection and learning effects [31, 32]. In the former, new enterprises are able to eliminate low-performing existing enterprises through market selection, thus obtaining the resources previously occupied by the inefficient enterprises. In the latter, new enterprises will maintain their accelerated growth through learning and continue to attract the flow of resources. Therefore, the reform of the administrative examination and approval system achieves the optimization of resource allocation by lowering market entry requirements.

A zombie company refers to an inefficient company that has failed to achieve consistent profitability but still refuses to exit the market [33]. A large number of zombie companies in an economy will create a multitude of negative impacts on economic development [34]. If a zombie company is unable to exit the market in a timely manner, it will continue to occupy a larger share of production resources and obstruct the flow of factors of production from enterprises with low productivity and return to those with high productivity and return. This is detrimental for the reallocation of production resources. Zombie companies are formed by numerous factors such as credit subsidies, loan extensions, and low-interest loans offered by banks to inefficient companies [33, 35, 36], as well as government intervention [34]. Low product quality, inadequate technological innovation, and lack of entrepreneurial spirit are all important microfactors influencing the creation of zombie companies. The reform of the administrative examination and approval system implies the streamlining of examination and approval procedures,
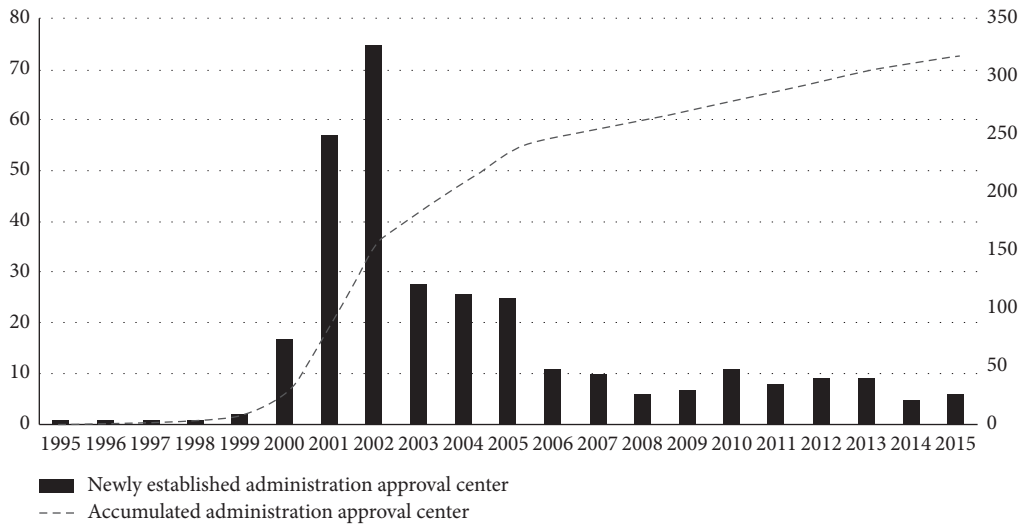
FIGURE 1: Distribution of the time of establishment of administrative examination and approval centers in Chinese cities.

decentralization of items, and integration of services so as to reduce rent-seeking activities caused by administrative examination and approval, lower institutional transaction costs, stimulate the entrepreneurial spirit of productive companies, increase enterprise innovation, and prevent companies from turning into zombie companies. Therefore, the reform is able to prevent the formation of zombie companies, thereby reducing the number of inefficient enterprises in the market and their possibility of resource occupancy, as well as enhancing the efficiency of resource allocation.

Based on the analysis above, the hypothesis of this study can be deduced. The reform of the administrative examination and approval system was able to enhance the efficiency of resource allocation by correcting market entry and exit mechanisms.

4. Research Design

4.1. Data Sources. The data in this study were obtained from the China Industry Business Performance Database from 1998 to 2007. The database offers details on Chinese state-owned enterprises and sizable nonstate-owned enterprises (with a sales revenue of over CNY¥5 million). The statistical indexes of the China Industry Business Performance Database consist of basic enterprise attributes (type of registration, number of employees, type of industry, geographical location, and so on) and detailed financial information (gross output of an industry, revenue, management fee, subsidized revenue, value of export delivery, and so on). In this study, nonmanufacturing enterprises in the database were removed in accordance with China's industrial classification for national economic activities. Then, the industrial classification around the year 2003 was arranged as a standardized tertiary industrial classification according to Brandt et al.'s [37] adjusted classification standards. Next, the administrative division codes of a sampled year in the data sample were adjusted in accordance with Lu and Tao's

[38] method. Lastly, the study by Brandt et al. [7] served as a reference for the omission of samples with missing information of industry gross output and annual average balance of net value of fixed assets, samples with a lower accumulated depreciation than the depreciation of the current year, samples with lower total assets than current assets, samples with lower total assets than the annual average balance of net value of fixed assets, and samples with less than eight employees.

4.2. Measurement Models. The degree of productivity dispersion among enterprises in an industry is an effective proxy variable for describing the industry's efficiency of resource allocation. According to market competition mechanisms, enterprises with lower productivity will exit the market. When a market is in equilibrium, the productivity of all enterprises should be similar and the degree of productivity dispersion should be zero [2]. If a market factor becomes distorted, enterprises with lower productivity will continue to occupy a share of resources and will be unable to exit the market in an orderly manner. Consequently, the degree of productivity dispersion among enterprises will no longer become zero. A higher degree of productivity dispersion indicates a lower efficiency of resource allocation. With these questions in mind, the effects of the administrative examination and approval system reform on the degree of productivity dispersion were utilized in this study to determine if the reform was able to promote improvements in an industry's efficiency of resource allocation. If the reform of the administrative examination and approval system was able to reduce the degree of productivity dispersion among enterprises in an industry, it can be concluded that the reform had also enhanced the industry's efficiency of resource allocation. Therefore, the methods developed by Bertrand et al. [39] served as a reference in this study with respect to the construction of the regression model for empirical analysis, as shown in the following equation:

$$\text{dispersion}_{ict} = \alpha_0 + \alpha_1 \text{AEASR}_{ct} + \sum_{p=1}^{n} \beta_p Ctr_{pict} \qquad (1)$$

$$+ \mu_i + \gamma_c + \nu_t + \varepsilon_{ict},$$

where t, c, and i are, respectively, the year, city, and three-digit standard industrial classification (SIC) code. Dispersion is the dependent variable representing the degree of productivity dispersion among enterprises in an industry i in a city c for the year t. AEASR is the core independent variable representing the degree of administrative examination and approval system reform in each city. Ctr is a set of control variables which could affect the degree of productivity dispersion of an enterprise. In this study, an industry's fixed cost (FC), sunk cost (SC), export percentage (Export), subsidy percentage (SOE), foreign investment enterprises percentage (FIE), and market competitiveness (HHI) were added into the control variables in accordance with the studies by Balasubramanian and Sivadasan [25], Ding et al. [27], and Melitz [40]. μ, γ, and ν are, respectively, the fixed effects of an industry, the fixed effects of a city, and the fixed effects of a year. ε is a random error term.

4.3. Measurement of Variables

4.3.1. Efficiency of Resource Allocation. To measure the dispersion of productivity among firms in the industry, we first measured the productivity of the firm. Since we estimated the productivity of enterprises based on the 1998–2007 China Industrial Enterprise Database, estimating the productivity with the traditional ordinary least squares (OLS) method will bring about the missing variables and considering the problem of sample selection bias caused by the entry and exit of enterprises, so to overcome these problems, the paper calculated the firm productivity by using the method of Olley and Pakes [41]. The estimated model is shown in the following equation:

$$Y_{ft} = \beta_0 + \beta_L L_{ft} + \beta_K K_{ft} + \beta_M M_{ft} + \eta_{ft} + \varepsilon_{ft}, \qquad (2)$$

in which Y, L, K, and M represent, respectively, the enterprise output, number of employees, capital stock, and log of intermediate inputs; η_{ft} are productivity shocks observable to the enterprise decision makers but not to the researchers; ε_{ft} are the productivity shocks unobservable to the enterprise decision makers and the researchers. The conventional OLS approach neglects the existence of η_{ft}. However, as an enterprise's labor employment decisions and investments are influenced by productivity shocks η_{ft}, and the direct use of the conventional OLS method for calculation will result in endogeneity and biased regression results. According to the core concept of Olley and Pakes's [41] approach, if a positive correlation exists between enterprise investments and productivity, the productivity shocks η_{ft} can be written as a function of enterprise investment. This function can be substituted to equation (2) to correct for the missing variables brought by the productivity shocks η_{ft}. Hence, the productivity level of each enterprise can be calculated on this basis[1].

This study utilized standard deviations to measure the degree of productivity dispersion among the enterprises in an industry. A smaller standard deviation indicates a lower degree of productivity dispersion and a higher efficiency of resource allocation. The measurement model for the degree of productivity dispersion is expressed in the following equation:

$$\text{dispersion}_{ict} = \sqrt{\frac{1}{N} \sum_{f=1}^{N} \left(\ln tfp_{fict} - \overline{\ln tfp}_{ict} \right)^2}. \qquad (3)$$

in which N corresponds to the number of enterprises in an industry; $ln\ tfp_{fict}$ represents the log productivity of an enterprise f in an industry i in a city c for the year t; and $\overline{\ln tfp}_{ict}$ represents the average of the log productivity levels of all enterprises in an industry i in a city c for the year t.

4.3.2. Administrative Examination and Approval System Reform. The administrative examination and approval center is the centerpiece for China's administrative examination and approval system reform. It allows for a simplified and centralized approach to the handling of administrative examination and approval items and reduces the institutional transaction cost of an enterprise. As it has collection, integration, and innovation functions, it is a key determinant for the comprehensive deepening of the reform. Therefore, the progress of the reform can be reflected by the establishment of administrative examination and approval centers. Based on the methodology developed by Zhu and Zhang [42], this study set the establishment of an administrative examination and approval center as a dummy variable to depict the progress of the reform in a city. If a city c establishes an administrative examination and approval center in the year t, the valuation of AEASR becomes 1 during year t and after the years after it, while the valuation of other years become 0.

4.3.3. Control Variables

(1) Fixed Cost. It is defined as the ratio between the management fee and the added value of an enterprise in an industry. In an industry with a higher fixed cost, a higher productivity level must be attained by an enterprise to become profitable. In other words, a fixed cost will enhance the critical productivity level of an industry, thus allowing enterprises with low productivity levels to exit the market and reducing the degree of enterprise productivity dispersion [25].

(2) Sunk Cost. It is defined as the ratio between the capital stock and the added value of an industry. When facing emerging enterprises, existing enterprises often employ strategic actions to increase the sunk cost of entering an industry, thus blocking the market entry of new enterprises and enhancing the market power of enterprises in an industry. Consequently, the critical productivity level of an industry is lowered, and existing low productivity

enterprises are still able to benefit from high prices and maintain their presence in the industry, which increases the degree of productivity dispersion among enterprises [43, 44].

(3) Export Percentage. It is defined as the ratio between the total export volume and the sales volume of an enterprise in an industry. The theory of heterogeneous firms and trade proposed by Melitz [40] stresses the decisive role of productivity in an enterprise's export operations. An export market is able to self-select high productivity enterprises to participate in the export market. According to the theory, exports are beneficial for markets to realize optimal configurations. Therefore, low productivity enterprises will exit the market while high productivity enterprises will account for a higher share, thereby reducing the degree of productivity dispersion.

(4) Subsidy Percentage. It is defined as the ratio between the total government subsidies and the sales volume of an enterprise in an industry. China's subsidy policies have a tendency to "protect the weak." Hence, low productivity enterprises are better able to obtain government subsidies and occupy production resources for a long period. In addition, as they are unable to exit the market, the degree of productivity dispersion of an industry is increased.

(5) Foreign Investment Enterprises Percentage. It is defined as the ratio between the number of foreign investment enterprises and the total number of enterprises in an industry. The extent of the influence of foreign investment enterprises on the degree of productivity dispersion of enterprises in an industry remains unclear. On the other hand, foreign investment enterprises have higher productivity levels [45]. When a foreign investment enterprise enters a host country and starts to operate, a large productivity gap is generated between the enterprise and the host country's enterprises, thereby extending the degree of productivity dispersion among enterprises. In addition, a foreign investment enterprise will create competition when it enters a local market. Under the pressure of competition from a foreign investment enterprise, low productivity enterprises will exit the market of the host country, thereby reducing the degree of productivity dispersion among the enterprises in an industry.

(6) Market Competitiveness. The Herfindahl–Hirschman Index was used in this study to estimate market competition. The equation is expressed as $\text{HHI}_{ict} = \sum_{f \epsilon I_i} (\text{Sale}_{fict} / \text{Sale}_{ict})^2$, in which Sale_{fict} and Sale_{ict} are, respectively, the sales volume of an enterprise f in an industry i in a city c for the year t and the sales volume of an industry i in a city c for the year t. The index ranges from 0 to 1, and a larger value indicates lower market competitiveness. An increase in market competitiveness will promote high productivity enterprises to enter the market and low productivity enterprises will exit the market, thereby reducing the degree of productivity dispersion among the enterprises in an industry.

5. Empirical Results and Analysis

5.1. Basic Regression Results. Based on the regression model in equation (1), the OLS approach was used in this study for the empirical validation of the impacts of the administrative examination and approval system reform on the degree of productivity dispersion. The basic regression results are shown in Table 1.

Table 1 reports the relevant estimation results obtained through the gradual addition of control variables. In Row (1), no control variables are added, and the fixed effects of an industry, city, and year are uncontrolled. The estimated coefficient of the variable AEASR is –0.1074, which is statistically significant at a 1% level. This suggests that the degree of productivity dispersion among enterprises is lower in cities with an administrative examination and approval center. In Row (2), the fixed effects of an industry, city, and year are controlled. The estimated coefficient of AEASR is −0.0186, which is statistically significant at a 1% level. It can be seen that the inhibitory effects of the administrative examination and approval system reform on the degree of productivity dispersion are comparatively lower than the estimation results in Row (1). Therefore, the omission of dependent variables will overestimate the policy effects of the administrative examination and approval system reform. Based on the results of Row (2), the control variables of fixed cost, sunk cost, export percentage, subsidy percentage, foreign investment enterprises percentage, and market competitiveness are added into Row (3). After controlling for these variables, the estimated coefficient of AEASR remains negative and statistically significant, and the value is similar to that in Row (2). The estimation results in Table 1 suggest that the establishment of administrative examination and approval centers was beneficial for reducing the degree of industry productivity dispersion, thereby delineating the improvement effects of the reform on the efficiency of resource allocation.

Next, the control variables in Table 1 are described in brief as follows. The estimated coefficients of fixed cost (FC) and export percentage (Export) are negative and statistically significant. This shows that the degree of productivity dispersion among enterprises is lower in an industry with a higher fixed cost and export percentage. On the other hand, the estimated coefficients of sunk cost (SC) and subsidy percentage (Subsidy) are positive and statistically significant. This shows that an industry with a higher sunk cost will receive more government subsidies, which increases the degree of productivity dispersion. However, the estimated coefficient of foreign investment enterprises (FIE) percentage had failed to pass the test of significance. As described in the measurement of variables section, the foreign investment enterprises percentage has negative and positive effects on the degree of productivity dispersion. Hence, as a whole, this variable is unable to generate significant effects with regard to the degree of productivity dispersion. The estimated coefficient of market competitiveness (HHI) is positive and statistically significant at a 1% level. This suggests that the degree of productivity dispersion is greater in an industry with lower market competitiveness.

TABLE 1: Basic regression results.

Variables	(1)	(2)	(3)
AEASR	−0.1074***	−0.0186***	−0.0188***
	(0.0029)	(0.0053)	(0.0053)
FC			−0.0008**
			(0.0004)
SC			0.0002*
			(0.0001)
Export			−0.0206**
			(0.0088)
Subsidy			0.2072**
			(0.0985)
FIE			−0.0067
			(0.0091)
HHI			0.1465***
			(0.0084)
Industry fixed effect	N	Y	Y
City fixed effect	N	Y	Y
Year fixed effect	N	Y	Y
R^2	0.0091	0.0935	0.0967
Sample size	148498	148498	148302

Note. Symbols *, **, and *** represented significance levels of 10%, 5%, and 1%, respectively; the numbers in brackets were robust standard errors of estimated coefficients.

Therefore, market competitiveness promotes the improvement of an industry's efficiency of resource allocation.

5.2. Robustness Tests

5.2.1. Treatment of Endogeneity. The estimation results in this study could be affected by endogeneity. To address this issue, the researchers first omitted samples with suspected endogeneity and then selected instrumental variables of the administrative examination and approval system reform. Next, the effects of the reform on the degree of productivity dispersion were re-estimated through a two-step OLS approach.

In 2001, China began its nationwide plan to reform the administrative examination and approval system. Under the context of a nationwide unified plan, many cities began to establish administrative examination and approval centers. For samples involving these cities, the establishment of administrative examination and approval centers can be regarded as exogenous. However, some cities had voluntarily implemented the reform by establishing an administrative examination and approval center even before the nationwide plan, possibly as to stimulate market vitality and eliminate excess capacity. Therefore, samples involving cities with an administrative examination and approval center established before 2001 are suspected to be endogenous. To avoid the impacts of endogeneity on the estimation results, samples involving cities with an administrative examination and approval center established before 2001 were omitted. The estimation results are shown in Row (1) in Table 2. The estimated coefficients of the variable AEASR are all negative, which suggests that the degree of productivity dispersion among enterprises is lower in cities with an administrative examination and approval center. As a result, an industry's efficiency of resource allocation can be improved effectively.

Next, instrumental variable regression was performed as a further treatment of endogeneity. The instrumental variable was set as the rate of administrative examination and approval center establishment in other cities within the same province. This instrumental variable was chosen as the innovation of China's local governments is proliferative [42]. If the administrative examination and approval system reform had taken place in a certain city, other cities will be prompted to implement the reform as well. Furthermore, the degree of productivity dispersion among the enterprises in an industry in a city has no direct effects on the rate of administrative examination and approval center establishment in other cities. Therefore, a two-step OLS estimation was performed in this study with the aforementioned instrumental variable so as to re-examine whether the administrative examination and approval system reform had improved the efficiency of resource allocation. The regression results are shown in Row (2) in Table 2. The estimated coefficient of AEASR is negative and statistically significant, which suggests that the reform had improved the efficiency of resource allocation Furthermore, the results of the test of under identification and the test of weak identification have passed the test of significance at a 1% level. Hence, the selected instrumental variable is appropriate, and the two-step OLS approach for regression analysis is effective.

5.2.2. Regression Analysis for Different Industries. In order to understand the impacts of the administrative examination and approval system reform on the efficiency of resource allocation, the basic regression in this study was performed according to the three-digit SIC codes. The three-digit SIC codes were chosen based on two considerations. If the two-digit SIC codes were to be used to calculate the degree of productivity dispersion among enterprises, the scope of industry classification may be overly large and the heterogeneity of industries would be unaccounted for. If the four-digit SIC does were to be used to calculate the degree of productivity dispersion, the number of enterprises in an industry may be too small, which affects the effectiveness of the degree of dispersion measurement. However, in order to support the robustness of the conclusions of this study, the researchers also examined the impacts of the administrative examination and approval system reform on the efficiency of resource allocation by using the two-digit and four-digit SIC codes. Table 2 shows the corresponding estimation results. The estimation results using the two-digit SIC codes are listed in Row (3); the estimation results using the four-digit SIC codes are listed in Row (4). The regression results indicate that regardless of two-digit or four-digit SIC codes, the estimated coefficients of AEASR are both negative and statistically significant. Hence, the conclusion that the administrative examination and approval system reform had enhanced the efficiency of resource allocation is robust.

5.2.3. Remeasurement of Efficiency of Resource Allocation. The measurement index of the efficiency of resource allocation is affected by two factors: one is the method for measuring productivity dispersion and the other is the

TABLE 2: Robustness test results.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AEASR	−0.0199***	−0.0435**	−0.0164**	−0.0152***	−0.0148**	−0.0219***	−0.0123**	−0.0032
	(0.0057)	(0.0206)	(0.0071)	(0.0049)	(0.0066)	(0.0084)	(0.0053)	(0.0062)
FC	−0.0008**	−0.0008**	−0.0051**	−0.0002	−0.0000	−0.0000	−0.0002	−0.0008**
	(0.0003)	(0.0004)	(0.0021)	(0.0008)	(0.0001)	(0.0001)	(0.0002)	(0.0004)
SC	0.0002*	0.0002*	0.0006***	0.0002	−0.0000*	−0.0000	0.0001	0.0002*
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Export	−0.0136	−0.0203**	−0.0575***	−0.0141*	0.0058	0.0748***	0.0732***	−0.0206**
	(0.0095)	(0.0088)	(0.0153)	(0.0074)	(0.0094)	(0.0115)	(0.0087)	(0.0088)
Subsidy	0.2108**	0.2191**	0.2566*	0.2335***	−0.0273*	−0.0322	−0.0292	0.2072**
	(0.1020)	(0.1005)	(0.1352)	(0.0682)	(0.0140)	(0.0198)	(0.1041)	(0.0986)
FIE	−0.0054	−0.0065	−0.0032	−0.0067	0.0340***	0.0295***	0.0075	−0.0070
	(0.0097)	(0.0091)	(0.0177)	(0.0073)	(0.0078)	(0.0091)	(0.0092)	(0.0091)
HHI	0.1545***	0.1464***	0.0902***	0.1878***	−1.2383***	−2.4056***	0.6683***	0.1465***
	(0.0089)	(0.0084)	(0.0134)	(0.0074)	(0.0061)	(0.0079)	(0.0085)	(0.0084)
Industry fixed effect	Y	Y	Y	Y	Y	Y	Y	Y
City fixed effect	Y	Y	Y	Y	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y	Y	Y	Y	Y
Under identification test		7837.9620***						
Weak identification test		9258.8740***						
R^2	0.0965	0.0966	0.1385	0.0876	0.2377	0.4333	0.1441	0.0966
Sample size	133509	148082	60237	190427	218971	218971	148302	148302

Note. Symbols *, **, and *** represent, respectively, the 10%, 5%, and 1% level of significance; the numbers in brackets are robust standard deviations; the under identification test refers to the Kleibergen–Paap rk LM statistic in the test, the null hypothesis is that an under identification problem exists in the instrumental variable regression; the weak identification test refers to the Kleibergen–Paap rk Wald F statistic in the test; the null hypothesis is that a weak identification problem exists in the instrumental variable regression.

method for calculating enterprise productivity level. Both factors will result in differences in the measurement of the efficiency of resource allocation. Therefore, the efficiency of resource allocation was remeasured in this stud so as to ensure the robustness of the conclusions.

First, the method for measuring the degree of productivity dispersion was changed. The basic regression results utilized standard deviations to measure the degree of productivity dispersion among enterprises as well as the efficiency of resource allocation of an industry. There are many methods for measuring the degree of productivity dispersion, such as quartile deviation and 90/10 percentile deviation. Therefore, in order to validate that the regression results were unaffected by the methods for measuring the degree of productivity dispersion, quartile deviation and 90/10 percentile deviation were used to remeasure the degree of productivity dispersion among the enterprises in an industry. In quartile deviation, the 25th percentile of productivity was subtracted from the 75th percentile; in 90/10 percentile deviation, the 10th percentile of productivity was subtracted from the 90th percentile. A larger deviation in both approaches indicates a higher degree of productivity dispersion and a lower efficiency of resource allocation. The regression results of the quartile deviation are listed in Row (5) in Table 2; the regression results of the 90/10 percentile deviation are listed in Row (6). The results indicate that regardless of indexes for measuring the degree of productivity dispersion, the estimated coefficients of AEASR are both negative and statistically significant. Hence, the

establishment of administrative examination and approval centers in a city is beneficial for reducing the city's degree of productivity dispersion among the enterprises in an industry, thereby improving the efficiency of resource allocation.

Next, the method for measuring enterprise productivity level was changed. This study re-estimated an enterprise's productivity level in accordance with the methodology of Levinsohn and Petrin [46]. Standard deviations were also used to measure the degree of productivity dispersion among enterprises. Row (7) in Table 2 reports the estimation results re-measured according to Levinsohn and Petrin's methodology. The results indicate that the estimated coefficients of AEASR are all negative and statistically significant. Hence, the conclusion that the administrative examination and approval system reform had enhanced the efficiency of resource allocation is unaffected by the measurement methods.

5.2.4. Remeasurement of the Reform of Administrative Examination and Approval. The basic regression model uses whether the administrative examination and approval center is established to measure the reform, which takes the reform of administrative examination and approval degree of the city with the administrative examination and approval center as the same, but in fact, the degree of the administrative examination and approval reform differs in different cities. Therefore, this section will use other indicators to

remeasure the reform of administrative examination and approval and then test the robustness of the research conclusion. The reform of administrative examination and approval will be analyzed from two aspects.

First of all, generally speaking, the higher the degree of administrative examination and approval reform is, the higher the degree of marketization will be, and the more the market can play its role in resource allocation so as to realize the effective allocation of resources. Therefore, this paper constructs an interactive item (AEASR—MARKET) between administrative examination and approval center and marketization degree to measure the degree of administrative examination and approval reform in each city. The marketization degree index includes five aspects: the relationship between government and market, the development of nonstate-owned economy, the development of product market, the development of factor market, and the development of market intermediary organizations and the legal system environment The composition of the index comes from "China's marketization index Report on the relative process of marketization in various regions in 2009." As shown in column (1) of Table 3, the estimated coefficient of AEASR—MARKET is –0.0021, which is significant at the significance level of 1%. It shows that the higher the degree of urban administrative examination and approval reform, the more helpful it is to reduce the degree of productivity dispersion among enterprises in the city, thereby improving the efficiency of resource allocation in the industry.

Secondly, this paper uses China's "Business Environment Report" released by the World Bank in 2008 to measure the degree of administrative examination and approval. The report investigates the administrative examination and approval procedures and time required to set up enterprises in four municipalities directly under the central government and 27 provincial capital cities in 2006. Based on these two indicators and cross-sectional data of 31 cities in 2006, this paper explores the impact of the reform of administrative examination and approval on resource allocation efficiency. The regression results in columns (2) and (3) of Table 3 show that the estimated coefficients of the administrative examination and approval procedure (Procedure) and administrative examination and approval time (Day) are both positive at the 1% significance level. This shows that in cities with stricter administrative examination and approval, the higher the dispersion degree of productivity among enterprises in the industry, the lower the efficiency of resource allocation. That is to say, cities with less administrative examination and approval procedures and time can improve the efficiency of resource allocation.

5.2.5. Placebo Test. A placebo test is a counterfactual test in which counterfactual hypotheses are proposed to determine the impacts of a policy or an event. In the placebo test of this study, the time of establishment of the administrative examination and approval center in each city was changed. If a spurious regression exists within the basic regression, changing the time of center establishment would indicate the conclusion that the reform had improved the efficiency of

resource allocation and is still significant. Hence, the time of center establishment was brought forward for three years so as to understand the impacts of the administrative examination and approval system reform on the degree of productivity dispersion. As shown in Row (8) in Table 2, AEASR has no longer significant. Therefore, if the time of center establishment was manually changed, the impacts of the administrative examination and approval system reform on the degree of productivity dispersion would no longer be significant. In other terms, based on the placebo test results, it can be highlighted that the reform had reduced the degree of productivity dispersion among the enterprises in an industry, as well as enhancing the industry's efficiency of resource allocation.

6. Tests of Heterogeneity

This study assumes that market entry and exit play an important role throughout the process of the reform's influence on the efficiency of resource allocation. Hence, a further test was performed to investigate the heterogeneity of the effects of the reform on the efficiency of resource allocation at different market entry and exit rates.

6.1. Influence of Administrative Examination and Approval Reform on Resource Allocation Efficiency in Different Samples. In general, new enterprises have higher productivity levels, and the production resources often flow from low productivity enterprises to high productivity enterprises, thereby achieving effective resource allocation. Following the continuous entry of new enterprises, market competitiveness is constantly intensified, which promotes low productivity enterprises to exit the market. Once these enterprises exit the market, they will release resources for use by high productivity enterprises, thereby optimizing the efficiency of resource allocation. If the administrative examination and approval system reform was able to improve the efficiency of resource allocation through the promotion of market entry and exit, then the positive effects of the reform on the efficiency of resource allocation would be more significant for industries with lower entry and exit rates. In order to determine the heterogeneity of this effect, a split sample regression was performed in this study according to the entry rates and exit rates of different industries, thereby examining the effects of the reform on the efficiency of resource allocation in different samples.

The entry and exit status of an enterprise is first examined to estimate an industry's entry rate and exit rate. For instance, if an enterprise exists in year t but not in year $t – 1$, it can be assumed that the enterprise had entered the market in year t; if an enterprise exists in year t but no longer exists in year $t + 1$, it can be assumed that the enterprise had exited the market in year t. Since this study uses data from the China Industry Business Performance Database from 1998 to 2007, the researchers are unable to determine if enterprises in 1998 existed in the year before nor if enterprises in 2007 existed in the year after. Hence, samples in 1998 were not considered when examining the market entry of

TABLE 3: Remeasurement results of the reform of administrative examination and approval.

	(1)	(2)	(3)
AEASR—MARKET	−0.0021*		
	−0.0005		
Procedure		0.0730***	
		(0.0110)	
Day			0.0077***
			(0.0018)
FC	−0.0008**	−0.0259	−0.0205
	−0.0004	(0.0370)	(0.0370)
SC	0.0002 *	0.0025	0.0020
	−0.0001	(0.0031)	(0.0031)
Export	−0.0444***	0.0237	0.0112
	−0.0086	(0.0664)	(0.0668)
Subsidy	0.2215**	−2.0272**	−2.1744**
	−0.0984	(0.8972)	(0.9043)
FIE	0.0588***	0.1004 *	0.1137 *
	−0.0083	(0.0594)	(0.0608)
HHI	0.0887***	−0.0385	−0.0776
	−0.0081	(0.0597)	(0.0603)
Industry fixed effect	Control	Control	Control
City fixed effect	Control	No	No
Year fixed effect	Control	No	No
Pseudo goodness of fit	0.07	0.136	0.132
Sample size	148300	2911	2911

Note. Symbols *, **, and *** represent 10%, 5%, and 1% significance levels respectively; the values in brackets are robust standard deviation.

enterprises nor were samples in 2007 when examining the market exit of enterprises. After examining the entry status and exit status of enterprises, this study defined industry entry rate as the percentage of the quantity of new enterprises entering the market to the total number of enterprises, while industry exit rate was defined as the percentage of the quantity of enterprises exiting the market to the total number of enterprises. In this study, an industry with a higher than average entry rate is categorized as an industry with a high entry rate, whereas an industry with a lower than average entry rate is categorized as an industry with a low entry rate. Industries with high and low exit rates are categorized in a similar fashion.

Table 4 shows the split sample regression results. Rows (1) and (2) are the estimation results of industries with different entry rates. For industries with high entry rates, the estimated coefficient of the reform is −0.0143, which is statistically significant at a 10% level; for industries with low entry rates, the estimated coefficient of the reform is −0.0278, which is statistically significant at a 1% level. Therefore, it can be seen that the reform was able to promote more industries with low entry rates to enter the market, thereby reducing the degree of productivity dispersion among enterprises and increasing the efficiency of resource allocation. As enterprises in industries with high entry rates are more proactive in market entry, the effects of the reform are less prominent. Rows (3) and (4) are the estimation results of industries with different exit rates. For industries with high exit rates, the estimated coefficient of the reform is −0.0153, which is statistically significant at a 5% level; for industries

with low exit rates, the estimated coefficient of the reform is −0.0287, which is statistically significant at a 1% level. In comparison, the positive effects of the reform on the efficiency of resource allocation are more prominent in industries with low exit rates. This is due to the high market exit tendency of enterprises in industries with high exit rates. These enterprises will exit the market even without the reform. Thus, the positive effects of the reform on the efficiency of resource allocation are less prominent in industries with high exit rates. In contrast, the reform promotes the entry of industries with lower exit rates, which stimulates market vitality and enhances the exit tendency of enterprises in these industries. This greatly reduces the degree of productivity dispersion among the enterprises in the industries.

6.2. Inhibitory Effect of Administrative Approval Reform on Zombie Companies. China's administrative examination and approval reform has significantly inhibited the formation of zombie companies, smoothed market entry mechanisms, and improved the efficiency of resource allocation. On the one hand, the reform of administrative examination and approval means the streamlining of examination and approval procedures, the decentralization of examination and approval matters, and the integration of examination and approval services, which can reduce rent-seeking activities caused by administrative examination and approval, reduce institutional transaction costs, stimulate productive entrepreneurship, and improve enterprises innovation level. On the other hand, incumbent companies that are about to go bankrupt usually have large assets and a large number of employees. If these companies go bankrupt, it will affect the local economy and employment to a certain extent. Therefore, the local government will help these enterprises so that they will not die. The reform of administrative examination and approval reduces entry costs and barriers and promotes market entry to achieve "dual growth" of economy and employment. With the increase in market vitality, the adverse effects of enterprises' withdrawal from the market will be weakened, which will reduce the motivation of government to rescue enterprises that should have withdrawn from the market, inhibit the formation of zombie companies, and improve the efficiency of resource allocation. In order to test this impact mechanism, this article first identifies zombie companies, then calculates the proportion of zombie companies in the industry, and uses this proportion to measure zombie companies. Finally, on this basis, we empirically test the inhibitory effect of administrative approval reform on the formation of zombie companies.

For the identification method of zombie companies, this article is based on the scholars such as Caballero et al. [33] and Fukuda and Nakamura [35] to identify and analyze the zombie companies in the Chinese industrial enterprise database. The specific identification methods are as follows:

① Calculate the interest rate difference of the enterprise in period t. The interest rate difference is defined as the difference between the actual interest paid by the

TABLE 4: Heterogeneity test results.

	(1) High entry rate	(2) Low entry rate	(3) High exit rate	(4) Low exit rate
AEASR	−0.0143 *	−0.0278 ***	−0.0153 **	−0.0287 ***
	(0.0076)	(0.0076)	(0.0072)	(0.0080)
FC	−0.0026	−0.0026 **	−0.0025	−0.0025 **
	(0.0017)	(0.0012)	(0.0017)	(0.0012)
SC	0.0002 *	0.0007 **	0.0002 *	0.0007 **
	(0.0001)	(0.0004)	(0.0001)	(0.0004)
Export	−0.0359 ***	−0.0019	−0.0341 ***	−0.0025
	(0.0121)	(0.0127)	(0.0115)	(0.0134)
Subsidy	0.3271 ***	0.0981	0.3016 ***	0.1258
	(0.1044)	(0.1567)	(0.1023)	(0.1622)
FIE	0.0209	−0.0267 **	0.0186	−0.0253 *
	(0.0133)	(0.0125)	(0.0127)	(0.0130)
HHI	0.1396 ***	0.1619 ***	0.1366 ***	0.1784 ***
	(0.0120)	(0.0118)	(0.0115)	(0.0124)
Industry fixed effect	Y	Y	Y	Y
City fixed effect	Y	Y	Y	Y
Year fixed effect	Y	Y	Y	Y
R^2	0.0996	0.1037	0.0979	0.1010
Sample size	72865	75437	78854	69448

Note. Symbols *, **, and *** represented significance levels of 10%, 5%, and 1%, respectively; the numbers in brackets were robust standard errors of estimated coefficients.

enterprise and the minimum interest payable and is standardized. The calculation formula is as follows:

$$R_{ft}^* = rs_{t-1} \times BS_{ft-1} + \left(\frac{1}{5} \sum_{j=1}^{5} rl_{t-j} \right) \times BL_{ft-1},$$

$$\mathrm{EIR}_{ft} = \frac{R_{ft} - R_{ft}^*}{R_{ft}^*}, \qquad (4)$$

where EIR represents the interest rate difference; R^* represents the minimum interest payable; R represents the actual interest payment; rs_{t-1} represents the lowest interest rate for short-term borrowing in t-1; rl_{t-j} represents the lowest interest rate for long-term borrowing in t-j; and BS_{ft-1} and BL_{ft-1} are the short-term loan balance and long-term loan balance of enterprise f in $t - 1$ year. In the database of Chinese industrial enterprises, the balance of short-term loans is represented by current liabilities, and the balance of long-term loans is represented by long-term liabilities.

② Calculate the company's income before interest and tax after deducting subsidy income in period t.

③ Calculate the company's debt-to-asset ratio in period $t - 1$ and the growth rate of its debt in period t.

According to the above calculations, if a company meets the following four conditions at the same time, it is considered as a zombie company. ① The interest rate differential of the company in period t is negative; ② The company's income before interest and tax after deduction of subsidies in period t is less than the minimum interest payable; ③ The debt to asset ratio of companies in $T - 1$ period is higher than 50%, and the debt growth rate in t period is greater than 0; ④ The company was identified as a zombie company in both period $t - 1$ and t. The first condition identifies companies that actually pay interest less than the minimum interest payable as zombie companies, indicating that companies that should have withdrawn from the market continue to exist in the market because of credit concessions and become zombie companies. The first condition is easy to identify companies with better qualifications and lower default risk that have obtained bank credit concessions as zombie companies or omit zombie companies that achieve "borrowing new loans to repay old debts" through loan extensions due to maturity. Therefore, adding the second condition and the third condition to solve the problem of identifying zombie companies in the first condition. The zombie companies identified based on the first three conditions are "one-off zombie companies." This may identify normal companies that only encountered short-term problems in their operation and management or suffered short-term external shocks as zombie companies. In order to exclude these companies, we add the fourth condition. It can be said that if a company meets the above four conditions at the same time, it can be considered as a zombie company to a large extent.

Table 5 reports the corresponding estimation results by gradually adding control variables. Column (1) does not include any control variables nor does it control the fixed effects of industry, city, and year. The estimated coefficient of the reform of administrative examination and approval (AEASR) is −0.0326, and it has passed the 1% significance test. However, it can only be inferred that there is a negative correlation between the reform of administrative approval and the formation of zombie companies. The conclusion of causality needs to be further controlled by relevant variables.

TABLE 5: Regression results of the inhibitory effect of the reform of administrative examination and approval on zombie companies.

	(1)	(2)	(3)
AEASR	−0.0326 ***	−0.0081 ***	−0.0072 ***
	(0.0013)	(0.0024)	(0.0024)
FC			0.0001
			(0.0001)
SC			0.0000 ***
			(0.0000)
Export			0.0114 ***
			(0.0035)
Subsidy			0.0614 ***
			(0.0229)
FIE			−0.0535 ***
			(0.0031)
HHI			0.0279 ***
			(0.0024)
Industry fixed effect	No	Control	Control
City fixed effect	No	Control	Control
Year fixed effect	No	Control	Control
Pseudo goodness of fit	0.004	0.064	0.067
Sample size	151855	151855	151190

Note. Symbols *, **, *** represent the significance level of 10%, 5%, and 1% respectively; the values in parentheses are the robust standard deviations.

On the basis of column (1), column (2) further controls the fixed effects of industry, city, and year. The estimated coefficient of the reform of administrative examination and approval is −0.0081, and it is significant at the 1% significance level. Compared with the estimated results in column (1), the inhibitory effect of the reform of administrative examination and approval on the formation of zombie companies is reduced. Therefore, omitting important explanatory variables will overestimate the policy effect of the reform of administrative examination and approval. On the basis of column (2), column (3) adds control variables such as fixed costs, sunk costs, proportion of exports, proportion of subsidies, proportion of foreign-funded enterprises, and degree of market competition. When these variables are controlled, the estimated coefficient of the reform of administrative examination and approval is significantly negative; compared with column (2), the difference of coefficient size is small. According to the estimation results in Table 5, it can be seen that in cities where administrative approval centers have been established, companies are less likely to become zombie companies. Therefore, it can be considered that the reform of administrative examination and approval has inhibited the formation of zombie companies, which indicates that the reform of administrative examination and approval can improve the efficiency of resource allocation.

7. Research Conclusions

The reform of the administrative examination and approval system is an important governmental approach aimed at easing regulations and reducing government intervention.

The concept of the reform is centered on optimizing the examination and approval procedure, so as to increase the efficiency of the administrative examination and approval system and to reduce the institutional transaction costs of enterprises. Based on the views of the degree of productivity dispersion, this study utilized the data of the time of establishment of administrative examination and approval centers in each administrative division in China, as well as data from the China Industry Business Performance Database from 1998 to 2007 to empirically determine whether the reform of the administrative examination and approval system was able to improve the efficiency of resource allocation. The empirical results show that the reform was beneficial for reducing the degree of productivity dispersion among enterprises and was able to improve the efficiency of resource allocation. This conclusion was further supported by a series of robustness tests which included a treatment of endogeneity, a regression analysis for different industries, a remeasured efficiency of resource allocation regression, and a placebo test. Furthermore, this study examined the heterogeneity of the effects of the reform on the efficiency of resource allocation for industries with different market entry and exit rates. The results revealed that the reform was able to reduce the degree of productivity dispersion among the enterprises in industries with lower entry rates, thereby increasing the efficiency of resource allocation, whereas this effect was less prominent for industries with higher entry rates. On the other hand, the positive effects of the reform on the efficiency of resource allocation are more prominent in industries with low exit rates, while the effects were less prominent in industries with high exit rates.

Data Availability

The data used in this study were obtained from the China Industry Business Performance Database from 1998 to 2007 and the time of establishment of the administrative examination and approval center in each administrative division in China.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Y. Z. and L. Z. conceptualized the study; L. Z. and Y. Z. developed methodology; L. Z. provided software; Y. Z. performed formal analysis; Y. Z. and L. Z provided the resources; Y. Z, Y. Z., and L. Z wrote and prepared original draft; Y. Z., Y. Z., and L. Z. wrote, reviewed, and edited the article; L. Z. and Y. Z. supervised the study; L. Z. was responsible for project administration; all authors have read and agreed to the published version of the manuscript.

Acknowledgments

References

[1] E. Bartelsman, J. Haltiwanger, and S. Scarpetta, "Cross-country differences in productivity: the role of allocation and selection," *American Economic Review*, vol. 103, no. 1, pp. 305–334, 2013.

[2] C.-T. Hsieh and P. J. Klenow, "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics*, vol. 124, no. 4, pp. 1403–1448, 2009.

[3] D. Restuccia and R. Rogerson, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, vol. 11, no. 4, pp. 707–720, 2008.

[4] B. Jovanovic and P. K. Goldberg, "Misallocation and growth," *American Economic Review*, vol. 104, no. 4, pp. 1149–1171, 2014.

[5] C. Syverson, "What determines productivity?" *Journal of Economic Literature*, vol. 49, no. 2, pp. 326–365, 2011.

[6] J. M. Arnold, G. Nicoletti, and S. Scarpetta, "Regulation, resource reallocation and productivity growth," *European Investment Bank Papers*, vol. 16, pp. 90–115, 2011.

[7] L. Brandt, J. Van Biesebroeck, and Y. Zhang, "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing," *Journal of Development Economics*, vol. 97, no. 2, pp. 339–351, 2012.

[8] D. Dollar and S. Wei, "Das (wasted) kapital: firm ownership and investment efficiency in China," NBER Working Paper, 2007.

[9] A. C. Pigou, *The Economics of Welfare*, Macmillan and Co, London, UK, 1938.

[10] S. Peltzman, "Toward a more general theory of regulation," *The Journal of Law and Economics*, vol. 19, pp. 245–248, 1976.

[11] R. A. Posner, "The social costs of monopoly and regulation," *Journal of Political Economy*, vol. 83, no. 4, pp. 807–827, 1975.

[12] G. J. Stigler, "The theory of economic regulation," *The Bell Journal of Economics and Management Science*, vol. 2, no. 1, pp. 3–21, 1971.

[13] H. De Soto, *The Other Path: The Invisible Revolution in the Third World*, Harper and Row, New York, NY, USA, 1989.

[14] S. Djankov, R. La Porta, F. Lopez-de-Silanes, and A. Shleifer, "The regulation of entry," *The Quarterly Journal of Economics*, vol. 117, no. 1, pp. 1–37, 2002.

[15] L. Branstetter, F. Lima, L. J. Taylor, and A. Venâncio, "Do entry regulations deter entrepreneurship and job creation? Evidence from recent reforms in Portugal," *The Economic Journal*, vol. 124, no. 577, pp. 805–832, 2014.

[16] M. Bruhn, "License to sell: the effect of business registration reform on entrepreneurial activity in Mexico," *Review of Economics and Statistics*, vol. 93, no. 1, pp. 382–386, 2011.

[17] D. S. Kaplan, E. Piedra, and E. Seira, "Entry regulation and business start-ups: evidence from Mexico," *Journal of Public Economics*, vol. 95, pp. 1501–1515, 2001.

[18] D. Rostam-Afschar, "Entry regulation and entrepreneurship: a natural experiment in German craftsmanship," *Empirical Economics*, vol. 47, no. 3, pp. 1067–1101, 2014.

[19] F. Schivardi and E. Viviano, "Entry barriers in retail trade," *The Economic Journal*, vol. 121, no. 551, pp. 145–170, 2011.

[20] S. E. Black and P. E. Strahan, "Entrepreneurship and bank credit availability," *The Journal of Finance*, vol. 57, no. 6, pp. 2807–2833, 2002.

[21] S.-H. Lee, Y. Yamakawa, M. W. Peng, and J. B. Barney, "How do bankruptcy laws affect entrepreneurship development around the world?" *Journal of Business Venturing*, vol. 26, no. 5, pp. 505–520, 2011.

[22] K. Nystrom, "The institutions of economic freedom and entrepreneurship: evidence from panel data," *Public Choice*, vol. 136, no. 3-4, pp. 269–282, 2008.

[23] B. Casu, B. Deng, and A. Ferrari, "Post-crisis regulatory reforms and bank performance: lessons from asia," *European Journal of Finance*, vol. 23, pp. 1–28, 2016.

[24] A. Guerrini, M. Molinos, and G. Romano, *Italian Regulatory Reform and Water Utility Performance: An Impact Analysis*, Utilities Policy S0957178717302023, 2018.

[25] N. Balasubramanian and J. Sivadasan, "Capital resalability, productivity dispersion, and market structure," *Review of Economics and Statistics*, vol. 91, no. 3, pp. 547–557, 2009.

[26] N. Guner, G. Ventura, and Y. Xu, "Macroeconomic implications of size-dependent policies," *Review of Economic Dynamics*, vol. 11, no. 4, pp. 721–744, 2008.

[27] S. Ding, W. Jiang, and P. Sun, "Import competition, dynamic resource allocation and productivity dispersion: micro-level evidence from China," *Oxford Economic Papers*, vol. 68, no. 4, pp. 994–1015, 2016.

[28] K. Desmet and E. Rossi-Hansberg, "Urban accounting and welfare," *American Economic Review*, vol. 103, no. 6, pp. 2296–2327, 2013.

[29] C. T. Hsieh and E. Moretti, "Housing constraints and spatial misallocation," CEPR Discussion Papers, 2018.

[30] P. D. Fajgelbaum, E. Morales, J. C. Suárez Serrato et al., "State taxes and spatial misallocation," *The Review of Economic Studies*, vol. 86, pp. 333–376, 2018.

[31] L. Foster, J. Haltiwanger, and C. J. Krizan, "Market selection, reallocation, and restructuring in the U.S. Retail trade sector in the 1990s," *Review of Economics and Statistics*, vol. 88, no. 4, pp. 748–758, 2006.

[32] L. Liu, "Entry-exit, learning, and productivity change Evidence from Chile," *Journal of Development Economics*, vol. 42, no. 2, pp. 217–242, 1993.

[33] R. J. Caballero, T. Hoshi, and A. K. Kashyap, "Zombie lending and depressed restructuring in Japan," *American Economic Review*, vol. 98, no. 5, pp. 1943–1977, 2008.

[34] M. A. Mcgowan, A. Dan, and V. Millot, "The walking dead ?: zombie firms and productivity performance in OECD countries," OECD Economics Department Working Papers, 2017.

[35] S.-I. Fukuda and J.-I. Nakamura, "Why did 'zombie' firms recover in Japan?" *The World Economy*, vol. 34, no. 7, pp. 1124–1137, 2011.

[36] M. Storz, M. Koetter, R. Setzer et al., "Do we want these two to tango? On zombie firms and stressed banks in europe," IWH Discussion Papers, 2017.

[37] L. Brandt, T. Tombe, and X. Zhu, "Factor market distortions across time, space and sectors in China," *Review of Economic Dynamics*, vol. 16, no. 1, pp. 39–58, 2013.

[38] J. Lu and Z. Tao, "Trends and determinants of China's industrial agglomeration," *Journal of Urban Economics*, vol. 65, no. 2, pp. 167–180, 2009.

[39] M. Bertrand, E. Duflo, and S. Mullainathan, "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, vol. 119, no. 1, pp. 249–275, 2004.

[40] M. J. Melitz, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, vol. 71, no. 6, pp. 1695–1725, 2003.

[41] G. S. Olley and A. Pakes, "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, vol. 64, no. 6, pp. 1263–1297, 1996.

[42] X. Zhu and Y. Zhang, "Political mobility and dynamic diffusion of innovation: the spread of municipal pro-business

administrative reform in China," *Journal of Public Administration Research and Theory*, vol. 26, no. 3, pp. 535–551, 2016.

[43] H. A. Hopenhayn, "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica*, vol. 60, no. 5, pp. 1127–1150, 1992.

[44] C. Syverson, "Product substitutability and productivity dispersion," *Review of Economics and Statistics*, vol. 86, no. 2, pp. 534–550, 2004.

[45] Z. Song, K. Storesletten, and F. Zilibotti, "Growing like China," *American Economic Review*, vol. 101, no. 1, pp. 196–233, 2011.

[46] J. Levinsohn and A. Petrin, "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies*, vol. 70, no. 2, pp. 317–341, 2003.

Research Article

Research on the Influence of Development Scenarios on the OLCOE of Wind Power: A Case Study of China

Chao Liang,[1,2] Jing Zhang [iD],[1] Yongqian Liu [iD],[1] Jie Yan,[1] and Wei He[3]

[1]*State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (NCEPU),*
 School of Renewable Energy, North China Electric Power University, Changping, Beijing 102206, China
[2]*Guodian Technology & Environment Group Corporation Limited, Beijing 100039, China*
[3]*State Grid Jiangxi Electric Power Research Institute, Nanjing, Jiangsu 330096, China*

Correspondence should be addressed to Yongqian Liu; yqliu@ncepu.edu.cn

To achieve a high penetration of renewable energy, wind power development in China has gradually moved to diverse manifestation (e.g., centralized onshore, low wind speed, and offshore wind power). However, preexisting studies regarding wind power cost neglect to consider the respective characteristics of different development scenarios. In this paper, the overall levelized cost of energy (OLCOE) model is established for different scenarios. Taking China's wind farm data as an example, the impact of development scenarios and wind power permeability on OLCOE and its cost components is quantitatively analysed. The results show that, (1) in the low penetration scenario, low wind speed power has the best economy and is beneficial to the conventional units; (2) the large-scale development of offshore wind power requires a reduction in the cost of offshore wind turbines and submarine cables; and (3) at present, onshore centralized wind power has economic advantages, but there is little room for its cost reduction.

1. Introduction

Rapid global energy transformation is bringing, within reach, an era with a high share of renewables. According to research by the International Energy Agency (IEA), the share of renewable electricity in global primary energy consumption will rise to 23.8% by 2030. In China, as of the end of 2018, the installed renewable capacity accounted for 38.3% of the total, up by 1.7% the previous year. Therefore, China is playing a very important role in the transformation process. As of the end of 2018, China had an installed renewable capacity of 728 million kW and an installed wind capacity of 184 million kW, up by 12% and 12.4%, respectively, over the previous year. Among various forms of clean energy, wind power is developing the fastest and is the most established commercial industry. However, the levelized cost of energy (LCOE) for wind power is still considerably higher than that for conventional forms of energy such as thermal power and hydropower [1, 2]. The need to reduce the costs of wind power is increasingly clear.

Significant progress has been made by researchers, with a focus on the costs of wind power. However, these studies only focus on one aspect of cost (e.g., only generation cost, or transmission cost, or integration cost) and do not give the overall cost of wind power in the process of generation, transmission, and integration.

The generation cost of wind power refers to the annual investment of a wind farm divided by the annual energy production. A modelling method for wind power production and its unpredictability is proposed by analysing the technical limitations of a series of wind turbines, and then the generation cost of wind power and its effect on carbon dioxide emission are discussed [3]. The LCOE for future wind power is estimated with the learning curve method, and the results show that the LCOE for onshore wind power will decrease by 13.91% from 0.40 yuan/kWh in 2016 to 0.34 yuan/kWh in 2025 [4]. An LCOE estimation method taking audit information into account is proposed [5]. The uncertainty of input variables in an LCOE calculation is

discussed, such as variables related to operation cost, initial investment, and power generation, and then the probabilities and their confidence interval of an LCOE larger than its set threshold are estimated by using the joint probability distribution of the LCOE obtained from the Monte Carlo simulation [6]. The research results on the generation cost of some European wind power projects are summarized, and the results show that the utilization hours and initial investment cost have the greatest effect on the generation cost of offshore wind power. In addition, rising raw-material prices have pushed up the generation cost by more than 20% in 3 years [7].

The transmission cost refers to the summation of the infrastructure construction costs caused by high-voltage power transmission lines. A new method for estimating the cost of long-distance electricity transmission (LDET) is proposed, which considers markup by power exporters and importers, and then the factors influencing LDET operation cost are assessed with a weighted least squares regression model [8].

The integration cost of wind power refers to the cost per unit of wind power electricity to be safely and stably consumed in the grid. It does not include transmission costs. The whole life cycle cost of wind power incorporating intermittency cost and the backup cost is analysed, and the competitiveness of wind power in four scenarios—natural gas, thermal power, wind power, and wind power coupled with an energy storage system—is assessed. The results show that the lower intermittency cost of wind power can drive down its cost to a level lower than that of thermal power [9]. The results in [10] show that the intermittency cost of wind power still exists when wind power penetration is low. The integration cost is quantified based on the marginal economic value of power and is decomposed according to the temporal variability and uncertainty of wind power. The results show that the reduction of the utilization rate of thermal power leads to a considerable integration cost [11]. Wind power forecasting creates uncertain factors in the power system due to the intermittency of wind generation [12] and brings about additional operation costs caused by forecasting uncertainty. In the context of this, new concepts and a general methodology are proposed to quantify wind power uncertainty incremental cost and wind power uncertainty dispatch cost based on probabilistic forecasting of wind power [13].

On this basis, the concept of overall levelized cost of energy (OLCOE) for wind power is introduced in the authors' previous work, covering generation cost, transmission cost, and integration cost. The integration cost model quantifies the cost of grid auxiliary services and the cost caused by changes in the operating characteristics of conventional units [14].

In addition to the lack of comprehensiveness of the wind power cost model, the differences between development scenarios are also worthy of attention. Most large wind-power bases in China are generally located in the northeast, northwest, and north of China, far away from load centers in the east. For these regions, the system LCOE for wind power

is high due to the cost of transmission and power curtailment. To improve integration, wind power sources in China are being diversified with offshore, low wind speed, and centralized onshore wind power. A rather balanced mix, instead of the dominance assumed by large-scale centralized onshore wind power, is forming.

Different development scenarios of wind power have their own characteristics and advantages in the process of generation, transmission, and integration. These characteristics lead to differences in the composition of system LCOE. Realizing a higher proportion of wind power integration requires a comprehensive understanding of the system LCOE in different scenarios. It can help guide researchers and policymakers on optimizing development strategies.

Many researchers work extensively on wind power cost under specific scenarios. A life cycle cost model (LCC) for offshore wind power by considering factors, such as location, water depth, and distance to shore, is developed to compare the economics of three different offshore wind farms [15]. The results in [16, 17] show that the location and type of supporting structure of offshore wind farms have a large effect on the overall cost of offshore wind power. A cost model for offshore wind power in the operating period is developed, and the sensitivity of the cost is assessed [18]. The LCC of offshore wind power with high-voltage alternating current (HVAC) and high-voltage direct current (HVDC) technologies is compared, respectively [19]. The LCOE of floating offshore wind power is analysed, and the results show that the LCOE is greatly affected by water depth and distance to shore [20]. The calculation method for balance cost and variability cost of onshore wind power is proposed [21–23]. The relationship between levelized cost and blade diameter of wind turbines at low wind speed is examined [24].

However, the above studies only focus on a single wind power scenario. They neither address structural differences in the system LCOE under the scenarios of onshore, offshore, and low wind speed, nor perform a comparison of the costs at different penetration levels. For the above areas still to be addressed, this work constructs OLCOE models for each of the scenarios based on characteristics of their cost components. To test the validity of the models, scenario-specific parameters are input into the models for comparing each of the cost components under different scenarios; and the relationship between the OLCOE for wind power and the penetration level is also quantified.

The study is structured as follows. Section 2 presents the modelling method for the OLCOE of wind power. Section 3 analyses the differences between different development scenarios for wind power. Section 4 compares the OLCOE for wind power and its components under different scenarios, including different penetration levels and different wind power development modes. A sensitive analysis is also conducted to identify how dependent the OLCOE is on the involved variables. Section 5 summarizes this work. Abbreviations and variables in this paper are shown in Table 1.

TABLE 1: Abbreviation.

Name	Meaning	Unit
AOE_1	Annual operating expenses of wind power	$/year
AOE_2	Annual operating expenses of thermal power	$/year
AEP	Annual energy production	$kW \cdot h$/year
b	Standard coal consumption for power supply	$g/kW \cdot h$
C_{wind}	Generation cost of wind power	$/kW \cdot h
C_{trans}	Transmission cost	$/kW \cdot h
C_{int}	Integration cost	$/kW \cdot h
C_{unit1}	Initial capital cost of wind power	$/kW
C_{unit2}	Initial capital cost of thermal power	$/kW
$C_{thermal}$	Generation cost of thermal power	$/kW \cdot h
C_{infra}	Annual total investment in power transmission and transformation	$
C_{loss}	Electricity loss cost of transmission lines	$/year
$C_{O\&M}$	Annual operation and maintenance cost of lines	$/year
C_{com}	Compensator investment of unit capacity	$/Mvar
$C_{service}$	Ancillary cost	$/kWh
$C_{profile}$	Profile cost	$/kWh
$C_{balance}$	Balance cost	$/kW \cdot h
C_{backup}	Backup cost	$/kW·h
CRF_1	Fixed charge rate of wind power	None
CRF_2	Fixed charge rate of thermal power	None
CRF_3	Fixed charge rate of transmission and transformation	None
I	Total investment in transmission and transformation	$
I_1	Total investment of converter station	$
I_2	Total investment of transmission lines	$
I_3	Total investment of reactive power compensation devices	$
I_{unit1}	Converter/substation investment of unit capacity	$/MVA
I_{unit2}	DC/AC line investment per unit length	$/km
k	Penetration level of wind power	None
L_e	Transmission distance	km
LLC	Land lease cost	$/year
LRC	Levelized replacement/overhaul cost	$/year
OLCOE	Overall levelized cost of energy	$/kW \cdot h
$O\&M$	Annual operation and maintenance cost of wind power	$/year
p_c	Coal price	$/kg
P_{max}	Maximum power of DC/AC lines	MW
P_{com}	Compensator capacity	Mvar
Q	Calorific value of standard coal	kcal/kg
q	Calorific value of coal	kcal/kg
r_{loss_t}	Power loss rate of converter/substation station	None
r_{loss_l}	Power loss rate of DC/AC lines	None
R	Resistance per unit length	Ω/km
r_{oper}	Annual operating maintenance rate of lines	None
T_{wind}	Utilization hours of wind power	h
$T_{thermal}$	Utilization hours of thermal power	h
U_n	Rated voltage of lines	kV
W	Annual electricity received by load	$kW \cdot h$/year
W_{loss}	Annual electricity loss	$MW \cdot h$/year
W_{loss_t}	Annual electricity loss of converter/substation stations	MWh/year
W_{loss_l}	Annual electricity loss of lines	MWh/year
W_{wind}	Electricity generated by wind power	$kW \cdot h$
$W_{thermal}$	Electricity generated by thermal power	$kW \cdot h$
τ	Maximum load loss hours	h
α	Capacity credit of power system	None
α_{wind}	Capacity credit of wind power	None

2. OLCOE Model for Wind Power

In this study, OLCOE for wind power is defined as total expenses incurred in the processes of producing, transporting, and integrating unit wind power [14], as shown in the following equation:

$$OLCOE = C_{wind} + C_{trans} + C_{int}, \qquad (1)$$

where C_{wind} is the levelized generation cost ($/kWh); C_{trans} is the levelized transmission cost ($/kWh); and C_{int} is the levelized integration cost ($/kWh).

Traditional LCOE for wind power only involves the generation cost. However, in reality, the overall cost of the power system in the integration process rises remarkably due to antipeak-shaving property for wind power and its unique variability and uncertainty. Therefore, this study introduces the concept of OLCOE to reflect the actual cost of wind power.

2.1. Generation Cost of Unit Electricity.

The generation cost of unit electricity for wind power is calculated as shown in equation (2). Offshore wind power includes the following additional capital costs: marinization, offshore support structure, offshore transportation, port and staging equipment, personnel access equipment, scour protection, surety bond, and offshore warranty premium. Different types of wind turbines have different computing formulas for the cost of gearbox, generator, and mainframe. The inputs include the following: rated power, rotor diameter, hub height, cut-in, cut-out, and rated wind speed. The details in the computing method can be found in [25]:

$$C_{\text{wind}} = \frac{C_{\text{unit1}} \times \text{CRF}_1}{T_{\text{wind}}} + \text{AOE}_1, \tag{2}$$

where C_{unit1} is the initial capital cost of the wind turbine (\$/kW); CRF_1 is the fixed charge rate of wind power; AOE_1 is the annual operating expenses of wind power; and T_{wind} is the utilization hours of wind power.

As the installed thermal capacity in China is relatively high, thermal power units are assumed as representative units for peak frequency modulation. The generation cost of unit electricity for thermal power units is calculated as shown in the following equation [26]:

$$C_{\text{thermal}} = \frac{C_{\text{unit2}} \times \text{CRF}_2}{T_{\text{thermal}}} + \frac{Q \times b \times p_c}{q} + \text{AOE}_2, \tag{3}$$

where C_{unit2} is the initial capital cost of the thermal power unit (\$/kW); CRF_2 is the fixed charge rate of thermal power; T_{thermal} is the utilization hours of thermal power; Q is the calorific value of standard coal (kcal/kg); q is the calorific value of coal (kcal/kg); b is the standard coal consumption for power supply (g/kWh); p_c is the coal price; AOE_2 is the annual operating expenses of thermal power.

Annual operating expenses (AOE) for the units is given by the following equation:

$$\text{AOE} = \text{LLC} + \frac{O\&M + \text{LRC}}{\text{AEP}}, \tag{4}$$

where LLC is the land lease cost (\$/year); $O\&M$ is the annual operation and maintenance cost of wind power (\$/year); LRC is the levelized replacement/overhaul cost (\$/year); AEP is the annual energy production (kW·h/year).

2.2. Transmission Cost of Unit Electricity.

The transmission cost of unit electricity refers to the annual transmission cost at given annual electricity received by load per W. The annual transmission cost of wind power is calculated as shown in the following equation [27]:

$$C_{\text{trans}} = \frac{C_{\text{infra}} + C_{\text{loss}} + C_{O\&M}}{W}, \tag{5}$$

where C_{infra} is the annual total investment in power transmission and transformation (\$/year); C_{loss} is the electricity loss cost of transmission lines (\$/year); $C_{O\&M}$ is the annual operation and maintenance cost of lines (\$/year); and W is the annual electricity received by load (kW·h/year).

2.2.1. Annual Total Investment in Transmission Systems.

In this study, C_{infra} is the annual value calculated from the total fixed investment in the transmission system:

$$C_{\text{infra}} = I \times \text{CRF}_3, \tag{6}$$

where I is the total investment in transmission and transformation (\$) and CRF_3 is the fixed charge rate of transmission and transformation.

The investment in transmission cost consists of the line, converter/substation, and reactive power compensation. The following equations are applied. As no reactive power needs to be compensated for DC transmission, the investment in DC transmission consists of the line and converter station only. For DC transmission, $I_3 = 0$. For AC transmission, reactive power compensation needs to be considered as high reactive power not only occupies the line capacity but also causes substantial active power loss. The investment in AC transmission consists of the line, substation, and reactive power compensation devices:

$$\begin{aligned}
I &= I_1 + I_2 + I_3, \\
I_1 &= 2P_{\text{max}} \times I_{\text{unit1}}, \\
I_2 &= L_e + I_{\text{unit2}}, \\
I_3 &= C_{\text{com}} \times P_{\text{com}},
\end{aligned} \tag{7}$$

where I_1 is the total investment of converter/substation (\$); I_2 is the total investment of transmission lines (\$); I_3 is the total investment of reactive power compensation devices (\$); P_{max} is the maximum power of DC/AC line (MW); I_{unit1} is the converter/substation investment of unit capacity (\$/MVA); L_e is the transmission distance (km); I_{unit2} is the DC/AC line investment per unit length (\$/km); C_{com} is the compensator investment of unit capacity (\$/Mvar); and P_{com} is the compensator capacity (Mvar).

2.2.2. Electricity Loss Cost of Transmission Lines.

Annual electricity loss cost C_{loss} refers to the loss cost calculated by multiplying the annual electricity loss and generation cost [14]. The electricity loss in transmission can be further divided into line loss and converter/substation loss, as shown in equations (8)–(13).

$$C_{\text{loss}} = C_{\text{wind}} \times W_{\text{loss}}, \tag{8}$$

$$W_{\text{loss}} = W_{\text{loss}\_t} + W_{\text{loss}\_l}, \tag{9}$$

$$W_{\text{loss}\_t} = 2r_{\text{loss}\_t} \times P_{\max} \times \tau, \tag{10}$$

$$W_{\text{loss}\_l} = P_{\max} \times r_{\text{loss}\_l} \times \tau, \tag{11}$$

where W_{loss} is the annual electricity loss (MW·h/year); W_{loss_t} is the annual electricity loss of converter/substation stations (MWh/year); W_{loss_l} is the annual electricity loss of lines (MWh/year); r_{loss_t} is the power loss rate of converter/substation stations; $P_{\max}$ is the maximum power of DC/AC lines (MW); τ is the maximum load loss hours (h); r_{loss_l} is the power loss rate of DC/AC lines.

The power loss rate of DC line:

$$r_{\text{loss}\_l} = \frac{P_{\max} \times R \times L_e}{2U_n^2}, \tag{12}$$

where R is the resistance per unit length (Ω/km) and U_n is the rated voltage of lines (kV).

The power loss rate of AC line:

$$r_{\text{loss}\_l} = \frac{P_{\max} \times R \times L_e}{U_n^2}. \tag{13}$$

2.2.3. Annual Operation and Maintenance Cost.

The annual operation and maintenance expense rate r_{oper} is calculated according to [28]; then the annual operation and maintenance cost $C_{O\&M}$ is given by the following equation:

$$C_{O\&M} = I \times r_{\text{oper}}. \tag{14}$$

2.3. Integration Cost Model of Unit Electricity.

Wind power integration cost C_{int} refers to the expenses incurred by a grid for consuming unit electricity in a safe and stable manner. It consists of the ancillary cost C_{service} and the profile cost C_{profile} [9], as shown in the following equation:

$$C_{\text{int}} = C_{\text{service}} + C_{\text{profile}}. \tag{15}$$

2.3.1. Ancillary Cost.

The ancillary cost of unit electricity consists of balance cost C_{balance} incurred for stabilizing grid fluctuations and adjusting units' ramping as well as backup cost C_{backup} due to the lower capacity reliability of wind power:

$$C_{\text{service}} = C_{\text{balance}} + C_{\text{backup}}. \tag{16}$$

(1) Balance Cost. The balance cost refers to the cost incurred in applying a series of measures to maintain the steady state of the power system and balance the active and reactive power in response to output variability of intermittent renewable energy. Previous studies find that the balance cost of wind power increases with the penetration level due to several factors: the addition of supply to a small control area, forecast errors, interhour variability, intrahour energy balance, and regulation [9]. Reference [3] shows the trend of balance cost in relation to the penetration. In this paper, the balance cost is generalized and fitted by the polynomial

fitting method as shown in equation (17) [14]. There may be a little inaccuracy in this method, but it can correctly reflect the general trend:

$$\begin{aligned} C_{\text{balance}} &= 1.341 \times 10^{-9} \times k^3 - 4.435 \times k^2 \\ &\quad + 14.63 \times k + 0.3105. \end{aligned} \tag{17}$$

(2) Backup Cost. Renewable energy, such as wind power, has lower capacity reliability than those of traditional thermal power units; therefore, the power system needs backup capacity to fill the gap between output and load, particularly under peak load, large variability, and prediction error. The cost incurred by such backup capacity is referred to as backup cost.

In this study, thermal power units are taken as backup capacity. Assume that the capacity of a wind farm is P_{wind}, and its capacity confidence is α_{wind}, while the capacity credit of power system is α. Then, the capacity of $P_{\text{wind}}(\alpha - \alpha_{\text{wind}})$ needs to be spared by thermal power units. The resulting backup cost is calculated as shown in equation (18). P_{wind} in the numerator and denominator are eliminated:

$$C_{\text{backup}} = \frac{(\alpha - \alpha_{\text{wind}}) \times C_{\text{unit2}} \times \text{CRF}_2}{T_{\text{wind}}}, \tag{18}$$

where α is the capacity credit of power system and α_{wind} is the capacity credit of wind power.

Previous studies find that the capacity reliability of wind turbines decreases with the penetration level [29]. Reference [30] shows the trend of capacity reliability in relation to the penetration. In this paper, the rational number fitting method is used to obtain the relationship between capacity reliability and penetration level. The relationship is shown in the following equation [14]:

$$\alpha_{\text{wind}} = 0.384 \times k^2 - 0.6905 \times k + 0.3975. \tag{19}$$

2.3.2. Profile Cost.

After the integration of wind power, with reduced full-load hours (FLH) of conventional units, particularly those for base load and intermediate load, the levelized generation cost of thermal power units will rise as they cannot run at an economical state [9, 31]. This will result in differences in terms of the marginal cost. Define C_{FLH} as the difference of marginal generation cost for conventional units timing their power production, and C_{profile} is the C_{FLH} of unit wind power production, as shown in the following equation [14]:

$$C_{\text{profile}}^{i+1} = \frac{C_{\text{FLH}}}{W_{\text{wind}}^{i+1}} = \frac{\left(C_{\text{thermal}}^{i+1} - C_{\text{thermal}}^{i}\right) \times W_{\text{thermal}}^{i+1}}{W_{\text{wind}}^{i+1}}, \tag{20}$$

where W_{wind} is the electricity generated by wind power (kW·h) and W_{thermal} is the electricity generated by thermal power (kW·h).

3. Three Wind Power Development Scenarios and Their Economic Differences

3.1. Centralized Onshore Wind Power.

Large centralized onshore wind power with long-distance transmission is

typical in China and is the most developed for wind power technology. However, centralized onshore wind power centers are in reverse distribution in relation to load centers, and the long-distance transmission involves high cost and substantial electricity loss. In this study, we assume that HVDC technology is used for transporting wind power from centralized onshore locations as it has the advantages of low construction cost, small loss, large capacity, long life, and unlimited distance of transmission. However, converters for DC transmission are more costly and the current conversion process involves more losses.

From the perspective of investment cost, most centralized onshore wind farms are built in open areas with flat terrain as well as better wind energy resources and weather conditions. Therefore, these locations have lower costs of installation, operation, and maintenance. The design of wind turbines for these locations is also relatively well established, and additional features for low wind speed and offshore turbines are not required, so the cost of the turbines is relatively low.

3.2. Low Wind Speed Power. The areas with low wind speed generally relate to those with an average annual wind speed of 5.3 m/s–6.5 m/s and annual utilization hours of wind resources below 2000 hours. Low wind speed power has the following advantages: (1) good distribution grid networks mean that wind power can be absorbed locally to improve utilization efficiency; (2) these areas are also load centers, and therefore grid loss is small as long-distance transmission is not required; and (3) additional line does not need to be erected and the transmission cost is low. In this work, we assume that overhead AC short-distance transmission is used for these locations; therefore, reactive power compensators are required. AC substations have a lower cost and electricity loss rate than DC converters.

However, the terrain in low wind speed areas is generally more complex with high wind shear exponents, particularly in hills and mountains. Special design features are needed for low wind speed to compensate for lower wind energy. The features mostly adopted at present are longer blades and higher towers that will increase the construction cost of turbines.

3.3. Offshore Wind Power. Offshore wind power has a very high generation cost in terms of both initial projects and later maintenance in their life cycle. Compared with ordinary onshore wind power, the additional costs to be considered in the offshore fields are as follows [32]: (1) expenses needed to endure extreme marine environment; (2) higher base cost of offshore wind turbines; (3) transportation and installation cost of offshore units; (4) shipping as well as loading and unloading of equipment; (5) offshore engineering planning and wind measurement are more complex and time-consuming; (6) protection against wave erosion; and (7) the replacement cost of offshore unit components and the operation and maintenance costs of the wind farms are higher.

Although offshore wind power does not have any advantages in terms of cost currently, it has a high annual average wind speed, small wind shear exponent, stable wind force, high utilization hours, and potential for large-scale development. The measures for reducing the OLCOE of offshore wind power include selecting deepwater locations instead of shallow ones, shifting from fixed to floating types, and increasing the capacity of each individual unit.

China's offshore wind power locations are mostly concentrated in the southeast coastal areas. Although they are close to load centers, the transmission cost is still high due to the expensive submarine cables. In this study, we assume that AC cables are used for transporting power from these locations. Compared with that of onshore substations, the cost of offshore substations is also high due to greater difficulty in their construction, operation, and maintenance.

4. Case Study

4.1. Model Parameters. Tables 2 and 3 present the model parameters used in each development scenario based on their characteristics as discussed in Section 3. Centralized onshore, low wind speed, and offshore turbine, respectively, refer to the actual parameters of GW136-4.2 MW, GW140-2.5 MW, and GW171-6.45 MW.

4.2. Changes of OLCOE for Wind Power and Its Cost Components under Different Penetration Levels. Figure 1 depicts the changes of OLCOE for wind power in relation to penetration levels for different generation modes. It shows that the OLCOE curve for each scenario generally follows an upward trend. Low wind speed has the lowest OLCOE, the curve of which rises quickly with a growth rate of 0.71 $/(MWh * 1\%)$. The OLCOE for onshore and offshore wind power increases almost linearly with penetration levels. The OLCOE for offshore wind power is much higher than that of the other two wind power scenarios, and it is 0.249 $/kWh at a penetration level of 5%.

Figure 2 shows the comparison of cost components of three scenarios at penetration levels of 5%, 25%, and 45%. In general, the generation cost and transmission cost are overwhelming, and the integration cost has an uptrend with the increase of penetration level. The cost compositions in different scenarios differ from each other greatly. Specifically, the results show the following:

(1) At different penetration levels, the relative size of the costs for different generation modes as well as cost components does not change materially. The balance cost, backup cost, and profile cost, as three components of the integration cost, increase slightly in relation to penetration levels.

(2) For centralized onshore wind power, the highest transmission cost among the three generation modes can be attributed to their long distance to load centers. The cost is nearly twice that of its generation cost. This also explains its high OLCOE even with the lowest generation cost.

TABLE 2: Wind turbine parameters.

Wind turbine parameter	Centralized onshore turbine	Low wind speed turbine	Offshore turbine
Rated power (kW)	4200	2500	6700
Rotor diameter (m)	136	139.5	154
Hub height (m)	110	140	103
Rated wind speed (m/s)	11.2	8.5	12.2
Cut-in wind speed (m/s)	2.5	2.5	3
Cut-out wind speed (m/s)	25	20	25

TABLE 3: Transmission cost parameters of onshore, low wind speed, and offshore wind power.

Parameter	Onshore [33, 34]	Low wind speed [35]	Offshore [19, 36, 37]
Line maximum capacity P_{max} (MW)	7200	300	800
Converter/substation investment of unit capacity I_{unit1} ($/MVA)	108180	44234	60847
Transmission line length L_e (km)	2191.5	50	50
Transmission line investment of unit length I_{unit2} ($/km)	706990	163370	1567400
Power loss rate of converter/substation r_{loss_t}	0.0075	0.004	0.004
Maximum load loss hours of line τ (h)	3000	3000	3000
Line resistance of unit length R (Ω/km)	0.0063	0.0205	0.0205
Rated voltage of line U_n (kV)	800	220	220
Annual operation and maintenance expense rate r_{oper}	0.018	0.018	0.012
Utilization hours of line T_{max_l} (h)	5000	5000	5000
Compensator cost of unit capacity C_{com} ($/Mvar)	—	74620	74620
Compensator capacity P_{com} (Mvar)	—	140	140
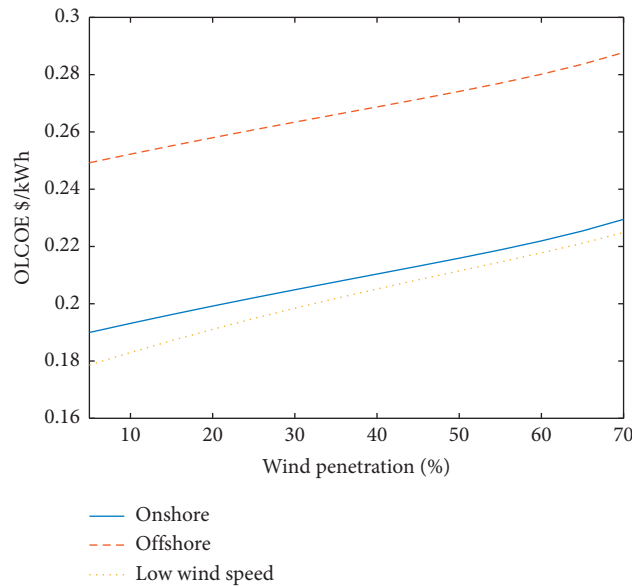
FIGURE 1: Changes of OLCOE in relation to penetration levels under three development scenarios.

(3) As the power generated can be absorbed locally, low wind speed power has the lowest OLCOE, largely due to the lowest transmission cost with short-distance AC lines used. However, with their small FLH and the design features of a high tower and long blades, they have a high generation cost.

(4) With high FLH offset by the disadvantages of high construction cost and operation and maintenance cost, offshore wind power has a high generation cost. Even with a short transmission distance, high cost of submarine cables and offshore substations renders their transmission cost almost as high as that of centralized onshore wind power.

(5) At a penetration level of 5%, only the low wind speed has a profile cost that is below zero. Therefore, only in the case of low wind speed, thermal power generation cost shows a downward trend near the penetration level. Since the backup cost is determined by the FLH of wind power, it is the highest for low wind speed.

Figure 3 shows the change of cost components in relation to penetration levels under three scenarios as follows:

(1) The generation and transmission costs do not vary with wind power penetration levels. This is because the generation cost for wind turbines is related to

(a)

(b)

(c)
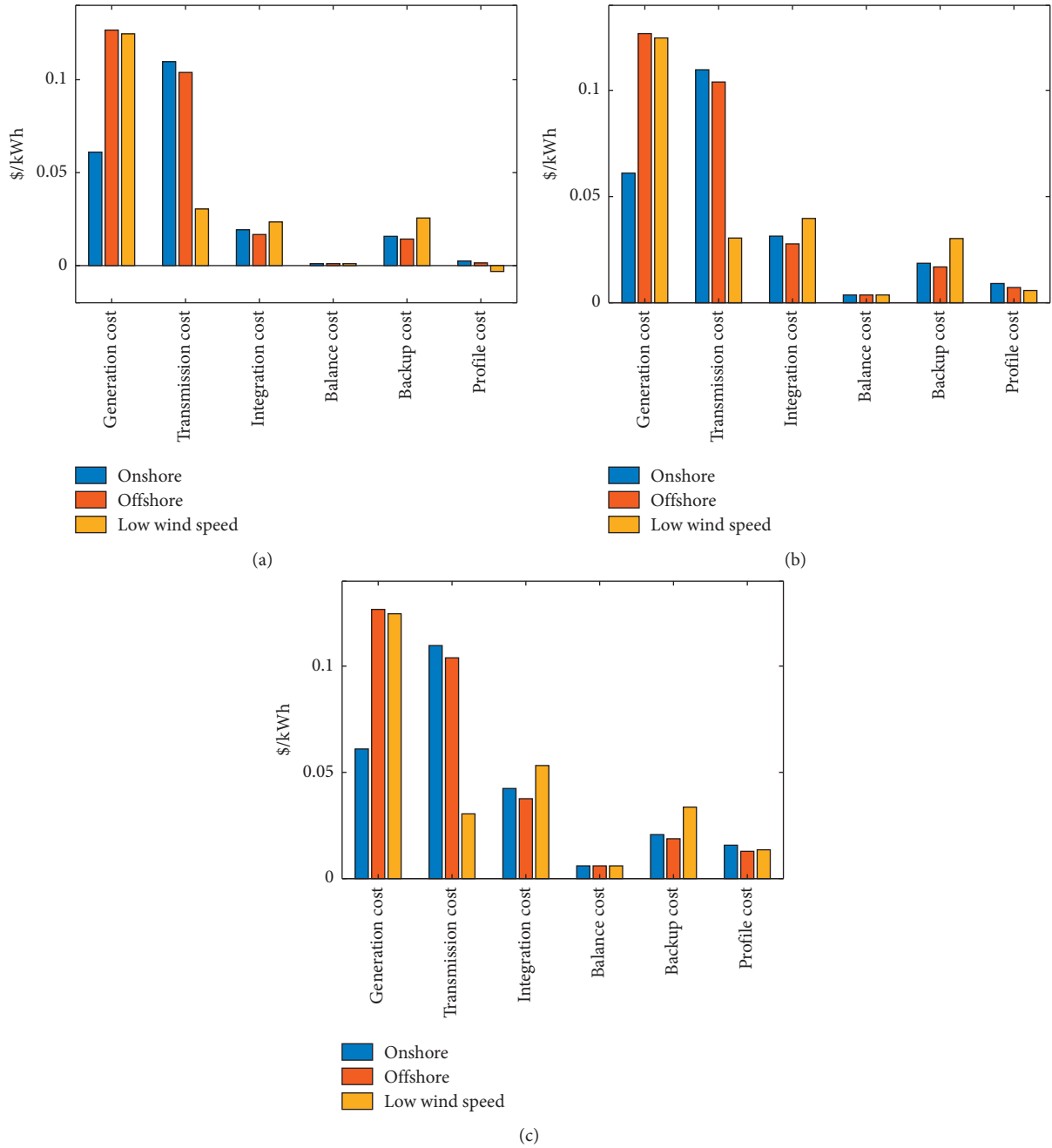
FIGURE 2: Comparison of OLCOE cost components for three scenarios under different penetration levels. (a) 5% penetration level. (b) 25% penetration level. (c) 45% penetration level.

their FLH, which in turn is determined by wind resources and control strategies at their locations. The transmission cost is not a function of penetration because both the power transported and its cost allotted according to the penetration levels.

(2) In centralized onshore wind power, the transmission cost is much higher than the generation cost, at 0.110 $/kWh and 0.061 $/kWh, respectively. The generation cost of the other two generation modes is greater than the transmission cost. For offshore wind power, the generation and transmission costs are 0.127

$/kWh and 0.104 $/kWh, respectively. Low wind speed provides a similar generation cost to that of offshore wind power.

(3) The integration cost rises sharply with increasing penetration level. For centralized onshore wind power, the integration cost is equal to the generation cost at a penetration level of 70%.

(4) The profile cost rises at the steepest rate among all the components of integration cost, particularly for low wind speed; its growth rate is 0.43 $/(MWh ∗ 1%) while that for offshore wind power is 0.33
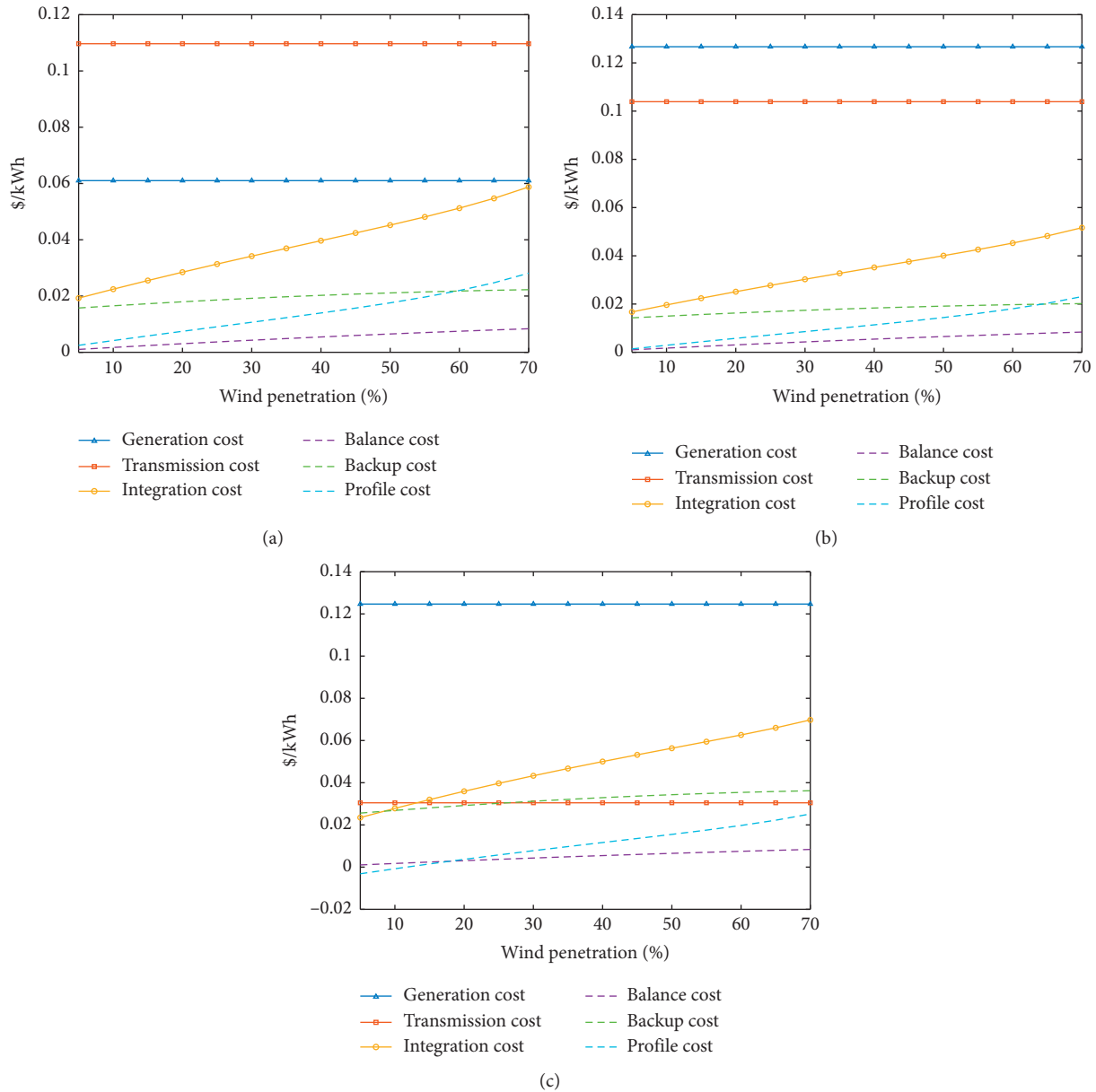
FIGURE 3: Changes of each cost component in relation to penetration levels under different scenarios. (a) Centralized onshore. (b) Offshore wind power. (c) Low wind speed.

$/(MWh * 1%). For low wind speed, the profile cost is negative when the penetration level is less than 10%.

(5) Although the base of backup cost is relatively high, it grows slowly with the increase of penetration levels. Due to low FLH of low wind speed power, more backup units of thermal power are required, and therefore the backup cost is the highest of the three scenarios at 0.036 $/kWh. The profile cost for centralized onshore and offshore wind power exceeds their backup cost at penetration levels of 55% and 65%, respectively, while that for low wind speed has never been higher than the backup cost.

4.3. Changes of Integration Cost in relation to Penetration Levels.
If the load demand remains unchanged, the integration of wind power will reduce the power generated by thermal plants, which will then result in lower FLH and high generation cost for the thermal plants. In addition, the reliability of wind power capacity decreases with the increase of penetration. The unreliable capacity must be compensated by thermal power. Therefore, the additional profile cost and backup cost will rise with increasing the penetration level.

Figure 4 shows the change in the proportion of integration cost in the OLCOE in relation to the penetration levels. The figure shows that the proportion of integration cost is increasing almost linearly. The integration cost of low
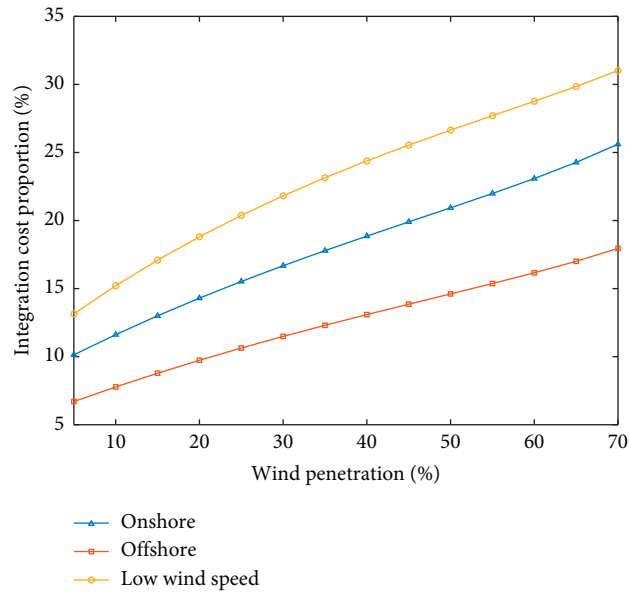
FIGURE 4: Proportion of integration cost in relation to penetration levels under different scenarios.

wind speed accounts for 14–31% of the OLCOE, the highest among the three scenarios, while that of offshore wind power is the lowest. As shown in Figure 2, this is mainly due to the fact that low wind speed needs more backup capacity to ensure the reliability of power supply.

4.4. Changes of Profile Cost and Their Effects on Conventional Units. Given a constant total load, wind power integration will replace part of the conventional capacity and reduce the power generated by conventional units. However, the relationship between the capacity reliability of wind power and the penetration levels is not linear. Therefore, the FLH of conventional units do not change linearly with penetration. Figure 5 shows the effects of wind power on conventional units under three different scenarios.

As for the influence of wind power on thermal power FLH, the following findings are made:

(1) With the increase in penetration, the FLH of conventional power is gradually reduced under the scenarios of offshore and centralized onshore wind power.

(2) With a high proportion of wind power, the FLH of thermal power will drop even faster. This is because wind power has lower capacity confidence at a high penetration, and more backup thermal power units are needed to absorb each kilowatt-hour of wind power. This part of reserve capacity is usually operated in a noneconomic state. In the offshore case, the decline rate of thermal FLH increases from 15 to 59 h per 1% penetration.

(3) As the FLH of offshore wind power are higher than those of the other two, they have the greatest effect on conventional units in the system. Therefore, the thermal FLH will drop at the highest speed. Offshore wind power has an average decreasing rate of 46.6 h

per 1% penetration, while that for centralized onshore wind power is 46.2 h per 1% penetration.

(4) In the case of low wind speed, thermal FLH shows a slightly increasing trend when the penetration level is low (below 10%), indicating that low wind speed power at low penetration levels is beneficial to reducing thermal generation cost. This may be due to the fact that, at low penetration in low wind speed scenarios, the capacity confidence of wind power is relatively high, while FLH of low wind speed are relatively low. The task of thermal power units is more to bear the base load, rather than as a backup. Therefore, when the proportion of wind power increases, the capacity of thermal power units as reserves will be reduced more, so the FLH of thermal power will increase slightly. However, at a relatively high penetration, the base load is transferred from thermal power plants to wind power plants. Thermal power units are more used as backups, resulting in a decrease of FLH.

As for the profile costs under different wind power ratios and different development scenarios, the following findings are made:

(1) It can be seen from equation (20) that the profile cost of wind power is positively correlated to the growth rate of the thermal generation cost. Due to the dropping FLH with an increasingly faster speed, the generation cost of thermal power gets higher and grows faster at high penetration. Therefore, the profile cost increases with the penetration level.

(2) Similar to the FLH, the greater the effect on conventional units, the greater the profile cost of wind power. Therefore, the profile cost for offshore wind power is the highest among the three scenarios.
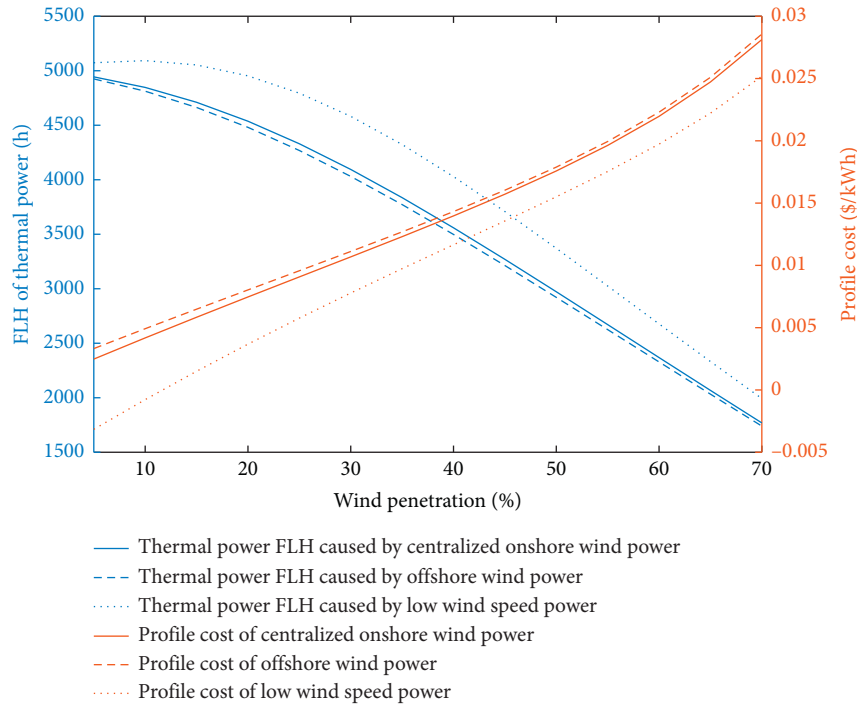
Figure 5: Changes in thermal FLH and profile cost with increasing penetration levels for various scenarios.

(3) Although the profile cost for low wind speed is negative at low penetration levels, it has a high growth rate.

4.5. Sensitivity Analysis. According to the above analysis, OLCOE of wind power is jointly determined by many variables. A sensitivity analysis is conducted to clearly identify how dependent the OLCOE is on various variables. The parameters that have a great impact on the generation cost, transmission cost, and integration cost, that is, the FLH of wind power, transmission distance, and capacity confidence of wind power, are selected for sensitivity analysis under the scenario of 30% wind power ratio.

Figures 6–8 show the variation trend of wind power OLCOE and its cost components with increasing FLH, transmission distance, and capacity confidence. In general, under the 30% ratio of wind power, the FLH and transmission distance has a greater impact on OLCOE.

It can be seen from Figure 6 that various costs are reduced to varying degrees with the growth of wind power FLH. In addition, the reduction in generation costs is the most obvious. This is because the generation cost is closely related to the annual power generation and FLH of wind power. The higher the FLH, the lower the investment as well as the operation and maintenance costs allocated to each kWh of wind power. However, as FLH of wind power becomes higher and higher, the downward trend of OLCOE will become more and more gentle. In the stage when the FLH increases from 1500 h to 2000 h, OLCOE decreased by 0.027$, while in the 3000 h–3500 h period, OLCOE only decreased by 0.008$. This is because with the gradual increase of FLH, although the levelized cost of
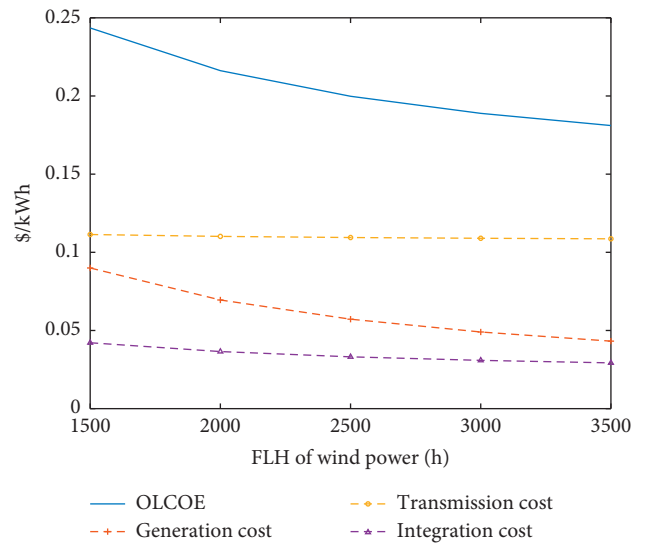
Figure 6: Changes of OLCOE and each cost component with increasing FLH of wind power.

wind turbines is reduced, the variable operating costs (i.e., the operating and maintenance costs required for generating each kilowatt-hour of wind power) remain unchanged.

It can be seen from Figure 7 that the transmission distance only affects the transmission cost. OLCOE is linearly related to the transmission distance. For every 500 km increase in transmission distance, OLCOE increases by 0.013$.

It can be seen from Figure 8 that the confidence level of wind power capacity only affects the integration cost. The
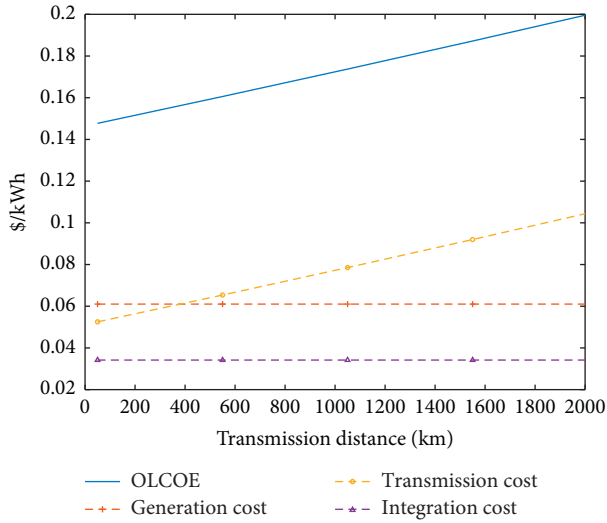
FIGURE 7: Changes in OLCOE and each cost component with increasing transmission distance.
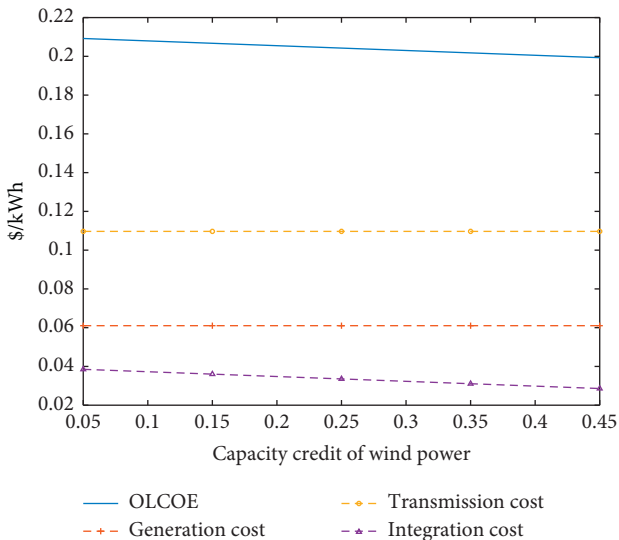
FIGURE 8: Changes of OLCOE and each cost component with increasing capacity credit of wind power.

capacity confidence of wind power indicates the capacity of conventional units that wind turbines can replace. In scenarios with high confidence in wind power capacity, fewer backup conventional units are required; thus, the integration costs will be lower. With the increase in capacity confidence, OLCOE declines but not significantly (the rate of decrease is 0.025$/unit capacity confidence).

5. Conclusions

In this work, for three development scenarios of wind power in China—large wind power bases that require long-distance transmission, offshore wind power that is under large-scale development, and those with low wind speed—we propose methods for calculating the OLCOE for the different scenarios. Given the differences between

various sources of wind power, we analyse the OLCOE and its components for various development modes and present the distribution of various costs at different penetrations.

The results indicate that, in the current scenario, low wind speed power has the best economy due to its advantage of being close to the load center. The development of low wind speed power at low penetration is conducive to economically reform the conventional units in the transformation of the power system. The large-scale development of offshore wind power requires a reduction in the cost of offshore wind turbines and submarine cables. At present, onshore centralized wind power has economic advantages. But its technology is relatively mature; hence, there is little room for its cost reduction.

The details of the findings are as follows:

(1) The OLCOE curve of wind power for each scenario generally follows an upward trend. The OLCOE for low wind speed is the lowest, while that for offshore wind power is the highest with a large gap between the other two cases.

(2) For centralized onshore wind power, the generation cost is the lowest (0.061 $/kWh) and the transmission cost is the highest (0.110 $/kWh).

(3) Offshore wind power has the highest generation cost of 0.127 $/kWh. Its transmission cost of 0.104 $/kWh almost catches up with that of centralized onshore wind power.

(4) Low wind speed power has the lowest transmission cost (0.030 $/kWh) and a high generation cost (0.125 $/kWh).

(5) With the increase of penetration level, the integration cost continues to increase. The integration cost for low wind speed accounts for 14–31% of the OLCOE, the highest proportion among the three wind power scenarios.

(6) Offshore wind power has the highest profile cost. Although the profile cost for low wind speed is negative at low penetration levels, it has a high growth rate.

(7) FLH of wind power and transmission distance influence the OLCOE to a large extent.

Data Availability

The 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

References

[1] Z. M. Wang and Z. N. Lu, "Techno-economic analysis on cost structure and operation value of the wind power project," *Science Management Research*, vol. 27, no. 2, pp. 51–54, 2009, in Chinese.

[2] H. Q. Yu, *China's Wind Power Development Study of the Cost and Wind Power Generators Selection*, North China Electric Power University, Beijing, China, 2009, in Chinese.

[3] E. D. Delarue, P. J. Luickx, and W. D. D'haeseleer, "The actual effect of wind power on overall electricity generation costs and CO2 emissions," *Energy Conversion and Management*, vol. 50, no. 6, pp. 1450–1456, 2009.

[4] Q. Tu, R. Betz, J. Mo, Y. Fan, and Y. Liu, "Achieving grid parity of wind power in China - present levelized cost of electricity and future evolution," *Applied Energy*, vol. 250, pp. 1053–1064, 2019.

[5] J. Aldersey-Williams, I. D. Broadbent, and P. A. Strachan, "Better estimates of LCOE from audited accounts-a new methodology with examples from United Kingdom offshore wind and CCGT," *Energy Policy*, vol. 128, pp. 25–35, 2019.

[6] M. I. Blanco, "The economics of wind energy," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 6-7, pp. 1372–1382, 2009.

[7] C.-T. Li, H. Peng, and J. Sun, "Life cycle cost analysis of wind power considering stochastic uncertainties," *Energy*, vol. 75, pp. 411–418, 2014.

[8] B. Lin and W. Wu, "Cost of long distance electricity transmission in China," *Energy Policy*, vol. 109, pp. 132–140, 2017.

[9] J. F. DeCarolis and D. W. Keith, "The costs of wind's variability: is there a threshold?" *The Electricity Journal*, vol. 18, no. 1, pp. 69–77, 2005.

[10] L. Hirth, F. Ueckerdt, and O. Edenhofer, "Integration costs and the value of wind power," *SSRN Electronic Journal*, 2012.

[11] F. Ueckerdt, L. Hirth, G. Luderer, and O. Edenhofer, "System LCOE: what are the costs of variable renewables?" *Energy*, vol. 63, pp. 61–75, 2013.

[12] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, "Reviews on uncertainty analysis of wind power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 52, pp. 1322–1330, 2015.

[13] J. Yan, F. Li, Y. Liu et al., "Novel cost model for balancing wind power forecasting uncertainty," *IEEE Transactions on Energy Conversion*, vol. 32, no. 1, pp. 318–329, 2016.

[14] Y. Liu, C. Liang, J. Yan et al., "Overall levelized cost of energy for a large-scale wind farm base with long-distance transmission and accommodation," *Proceedings of the CSEE*, vol. 40, no. 2, pp. 500–509, 2020.

[15] M. Varvara and A. J. Kolios, "Techno-economic optimization of offshore wind farms based on life cycle cost analysis on the UK," *Renewable Energy*, vol. 132, pp. 439–454, 2019.

[16] V. Mytilinou, A. J. Kolios, and G. D. Lorenzo, "A comparative multi-disciplinary policy review in wind energy developments in Europe," *International Journal of Sustainable Energy*, vol. 36, no. 8, pp. 1–21, 2015.

[17] S.-Y. Lin and J.-F. Chen, "Distributed optimal power flow for smart grid transmission system with renewable energy sources," *Energy*, vol. 56, pp. 184–192, 2013.

[18] H. Y. Chen, K. Tan, S. T. Xi et al., "A model for calculating operation period cost of offshore wind power," *Automation of Electric Power Systems*, vol. 38, no. 13, pp. 135–139, 2014, in Chinese.

[19] Z. C. Zhang, J. T. Guo, H. J. Zhu et al., "Optimization scheme of offshore wind power grid connection based on LCC model," *Power System Protection and Control*, vol. 45, no. 21, pp. 51–57, 2017, in Chinese.

[20] A. Myhr, C. Bjerkseter, A. Ågotnes, and T. A. Nygaard, "Levelised cost of energy for offshore floating wind turbines in a life cycle perspective," *Renewable Energy*, vol. 66, pp. 714–728, 2014.

[21] H. Holttinen, P. Meibom, A. Orths et al., "Impacts of large amounts of wind power on design and operation of power systems, results of IEA collaboration," *Wind Energy*, vol. 14, no. 2, pp. 179–192, 2011.

[22] C. S. Chen, *Research on System Cost and Block Value of Wind power*, Southeast University, Nanjing, China, 2015, in Chinese.

[23] L. Li, Y. X. Zhang, C. S. Chen et al., "Calculation of wind power variation costs and its application research," *Proceedings of the CSEE*, vol. 36, no. 19, pp. 5396–5396, 2016, in Chinese.

[24] J. Y. Chai and S. Y. Zhang, "Total cost and energy cost of wind turbine generation systems for low wind speeds," *Journal of Tsinghua University Science and Technology*, vol. 51, no. 3, pp. 356–360, 2011, in Chinese.

[25] L. Fingersh, M. Hand, and A. Laxson, *Wind Turbine Design Cost and Scaling Model*, National Renewable Energy Lab (NREL), Golden, CO, USA, 2006.

[26] R. B. Shen, L. Z. Zhang, H. Zhang et al., "Comparative research on the economy of coal transportation and power transmission," *Electric Power*, vol. 46, no. 10, pp. 133–139, 2013, in Chinese.

[27] H. Zhao, S. Guo, and H. Li, "Economic impact assessment of wind power integration: a quasi-public goods property perspective," *Energies*, vol. 8, no. 8, pp. 8749–8774, 2015.

[28] R. Karki and R. Billinton, "Cost-effective wind energy utilization for reliable power supply," *IEEE Transactions on Energy Conversion*, vol. 19, no. 2, pp. 435–440, 2004.

[29] H. Weigt, "Germany's wind energy: the potential for fossil capacity replacement and cost saving," *Applied Energy*, vol. 86, no. 10, pp. 1857–1863, 2009.

[30] P. Heptonstall, R. Gross, and F. Steiner, *The Costs and Impacts of Intermittency-2016 Update: A Systematic Review of the Evidence on the Costs and Impacts of Intermittent Electricity Generation technologies*, UK Energy Research Centre (UKERC), London, UK, 2017.

[31] Z. Liu, W. Zhang, C. Zhao, and J. Yuan, "The economics of wind power in China and policy implications," *Energies*, vol. 8, no. 2, pp. 1529–1546, 2015.

[32] J. Xu, Y. Jin, C. C. Hu et al., "Research on combined power transmission scheme for offshore windfarm cluster," *Advances of Power System & Hydroelectric Engineering*, vol. 32, no. 11, pp. 107–113, 2016, in Chinese.

[33] Y. Zhao, B. Han, and G. L. Fang, "A survey of condition monitoring and fault diagnosis for wind power generators," *Thermal Power Generation*, vol. 45, no. 10, pp. 1–5, 2016, in Chinese.

[34] B. Li and X. J. Du, "The comprehensive cost analysis of HVDC transmission project based on different voltage levels," *Construction & Design for Project*, no. 1, pp. 128–130, 2017, in Chinese.

[35] H. Yao, X. N. Huang, G. Yu et al., *Cost of Power Engineering Projects Put into Operation during the 11th Five-Year Plan period*, State Electricity Regulatory Commission, Beijing, China, 2012, in Chinese.

[36] B. J. Cheng, Z. Xu, Y. W. Xuan et al., "Economic comparison of AC/DC power transmission system for submarine cables," *Electric Power Construction*, vol. 35, no. 12, pp. 131–136, 2014, in Chinese.

[37] M. Zheng and C. H. Wang, "Research on the transmission mode of offshore wind farm," *Southern Energy Construction*, no. 2, pp. 99–108, 2018, in Chinese.

Research Article

Assessment of Uncertainties in Energetic and Exergetic Performances of a Flat Plate Solar Water Heater

Intissar Harrabi [ID],[1] Mohamed Hamdi [ID],[2] and Majdi Hazami[1]

[1]*Thermal Process Laboratory, Research and Technology Centre of Energy, BP 95, Hammam-Lif 2050, Tunisia*
[2]*Laboratory of Wind Power Control and Waste Energy Recovery, Research and Technology Centre of Energy, BP 95, Hammam-Lif 2050, Tunisia*

Correspondence should be addressed to Intissar Harrabi; harrabiintissar55@gmail.com

This paper aims to quantify sensitivities of energy and exergy performances of Flat Plate Solar Water Heaters (FPSWHs) with respect to measurement parameters. For that purpose, a computational tool is developed and validated by using outdoor conditions according to the test standard EN 12975. First of all, numerical simulations are compared with experimental results and available data in the literature, and the comparison shows a good agreement. Then, we apply the proposed model to the quantification of uncertainties associated with transient simulation. Results show that ambient temperature is the main relevant factor in operating conditions, and its effect reaches 13.7% and 3.89% on energy and exergy efficiencies, respectively, when the deviation in the sensor measurement is about ±1°C. When 0.15 v% multiwall carbon nanotubes (MWCNT)-Ethylene-Glycol (E-G) (30 : 70) nanofluid is used as working fluid, results show that a suitable choice of nanofluid properties achieves 84.7% of the thermal efficiency during the zero reduced temperature conditions compared to 75.4% when the collector works with E-G. Using common empirical correlations affects substantially the accuracy of the fitting parameters, and the deviation in exergy efficiency reaches 1.18%.

1. Introduction

There is evidence that the investment in solar water heater installations is reversed by gaining electricity for the first 4 to 5 years, and then the hot water becomes free during the life time of equipment over 15 years. In addition, governments provide subsidized loans to finance equipment costs. For our country, Tunisia, the solar water heater market has grown significantly in recent years thanks to the subsidy granted by the program to promote the use of the solar water heater (35% of the purchase price is refunded), which is among the projects eligible for the Clean Development Mechanism (CDM). This mechanism represents one of the three flexibility mechanisms provided in Article 12 of the Kyoto Protocol of Climate Change Convention signed by Tunisia [1]. To reach this goal, prior to the sale of a Solar Water Heater (CES), each supplier must submit to the National

Agency of Energy Management (ANME) a specific eligibility file for each proposed model. Several studies have been carried out in the direction of perfecting solar water heaters, and currently, there are several international standards that set the test methods for solar water heaters in order to predict their performance. Among the global testing methods, we mention the DST test procedure that was developed in Germany, and that has been registered as ISO 9459-5. Moreover, it is for this reason that about hundreds of technical papers and reports of Performance Test according to the DST procedure have been elaborated on request in Research and Technology Centre of Energy (CRTEn). The second test procedure (Input/output) does not require a study of the whole system, i.e., without the storage tank.

In fact, all test procedures aim to evaluate the thermal efficiency of solar collectors; a few investigations [2] are found in the literatures that are interested in studying their

exergy efficiencies. Kalogirou et al. [3] reviewed the exergy analysis of solar thermal collectors. They gave a clear methodology of exergy analysis of solar collectors. They highlighted that exergy is more representative of performance evaluation. Farahat et al. [4] carried out a numerical study to optimize the exergy performance of flat plate solar collectors by using design parameters. They validated their code experimentally under given design and operating conditions. Their results showed that there exist optimum values of the maximum exergy efficiency for a specific mass flow rate and absorber area.

In general, numerical tools are used in order to validate the experimental results. The latter often shows several uncertainties that directly affect the collector's performances. The physics community usually defined an interval of "probable" values of the quantity to be measured in which we have a 95% chance of finding the "true value." We are talking about 95% confidence interval. Several authors have been interested in uncertainty tests for solar water heaters. Tang et al. [5] used a test standard published by the National Technology Supervision Bureau of China in 1997 to assess the uncertainty of heat loss coefficient, U_L, of a glass evacuated solar collector tube. They found that the choice of the test procedure is the most influencing parameter on the evaluation of U_L. Their results also showed that measuring errors of instruments contributed little to U_L. Kicsiny [6] validated a model based on Multiple Linear Regression (MLR) for solar collectors. They showed that the MLR-based model has a lot of precision than models working on regression in the literature. Their results showed that the proposal based on the MLR model is considered to be the simplest black box with a precision greater than 5%. They also showed that the MLR-based model gives collector output temperature results with higher accuracy than the physically based model. Facão and Oliveira [7] presented a new method in order to evaluate heat pipe solar collector efficiency. They analyzed the uncertainty by separating systematic and random uncertainties. Their results showed that the uncertainty of efficiency is higher for smaller values of $(T_{fm} - T_a)/G$, where T_{fm}, T_a, and G are the fluid mean temperature, the ambient temperature, and the solar incident irradiation, respectively. Sowmy et al. [8] evaluated the uncertainty in the efficiency tests of conventional flat plate solar collectors by using a solar simulator. They performed indoor tests for the uncertainty estimation associated with the determination of efficiency. They evaluated the contribution of the errors in the instrumentation, the test conditions, and the statistical regression. Their results showed that using an artificial solar simulator is primordial in order to obtain good accuracy. Their study also showed that regression uncertainties also have a significant effect as measurement uncertainty on the results. Mathioulakis et al. [9] performed a systematic analysis of the contribution of all components of uncertainty based on the ISO 9806-1 procedure test to determine the final uncertainty in the parameters of the characteristic equation and the instantaneous efficiency of the collector. They proposed a step-by-step methodology

that uses specific statistical tools to assess the reliability of the test procedure and also to quantify the quality of the adjustment. Their results showed that the stepwise methodology is a reliable and efficient tool to predict the performance of the collector by considering the efficiency of the collector a key parameter. Unlike the ISO 9806-1 standard which does not provide the uncertainty of the parameters of the efficiency curve, Sabatelli et al. [10] conducted an experimental study in order to find the uncertainty of curve fitting parameters. In their study, the authors proposed a methodology to solve this problem and to evaluate not only the parameters and their uncertainties, but also the reliability of the test procedure and its effectiveness aptitude. They performed a sensitivity analysis to evaluate the effects of measurement errors on the uncertainty values in the estimates settings.

Regression Models and Sensitivity Analysis for the Thermal Performance of solar collectors were also studied by Rehman et al. [11]. The authors used the water, 20% glycol-water, and 40% glycol-water as working fluids. They found that the coefficients of determination exceed 97% for the efficiency models and 93% for the fluid outlet temperature. They also performed a sensitivity analysis to determine the most influential design, operational and natural parameters on the collector's efficiency. Karwa and Baghel [12] used a mathematical model to study the effect of measurement uncertainties on the thermohydraulic performance of solar plate heaters. They showed that, based on thermohydraulic consideration, the driving height is 10 mm is preferable. Their results also showed that uncertainty in the wind heat transfer coefficient could affect the thermal efficiency by 2.6%. The efficiency varies by 1.3% when the sky temperature varies by 5 K. This efficiency was affected by approximately −1.5 to 1.3% by varying the solar irradiation from 500 to 1000 Wm^{-2}. Their study also showed that by increasing the flow, these effects decrease. They are lower for the collector with a selective coating on the surface of the absorbing plate.

As described above, the determination of the influence input parameters on the efficiency of solar collectors attracted the intention of several studies in recent years. The determination of tolerances and the effect of uncertainties measurement on the results is essential for designing solar collectors and influences the manufacturing costs. This also helps in evaluating and comparing possible configurations of these systems. The studies carried out in this area are limited and require further investigation essentially in terms of exergy performances. The aim of this study is to quantify some pertinent parameter sensitivities on both energy and exergy performances of a FPSWH. We focus on better understanding the origin of different sources of deviation in energy and exergy efficiencies estimation during solar collector test motivated by applications that aim to improve solar water heaters. In the following section, we describe the solar collector tested in the Research and Technology Centre of Energy (CRTEn), the mathematical model, and the solution procedure taken to analyze the effect of uncertainties of mentioned parameters on thermal behavior and exergy performance of the system.

2. Mathematical Model

As described previously, the aim of this work is to quantify the effects of several operating parameters on the energy and exergy efficiencies of a FPSC. For this reason, it is essential to develop an efficient numerical tool allowing quantifying the uncertainties in the FPC's performances. For this purpose, a validation step is necessary. The validation of the code helps to determine if the mathematical model is able to investigate the heat transfer and control the exchange of entropy in the system. The test of the solar collector is carried out according to EN 12975 procedure. Solar collector testing determines the reliability and efficiency characteristics of a collector. Knowledge about efficiency makes it possible to optimize the cost/benefit ratio and to classify the performance of collectors in the productive sector. As a result of the test, it is possible to compute the energy that the collector will produce in a specific climate region.

The model used in this study is the one-dimensional mathematical model previously used by Zima and Dziewa [13], Saleh [14], Saleh et al. [15], and Cadafalch [16]. This model consists of five nodes and thus a system of five equations. It has been shown that this model is useful for analyzing the influence of the mass flow rate on the performance of the collector and also the analysis of the determination of the real time collector's performances. This model can be applied to the collectors in a serpentine or parallel tube arrangement configuration with single or double covers [13]. The energy balance equation for the glass cover is written as follows:

$$C_g \rho_g V_g \frac{dT_g}{dt} = \left[\left(h_{r_{g-am}} + h_{c_{g-am}}\right)\left(T_{am} - T_g\right) + h_{r_{ab-g}}\left(T_{ab} - T_g\right) \right. $$
$$\left. + h_{c_{g-a}}\left(T_a - T_g\right) + \alpha G\right] p\Delta z, \tag{1}$$

where C_g is the specific heat of the glass cover, ρ_g and V_g are its density and volume, respectively. T_{am}, T_g, T_{ab} and T_a are the ambient, the glass, the absorber, and the air gap temperature, respectively. α is the absorption coefficient. The radiation heat transfer coefficient between the outer surface of the glass cover and the ambient is computed as follows:

$$h_{r_{ab-g}} = \frac{\sigma\left(T_{ab}^2 + T_g^2\right)\left(T_{ab} + T_g\right)}{\left(1 - \varepsilon_{ab}/\varepsilon_{ab}\right) + \left(1/F_{ab-g}\right) + \left(\left(1 - \varepsilon_g\right)/\varepsilon_g\right)\left(A_{ab}/A_g\right)}, \tag{2}$$

where F_{ab-g} is the view factor between two surfaces of absorber and glass. The radiation transfer coefficient between the glass and the ambient is computed as the following:

$$h_{r_{g-am}} = \frac{\sigma \varepsilon_g\left(T_g^4 - T_{sky}^4\right)}{T_g - T_{am}}. \tag{3}$$

$h_{c_{g-a}} = \mathrm{Nu}_a \lambda_a/\delta_a$ is the convective heat transfer coefficient between the absorber and the air gap, where Nu_a is the average Nusselt number for the air gap computed as follows [17]:

$$\mathrm{Nu}_a = \max\left\{0, \left(1 + 1.44\left[1 - \frac{1708\left[\sin\left(1.8\gamma\right)\right]^{1.6}}{\mathrm{Ra}\cos\left(\gamma\right)}\right]\left(1 - \frac{1708}{\mathrm{Ra}\cos\left(\gamma\right)}\right) + \left[\left(\frac{\mathrm{Ra}\cos\left(\gamma\right)}{5830}\right)^{0.33} - 1\right]\right)\right\}, \tag{4}$$

γ is the collector inclination angle and $\mathrm{Ra} = g\beta\Delta TL^3/\nu\alpha$ is the Rayleigh number. α, λ_a, δ_a, A_{ab}, ε_g and A_g are the absorption coefficient, the thermal conductivity, the air gap thickness, the absorber area, and the emissivity of the cover glass and its area, respectively.

$h_{c_{g-am}} = \mathrm{Nu}_{am}\lambda_{am}/\delta$ is the convective heat transfer coefficient between the glass and the ambient. Nu_{am} is the average Nusselt number for the ambient defined as follows [18]:

$$\mathrm{Nu}_{am} = 0.86\,\mathrm{Re}_{am}^{0.5}\mathrm{Pr}_{am}^{0.33}. \tag{5}$$

Re_{am} is the ambient Reynolds number, Pr_{am} is the Prandtl number of the ambient air. G is the heat flux of solar radiation, p is the tube pitch, Δz is the spatial size of control volume, σ is the Stefan-Boltzmann constant, ε_g is the emissivity of the glass, and $T_{sky} = 0.0552T_{am}^{1.5}$ [19] is the sky temperature. The energy balance on an elemental volume of the air gap gives the following:

$$C_a \rho_a V_a \frac{dT_a}{dt} = \left[h_{c_{g-a}}\left(T_g - T_a\right) + h_{c_{ab-a}}\left(T_{ab} - T_a\right)\right] p.\Delta z, \tag{6}$$

where C_a, ρ_a, V_a, T_a, T_{ab} are, respectively, the specific heat of the air, its density, its volume, its temperature, and the absorber temperature.

The energy balance equation for the absorber is written as follows:

$$C_{ab}\rho_{ab}V_{ab}\frac{dT_{ab}}{dt} = \left[\tau\alpha G + h_{r_{ab-g}}\left(T_g - T_{ab}\right) + h_{c_{g-a}}\left(T_a - T_{ab}\right)\right.$$
$$\left. + \frac{\lambda_{is}}{\delta_{is}}\left(T_{is} - T_{ab}\right)\right]p.\Delta z$$
$$+ \pi d_{in}h_f\Delta z\left(T_f - T_{ab}\right), \tag{7}$$

where $h_{c_{ab-a}} = h_{c_{g-a}}$ and $h_f = \mathrm{Nu}_f \lambda_f/d_{in}$ is a transfer on the internal surface of the collector tube. The empirical formula of Heaton et al. [20] using the assumption that the flow is fully developed and is used as follows:

$$\mathrm{Nu_f} = 4.4 + \frac{0.00236 \left(\mathrm{Re_f Pr_f}\left(d_{\mathrm{in}}/L\right)\right)^{1.66}}{1 + 0.00857 \left(\mathrm{Re_f Pr_f}\left(d_{\mathrm{in}}/L\right)\right)^{1.13}}; \quad 1 < \mathrm{Re_f Pr_f d_{in}}/L \le 1000. \tag{8}$$

λ_{f} is the fluid thermal conductivity, d_{in} is the inner diameter of the collector tube, L is the length of the analyzed collector, λ_{is} is the thermal conductivity of insulation, δ_{is} is its thickness, σ is Stefan-Boltzmann constant, $\varepsilon_{\mathrm{ab}}$ is the emissivity on absorber, and ε_{g} is the glass emissivity.

The energy balance equation for the insulation:

$$C_{\mathrm{is}}\rho_{\mathrm{is}}V_{\mathrm{is}}\frac{dT_{\mathrm{is}}}{dt} = \frac{\lambda_{\mathrm{is}}}{\delta_{\mathrm{is}}}\left(T_{\mathrm{ab}} - T_{\mathrm{is}}\right) + \left(h_{r_{\mathrm{is-am}}} + h_{c_{\mathrm{is-am}}}\right)\left(T_{\mathrm{am}} - T_{\mathrm{is}}\right), \tag{9}$$

where C_{is} is the specific heat of the insulation, ρ_{is} is its density, V_{is} is its volume, and T_{is} is the insulation temperature.

The radiation transfer coefficient between the insulation and the ambient is computed as follows:

$$h_{r_{\mathrm{is-am}}} = \frac{\sigma \varepsilon_{\mathrm{is}}\left(T_{\mathrm{is}}^4 - T_{\mathrm{sky}}^4\right)}{T_{\mathrm{is}} - T_{\mathrm{am}}}. \tag{10}$$

The energy balance equation for working fluid (water):

$$C_{\mathrm{f}}\rho_{\mathrm{f}}p.\Delta z\frac{dT_{\mathrm{f}}}{dt} = \pi d_{\mathrm{in}}h_{\mathrm{f}}\left(T_{\mathrm{ab}} - T_{\mathrm{f}}\right) - \dot{m}_{\mathrm{f}}c_{\mathrm{f}}\frac{\partial T_{\mathrm{f}}}{\partial z}, \tag{11}$$

where p, Δz, d_{in} and $h_{\mathrm{f}} = \mathrm{Nu_f}\lambda_{\mathrm{f}}/d_{\mathrm{in}}$ are the tube pitch, the spatial size of the control volume, the inner diameter of the collector tube, and the heat transfer on the internal surface of the collector tube, respectively. Based on the above equations, we can calculate the thermal efficiency of the collector as follows:

$$\eta_{\mathrm{en}} = \frac{\dot{Q}_{\mathrm{u}}}{G} = \frac{\dot{m}C_{\mathrm{p}}\left(T_{\mathrm{out}} - T_{\mathrm{in}}\right)}{G}, \tag{12}$$

where $\dot{Q}_{\mathrm{u}}$ is the useful heat rate to the fluid. The exergy efficiency of the solar collector measures the deviation of the real process from the ideal one. It is given as follows [3]:

$$\eta_{\mathrm{ex}} = \frac{\dot{Ex}_{\mathrm{out,f}} - \dot{Ex}_{\mathrm{in,f}}}{\dot{Ex}_{\mathrm{s}}} = \frac{g\left(T_{\mathrm{out}}, T_{\mathrm{in}}, T_{\mathrm{am}}\right)}{f\left(T_{\mathrm{s}}, T_{\mathrm{am}}\right)}\left(\frac{\dot{Q}_{\mathrm{u}}}{\dot{Q}_{\mathrm{s}}}\right), \tag{13}$$

where g and f are computed using the Farahat's proposal [4], which assumes zero pressure drop in the collector. In equation (13), the solar exergy is obtained using the Jeter expression [21] and $T_{\mathrm{s}} = 5770\,\mathrm{K}$ is the apparent black body temperature of the sun, where T_{out} is the output temperature, T_{in} is the input temperature, and T_{am} is the ambient one and where the two functions $f = \dot{Ex}_{\mathrm{s}}/\dot{Q}_{\mathrm{s}}$ and $g = (\dot{Ex}_{\mathrm{out,f}} - \dot{Ex}_{\mathrm{in,f}})/\dot{Q}_{\mathrm{u}}$ express the ratio of the solar radiation exergy rate to the solar radiation heat rate on the collector and the ratio of the exergy increase of the fluid to the useful fluid heat rate, respectively. $\dot{Ex}_{\mathrm{s}}$ is the solar radiation exergy and $\dot{Q}_{\mathrm{s}}$ is the solar radiation heat rate on the collector.

In order to evaluate the effects of the uncertainty of the experimental measurement equipment on the precision of the energy and exergy efficiencies, the theoretical uncertainty of the functioning parameters, F, was evaluated assuming that it depends on the operating parameters $\{x_1, x_2, \ldots, x_n\}$ and according to the following relationship:

$$u_{\mathrm{F}}^2 = \left(\frac{\partial F}{\partial x_1}u_{x_1}\right)^2 + \left(\frac{\partial F}{\partial x_2}u_{x_2}\right)^2 + \cdots + \left(\frac{\partial F}{\partial x_n}u_{x_n}\right)^2, \tag{14}$$

where u_{x_1}, u_{x_2}, u_{x_n} are the independent uncertainties of parameters $\{x_1, x_2, \ldots, x_n\}$. The derivative of F is computed by using first-order finite difference approximation as follows:

$$\frac{\partial F}{\partial x_i} = \frac{F\left(x_i + \Delta\right) - F\left(x_i - \Delta\right)}{2\Delta} + \mathcal{O}\left(\Delta\right), \tag{15}$$

where Δ is the standard deviation of the value of the operating parameter x_i.

3. Model Validation

3.1. Validation with Existing Literature Data. Numerical simulation enables a better understanding of the physical phenomena involved in heating the solar collectors' working fluid and also assists in the proper sizing of thermal systems. It also provides an advantage over experimental testing as it is a lower cost tool as there is no need for sophisticated instruments and prototypes. Saleh [15] developed a mathematical model for simulating the transient behavior of a flat plate solar collector. The mathematical model was solved numerically by using an implicit scheme. He validated his model experimentally and found that the calculation scheme is useful for transient simulations of energy systems. Hilmer et al. [22] developed a model of first-order partial differential equations in order to calculate the short-term dynamic behavior of solar collectors for various mass flow rates. This model is applied for a large unglazed collector for heating a public outdoor pool in Marburg (Germany). Their results showed that the simple single-capacity model gives a high accuracy even for very wide flow rates by using average hourly values of the useful energy gain. They showed that in a steady state with a constant flow rate, the model gives better results.

Results of our computational tool were compared to the values collected from both the numerical and experimental results of Saleh [14] according to the operating parameters, specifically for the steady state tests. The flowchart of the code solves the mathematical model equations ((1)–(5)) of the solar collector for each operating configuration involved in the heat transfer process. Each component of the system is simulated using the same weather data as the experimental test. The parameters of the experimental system, as well as initial conditions, are entered as input in the simulation program. In the first step, the preprocessing provides a database with the physical properties of the water and air relevant to the problem, such as thermal conductivity, specific mass, and other parameters. The implementation also involves physical heat transfer and turbulence models. Therefore, the input data includes the total water flow, the solar radiation, and the ambient temperature. After the

stabilisation of system temperatures , the simulation program provides a zero solar radiation flow over a period of time to simulate the cooling of the experimental model of reservoirs disconnected from solar collectors. Numerical results of Saleh [14] were found by using an adapted numerical model for a flat solar collector and performed by comparing the input and output temperatures with experimental results. Data collection was recorded every minute for inlet, outlet, and ambient temperatures. In the case of solar radiation, an average was recorded every 15 minutes. To verify the sensitivity of the mathematical model, when subjected to heating and then thereafter, the same input conditions were considered, such as dimensions and materials of the collector. Operating conditions were as follows: Solar Radiation: 800 W/m^2 constant for 1 hour; then zero radiation, ambient temperature is 20°C, the collector inclination is 45°, the time interval is 120 minutes and the volumetric flow is taken 1.5 GPM.

In order to verify if results of the numerical model of the proposed model behaved according to results of the reference, the same collector was considered, that is, the flat type with glass cover, for water heating applications, with an external area of 2 m^2. The collector is lined with aluminum frame and 5″ thick rock wool insulation material at the bottom. Looking at Figure 1, we can see a comparison between the inlet and outlet temperature curves of numerical and experimental results obtained by Saleh [14] as well astemperature curves of numerical results obtained by the author. It is possible to verify a strong correlation between results.

3.2. Validation with Tested Model in the LPT Laboratory.
The second step in the validation process is the comparison of our results with those obtained experimentally in our laboratory. The FPC under investigation was tested in the Research and Technology Centre of Energy (CRTEn), Borj-Cedria. It is, as displayed in Figure 2, detailed as follows: The collector has overall dimensions of $2.004 \times 1.004 \times 0.95$ m^3, a surface area of 2.12 m^2 and the capacity of the absorber is 2.85 l. The painting is highly selective with an absorption coefficient 95.3% and an emission factor 4.7%. The solar collector was fixed so that the angle of incidence of the radiation is 45°. The solar collector test bench consists of the following:

(1) Heat pump to set the solar collector water inlet temperature.

(2) An anemometer to measure the wind speed.

(3) A flow meter for measuring the flow rate inside the sensor.

(4) A microcomputer for the acquisition of data as an Excel file.

During the test, the global solar radiation is measured by the calibrated first class pyranometer (Kipp & Zonen) and the ambient temperature so that the probe was placed in a protected and ventilated cage placed at the height of 1.25 m from the surface of the trial. Finally, PT100 probes for
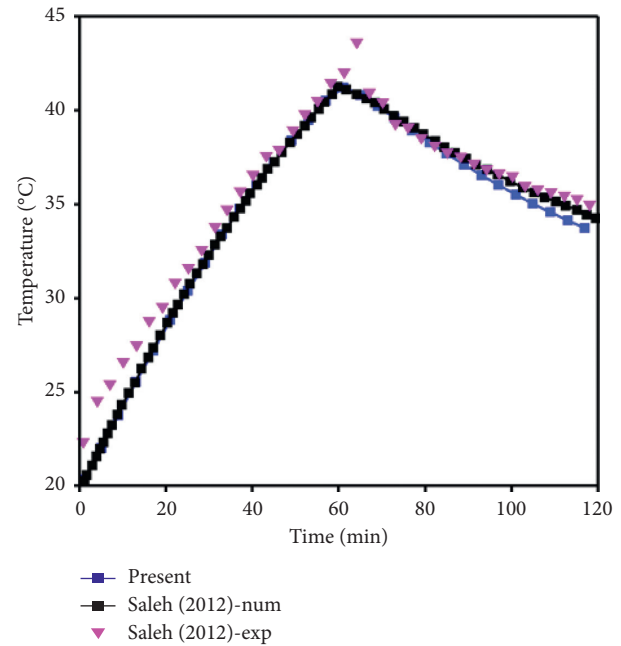
FIGURE 1: First step of the code validation: Comparison with results of Saleh [14].

measuring different temperatures (input, output, and ambient) were used. A multichannel digital data acquisition unit (type Agilent HP34970 A) was connected to the FPSC device to collect the data obtained during measurements. For the numerical investigation, ASHRAE data is used to obtain thermophysical properties of water.

In order to show the precision of the experimental investigation, we summarize in Table 1 measurement errors attributed to the uncertainly of measurements. In fact, uncertainties are mainly attributed to errors occurring during measurement and the errors of the measurement apparatus, the sensitiveness of the data acquisition system, and Pt100 temperature sensors.

Figure 3 is a representation of the instantaneous thermal efficiency curve of the solar collector η_{en} as a function of the reduced temperature $T_r = (T_m - T_a)/G$ where $T_m = (T_{out} + T_{in})/2$ and T_{in} is the inlet temperature. As it is predicted by a linear equation $\eta_{en} = \eta_0 - U_L T_r$, where U_L is the heat loss coefficient, η_{en} is almost linear, and slopes of curves are about −5.44 and -5.06 obtained numerically and experimentally, respectively. If we take into consideration the sensitivity of the equipment and the error occurring during the experimental investigation presented by error bars in Figure 2, we can affirm the accuracy of the proposed numerical model. Hence, the simulation program could reproduce with acceptable accuracy the real behavior of the FPC system.

Table 2 summarizes results of the measured ambient T_a, the solar radiation G, the inlet temperature, and the rate of temperature increase $\Delta T = T_{out} - T_{in}$. The table also shows a comparison between computed values of energy and exergy efficiencies η_{en} and η_{ex} obtained numerically and experimentally. In general, results are encouraging and the
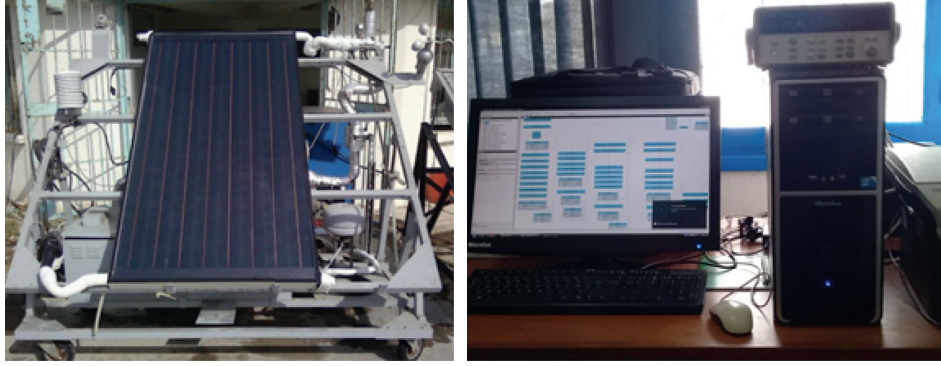
Figure 2: Experiment setup (tested FPSC and data acquisition toolbox).

Table 1: Uncertainty of sensors at a typical test condition.

Parameters	Units	Values
Water temperature sensor	$^\circ$C	± 0.05
Ambient air temperature	$^\circ$C	± 0.1
Flow meter	Kg/s	± 0.01
Pyranometer	W/m^2	± 50
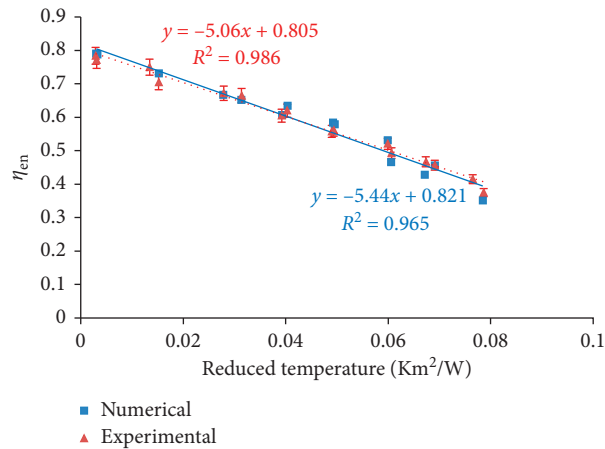
Figure 3: Present numerical and experimental results of the tested collector in CRTEn.

Table 2: Performance test according to EN 12975 obtained numerically (thin line) and experimentally (bold line).

Experimental time	T_a ($^\circ$C)	G (W/m^2)	T_{in} ($^\circ$C)	ΔT ($^\circ$C)	η_{en} (%)	η_{ex} (%)
10 : 34 : 34	18.14	820.77	65.61	4.41–**4.60**	46.61–**49.3**	7.17–**7.58**
11 : 18 : 43	19.51	874.17	75.99	4.32–**4.75**	42.88–**46.7**	7.56–**8.24**
11 : 57 : 53	20.57	856.54	86.03	3.47–**3.79**	35.21–**37.5**	6.92–**7.37**
12 : 44 : 21	18.24	874.90	17.09	7.93–**7.73**	78.73–**77.1**	0.83–**0.75**
12 : 05 : 12	18.80	866.86	17.48	7.89–**7.66**	79.02–**78.4**	0.78–**0.70**
12 : 34 : 54	16.12	929.51	26.37	7.83–**7.80**	73.09–**70.5**	3.63–**3.46**
13 : 19 : 33	17.63	949.10	40.33	7.29–**7.59**	66.72–**67.2**	5.87–**5.91**
11 : 39 : 44	17.38	890.23	50.02	6.51–**6.46**	63.48–**62.0**	7.38–**7.18**
12 : 14 : 52	17.66	924.78	50.76	6.47–**6.32**	60.72–**60.5**	7.13–**7.06**
11 : 35 : 56	19.02	910.34	60.73	6.13–**5.79**	58.44–**55.8**	8.21–**7.79**
12 : 46 : 42	17.47	847.65	65.67	5.18–**5.18**	53.10–**52.1**	8.34–**18.16**
13 : 19 : 07	16.07	852.37	39.59	6.40–**6.46**	65.21–**66.5**	5.84–**5.93**
13 : 59 : 14	17.75	929.24	60.68	6.20–**5.86**	57.97–**56.2**	8.37–**8.06**
10 : 57 : 02	17.35	877.09	75.61	4.61–**4.70**	45.59–**45.7**	8.30–**8.31**

proposed program was shown to be a reliable method for evaluating both energy and exergy performances of the FPC. The slight discrepancy between η_{en} and η_{ex} obtained by using the experimental set-up and simulation program can be explained by errors due to the misplacing of the temperature sensor or a faulty connection wire is suspected. Other reasons for the discrepancy are the used 1D mathematical model to compute the temperatures at each node and the simplified theoretical formula of the exergy efficiency, which does not take into account the pressure drops in the collector. The table shows that by increasing ΔT, the energy efficiency decreases while the exergetic efficiency increases. When ΔT is high, the useful exergy rate decreases quickly.

4. Results and Discussion

Figure 4 shows the ±50 W/m^2 of uncertainty in solar irradiation effects on the thermal efficiency of the tested collector. Note that 100 W is the standard uncertainty of the major calibrated first class pyranometer which is required by the EN 12975. As displayed in the figure, the uncertainty in the irradiance has an effect on both reduced temperature's results and thermal efficiency's values. More importantly, a more noticeable influence on the reduced temperature than on the efficiency of the collector is observed. More precisely, an uncertainty of ±50 W/m^2 in G affects the reduced temperature about ±0.004 Km^2/W. This uncertainty produces a deviation in efficiency results of about 5.5%.

Figure 5 shows the effects of $\pm1°$C uncertainty in thermocouples measuring the ambient temperature on the thermal performance of the collector. It is clear from the figure that when the reduced temperature is small, i.e., the collector is at near-ambient temperature, the 2°C of uncertainty in the ambient temperature has no significant effect on the thermal efficiency. When the reduced temperature increases, the effect of the uncertainty appears on both Tr and η_{en}. The figure also shows that when the collector is at far away-ambient temperature, uncertainty in the ambient temperature has an important effect on the result of the reduced temperature which decreases substantially. Since the reduced temperature gives rise to the difference between the mean collector and the ambient temperature, one can explain the accentuated effect of the ambient temperature on T_r. The deviation reaches 0.033 Km^2/W and 9.9% in T_r and η_{en}, respectively. Therefore, the collector efficiency is significantly influenced by the ambient temperature. This trend of the results shows that both uncertainties measurements in the ambient temperature must be taken into account as for regression parameters [11].

Figure 6 shows the effect of $\pm1°$C of uncertainty when using Pt100 Thermometer at the inlet of the collector on its thermal performance. As shown in the figure, the 2°C of uncertainty in the inlet temperature has a slight effect on the result of the reduced temperature but has no effect on the thermal efficiency. In fact, this result is expected because the inlet temperature has no requirement in the test standard. More importantly, the uncertainty of $\pm1°$C in T_{in} affects the reduced temperature by ±0.001 Km^2/W.
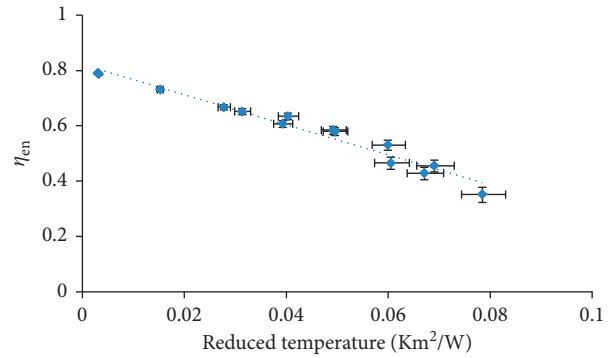
Figure 4: ±50 W/m^2 uncertainty in irradiation effects on the thermal performance of the collector.
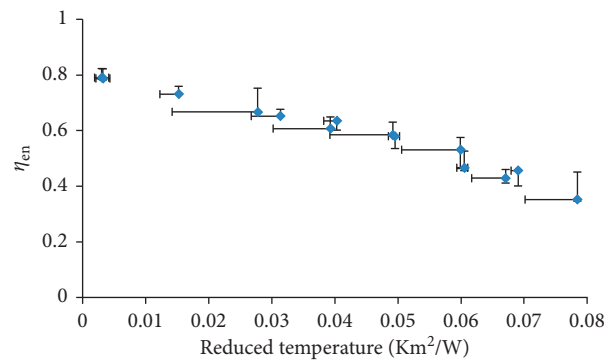
Figure 5: $\pm1°$C of deviation in ambient temperature effects on the thermal performance of the collector.
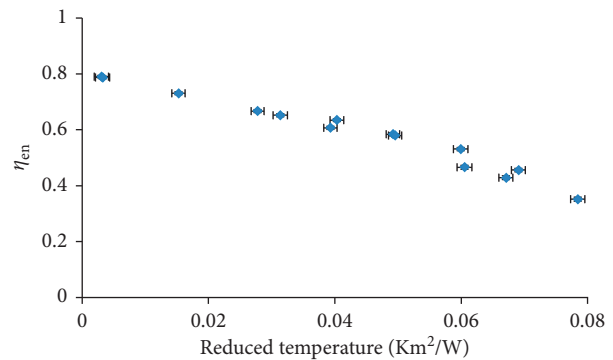
Figure 6: Effect of $\pm1°$C of uncertainty in temperature inlet effects on the thermal performance of the collector.

The flow rate uncertainty effect of $\pm10\%$ on thermal efficiency is shown in Figure 7. As displayed in the figure, the gap between measured and mean flow rate values has no effect on the η_{en} and it has a slight effect on the reduced temperature. When the deviation in mass flow rate is about $\pm10\%$, the uncertainty in T_r decreases about 0.002 Km^2/W.

Figure 8 shows the effect of a deviation of about $\pm10\%$ in flow rate on the collector performance for a test sample. According to the figure, the uncertainty in the flow rate has a small effect on the temperatures of the absorber, the air gap, and the glass. Importantly, these deviations are about 0.9°C,

FIGURE 7: Effect of ±10% flow rate uncertainty on the thermal efficiency.

FIGURE 8: Effect of ±10% deviation in the flow rate on the temperature histories for all the analyzed cross sections of the collector and its power output.

TABLE 3: Effects of uncertainties in input parameters on results of regression analysis.

Parameter	η_0	a_1 (W/m²K)	a_2 (W/m²K²)	R^2
Mean value	0.792	3.064	0.034	0.980
$G \pm 50$ W/m²	±0.5%	±0.01%	±5.93%	0.979–0.981
$T_{am} \pm 1$°C	±4.1%	±84.79%	64.31%	0.976–0.999
$T_{in} \pm 1$°C	±0.2%	±1.86%	±0.62%	0.980

and 2.6 Wm^{-2}K^{-2} for η_0 and a_1. This trend of the results means that regression uncertainties do not have the same order of relevance as ambient temperature uncertainty.

Figure 9 reports the variation of the exergy efficiency versus the inlet temperature. The figure shows that the exergy efficiency increases with the inlet temperature. When the water is heated from 17.48°C to 26.37°C, the exergy increases by 2.8%. The highest exergy observed is equal to 8.3% for $T_{in} = 75.61$°C. It is important to note that an optimum in exergy rate is observed by Ge et al. [23] and Farahat et al. [4].

Figure 10 shows the contribution of several operating parameters on the deviation in instantaneous energy efficiency values. The effect of uncertainty in $\dot{m}$ and T_{in} is around 1.09% and 1.42%, respectively. However, the ambient temperature was found to be the most influencing parameter, and its effect reaches 13.7%. The second important parameter is the solar irradiance, and its value may affect the energy efficiency by 5.47%.

Figure 11 shows the contribution of several operating parameters on the deviation in exergy efficiency. The effect of uncertainty in $\dot{m}$ and T_{in} is around 0.34% and 0.53%, respectively. However, the ambient temperature was found to be the most influencing parameter and its effect reaches 3.89%. The second important parameter is the solar irradiance and its value may affect the exergy efficiency by 1.11%.

The aim of this section is the analysis of the effect of a deviation in thermophysical properties of the working fluid on energetic and exergetic efficiencies of the collector. Therefore, we consider a new class of heat transfer fluids composed of metal nanometer-sized conductive particles dispersed in the base fluid often called nanofluid. The majority of published works highlight that classical correlations are not able to predict the superior characteristics of convective heat transfer of nanofluids. Figure 12 shows the efficiency of the collector with ethylene-glycol mixture (E-G) (30 : 70) with and without MCNT nanoparticles at 0.15v% versus the reduced temperature parameters. Experimental conditions are kept the same as the previous section and thermophysical properties data of E-G + MCNT nanofluid are given by Kumaresan and Velraj [24]. These properties depend on both the volume fraction and the temperature. The figure clearly shows that the efficiency of the collector using E-G + MCNT nanofluid is highest and the η_0 values when using pure E-G are the lowest. This result agrees well with previous investigations found in the literature [25, 26]. The η_0 values for E-G with and without MCNT nanoparticles are 84.7% and 75.4%, respectively. Therefore adding nanoparticles enhances energy parameter η_0 value by about 9.3%.

0.8°C, and 0.6°C for T_{ab}, T_a, and T_g, respectively. A remarkable effect of uncertainty is observed in outlet temperature, which deviates by 1.5°C. The effect is then accentuated in the power output of the collector which deviates by 6.8 W.

It is important to note that apart from uncertainties in operating parameters, regression uncertainties should be taken into account and should not be neglected, as confirmed by Facão and Oliveira [7] and Sowmy et al. [8]. Table 3 reports the effect of measurement errors in several operating parameters on regression coefficients η_0 , a_1 and a_2. The table reveals that the estimation in conversion factor mean value η_0 (efficiency for $T_m = T_{am}$) is affected by 1%, 8.2%, and 0.4% when there is an uncertainty in G, T_{am}, and T_{in}. The heat loss coefficient a_1 is shifted by ±84.79% when the uncertainty of measurement is about ±1°C. This trend of results means that other operating parameters do not have the same order of relevance as ambient temperature. With the same uncertainty, the heat loss coefficient dependent on temperature a_2 is moved by 0.022 Wm^{-2}K^{-2}. Mean uncertainties for linear analysis procedure were, respectively, 4.1%

Figure 9: Variation of the exergy efficiency versus the inlet temperature.
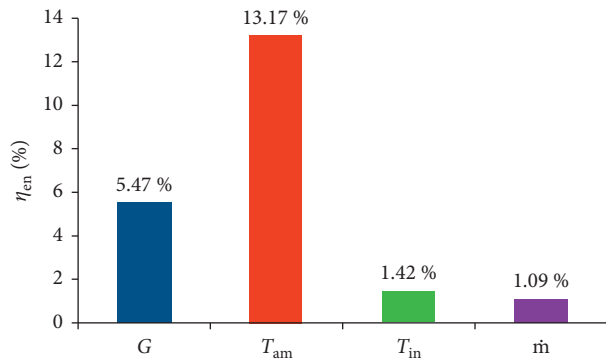
Figure 10: Change in energy efficiency per 10% in parameter values for a mean value.
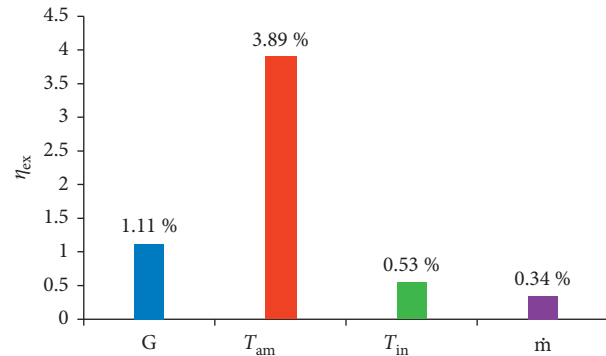
Figure 11: Change in exergy efficiency per 10% in operating parameter values for a mean value.

In order to reveal the effect and the importance of the determination of thermophysical properties of the working fluid on the thermal performance of the collector, numerical investigations were carried out as follows:

(1) Investigation I: The working fluid is the Ethylene-Glycol.

(2) Investigation II: The working fluid is at 0.15v% Carbon Nanotubes + Ethylene-Glycol (30 : 70). Thermophysical properties are extracted from data of Kumaresan and Velraj [24]. Other investigations are performed using the same nanofluid as follows:
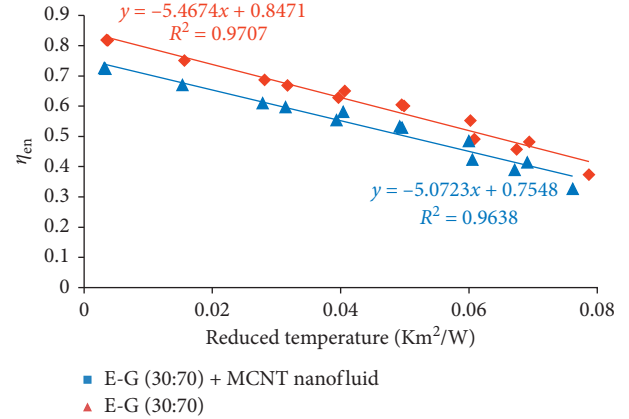
Figure 12: Efficiency of the collector with ethylene-glycol mixture (30 : 70) with and without CNT nanoparticles at 0.15v%.

(i) Investigation II-a: Only λ_f is computed by using Maxwell's model as follows [27]:

$$\lambda_f = \lambda_{bf} \frac{(\lambda_s + 2\lambda_{EG}) - 2\varphi(\lambda_{EG} - \lambda_s)}{(\lambda_s + 2\lambda_{EG}) + \varphi(\lambda_{EG} - \lambda_s)}, \tag{16}$$

where $\varphi = 0.15\%$ is volume fraction, λ_s and λ_{EG} are the thermal conductivities of the MWCNT nanoparticles and the E-G fluid, respectively.

(ii) Investigation II-b: Only μ_f is computed by using Brinkman's model [28] by using the following equation:

$$\mu_f = (1 - \varphi)^{-2.5} \mu_{EG}, \tag{17}$$

where μ_{EG} is the E-G dynamic viscosity.

(iii) Investigation II-c: C_f is computed by using Xuan and Roetzel's model which is claimed to be the fittest for getting specific heat of nanofluid [29, 30]:

$$\left(\rho C_p\right)_f = (1 - \varphi)\left(\rho C_p\right)_{EG} + \varphi\left(\rho C_p\right)_s. \tag{18}$$

(iv) Investigation II-d: μ_f, λ_f, and C_p are computed from prescribed theoretical correlations and combined together.

Figure 13 presents the effect of correlation determining the viscosity μ_f, the thermal conductivity λ_f and the specific heat C_p of nanofluid on the thermal efficiency of the collector. Correlations used are Maxwell's model, Brinkman's model, and Xuan and Roetzel's model for μ_f, λ_f, and C_p, respectively. Reference characteristic parameters of the flat plate solar collector use thermophysical properties of E-G + MCNT nanofluid published by Kumaresan and Velraj [24]. The figure indicates that when we use Brinkman's model for the viscosity, η_0 increases about 6.3%. By using Maxwell's model for the thermal conductivity, η_0 increases about 1.4% and the specific heat formulae of the nanofluid
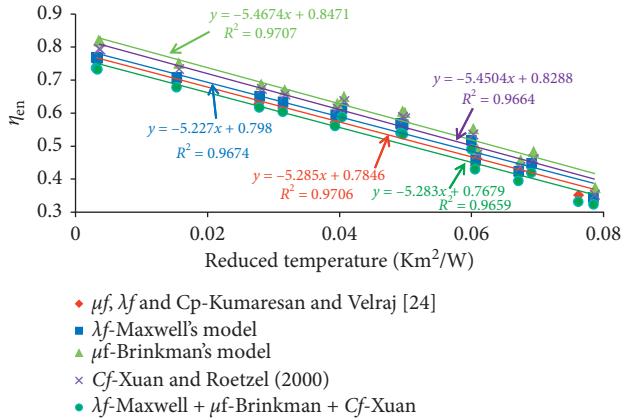
FIGURE 13: Effect of correlation determining thermophysical properties of nanofluid on the thermal efficiency of the collector.

deviate the thermal efficiency result by about 4.4%. This trend of results reveals the importance of the determination of thermophysical properties of the nanofluid (and hence the preparation step of the nanofluid) for the accuracy of fitting parameters.

As expected, the specific heat decreased when using the correlation prescribed while thermal conductivity and viscosity increased. The increase in viscosity is due to the fact that nanoparticles are at low concentrations (0.15%). The increase in viscosity has exceeded the improvement of the thermal conductivity which is manifested in the figure by the remarkable effect of viscosity compared to other thermophysical parameters effects. This also explains some inconsistencies among experimental evaluations of the thermal performance of nanofluids in thermal systems [31, 32]. Especially when the uncertainty in the heat specific capacity of the nanofluid increases hugely, the nanofluid improves the outlet temperature without increasing the efficiency.

Figure 14 shows the variation of the exergy efficiency versus the inlet temperature with E-G and E-G + MCNT nanofluid at 0.15v% using thermophysical properties found by Kumaresan and Velraj [24] and by using Maxwell's model, Brinkman's model, and Xuan and Roetzel's correlation. By using prescribed theoretical correlations, the exergy efficiency takes values close to those of the collector in the absence of nanoparticles. By using values obtained experimentally by Kumaresan and Velraj [24], the exergy increases substantially. The maximum enhancement in exergy efficiency is obtained for $T_{in} = 75.76°C$ and is equal to 1.18%. This result reflects the fact that theoretical correlations of thermophysical properties have a great influence on heat transfer characteristics and hence on the energy and exergy efficiency of the collector.

5. Conclusions

In this work, a 1D mathematical model was developed and implemented in order to assess the sensitivity of energy and exergy performances of a Flat Plate Solar Collector (FPSC) according to EN 12975 procedure. The proposed method allows quantifying effects of uncertainties of both operating
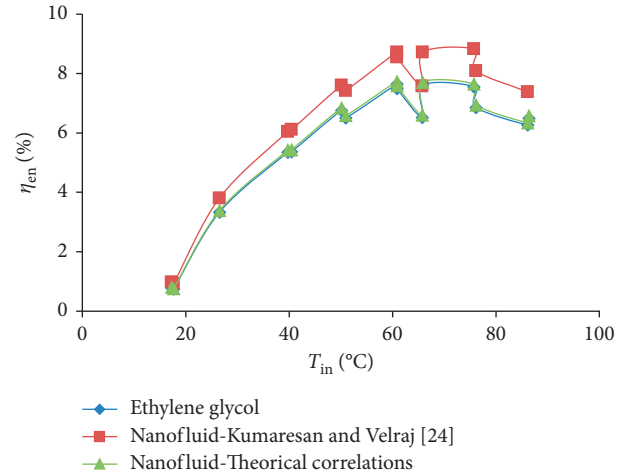
FIGURE 14: Variation of the exergy efficiency of the collector versus the inlet temperature with E-G and E-G + MCNT at 0.15v% using thermophysical properties obtained by Kumaresan and Velraj [24] and by using empirical correlations.

conditions and thermophysical properties of working fluid on the performance estimation of the collector. The main results of numerical and experimental analysis gave the following conclusions:

(1) Results of the computational program are in good agreement with experimental measurements and benchmark solutions available in literature in terms of collector efficiency curve and outlet temperature, respectively. The proposed program was shown to be a reliable method for understanding major measurement parameter sensitivities in efficiencies and their uncertainties.

(2) During the test procedure, the ambient temperature is found to be the most influencing parameter on the performance estimation and a deviation about ±1°C may affect the energy efficiency by 13.7% and the exergy efficiency by 3.9%.

(3) When 0.15v% MWCNT-E-G (30 : 70) nanofluid is used as working fluid, uncertainties in its thermophysical properties lead to important deviations in both energy and exergy efficiencies, and results show that common correlations are not able to predict the collector's performance neither in terms of energy nor in exergy. Especially, the deviation in exergy efficiency reaches 1.18%.

Nomenclature

C: Specific heat (J/(kg·K))
ρ: Density (kg/m^3)
V: Volume (m^3)
T: Temperature (K)
h: Heat transfer coefficient (W/(m^2·K))
α: Absorption coefficient
G: Heat flux of solar radiation (W/m^2)
p: Tube pitch (m)
Δz: Spatial size of control volume (m)

σ: Stefan–Boltzmann constant (W/(m^2 K^4))
ε: Emissivity
F: View factor
A: Pipe cross sectional area (m^2)
Nu: Nusselt number
λ: Thermal conductivity (W/(m·K))
δ: Thickness (m)
γ: Collector inclination angle (rad)
Ra: Rayleigh number
g: Gravity acceleration (m/s^2)
β: Thermal expansion coefficient
ΔT: Time step (s)
L: Length of the analyzed collector (m)
ν: Kinematic viscosity (m^2/s)
α: Thermal diffusivity (m^2/s)
Re: Reynolds number
Pr: Prandtl number
t: Time (s)
$(\tau\alpha)$: Effective transmittance-absorption coefficient
d: Diameter of the collector tube (m)
$\dot{m}$: Mass flow rate (Kg/s)
η: Thermal efficiency
am: Ambient
g: Glass cover
a: Air gap between glass cover and absorber
ab: Absorber
r: Reduced, radiation
c: Convection
is: Insulation
f: Fluid
in: Inlet
out: Outlet
s: Sun
en: Energetic
ex: Exergetic
sky: Sky
u: Useful
m: Mean.

Data Availability

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

Disclosure

The research was performed as part of the PhD thesis of Intissar Harrabi and the employment of Mohamed Hamdi in the Research and Technology Centre of Energy of Borj-Cedria, Tunisia.

Conflicts of Interest

The authors declare that there are no conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] M. B. A. Sghari and S. Hammami, "Energy, pollution, and economic development in Tunisia," *Energy Reports*, vol. 2, pp. 35–39, 2016.

[2] A. Ege and H. M. Şahin, "Determination of uncertainties in energy and exergy analysis of a power plant," *Energy Conversion and Management*, vol. 85, pp. 399–406, 2014.

[3] S. A. Kalogirou, S. Karellas, K. Braimakis, C. Stanciu, and V. Badescu, "Exergy analysis of solar thermal collectors and processes," *Progress in Energy and Combustion Science*, vol. 56, pp. 106–137, 2016.

[4] S. Farahat, F. Sarhaddi, and H. Ajam, "Exergetic optimization of flat plate solar collectors," *Renewable Energy*, vol. 34, no. 4, pp. 1169–1174, 2009.

[5] R. Tang, Z. Li, H. Zhong, and Q. Lan, "Assessment of uncertainty in mean heat loss coefficient of all glass evacuated solar collector tube testing," *Energy Conversion and Management*, vol. 47, no. 1, pp. 60–67, 2006.

[6] R. Kicsiny, "Multiple linear regression based model for solar collectors," *Solar Energy*, vol. 110, pp. 496–506, 2014.

[7] J. Facão and A. C. Oliveira, "Experimental uncertainty analysis in solar collectors," *International Journal of Ambient Energy*, vol. 27, no. 2, pp. 59–64, 2006.

[8] D. S. Sowmy, P. J. Schiavon Ara, and R. T. A. Prado, "Uncertainties associated with solar collector efficiency test using an artificial solar simulator," *Renewable Energy*, vol. 108, pp. 644–651, 2017.

[9] E. Mathioulakis, K. Voropoulos, and V. Belessiotis, "Assessment of uncertainty in solar collector modeling and testing," *Solar Energy*, vol. 66, no. 5, pp. 337–347, 1999.

[10] V. Sabatelli, D. Marano, G. Braccio, and V. K. Sharma, "Efficiency test of solar collectors: uncertainty in the estimation of regression parameters and sensitivity analysis," *Energy Conversion and Management*, vol. 43, no. 17, pp. 2287–2295, 2002.

[11] N. Rehman, M. Uzair, M. A. Siddiqui, and M. Khamooshi, "Regression models and sensitivity analysis for the thermal performance of solar flat-plate collectors," *Arabian Journal for Science and Engineering*, vol. 44, no. 2, pp. 1119–1127, 2019.

[12] R. Karwa and S. Baghel, "Effect of parametric uncertainties, variations, and tolerances on thermohydraulic performance of flat plate solar air heater," *Journal of Solar Energy*, vol. 2014, pp. 1–18, 2014.

[13] W. Zima and P. Dziewa, "Modelling of liquid flat-plate solar collector operation in transient states," *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, vol. 225, no. 1, pp. 53–62, 2011.

[14] A. M. Saleh, "Modeling of flat-plate solar collector operation in transient states," Master Thesis in Science Engineering, Purdue University Fort Wayne, Fort Wayne, IN, USA, 2012.

[15] A. Saleh, DW. Mueller, and H. I. Abu-Mulaweh, "Flat-Plate solar collector in transient operation: modeling and measurements," *Journal of Thermal Science and Engineering Applications*, vol. 7, no. 1, pp. 15–21, 2013.

[16] J. Cadafalch, "A detailed numerical model for flat-plate solar thermal devices," *Solar Energy*, vol. 83, no. 12, pp. 2157–2164, 2009.

[17] K. G. T. Hollands, T. E. Unny, G. D. Raithby, and L. Konicek, "Free convective heat transfer across inclined air layers," *Journal of Heat Transfer*, vol. 98, no. 2, pp. 189–193, 1976.

[18] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*, John Wiley & Sons, New York, NY, USA, 3rd edition, 2006.

[19] L. Adelard, F. Pignolet-Tardan, T. Mara, P. Lauret, F. Garde, and H. Boyer, "Sky temperature modelisation and applications in building simulation," *Renewable Energy*, vol. 15, no. 1-4, pp. 418–430, 1998.

[20] H. S. Heaton, W. C. Reynolds, and W. M. Kays, "Heat transfer in annular passages. Simultaneous development of velocity and temperature fields in laminar flow," *International Journal of Heat and Mass Transfer*, vol. 7, no. 7, pp. 763–781, 1964.

[21] S. M. Jeter, "Maximum conversion efficiency for the utilization of direct solar radiation," *Solar Energy*, vol. 26, no. 3, pp. 231–236, 1981.

[22] F. Hilmer, K. Vajen, A. Ratka, H. Ackermann, W. Fuhs, and O. Melsheimer, "Numerical solution and validation of a dynamic model of solar collectors working with varying fluid flow rate," *Solar Energy*, vol. 65, no. 5, pp. 305–321, 1999.

[23] Z. Ge, H. Wang, H. Wang, S. Zhang, and X. Guan, "Exergy analysis of flat plate solar collectors," *Entropy*, vol. 16, no. 5, pp. 2549–2567, 2014.

[24] V. Kumaresan and R. Velraj, "Experimental investigation of the thermo-physical properties of water-ethylene glycol mixture based CNT nanofluids," *Thermochimica Acta*, vol. 545, pp. 180–186, 2012.

[25] S. K. Verma, A. K. Tiwari, and D. S. Chauhan, "Experimental evaluation of flat plate solar collector using nanofluids," *Energy Conversion and Management*, vol. 134, pp. 103–115, 2017.

[26] M. Eltaweel and A. A. Abdel-Rehim, "Energy and exergy analysis of a thermosiphon and forced-circulation flat-plate solar collector using MWCNT/Water nanofluid," *Case Studies in Thermal Engineering*, vol. 14, Article ID 100416, 2019.

[27] J. C. Maxwell-Garnett, "Colours in metal glasses and in metallic films," *Philosophical Transactions of the Royal Society A*, vol. 203, pp. 385–420, 1904.

[28] H. C. Brinkman, "The viscosity of concentrated suspensions and solutions," *The Journal of Chemical Physics*, vol. 20, no. 4, p. 571, 1952.

[29] Y. Xuan and W. Roetzel, "Conceptions for heat transfer correlation of nanofluids," *International Journal of Heat and Mass Transfer*, vol. 43, no. 19, pp. 3701–3707, 2000.

[30] I. M. Shahrul, I. M. Mahbubul, S. S. Khaleduzzaman, R. Saidur, and M. F. M. Sabri, "A comparative review on the specific heat of nanofluids for energy perspective," *Renewable and Sustainable Energy Reviews*, vol. 38, pp. 88–98, 2014.

[31] R. Mondragón, D. Sánchez, R. Cabello, R. Llopis, and J. E. Juliá, "Flat plate solar collector performance using alumina nanofluids: experimental characterization and efficiency tests," *PLoS One*, vol. 14, no. 2, pp. 1–18, 2019.

[32] N. Purohit, S. Jakhar, P. Gullo, and M. S. Dasgupta, "Heat transfer and entropy generation analysis of alumina/water nanofluid in a flat plate PV/T collector under equal pumping power comparison criterion," *Renewable Energy*, vol. 120, pp. 14–22, 2018.

Research Article

GPS Data in Urban Online Car-Hailing: Simulation on Optimization and Prediction in Reducing Void Cruising Distance

Yuxuan Wang [iD],[1] **Jinyu Chen** [iD],[1] **Ning Xu** [iD],[2] **Wenjing Li** [iD],[1] **Qing Yu** [iD],[1,3] **and Xuan Song** [iD][1,4]

[1]*Center for Spatial Information Science, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa Chiba 277-8563, Japan*
[2]*Beijing Key Laboratory of Urban Oil and Gas Distribution Technology, China University of Petroleum-Beijing, Fuxue Road No. 18, Changping District, Beijing 102249, China*
[3]*Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Cao'an Road, Shanghai 201804, China*
[4]*SUSTech-UTokyo Joint Research Center on Super Smart City, Department of Computer Science and Engineering, Southern University of Science and Technology (SUSTech), Shenzhen, China*

Correspondence should be addressed to Xuan Song; songxuan@csis.u-tokyo.ac.jp

Ride-hailing, as a popular shared-transportation method, has been operated in many areas all over the world. Researchers conducted various researches based on global cases. They argued on whether car-hailing is an effective travel mode for emission reduction and drew different conclusions. The detailed emission performance of the ride-hailing system depends on the cases. Therefore, there is an urgent demand to reduce the overall picking up distance during the dispatch. In this study, we try to satisfy this demand by proposing an optimization method combined with a prediction model to minimize the global void cruising distance when solving the dispatch problem. We use Didi ride-hailing data on one day for simulation and found that our method can reduce the picking up distance by 7.51% compared with the baseline greedy algorithm. The proposed algorithm additionally makes the average waiting time of passengers more than 4 minutes shorter. The statistical results also show that the performance of our method is stable. Almost the metric in all cases can be kept in a low interval. What is more, we did a day-to-day comparison. We found that, despite the different spatial-temporal distribution of orders and drivers on different day conditions, there are little differences in the performance of the method. We also provide temporal analysis on the changing pattern of void cruising distance and quantity of orders on weekdays and weekends. Our findings show that our method can averagely reduce more void cruising distance when ride-hailing is active compared with the traditional greedy algorithm. The result also shows that the method can stably reduce void cruising distance by about 4000 to 5000 m per order across one day. We believe that our findings can improve deeper insight into the mechanism of the ride-hailing system and contribute to further studies.

1. Introduction

Ride-hailing, referring to the activity of calling a vehicle or driver to go to a destination, rises in many metropolises and takes a considerable part of mobility services [1]. The studies on the behavior of ride-hailing mount to a peak in recent years [2]. Some discussed the impact of ride-hailing on urban transportation behavior. Tang et al. [3] did an online app survey on 9762 respondents and found that about 35% of respondents are attracted from traditional taxis; 37% are from public transportation. Some tried to discover how ride-hailing influences urban sustainability. Li et al. [4] employed annual ride-hailing data in the United States from Uber Google researches and found that the usage of ride-hailing can help reduce traffic congestions in big cities. Some put efforts into revealing the carbon footprint of ride-hailing. Sui et al. [5] operated spatial analysis on empirical evidence in Chengdu area. They concluded that ride-hailing saves more energy and emits less GHG on per passenger on kilometer basis during the service compared with the traditional taxi industry.

Among all these studies, some scholars explore the efficiency of novel technologies in the ride-hailing system. Bischoff, Kaddoura [6] proposed an agent-based simulation method to optimize the service area of ride-hailing. Feng et al. [7] built a stylized model of a circular road and compared the average waiting times of passengers under different matching mechanisms of ride-hailing. Calderón and Miller [8] proposed a method for modeling the within-day service provision process of ride-hailing service providers with limited data availability. However, the efficiency of real-time scheduling should be further considered in practical application. Korolko et al. [9] indicated that bipartite matching with time window batching and dynamic pricing can lower waiting time for both riders and drivers as well as capacity utilization, trip throughput, and total welfare. However, they only consider the dispatching in the real-time time window and did not consider the travel demand in the future. What is more, Afeche et al. [10] pointed out that the interference from the service platform to avoid dispatching drivers to the area with low travel demand can be optimal. These two conclusions inspire us with the idea that whether we can predict the distribution of travel demands in the future and whether it can help us optimize the dispatching of the ride-hailing system, especially improving the utility of energy. Among current studies, there is no existing literature quantifying and discovering this improvement. This gap needs to be fulfilled and further instructs the development of the ride-hailing system.

Therefore, there are some gaps in current studies:

(1) Considering the real-time matching of driver and passenger with the combination of optimization method and prediction model to minimize the void cruising distance as well as maximizing the energy utility.

(2) The simulation of the method on real-world data to prove the performance and applicability as well as an analysis of the spatial-temporal pattern of emission behavior of two different dispatch strategies.

Operating such kind of research is not an easy task. Solid and real travel demand data are required as the basis of simulation and assessment. Next, a reliable prediction model is in a dominant position in the whole simulation as imprecise prediction can bring misjudge to dispatching decisions. In addition, the dispatching methods should be designed carefully and the performance and applicability should be ensured.

With the development of urban data mining [11, 12], the simulation method based on historical GPS data allows us to develop and analyze the performance of methods. This paper will overcome these difficulties by employing and combing empirical ride-hailing data from the real world, reliable prediction model, and dispatching simulation with optimization methods. Also, we will give a comprehensive analysis of the behavior of the system on improving the energy utility of ride-hailing. We believe that our work can provide the guideline for future decision-making and development of ride-hailing.

In this study, we simultaneously employ a performance-proved prediction model and optimization-based matching strategy as the methodology and apply it to the simulation based on millions of empirical ride-hailing data from the real world to quantify this improvement. Then, we will also provide a comprehensive spatial-temporal analysis. The contribution of this work can be listed as follows:

(1) Proposing a simulation framework combining the optimization and prediction to improve the efficiency of the ride-hailing system.

(2) A simulation sample on millions of Didi ride-hailing record data to provide persuasive evidence of the utility of methodology.

(3) A comprehensive spatial-temporal pattern analysis and comparison that will be provided.

2. Materials and Methods

2.1. Research Framework. Different dispatch programs cause different energy behaviors. Among different dispatch algorithms, optimization is a hot topic [13]. However, the performance of the optimization algorithm depends heavily on current knowledge. In real cases, time window batch is a common method for processing of matching pool in real-time dispatch [14]. The longer the time window is, the more optimal the solution is, but it is lesser applicable because it will cause longer waiting times for users. It is very hard to balance the performance of the solution and the length of the time window. Therefore, merging prediction in the future is a good choice to enrich the knowledge and improve the performance of optimization. In this study, we propose a new dispatch framework and compare the energy behavior of it with one of the original dispatch plans. We use the following framework in this study (see Figure 1).

The research framework used in this part is shown in Figure 1. The Didi apps in users' (both passenger and driver) phones collect users' GPS data. The data source is from Chengdu, China, where the ride-hailing service has been operated for over 6 years. According to the report made by Didi Media Research Institute [15], the times of ride-hailing services are beyond the local taxi services and the ride-hailing system serves over 1.4 million times one day. Thus, the dataset is suitable for study. After receiving these data, we distract the position and timestamp of the appearance of the driver as well as the origin and destination of the passenger. These data also include the timestamps. We added one more step, which is to preprocess the OD data of orders into the desired format of the prediction model. The detailed steps will be elaborated in the methodology part. The proposed methodology can be divided into two parts: the prediction part and the dispatching part. The prediction part is mainly responsible for predicting the distribution of travel demand in the future based on a deep-learning method. The input of this deep-learning model requires historical observation and corresponding metadata; the output is the predicted spatial distribution of travel demand. The dispatching part focuses on optimizing the dispatch under the consideration of minimizing void cruising distance proportion based on the predicted distribution of travel demand in the future. To
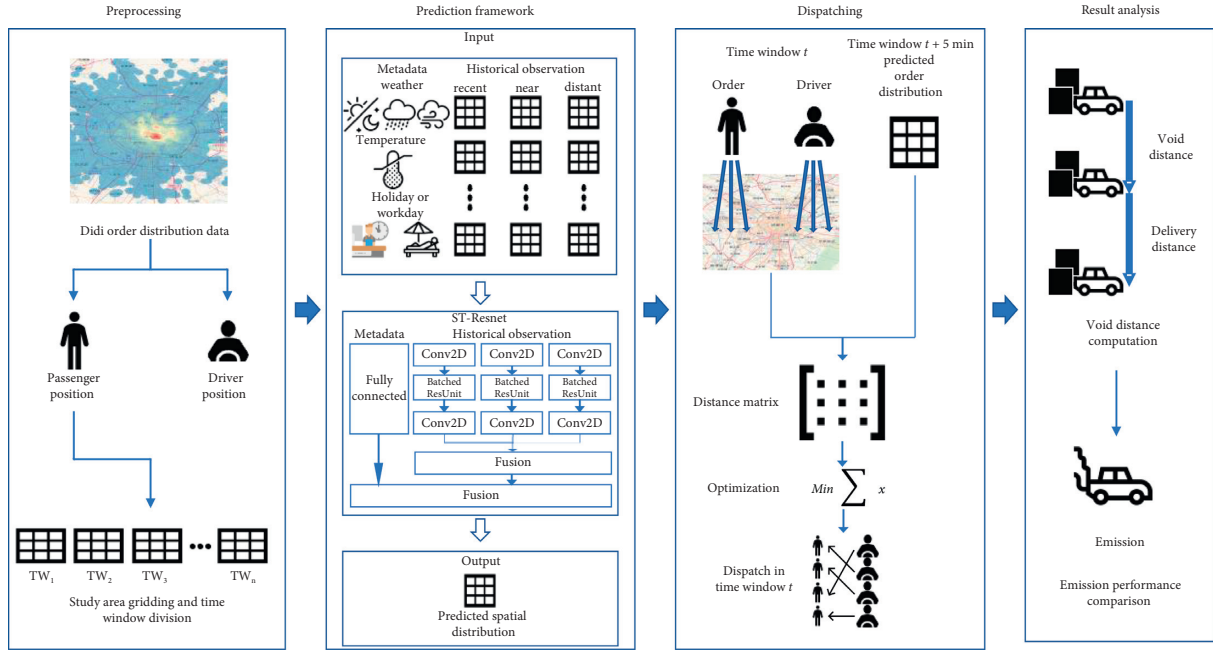
Figure 1: The framework used in this study.

better combine the two parts, we adopt the time window division method. The order and driver will be divided by serial time windows as input [16]. For each time window, the prediction and optimization method will be separately operated once to decide the assignment of the driver to order. To better show the utility of the proposed algorithm, we will use the greedy algorithm operated in the time window and dispatch in the original dataset as a baseline and compare the performance in result analysis.

The assumptions made in this part are shown as follows:

(1) The cost of each passenger and driver to adopt the ride-hailing will be considered. If a passenger waits for a driver to take the order longer than 15 minutes [17], the order will be canceled.

(2) The rejection of order out of the driver's personal issue is not allowed [18].

(3) The participation of the driver is in long term, which means that no driver will quit the system until the simulation is over [6].

(4) During idle time, the driver will park their cars nearby the drop-off location [19].

Finally, we give the metric to compute efficiency. As reviewed in literature reviews, a high percentage of void cruising distance traveled in ride-hailing operation was observed and reducing void cruising distance is an urgent effort. Thus, we choose the void cruising distance proportion as the main metric to measure energy behavior. This definition of void cruising distance proportion is defined as

$$P_v = \frac{d_v}{d_v + d_d}, \tag{1}$$

where P_v is the void cruising distance proportion; d_v is void cruising distance; d_d is the delivery distance.

As it can be seen from the definition, the lower void cruising distance proportion means lower invalid energy usage proportion in operation. In result analysis, we will compare the overall void cruising distance proportion between the former dispatch strategy and our proposed dispatch methodology. We will also provide a spatial-temporal analysis of the dispatch result. In work by Korolko et al. [9], they also took the waiting time of passengers into consideration. Thus, in this study, we will also consider the waiting time of passengers and the cancelation rate of orders. The definition of the waiting time is defined as

$$T_{\text{wait}} = T_{\text{match}} + T_{\text{pitch}}, \tag{2}$$

where T_{wait} is the waiting time of a passenger; T_{match} is the period from the time when the passenger places his or her order to the time when the order is dispatched successfully to a specified driver. It is calculated according to the length of the dispatch algorithm time window, which will be elaborated in Optimization in Dispatch section; T_{pitch} is the time between a driver driving to a pick-up location and a user boarding. It is calculated according to the spatial distances of a specified driver to pick up the passenger.

2.2. Case Study. The raw dataset we use in this study is the ride-hailing record data from Didi cooperation. The dataset is collected by users' mobile phones and includes the information of order ID, driver ID, start time, end time, start location, and end location. The dataset describes the city of Chengdu, a super city with high traffic volume, emission, and energy consumption [20]. The date range of the dataset is from November 2, 2016, to November 30, 2016, without November 10, 2016. There is also an obvious data vacancy on November 8. We plot the heatmap of the spatial distribution of the dataset on one day as examples (see Figure 2).

(a)   (b)
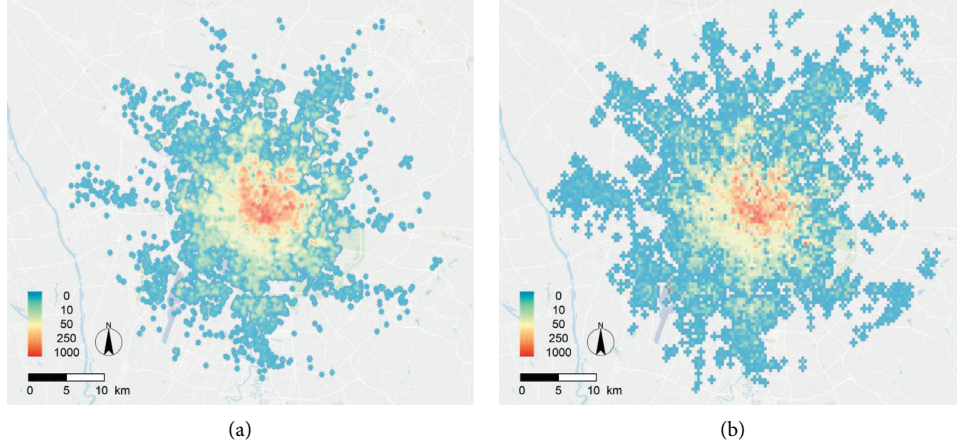
FIGURE 2: Thermal map of order OD data on one day in Chengdu. (a) Origins of all orders. (b) Destinations of all orders.

The study area covers the main areas of Chengdu city and includes some suburban areas. From an overall view, the travel demand is mainly distributed in the city center of Chengdu and the concentration decays with the radical extension.

2.3. Prediction Model. Travel demand prediction is currently a hot research topic in the field of computer science [21, 22]. The mainstream of methods is deep learning. Many researchers developed various deep-learning neural networks that concern this problem. In recent years, there exist a lot of achievements, like convolutional LSTM neural network [23]. The prediction model we use in this study is the ST-Resnet [24]. This is a deep-learning neural network based on the residual unit.

The input of this prediction model is divided into two parts, which are separately historical observation and metadata. The output is the spatial distribution in the future. The desired input format of historical observation is a matrix in essence. The preprocessing is needed to convert spatial data to the matrix. Firstly, we extract the order data in each time window. Next, we apply the regional grid method [25]. The concept is to convert the spatial distribution data to image-like data, which is in the form of matrixes. We divided the study area with a grid size of 40×40. The spatial size of each cell is $3690\,\text{m} \times 3690\,\text{m}$. If we set a larger grid size, although the number of predicted orders may be more accurate, more errors will be introduced to the position of predicted orders and affect the accuracy of the position of predicted orders. Next, this will affect the result of

optimization. However, a small grid size is not necessarily good. A small grid size, which means more cells, will make many cells zero. So, matrixes will be very sparse. Too sparse input and output will corrupt the performance of the prediction model. An example is that the gird with low quantity may be probably predicted as zero and the total quantity of orders will be far from the ground truth. We did some experiments on the grid size and finally found the size of 40×40 as the determined one.

Then, we count the number of orders in the area of each cell in each time window. The quantity of orders is the element of each matrix. Finally, we can get a sequence of matrixes abstracted from spatiotemporal data. We choose 5 minutes as the time length of prediction because the time scale of prediction with minute magnitude can help provide enough future knowledge for the decision. However, a long time scale may keep some drivers waiting a long time for the next order. From preprocessing of the data in one month, we can get totally 8352 matrixes, out of the concern of stability of the training process and shortening the training time. In the prediction model, the input of observation is divided into three sequences, separately recent, near, and distant: recent: a sequence of continuous matrixes of historical observation closely before the time window we want to predict; near: a sequence of continuous matrixes of historical observation that is one day before near; distant: a sequence of continuous matrixes of historical observation that is one week before near. We choose the length of the sequence to be 6. Thus, if we use $x_n, n \in [2022, 8351]$ to represent the output, the input can be represented as

$$X \begin{cases} \{x_i | i \in [n-6, n-5, n-4, n-3, n-2, n-1]\}, \\ \{x_j | j \in [n-288-6, n-288-5, n-288-4, n-288-3, n-288-2, n-288-1]\}, \\ \{x_z | z \in [n-2016-6, n-2016-5, n-2016-4, n-2016-3, n-2016-2, n-2016-1]\}. \end{cases} \tag{3}$$

The input dimension of historical observation is $40 \times 40 \times 6 \times$ batch size.

There are three individual input channels separately for recent, near, and distant. We will introduce the structure of

one channel as all three channels are the same. The first layer is a 2D convolutional layer with a kernel size of 3×3 and 64 filters that extract the feature of the input sequence to the matrix of size $40 \times 40 \times 64$. Then, the following part is a sequence of residual units. The job of each residual unit is to deeper analyze the features. The longer the sequence is, the deeper mechanism can be extracted.

After an iteration in the sequence, the result will go through the final 2D convolutional layer and fuse together. The method of fusion is to add the matrixes from three channels together to one. Then, this one matrix will be added with the reshaped output feature of metadata. Finally, the summed-up feature will be handled by a Tanh function and turned out to be the output of the prediction result.

During the training process, we use the former 20% of the dataset as the test set and the latter 80% as the training set. The optimizer for the gradient descent is Adam [26], which has shown a better performance among all the optimizers.

Another part of the input is the metadata. Generally speaking, metadata includes all the information that can have an impact on the spatial distribution of order. In the original paper, the author used the weather data and date information as the metadata. For compliance with the original model, in this study, we marked the hour that the time window is located in one day, the day in one week, and the week in one year of each time window, the mark of whether the day is a holiday or not (holiday is marked as 1; workday is marked as 0) as the date information. Meanwhile, we also use the weather and temperature data as the weather data. The table of detailed weather data is shown as follows (see Table 1).

The temperature data will be rescaled to $[-1, 1]$ by

$$\frac{\text{tem} - \text{tem}_{\min}}{\text{tem}_{\max} - \text{tem}_{\min}} \times (t_{\max} - t_{\min}) + t_{\min}, \qquad (4)$$

where tem is the temperature we want to rescale; $\text{tem}_{\min}$ is the minimal observed temperature; $\text{tem}_{\max}$ is the maximal observed temperature; $t_{\max}$ is the upper bound of the rescaled interval; $t_{\max}$ is the lower bound of the rescaled interval.

The date information and weather information will be separately turned to numerical data by a one-hot encoding method [27]. Then, two one-hot encoded data will be concatenated together as one matrix. The feature of metadata information will be extracted by a fully connected layer with a Relu function shown as

$$\text{Relu}(x) = \begin{cases} 0, & x < 0, \\ x, & x \geq 0. \end{cases} \qquad (5)$$

The output feature will be then reshaped to the size of $40 \times 40 \times 6$ as the same size of historical observation.

By far, we have completed the introduction of the prediction model. In the result analysis, we will illustrate the accuracy of the prediction model.

2.4. Optimization in Dispatch. Optimization is a classical mathematical method used in many research fields including the ride-hailing [28]. Generally speaking, the concept of optimization is to optimize the objective function and find a global solution.

In the dynamic ride-hailing dispatch problem, a widely used processing method is the time window division [29] (see Figure 3).

The process of time window division can be treated as a group of timeline. The end of each time window is also the beginning of the last one. Suppose that the length of the time window is l, the number of the time window is N, and the start time of simulation is 0. During the period of time window $n, n \in [1, 2, \ldots, N]$, the orders and available drivers given between time $(n-1)l$ and nl will be collected in the matching pool. Then, at the time nl, the simulation algorithm will be operated to give the matching result.

Baseline algorithm, greedy algorithm: the greedy algorithm is a classical algorithm used in many real-time dispatch studies of pick-up and delivery problems [18, 30]. In principle, the algorithm will iterate over every travel demand and find the closest driver who can pick up the order or follow the rule of first-come-first-serve [31]. Generally speaking, the algorithm only considers the optimized solution for each single object. Although this method is easy to implement and manage, it is naturally uncoordinated and tends to prioritize immediate passenger satisfaction over the global supply utilization. In the result and analysis part, we will illustrate the performance of the greedy algorithm.

In this study, we propose a dispatch strategy that integrates both optimization and prediction. Different from the strategy introduced before, the core of the algorithm is that when we consider the dispatch problem in time window n, the predicted distribution of orders in the next 5 minutes will also be taken into account.

At each time of execution, in addition to the orders that really exist in the current time window, the algorithm will add the orders predicted in the next 5 minutes to the matching pool. Then, the optimization algorithm will decide which and how orders in the current time window will be dispatched.

However, this does not mean that we simultaneously complete the dispatch problem in both the current time window n and the next 5 minutes. The dispatching problem in each individual time window is supposed to be solved independently. The prediction on the spatial distribution of orders in the future serves the purpose of enriching the knowledge in the current optimization problem.

Let us consider a dispatch problem with a low dimensionality of 2×2. In Figure 4, D refers to the driver; P refers to the passenger. The dashed object in the figure means that it is predicted; the solid one means it is existing. The distance marked near the arrow is the probable picking up distance. Without the preknowledge of the possible existence of passenger 2, it is obvious that driver 2 will be dispatched to passenger 1 and driver 1 to passenger 2, while if we can predict the appearance of passenger 2, it will be the opposite case. The global picking up distance will be reduced from 1200 m to 800 m. In the simulation, we will quantity this benefit. This also explains why we do not set the time scale of prediction too long. In case there is one predicted order, which is closer than any other order to the driver, the dispatch algorithm may keep the driver waiting until the

Weather	The highest temperature of the day	The lowest temperature of the day
Partly cloudy, cloudy, sunny, and little rainy	[9, 22]	[4, 13]
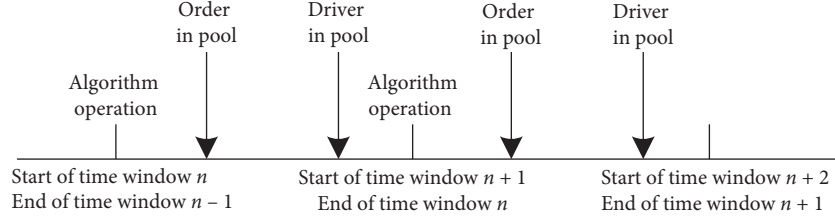
FIGURE 3: Time window division.
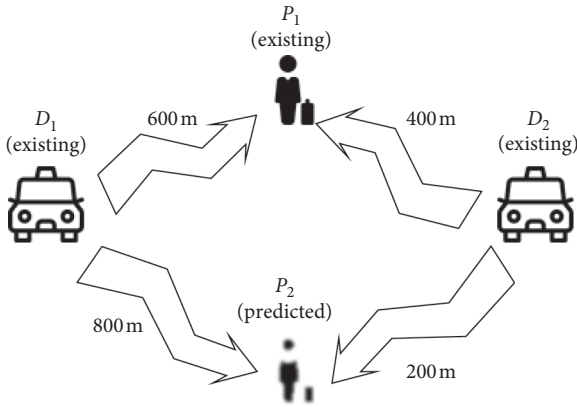
FIGURE 4: Sample of the benefit of prediction.

order appears. The longest possible waiting time is the time scale of prediction.

The target function of the optimization algorithm is

$$\min\left(f = \sum_i \sum_j S_{i,j} D_{i,j} \right), \quad (6)$$

subject to

$$\sum_j S_{i,j} \leq 1, \quad (7)$$

$$\sum_i S_{i,j} \leq 1, \quad (8)$$

where $S_{i,j}$ is the decision variable that decides whether driver i picks up passenger j or not; $D_{i,j}$ is the distance of driver i to pick up passenger j.

Constraint (7) aims to ensure that one driver can be maximally assigned with one order; constraint (8) aims to assure that one order can be maximally assigned with one driver. Here, we can choose to impose one more constraint:

$$\sum_i \sum_j S_{i,j} = \max(I, J). \quad (9)$$

This constraint can serve the purpose of trying best to satisfy orders in time window n with current available

drivers. The difference is that, without the constraint, a part of drivers will not be dispatched to the order in the optimized solution because there may be an order that is much closer to him, while with the constraint, if there is no other candidate driver for the order, the driver will be dispatched in the current time window. In this study, we will also compare the performance of the algorithm with and without the constraint.

From the target function, we can observe that the problem is an ILP (integer linear programming) problem. The basic method to solve ILP is Simplex algorithm [32]. Its basic concept is to firstly construct an initial solution, which is a feasible and finite solution. If the initial solution is not the globally optimal one, then the algorithm will introduce nonbase variables to replace a base variable for a better solution. The iteration is repeated until the globally optimal one is found. Here, we will explain the whole process of solving the optimization algorithm (see Figure 5).

In the first step, we construct the distance matrix of each pair of driver and order. The driver list only contains the available drivers in the current time window n; the order list contains orders in both current time window n and future time window $n + 1$. The column of the matrix refers to the list of drivers and the row refers to the list of orders. The element $D_{i,j}$ means the distance between driver i and order j. Because we can only predict the number of orders in each cell, we lack information on the exact spatial distribution of orders in each cell. Therefore, it is hard to compute the exact distance between predicted orders and existing drivers. To solve this problem, we furtherly divide each cell of the grid into a 10×10 grid (see Figure 6).

Each cell of a larger grid can be described by a 10×10 grid. Then, we do statistics on the historical spatial distribution of orders in each 10×10 grid. Based on the statistic result, we can estimate where the orders may be located if there are orders predicted in the larger cell. We assume the location of predicted orders at the spatial center of the cell of 10×10 grid. After that, we can compute the distance between the driver and predicted orders.

The following part in the process optimization solution is more like a greedy algorithm, where we pick out the pair of order and driver in the ascending order of distance to construct an initial solution of matching. In the next step, we
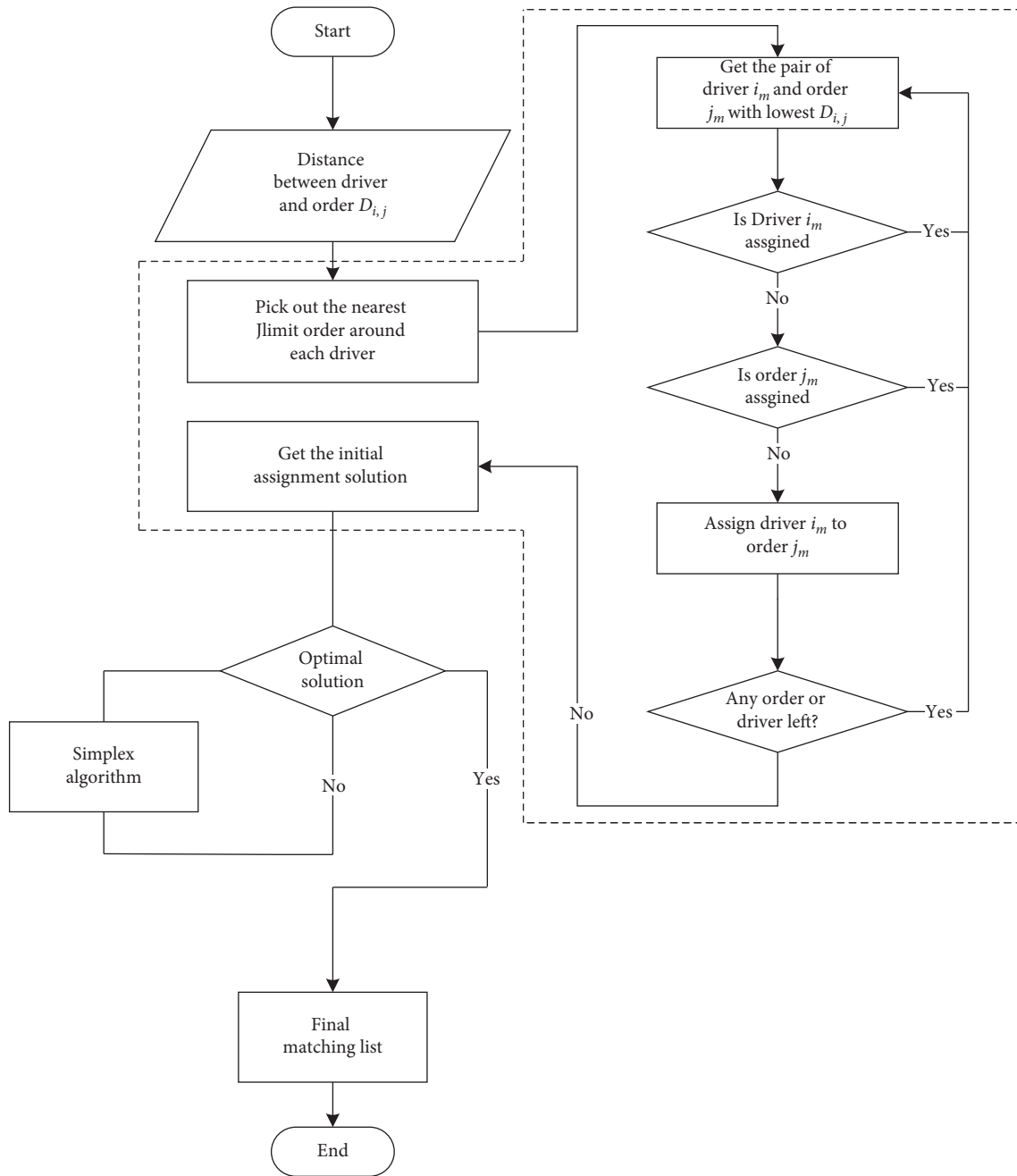
FIGURE 5: Flowchart of the optimization algorithm.

improve this solution by applying the simplex algorithm until we get the optimal solution. In our proposed method, one dispatch process would be finished within several seconds. This CPU compute time is efficient enough for the practical application.

3. Results and Discussion

3.1. Prediction Model Verification. An important parameter of ST-Resnet is the number of residual units used in the model. In the original paper, the author indicated that the more residual the units in the neural network, the deeper the neural network is, the accurate the prediction is. However, more residual units mean more memory usage during training and slower training. To find a balance between the accuracy and computation resources, we choose the number of residual units to be 21. There are a total of over 4.4 million trainable parameters in the model, which is quite a large quantity.

In the field of computer science, multiple forms of losses are used to evaluate the performance of the prediction model like *mse* (mean squared error) and *mae* (mean absolute error). However, these losses are usually used to compare the performance among different prediction models and hardly
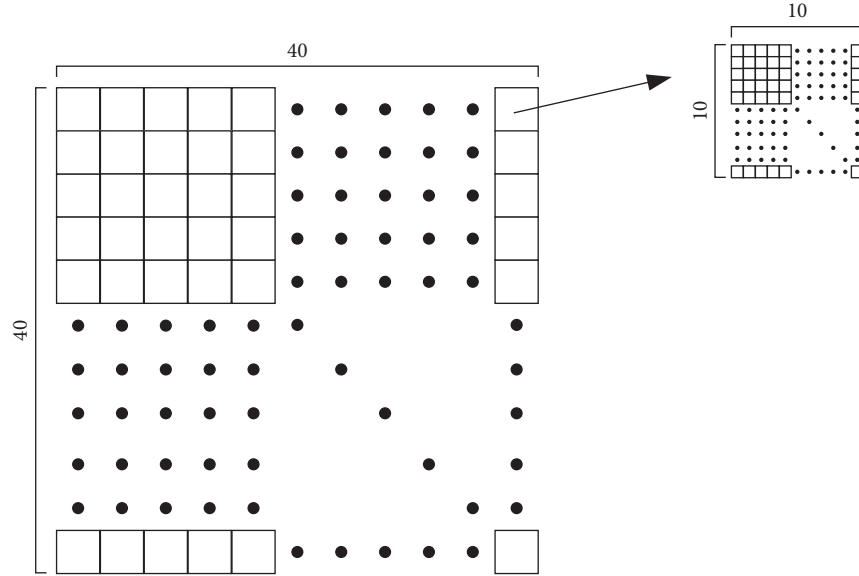
Figure 6: Cell division.

give a direct impression on the accuracy. In this study, we adopt the mape (mean absolute percentage error) as the metric of accuracy, which can directly show the differences.

After completing the training, we compute the mape of the test set to be 0.01169, which means that for each cell of the grid, the difference between prediction result and ground truth is about 1.169%. In addition, we also visualize the prediction result and the ground truth of two samples (see Figure 7).

The figures on the left side are the ground truth matrixes of the spatial distribution of order distribution; the figure in the middle is the corresponding prediction result; the figure on the right is the heatmap of the difference between ground truth and prediction result, which is computed by

$$A_{\text{diff}} = \left| A_{\text{GT}} - A_{\text{PR}} \right|, \tag{10}$$

where A_{diff} is the difference matrix; A_{GT} is the ground truth; A_{PR} is the prediction result.

We can see from the figure that there is little difference between the ground truth and prediction result. The prediction is relatively accurate and enough to put into the simulation.

3.2. Comparison of the Performance of Different Dispatch Strategies.

Here, we will separately introduce the performance of different dispatch strategies. To provide a clearer pattern of performance comparison, we start from the difference between the greedy algorithm and pure optimization without prediction. We randomly selected one day in the dataset and operated the simulation. The performance of the two algorithms is shown in Table 2.

From Table 2, we can see that the greedy algorithm shows a poorer performance. Though the proportion of canceled order of greedy algorithm is not much different compared with the proposed methods, the average waiting time of passengers of the greedy algorithm (over 7 minutes)

is twice as long as the proposed methods. This is mainly because the baseline algorithm does not provide priority to the orders that have been waiting long enough. In the real-world application, for commercial purposes, the ride-hailing dispatch platform may provide priority to the orders that have been waiting for a long time. Besides, when operating the dispatch, the Didi dispatch platform will provide each order to several candidate drivers to raise the chances that the order will be taken. Thus, the waiting time should be shorter in the real application. What is more, we did statistics on the void cruising distance proportion in the original record data. We found that the average value is 29.55%. This indicates a lever principle between the void cruising distance and the satisfaction of orders in former dispatch strategies. In the real application, the Didi company tried to assign the orders to more candidate drivers and caused more pick-up distance. In the simulation, the average waiting time of passengers in the optimization method is only 1/2 times of the greedy algorithm. This shows that the optimization algorithm can more easily and quickly answer the travel demand from customers. What is more, from the perspective of void cruising distance proportion, the optimization method shows a better performance, which is lesser than 1/2 times of the greedy algorithm. The overall result shows that the proposed algorithm surpasses the traditional greedy algorithm. We also operated a probability density statistic on the metrics of each order in the simulation (see Figure 8).

From left to right, the figures show the result of the separately greedy algorithm and optimization method without optional constraint and with optional constraint. The upper figures show the waiting time of passengers and the lower ones show the void cruising distance proportion. The x-axis is the value of the metric and the y-axis can be treated as the "probability." The larger the y value is, the higher probability is. The integration of the result of the greedy algorithm is small mainly because of the high
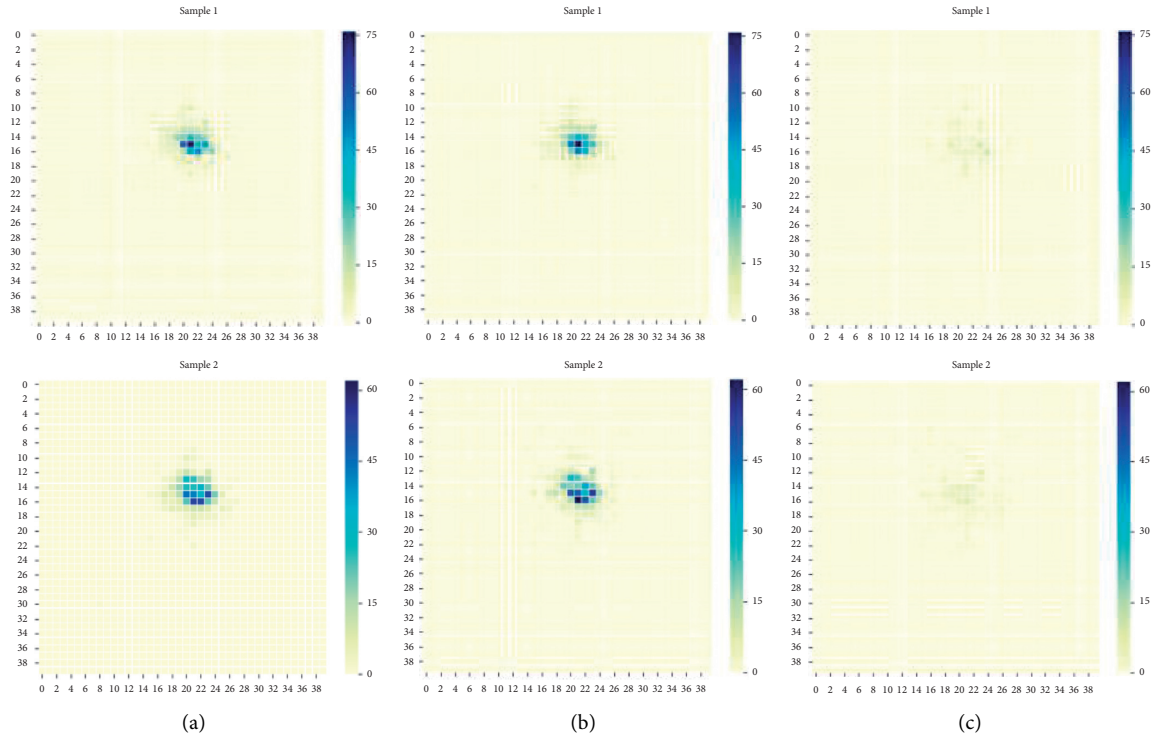
FIGURE 7: Visualization of prediction result. (a) Ground truth matrix. (b) Predicted matrix. (c) The difference between ground truth and prediction.

TABLE 2: Comparison of the greedy algorithm and proposed method by metric.

	Statistic	Simulation		
Metric	Original dataset	Baseline greedy algorithm	Proposed method without optional constraint (s)	Proposed method with optional constraint
AWT		466.15	219.23	217.02
AVCDP	29.55%	11.32	2.66	2.57
PCO	0.00%	1.61	0.00149	0.00

AWT: average waiting time of passenger; AVCDP: average void cruising distance proportion; PCO: proportion of canceled orders.

cancelation proportion. We can have a clear vision that the proposed algorithm can effectively suppress the waiting time and void distance proportion of most cases into a low interval. The waiting time in all of the cases is under 1000 seconds, which is about 16 minutes; the highest void cruising distance proportion is about 25%, while on the other hand, there are some extreme cases in the greedy algorithm. Some have been waiting for over 2500 seconds. In a small part of cases, the picking up distance is near half of the delivery distance. This furtherly proves the stability of the performance of the proposed algorithm.

If we do a transverse comparison between the optimization method with and without the optional constraint, we can find that the optimization algorithm with optional constraint performs better than that without constraint. A smaller proportion of canceled orders and shorter waiting times is natural. We also notice that it accidentally brings a lesser void cruising distance proportion. To better understand the mechanism behind this, we conduct an experiment under the perfect prediction, which means that there is no

error in the prediction result. The result of metrics is separate, the average waiting time of passenger is 215.11 s; the average void cruising distance proportion is 2.50%; the proportion of canceled orders is 0.00%. What may cause the difference between the result of the optimization algorithm with and without optional constraint is probably the uncertainty of distance in the prediction. In the process of solving the optimization problem, we compute the distance between each order and driver in both the current and predicted time window and construct a distance matrix, because we lack the information of the exact spatial distribution of orders in each cell. This causes many uncertainties in the simulation and affects the performance of optimization. Since currently, most of the prediction of travel demand is mainly based on the area gridding method, we recommend more to pay main efforts on optimizing the current dispatch problem. The prediction result aims to provide guidance on which group of drivers are better choices of dispatching at present. We can also judge from the result that there is still a space of 0.07% of void cruising
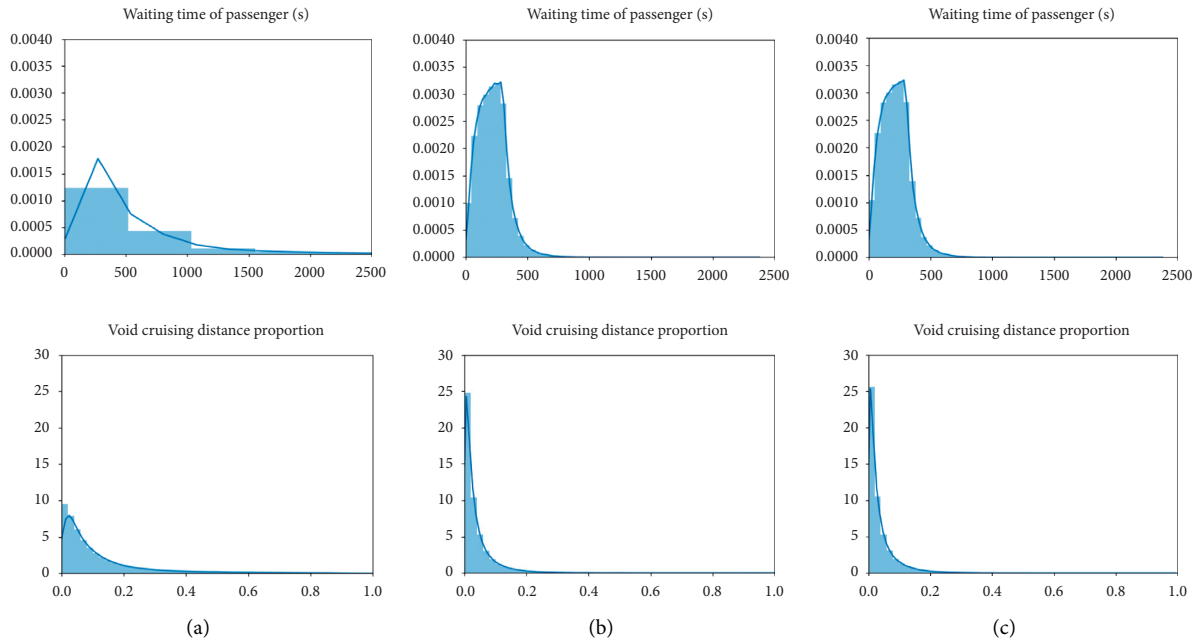
FIGURE 8: The probability density result of metrics. (a) Greedy algorithm. (b) Proposed method without optional constraint. (c) Proposed method with optional constraint.

distance proportion. This also indirectly proves the accuracy of the prediction model.

We further test the performance of different dispatch strategies in the region with high travel demand (Table 3). We operated a statistic on the metrics of the grids with more than 3000 orders. The area of the high travel demand region accounts for 7.38% of the total area, while the order's amount of the high travel demand region accounts for 96.68% of the total orders. The result in the high travel demand region is similar to the result of the whole research area.

In the next subsection, we will simulate the dataset on different days with different metadata to test the sensitivity of performance.

3.3. Comparison of Performance under Different Travel Distributions.

As mentioned in the methodology part, we use the weather and mark of holiday and workday as the input of metadata. Therefore, these two factors can have their own impact on the spatial and temporal distribution of travel demand in the study area. In this section, we will discuss and analyze the sensitivity of performance.

The random day we have chosen in the previous subsection is November 25, which is a cloudy workday. We also choose data in the other three days for test and comparison; the attribute and simulation results are shown in Table 4.

In the table, there is only a little difference among the performances on different days, especially on workdays. All the values of metrics are nearly the same. We found a higher waiting time and void cruising distance proportion on weekend. We also operated simulation on other holidays (see extended Table 1 in Supplementary Materials) and concluded that it is not related to the holiday. Some holidays

also show lower values of metrics. The overall performance of the method among different days is stable despite different daily conditions.

3.4. Temporal Analysis of Simulation Results.

In this subsection, we will give an analysis of the temporal change pattern of void cruising distance. Generally speaking, the average void cruising distance becomes low when the number of orders rises because of high density (see extended Figure 1 in Supplementary Materials). We plot the temporal change of order numbers and reduced average void cruising distance compared with the greedy algorithm and original dataset (see Figure 9).

Figure 9 shows the temporal change of averagely reduced void cruising distance. The result shows that, in simulation, the averagely reduced void cruising distance shares a similar pattern with average void cruising distance. The averagely reduced void cruising distance remains at a low level when the quantity of orders is high and high when quantity is low. From 9 am to 9 pm, the ride-hailing activity becomes active suddenly. Orders appear in the study area with high density. During this period, the pick-up distance of each order can be reduced by about 1000 to 1500 m on average. From 3 am to about 7 am, the travels in ride-hailing are at a low level. The sparse travels in the urban area can cause high pick-up distance and thus cause a higher reduction. The effectiveness of the algorithm shows the best performance during this period. In Figures 9(c) and 9(d), the reduced void cruising distance is relatively stable across one day. The original dispatch contains many other factors that affect the performance; for example, drivers can choose whether to pick up or reject the assignment and the platform will provide the order for as many candidate drivers as possible to assure the

TABLE 3: Comparison of the greedy algorithm and proposed method by metric in high travel demand region.

Statistic Metric	Original dataset (%)	Baseline greedy algorithm (s)	Proposed method with optional constraint (s)
AWT		447.47	215.41
AVCDP	29.34	11.24	3.73
PCO	0.00	1.64	0.00

AWT: average waiting time of passenger; AVCDP: average void cruising distance proportion; PCO: proportion of canceled orders.

TABLE 4: The metadata and simulation result on different days.

Date attribute	November 14	November 16	November 25	November 27
Weather	Sunny	Rainy	Cloudy	Cloudy
Holiday or workday	Workday	Workday	Workday	Holiday
AWT	217.42 s	217.64 s	217.02 s	225.23 s
AVCDP	2.56%	2.54%	2.57%	2.77%
PCO	0.00%	0.00%	0.00%	0.00%

AWT: average waiting time of passenger; AVCDP: average void cruising distance proportion; PCO: proportion of canceled orders.

(a)

(b)
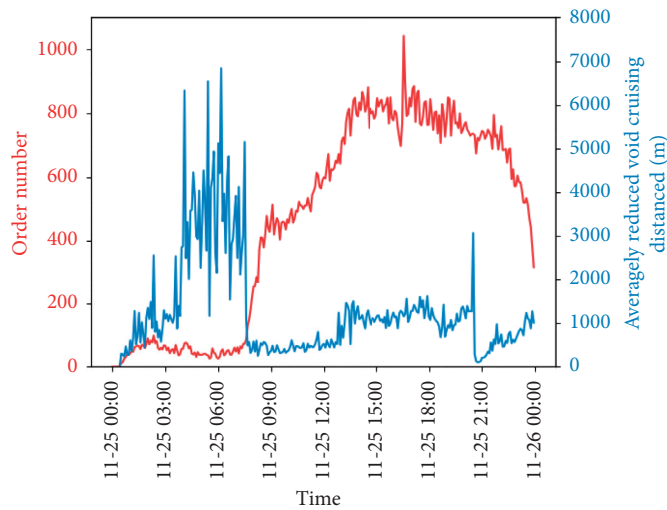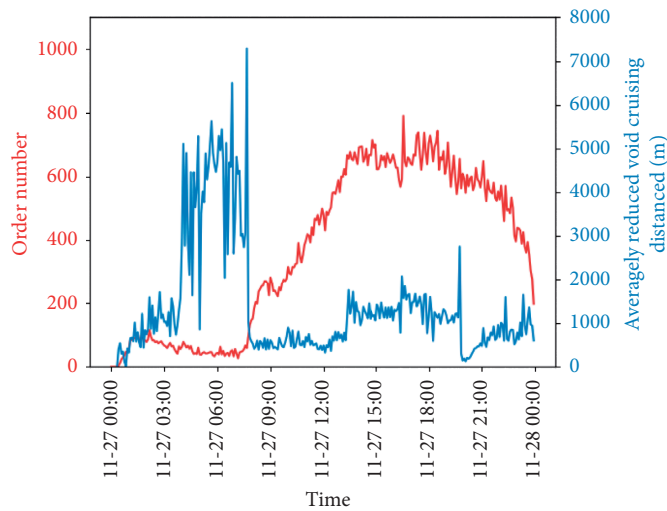
FIGURE 9: Continued.
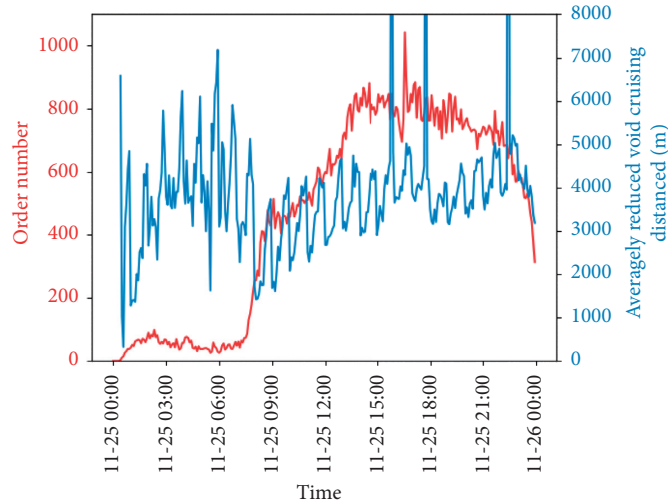
(c)
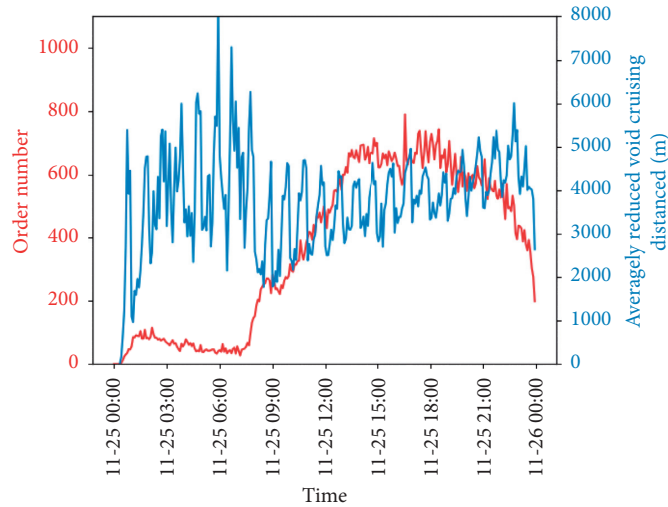
(d)

Figure 9: The temporal change pattern of the order number and averagely reduced void cruising distance compared with (a) greedy algorithm on November 25; (b) greedy algorithm on November 25; (c) original dispatch on 25; (d) original dispatch on 27.

acceptance. We can conclude that there is a lot of potential in the reduction of void cruising distance in the dispatch process. The platform can try to propose some policies to maximally ensure the optimal assignment if necessary.

4. Conclusions

Under the background of energy-saving and emission reduction, many studies focus on traffic emissions [33]. It is pointed out that void cruising for the next passenger caused a lot of unnecessary exhaust emission [34].

The main idea of this study is to propose a dispatch method based on both prediction and optimization methods to improve the efficiency of the ride-hailing system. We use the same dataset for simulation. We firstly preprocess the data into the desired input format of historical observation needed by the prediction model and collect the metadata for additional input. Next, we adopt the ST-Resnet as the deep-learning neural network for prediction and successfully train

the prediction model. The rescaled MAPE is rather little and enough for simulation of the dispatch algorithm. Then, we introduce the dispatch algorithm based on the optimization method. We state the target function to optimize and the constraint including the optional one which can impose a full satisfaction of orders. We use the greedy algorithm which is used widely in the current real-time dispatch system as the baseline and compare the performances. We find that the proposed method outperformances the baseline and shows a good stability of performance on the evaluation metrics, which proves a great potential in real-time application. In addition, we also find that the algorithm shows better performance with the optional constraint. This is mainly because of the uncertainties in the location of predicted orders. Thus, we suggest imposing the constraint that maximizes the number of served orders in solving the current dispatch problem. By far we have successfully answered and filled research gaps mentioned in the literature review.

There are certainly some limitations in this work. For example, we mainly use the statistical result to estimate the location of predicted orders in the future. Adopting a smaller cell can help get a more accurate location. However, it will inevitably make the spatial distribution matrix sparser, thus making the model hard to be trained. In the future, if a better prediction method like GCN (graphical convolutional network) can be developed better, we can improve this limitation by adopting them.

Data Availability

The Didi ride-hailing record data used to support the findings of this study have been deposited in the GAIA Open Dataset repository https://outreach.didichuxing.com/research/opendata/en/#.

Conflicts of Interest

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

Supplementary Materials

Extended Table 1: simulation result on holidays. Extended Figure 1: temporal change of order number and average void cruising distance. (a) November 25; (b) November 27. (*Supplementary Materials*)

References

[1] P. S. Lavieri and C. R. Bhat, "Investigating objective and subjective factors influencing the adoption, frequency, and characteristics of ride-hailing trips," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 100–125, 2019.

[2] A. Tirachini, "Ride-hailing, travel behaviour and sustainable mobility: an international review," *Transportation*, vol. 47, pp. 1–37, 2019.

[3] B.-J. Tang, X.-Y. Li, B. Yu, and Y.-M. Wei, "How app-based ride-hailing services influence travel behavior: an empirical study from China," *International Journal of Sustainable Transportation*, vol. 14, pp. 1–15, 2019.

[4] Z. Li, Y. Hong, and Z. Zhang, "An empirical analysis of on-demand ride sharing and traffic congestion," in *Proceedings of the International Conference on Information Systems*, Dublin, Ireland, 2016.

[5] Y. Sui, H. Zhang, X. Song et al., "GPS data in urban online ride-hailing: a comparative analysis on fuel consumption and emissions," *Journal of Cleaner Production*, vol. 227, pp. 495–505, 2019.

[6] J. Bischoff, I. Kaddoura, M. Maciejewski, and K. Nagel, "Simulation-based optimization of service areas for pooled ride-hailing operators," *Procedia Computer Science*, vol. 130, pp. 816–823, 2018.

[7] G. Feng, G. Kong, and Z. Wang, "We are on the way: analysis of on-demand ride-hailing systems," Research Paper, 2017.

[8] F. Calderón and E. J. Miller, "Modelling within-day ride-hailing service provision with limited data," *Transportmetrica B: Transport Dynamics*, vol. 67, no. 8, pp. 1–24, 2020.

[9] N. Korolko, D. Woodard, C. Yan, and H. Zhu, "Dynamic pricing and matching in ride-hailing platforms," Research Paper, 2018.

[10] P. Afeche, Z. Liu, and C. Maglaras, "Ride-hailing networks with strategic drivers: the impact of platform control capabilities on performance," pp. 18-19, 2018, Columbia Business School Research Paper.

[11] H. Zhang, X. Song, T. Xia et al., "Battery electric vehicles in Japan: human mobile behavior based adoption potential analysis and policy target response," *Applied Energy*, vol. 220, pp. 527–535, 2018.

[12] H. Zhang, X. Song, Y. Long et al., "Mobile phone GPS data in urban bicycle-sharing: layout optimization and emissions reduction analysis," *Applied Energy*, vol. 242, pp. 138–147, 2019.

[13] T. Sühr, A. J. Biega, M. Zehlike, K. P. Gummadi, and A. Chakraborty, "Two-sided fairness for repeated matchings in two-sided markets: a case study of a ride-hailing platform," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3082–3092, Anchorage, AK, USA, 2019.

[14] C. Yan, H. Zhu, N. Korolko, and D. Woodard, "Dynamic pricing and matching in ride-hailing platforms," *Naval Research Logistics (NRL)*, vol. 67, 2019.

[15] D. M. R. Institute, "Report on big data of intelegent travel in chengdu," 2016.

[16] L. Zheng, L. Chen, and J. Ye, "Order dispatch in price-aware ridesharing," *Proceedings of the VLDB Endowment*, vol. 11, no. 8, pp. 853–865, 2018.

[17] M. M. Vazifeh, P. Santi, G. Resta, S. H. Strogatz, and C. Ratti, "Addressing the minimum fleet problem in on-demand urban mobility," *Nature*, vol. 557, no. 7706, p. 534, 2018.

[18] A. M. Arslan, N. Agatz, L. Kroon, and R. Zuidwijk, "Crowdsourced delivery—a dynamic pickup and delivery problem with ad hoc drivers," *Transportation Science*, vol. 53, pp. 222–235, 2018.

[19] T. ICfEa, "Online ride-hailing network and fuel consumption: a driver's perspective," *Innovation Center for Energy and Transportation*, vol. 20, 2017.

[20] K. Shi, B. Di, K. Zhang, C. Feng, and L. Svirchev, "Detrended cross-correlation analysis of urban traffic congestion and NO_2 concentrations in Chengdu," *Transportation Research Part D: Transport and Environment*, vol. 61, pp. 165–173, 2018.

[21] D. Wang, Y. Yang, and S. Ning, "Deepstcl: a deep spatio-temporal convlstm for travel demand prediction," in *Proceedings of the 2018 international Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, 2018.

[22] H. Yao, F. Wu, J. Ke et al., "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, Toronto, Canada, 2018.

[23] K. F. Chu, A. Y. Lam, and V. O. Li, "Travel demand prediction using deep multi-scale convolutional LSTM network," in *Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1402–1407, IEEE, Maui, HI, USA, 2018.

[24] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, 2017.

[25] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.

[26] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014, http://arxiv.org/abs/14126980.

[27] M. Cassel and F. Lima, "Evaluating one-hot encoding finite state machines for SEU reliability in SRAM-based FPGAs," in *Proceedings of the 12th IEEE International On-Line Testing Symposium (IOLTS'06)*, p. 6, IEEE, Lake of Como, Italy, 2006.

[28] L. Al-Kanj, J. Nascimento, and W. B. Powell, "Approximate dynamic programming for planning a ride-hailing system using autonomous fleets of electric vehicles," *European Journal of Operational Research*, vol. 284, no. 3, pp. 1088–1106, 2020.

[29] X. Yu and S. Shen, "An integrated decomposition and approximate dynamic programming approach for on-demand ride pooling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, 2019.

[30] Y. Khazbak, J. Fan, S. Zhu, and G. Cao, "Preserving location privacy in ride-hailing service," in *Proceedings of the 2018 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–9, IEEE, Beijing, China, 2018.

[31] R. Zhang and M. Pavone, "Control of robotic mobility-on-demand systems: a queueing-theoretical perspective," *The International Journal of Robotics Research*, vol. 35, no. 1-3, pp. 186–203, 2016.

[32] V. Klee and G. J. Minty, "How good is the simplex algorithm," *Inequalities*, vol. 3, pp. 159–175, 1972.

[33] Z. Huang, F. Cao, C. Jin, Z. Yu, and R. Huang, "Carbon emission flow from self-driving tours and its spatial relationship with scenic spots - a traffic-related big data method," *Journal of Cleaner Production*, vol. 142, pp. 946–955, 2017.

[34] X. Luo, L. Dong, Y. Dou et al., "Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China," *Journal of Cleaner Production*, vol. 142, pp. 926–935, 2017.

Research Article

An Empirical Study on Travelers' Acceptance Intention of Travel Information on Social Networks

Jian Chen [ID],[1] **Rui Li** [ID],[1] **Zhiyan Fu**,[2] **Chi Zhang**,[1] **and Fatao Yuan**[1]

[1]*College of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China*
[2]*School of Economics and Business Administration, Chongqing University of Education, Chongqing 400067, China*

Correspondence should be addressed to Jian Chen; chenjian525@126.com

Social networks are new channels for travelers to obtain or share travel information, which has important impacts on their travel decision-making behavior. Therefore, the psychological feelings of travelers and their acceptance intention (AI) of this type of travel information should be explored. In this study, certain psychological latent variables were incorporated into a technology acceptance model to construct an extended model that explores the factors influencing the travelers' AI of travel information on social networks. This model was validated using survey data collected in Chongqing, China. The influence of each factor on the AI and the interaction between factors were quantitatively described using the structural equation modeling method. The results showed that the perceived risk, perceived trust, and perceived usefulness are the most important factors affecting travelers' AI; the subjective norm, hedonic motivation, and perceived ease of use also exert a certain degree of influence; the proposed research model has a good interpretation ability for AI, and the explanatory power has reached 52%. This study confirmed the applicability of the constructed model in this research field on the basis of survey data and provided a theoretical reference for ascertaining the attitude of travelers toward travel information available on social networks.

1. Introduction

With the development of information technology and the popularization of smart terminal devices, the way people use the Internet has gradually shifted from simple information search and web browsing to information interaction based on social media. Thus, social media has become an important service platform for people to publish, acquire, and disseminate information. According to the Global Digital Report of 2020 [1], the number of Internet users worldwide is 4.54 billion and the number of active social media users is 3.80 billion. The social media platforms Facebook, YouTube, and WhatsApp are ranked as the top three globally, with 2.45, 2, and 1.6 billion active users, respectively. Chinese social media platforms WeChat, QQ, Qzone, and Sina Weibo are also ranked among the top ten in the world by a number of active users. Social networks are related to the Internet based on the social relationship between social media users, that is, "Social + Internet". The rapid

development of social networks has shortened the space-time distance of information, endowed it with higher value, and gradually created an ecosystem that "connects everything". People can express opinions, make friends, interact with them, and obtain information conveniently and quickly on the social network, thus generating a large amount of information that has an important influence on people's daily life behavior patterns [2, 3].

The mass information on social networks also involves a significant amount of travel information. The content of travel information on social networks mainly includes two types: one is the information obtained by the traveler via social media during the travel process, concerning the traffic system, such as road conditions, transit time, accidents, and weather information; the other concerns information on related experiences shared by other travelers regarding travel destinations, travel modes, and travel route decisions [4–7]. This information can be sent out in the form of a Twitter message, discussion post, WeChat moments, or Sina Weibo

message. The information content can be presented in various forms, such as text, pictures, video, or voice. The source of the travel information can be the official social media account of the traffic management department or the personal social account of the traveler. This travel information has an important influence on the travel decision-making behavior of travelers. It can affect people's attitudes regarding travel plans, allowing them to prejudge the travel environment before travel. It can also affect the optimization and adjustment of travel routes during travel and update and share the travel experience after travel. Travel information on social networks is qualitatively different from the travel information published on traditional channels such as TV and radio. It has a faster propagation speed, higher timeliness, wider coverage, and stronger interactive function between the publisher and the receiver. It also has obvious advantages compared with location-based service map applications, such as Google Maps, Gaode Maps, and Baidu Maps. The information content is richer, the release form is more diverse, the timeliness is higher, the trust of the traveler is higher, and the interactive function between the publisher and recipient of the travel information is also stronger. It should be noted that social networks are used as supplementary channels for obtaining travel information other than traditional channels, such as television and radio, and location-based map applications, such as Google Maps and Gaode Maps, rather than an alternative channel. Travel information on social networks has new information content, new release modes, and new dissemination efficiency, which makes residents' travel decision-making into a multidimensional information environment.

In recent years, scholars have gradually discovered that there exists a significant relationship between travel information on social networks and decision-making behavior. De Abreu e Silva et al. conducted an online survey of college students in the three cities of Lisbon, Granada, and Zagreb and explored the relationship between information and telecommunication technologies (ICT), social media use, and travel mode choice behavior. Then, they analyzed the cross-cultural differences within the surveyed population [8]. Khan et al. explained the impact of the use of transport-support applications and social networks on weekend travel behavior through online survey data collected in Halifax, Canada. They also considered the moderating effect of travelers' personal attributes such as age and gender [9]. Chen and Deng conducted a survey on residents in some cities in China and used the hierarchical clustering analysis method to find the relationship between information use, social networks, and cooperation awareness in the traveler's choice of travel mode [10].

The premise for travel information to be widely disseminated on social networks and influence travel decision-making behavior is that travelers must be willing to share and accept this information. Researchers have begun to explore the influencing factors on travel information sharing behaviors on social networks in recent years, but they have rarely considered the influencing factors on the travelers' acceptance behavior. Bilgihan et al. considered that factors such as subjective norms, utilitarian beliefs, and perceived

ease of use have a significant impact on travelers' willingness to share travel information on social networks [11]. Oliveira et al. found that perceived enjoyment, security, and privacy are important influencing factors for travelers to share travel information and experience on social networks [12]. On this basis, Oliveira et al. further pointed out the relationship between altruistic motivations, personal fulfillment, and self-actualization, and their impact on travelers' willingness to share travel information on social networks [13]. Information sharing and information reception are both important links of travel information dissemination on social networks; thus, the factors influencing their acceptance behavior are also worth discussing. Chung et al. found that argument quality, source credibility, perceived usefulness, and social relations are the decisive factors affecting travelers' acceptance of travel information on social networks [14]. Berhanu and Raj investigated the degree of international tourists' trust in travel information on social networks and analyzed the demographic attributes of age, gender, and education [15].

Although a small number of studies have begun to focus on the acceptance of travel information on social networks, these studies are still in their infancy, and the impact factors considered therein are limited. The subjective attitude of travelers to the acceptance of travel information on social networks is not yet clear. Thus, in this study, an analysis model of the intention to accept travel information on social networks was constructed based on the technology acceptance model. It aims to quantitatively describe the psychological feelings of travelers toward the travel information on social networks and explain the influence process of various influencing factors on the acceptance intention. The research outcomes are expected to provide decision-making references for publishers, policy makers, and operation managers of travel information on social networks, thereby aiding them in the provision of better, more accurate, and more acceptable travel information for travelers. This is one of the few studies in this field.

The remainder of this paper is organized as follows. In Section 2, the research hypotheses of this study are proposed on the basis of the technology acceptance model (TAM) and related theories. In Section 3, the research survey design, variable measurement, and data analysis methods are introduced. In Section 4, the data analysis results of the research model are presented. In Section 5, the impacts of model variables on the AI of travel information on social networks are compared and analyzed against the research findings of other studies. In Section 6, the theoretical and practical significance of this study is summarized, and the limitations of the current study and the future research directions are also noted.

2. Theoretical Model and Research Hypothesis

2.1. Basic Theory and Model

2.1.1. TAM Model. Many theoretical models have been developed in the research fields of sociology and psychology to analyze and explain the public's acceptance of certain

technology or information. The most widely used model is the TAM. In 1989, Davis proposed the TAM [16] in conjunction with the theory of reasoned action (TRA) while exploring the factors that influence users' usage behavior of information systems. The TAM considers that the behavioral intention (BI) plays a decisive role in the actual behavior (AB) of the user, and the BI is affected by the attitude, perceived usefulness (PU), and perceived ease of use (PEU). The model conceptual framework of the TAM is shown in Figure 1. PU reflects the degree to which an individual believes that using the information system would improve his/her job performance. PEU reflects the degree of ease that an individual feels while using an information system. Attitudes refer to the individual's positive or negative subjective feelings when using an information system. The TAM applied the TRA to the field of technology acceptance research for the first time. It deeply analyzed the process of users' intention and the usage behavior of information systems. It has a strong ability to explain the relationship between core variables such as attitude, BI, and AB.

The TAM is a powerful theoretical tool for predicting and explaining the acceptance of information technology by individuals or organizations. It has been widely used in research for exploring various aspects such as social behavior, learning behavior, and business behavior. However, because the original variables were designed to analyze users' acceptance of information systems, the research in other fields has to expand the influencing factors when utilizing the TAM, thereby enhancing its applicability in the new research field [17–20]. The social media platform that can obtain travel information is essentially an information system, and hence, the TAM can be used as the basic theoretical framework for this study to analyze the AI of travel information on social networks.

2.1.2. Other Classical Theories.
Other classic theories in the field of personal behavior research also provide support for the variable design of this research theoretical model, such as perceived risk, perceived trust, hedonic motivation, and rational behavior theory.

Perceived risk was originally a concept extended from the field of psychology by Bauer [21] of Harvard University in 1960. Bauer believes that an individual's behavior may not be able to determine whether the expected result is correct and certain results may make him feel unpleasant. Therefore, this implicit uncertainty in the outcome of a decision is the initial concept of perceived risk. When Bauer introduced this concept, he emphasized that perceived risk concerns subjective (perceived) rather than objective (real world) risk and further divided perceived risk into the uncertainty of decision results and severity of consequences of wrong decisions.

Perceived trust is derived from trust theory. Mayer et al. [22] defined trust as the belief that an individual is willing to believe the corresponding characteristics of other people or things and is willing to take risks. The extended TAM model proposed by Gefen et al. [23] also takes trust as a core variable, laying a foundation for the combination of trust and TAM. Since the social network is a virtual environment, the degree of uncertainty and security risk of travel information in it is higher than that in the traditional environment; thus, trust has become an important factor to consider.

Subjective norm is derived from the theory of rational behavior [24], which refers to the social pressure that an individual feels regarding whether to take a particular behavior or refers to the influence of the individuals or groups on the individual's behavioral decision-making. Venkatesh and Davis [25] introduced subjective norm variables in the extended model of the TAM.

Hedonic motivation refers to the user's willingness to initiate behaviors that can enhance positive experiences (pleasure, happiness, etc.) and reduce negative experiences (frustration, sadness, etc.). It is used to explain the general principle of personal behavior. The principle is that individuals are more likely to initiate behaviors that lead to rewards or away from punishment [26]. Hedonic motivation has been proven by research for its important role in the field of technology acceptance [27] but, in the field of technology acceptance, Venkatesh et al. conceptualized hedonic motivation as the perceived fun and joy of using a certain information technology system [28].

2.2. Research Hypothesis.
In this study, to better explain the traveler's AI of travel information in the social network and explore the corresponding influencing factors, an analysis model of the AI for travel information in the social network is constructed based on the TAM. In this model, the AI, PU, and PEU are borrowed from the original TAM, and the subjective norm (SN) is borrowed from the TRA and the extended TAM [24, 25]. On considering the new mode and content of travel information on social networks, travelers will make subjective cognitive judgments about the risk, trust, and interest in it. Therefore, the three variables of perceived risk (PR), perceived trust (PT), and hedonic motivation (HM) are added in this study, after referring to the relevant theories, to enhance the applicability of the constructed model for the analysis of the AI of travel information on social networks. The model research hypothesis and conceptual framework are shown in Figure 2.

2.2.1. Acceptance Intention (AI).
An important concept of the TAM is to reflect the AB with BI, and its effectiveness has been extensively demonstrated [29–31]. The AI in this study indicates the traveler's willingness to use or recommend others to use the travel information in the social network. It is a behavioral intention that reflects the willingness to accept. Thus, AI can be used as a dependent variable for various influencing factors in the study to explore the traveler's acceptance of travel information on social networks.

2.2.2. Perceived Usefulness (PU) and Perceived Ease of Use (PEU).
Both PU and PEU come from the TAM [16]. PU here indicates the travelers' perception of job performance
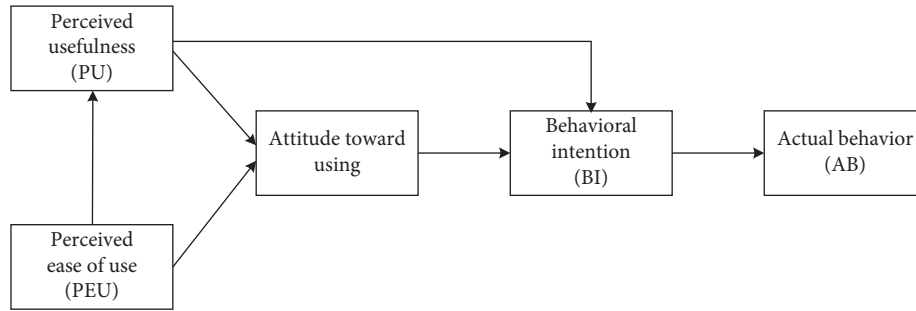
FIGURE 1: Conceptual framework of TAM.
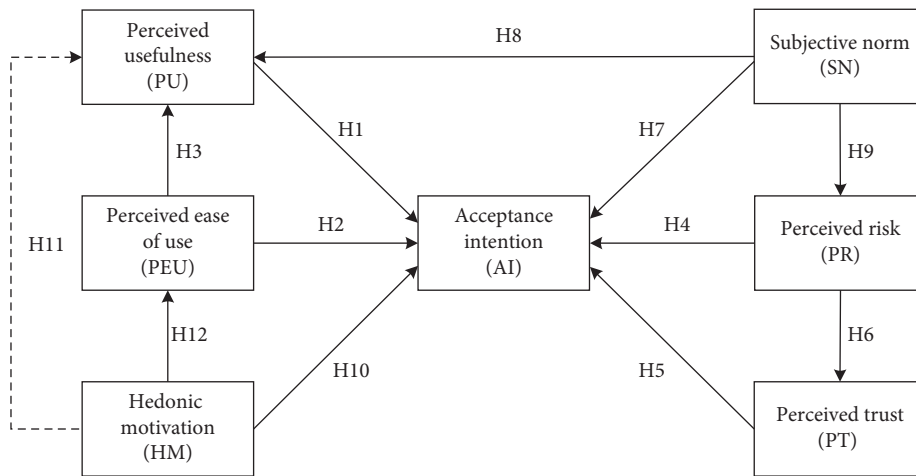
FIGURE 2: Research hypothesis and the conceptual framework of the model.

improvement using travel information on social networks, and PEU refers to the travelers' cognition of the ease of using a social network media platform (Facebook, YouTube, Sina Weibo, WeChat, etc.) to obtain the travel information. The greater the benefits of accepting travel information on social networks and the simpler the operation process of using social network media platforms to obtain travel information, the more likely the travelers will accept this information. At the same time, the TAM pointed out that the PEU will positively affect the PU. Therefore, the following hypotheses are formulated:

> H1: PU has a positive impact on the AI of travel information on social networks.
>
> H2: PEU has a positive impact on the AI of travel information on social networks.
>
> H3: PEU has a positive impact on PU.

2.2.3. Perceived Risk (PR) and Perceived Trust (PT).
PR refers to the individual's perception of uncertainty about the outcome of a decision [21]. It here refers to the traveler's awareness of the risks in terms of time, cost, and personal privacy created by travel information on social networks. Travelers show significant risk aversion when making behavioral decisions. The higher the risk of travel information on social networks is, the less willing they are to accept it. PT

comes from the expansion of the TAM by Gefen et al. [23]. It here refers to travelers' belief that the travel information on social networks will not affect their own interests. The higher the degree of trust in the travel information on social networks is, the higher the travelers' willingness to accept it. At the same time, the higher is the risk perceived by travelers of the travel information on social networks, the lower their degree of trust in the information is. Therefore, the following hypotheses are formulated:

> H4: PR has a negative impact on the AI of travel information on social networks
>
> H5: PT has a positive impact on the AI of travel information on social networks
>
> H6: PR has a negative impact on PT

2.2.4. Subjective Norm (SN) and Hedonic Motivation (HM).
SN is considered from the TRA proposed by Fishbein and Ajzen and the extended TAM proposed by Venkatesh and Davis [24, 25]. It here represents the influence of relatives, friends, and colleagues close to the travelers on their own behavior decisions. Social networks are essentially built on the basis of social relationships among users; thus, travelers' decision-making behavior will be influenced by social groups such as relatives, friends, and colleagues. The greater the positive impact is, the easier it is for travelers to accept

the travel information on social networks. At the same time, the PU and PR of the travel information on social networks by relatives, friends, and colleagues in the surroundings will also directly affect the traveler's perception of these two variables. HM refers to the user's expectation of using information or technology to enhance positive experiences (pleasure, happiness, etc.) and reduce negative experiences (frustration, sadness, etc.). The importance of the role of HM has been proven in the field of technology acceptance [26, 27]. However, Venkatesh et al. have conceptualized HM as a perception of the fun and joy of using an information technology system in the field of technology acceptance research [28]. HM here refers to the traveler's perception of the degree of interest and joy of using social networks to obtain travel information. The higher the level of interest and joy the information provides, the stronger the AI of the traveler will be. At the same time, if the traveler is in a happy mood, the stronger the PU and PEU of the travel information on social networks will be. Consequently, the following hypotheses are formulated:

> H7: SN has a positive impact on the AI of travel information on social networks
>
> H8: SN has a positive impact on PU
>
> H9: SN has a positive impact on PR
>
> H10: HM has a positive impact on the AI of travel information on social networks
>
> H11: HM has a positive impact on PU
>
> H12: HM has a positive impact on PEU

3. Methods

3.1. Survey Design. In this study, the residents of Chongqing were selected as the survey targets, and a combination of online and offline questionnaires was randomly distributed to investigate the AI of travel information on social networks. The online questionnaire is distributed through the "Questionnaire Star" network platform, and the network link is shared through social software, such as Sina Weibo, WeChat, and QQ. The area of the offline survey was selected in the business districts, bus stations, rail stations, etc., of the central city's administrative districts, where there were many people and travel, and the samples were obtained in the form of random interviews by investigators to ensure the randomness of the sample collection and reduce bias.

In the questionnaire, a brief textual introduction of travel information on social networks and the corresponding characteristics of their access platforms (Facebook, YouTube, Sina Weibo, WeChat, etc.) was given at the beginning. The survey comprised two parts. The first part gathered the demographic characteristics, including gender, age, education level, monthly income, and ownership of a private car. The second part included a measurement scale for the travelers' subjective perceptions of travel information on social networks, covering the measurement of seven variables, namely, PU, PEU, PR, PT, SN, HM, and AI. On the basis of the designed measurement scale, the reliability and validity of the questionnaires were examined using the data

collected via a small sample pretest. The measurement items that failed in the test were removed to obtain the final and formal questionnaires. After the testing process, it took approximately 5 to 10 min to complete one questionnaire. The survey was conducted in October 2019 in Chongqing, where 500 copies (100 copies of the online questionnaires and 400 copies of the offline) of the formal questionnaire were issued within two weeks.

3.2. Research Measurement. The five-point Likert scale was adopted in this study to measure the latent variables that cannot be directly observed in the model. To reflect the travelers' psychological feelings about travel information on social networks, a 5-point score—from "strongly disagree (=1)" to "strongly agree (=5)"—was used to rank the respondent's degree of approval toward the relevant description given in each questionnaire measurement item. The measurement items for each variable were adapted accordingly based on previous research in the field of technology acceptance. The specific measurement items are detailed in Table 1.

3.3. Data Analysis

3.3.1. Introduction of Method. To simultaneously explore the influence relationships between latent variables—that are relatively abstract in concept and cannot be measured directly—the structural equation modeling (SEM) method, a multivariate statistical analysis technique, was applied to analyze the data. The parameter estimation method used in this study is the maximum likelihood method, and the package software Mplus7.4 is used for the corresponding calculation. The SEM method can analyze the relationship between multiple dependent variables at the same time, and it has been widely applied in psychology, pedagogy, sociology, and other research fields [32–34]. The structural equation model is divided into two parts: a measurement model and a structural model [35, 36]. The measurement model is used to describe the relationship between explicit variables and latent variables, and the structural model is used to describe the relationship between latent variables. The basic equations are as follows:

$$X = \Lambda_X \xi + \delta, \tag{1}$$

$$Y = \Lambda_Y \eta + \varepsilon, \tag{2}$$

$$\eta = B\eta + \Gamma\xi + \zeta. \tag{3}$$

Among them, equations (1) and (2) represent the measurement model, and equation (3) represents the structural model. ξ and η are vectors composed of exogenous latent variables (SN, HM) and endogenous latent variables (PU, PEU, PR, PT, AI), and X and Y are the observed variables of ξ and η (measurement item). Λ_X and Λ_Y are the factor loading coefficient matrices of X and Y, respectively; B is the path coefficient matrix between endogenous latent variables; Γ is the path coefficient matrix between exogenous

TABLE 1: Measurement items for model variables.

Latent variable	Item no.	Measurement item	Adapted source/ reference
PU	PU1	The travel information obtained through social media platforms is very accurate.	Davis [16]
	PU2	The travel information content obtained through the social media platform is very comprehensive.	
	PU3	The travel information obtained through the social media platform is very helpful for my travel decision.	
PEU	PEU1	Downloading social media software and registering an account is an easy task.	Davis [16]
	PEU2	The method of using social media platforms is easy to learn.	
	PEU3	The operation interface of the social media platform is friendly in design and smooth in experience.	
PR	PR1	I am worried that using social media platforms to obtain travel information will divulge my privacy.	Bauer [21]
	PR2	I am worried that using social media platforms to obtain travel information will add unknown fees to me.	
	PR3	I am worried that using social media platforms to obtain travel information will require a significant amount of my time."	
	PR4	I am worried that the wrong travel information on social media platforms will adversely affect my travel decision.	
PT	PT1	I believe most of the travel information on the social media platform is true.	Gefen et al. [23]
	PT2	I trust the friends I make on social media platforms	
	PT3	The operators of social media platforms are very trustworthy.	
SN	SN1	An increasing number of people around me use social media platforms to obtain or share travel information.	Venkatesh and Davis [25]
	SN2	Friends around me often share travel information with me through social media platforms.	
	SN3	I am increasingly concerned about the travel information on social media platforms under the influence of others.	
HM	HM1	Sharing travel information on social media platforms is fun.	Venkatesh and Thong [28]
	HM2	Travel information on social media platforms is always presented in a very interesting way.	
	HM3	I am very happy to be able to help other travelers on social media platforms.	
AI	AI1	I am willing to use social media platforms to obtain travel information.	Davis [16]
	AI2	I plan to use social media platforms often to obtain travel information.	
	AI3	I would recommend friends around me to use social media platforms to obtain travel information.	

latent variables and endogenous latent variables, δ and ε are the measurement errors of the observed variables, and ζ is the model residual, which represents the part that ξ cannot explain or predict for η. It is worth noting that δ has no correlation with ξ, η, or ε. Furthermore, ε has no correlation with ξ, η, or δ.

The reliability and validity of the model need to be tested using the confirmatory factor analysis (CFA) method [37]. CFA is often tested using the SEM method. The analysis process is the quality inspection process of the measurement model in the structural equation model. The main inspection indicators include factor loading coefficient λ, construct reliability (CR), and average variance extracted (AVE). The calculation equations for CR and AVE are as follows:

$$CR = \frac{(\sum \lambda)^2}{[(\sum \lambda) + \sum \theta]}, \tag{4}$$

$$AVE = \frac{\sum \lambda^2}{[\sum \lambda^2 + \sum \theta]}, \tag{5}$$

where λ is the factor loading coefficient of the observed variable and θ is the measurement error of the observed

variables (δ and ε). Both λ and θ can be directly calculated via Mplus software, as well as CR and AVE can be calculated via (4) and (5).

3.3.2. Evaluation Criteria of Quality.

At the same time, to ensure the validity of the questionnaire design and the adaptability of the constructed model, the reliability and validity of the model and its overall goodness-of-fit need to be tested through the data obtained from the questionnaire survey.

The reliability and validity of the questionnaire reflect the inherent quality of its model. The reliability test typically examines the internal consistency of the questionnaire measurement items, which is usually evaluated using the construct reliability (CR). The evaluation criteria (threshold) adopted by Kline and many other researchers [35, 38, 39] are as follows: a CR value of above 0.9 indicates the best results, 0.8-0.9 indicates very good results, 0.7-0.8 indicates moderate results, and 0.5–0.7 indicates acceptable results. The validity test generally examines the two aspects of convergent validity and discriminant validity. If the factor loading coefficient of each measurement variable exceeds 0.6, and the average variance extracted (AVE) is greater than 0.5, the

model has a good convergent validity [40]. According to the evaluation criteria proposed by Fornell and Larcker [37], if the square root of the AVE of each latent variable is greater than the Pearson correlation coefficient (PCC) between this variable and the other variables, the model has good discriminant validity.

The covariance matrix of measurement variables generated via quantitative parameter estimation was tested for its closeness to the sample covariance matrix. The better the model fits, the more accurate the parameter estimation is. The model's overall goodness-of-fit indices [35, 41, 42] include the chi-square degree of freedom ratio (χ^2/df, also known as standard chi-square), root mean square error of approximation (RMSEA), comparative fit index (CFI), nonnormed fit index (NNFI, also known as the Tucker-Lewis Index, TLI), and standardized root mean square residual (SRMR). Kline pointed out that for a good model fit, the goodness-of-fit indices should meet the following criteria: χ^2/df is in the range of 1–3, CFI and TLI are greater than 0.90, and RMSEA and SRMR are less than 0.08 [35, 43].

4. Results

4.1. Questionnaire Statistics. After invalid questionnaires—i.e., those that had not been filled in carefully, had more than three missing values, and selected more than five extreme values consecutively—were eliminated, a total of 429 valid questionnaires (93 online questionnaires and 336 offline questionnaires, with a ratio of approximately 1 : 3.6) were recovered, with an effective recovery rate of 85.80%. The number of valid samples was more than ten times the number of questions in the questionnaire (28 questions), which conforms with the empirical criteria proposed by Kline [35, 44]. The basic information of the recovered 429 valid questionnaires was categorized. The proportions of both male and female respondents in this survey are approximately 50%, and no extreme gender difference is prevalent. The respondents are primarily between 20 and 40 years old (approximately 61%). Most of the respondents possess a bachelor's degree in terms of the level of education (approximately 38%). The monthly income of the respondents is mainly in the 3000–5000 Yuan range (approximately 43%). Approximately 32% of respondents have a private car. The specific personal attributes data are detailed in Table 2.

4.2. Measurement Model Evaluation. The CFA results of the measurement model are listed in Tables 3 and 4. The CR values of all latent variables are greater than 0.7, indicating that the internal consistency or isomorphism between the measurement variables of the same latent variable is good. As reflected by the measurement variables, the construct (latent trait) has a good consistency, and the reliability is at the "moderate" level. The factor loading coefficients of all measurement variables are greater than 0.6, and the AVE values are greater than 0.5, indicating that the measurement

variables have a high degree of internal correlation, which can effectively reflect the latent traits of the latent variables. Further, the proportion of variance explained by the latent variable is much higher than that explained by the measurement error, thus indicating that the measurement model has good convergent validity. The AVE value of each of the latent variables is always greater than the PCC value between this latent variable and the other ones, indicating that there are significant differences between the constructs of different latent variables, and the model has good discriminant validity. It can be observed that both the reliability and validity of the model have satisfied the requirements, thereby showing that the model is of good internal quality.

4.3. Structural Model Evaluation. The overall goodness-of-fit test results of the model are listed in Table 5. All the goodness-of-fit evaluation indices reached the test criteria, indicating that the model has good adaptability and external quality. The standardized path coefficient β can reflect the magnitude of the interaction between variables. The square multiple correlation (SMC) of the endogenous latent variables, namely, R^2, can reflect the explanatory power of the model. The standardized path analysis results of the structural model are presented in Table 6 and Figure 3.

The impact of each latent variable on AI is ranked from the highest to the lowest as follows: PR ($\beta = -0.71$, $P < 0.001$), PT ($\beta = 0.66$, $P < 0.001$), PU ($\beta = 0.65$, $P < 0.001$), SN ($\beta = 0.53$, $P < 0.001$), HM ($\beta = 0.42$, $P < 0.001$), and PEU ($\beta = 0.38$, $P < 0.01$). All the latent variables have shown a significant impact on the AI with a confidence interval of 95% (corresponding to $P < 0.05$). The only exception is that HM did not present a significant impact on PU ($\beta = 0.23$, $P > 0.05$), and PR has the most significant negative impact on PT ($\beta = -0.78$, $P < 0.001$) among all variables. Therefore, the research hypotheses pertaining to the latent variables, H1–H10 and H12, are supported, but H11 is rejected. In the model, the proportion of variance in AI that can be explained by all the other variables is 52% ($R^2 = 0.52$), which reaches the minimum standard of 0.4 proposed by Cohen et al. [45], and the model has a good interpretation ability.

5. Discussion

According to the results of the empirical analysis, the reliability and validity of the measurement model have reached a good level and the overall fit of the structural model has also reached the test criteria. The theoretical model constructed in this study has good internal and external quality. In the structural model, all research hypotheses have been verified, except for hypothesis H11, which is not supported. Overall, the newly added PR and PT variables in this study have the most significant influence on acceptance intention in the model, exceeding the influence of other variables in the model on acceptance intention. PEU has the weakest but most significant impact on acceptance intention; PR has a significant negative impact on PT and this negative impact is

TABLE 2: Description of questionnaire statistics ($N = 429$).

Personal attribute	Classification	Sample size (N)	Percentage
Gender	Male	209	48.80%
	Female	220	51.20%
Age (years)	<20	27	6.37%
	20–30	138	32.23%
	31–40	125	29.16%
	41–50	76	17.60%
	51–60	34	7.99%
	>60	29	6.65%
Education level	High school and below	64	15.03%
	College degree	118	27.56%
	Bachelor's degree	164	38.26%
	Master's degree and above	82	19.15%
Monthly income (yuan)	<3000 yuan	91	21.32%
	3000–5000 yuan	186	43.29%
	5000–10000 yuan	108	25.18%
	>10000 yuan	44	10.21%
Ownership of a private car	Own a private car	138	32.17%
	Not own a private car	291	67.83%

TABLE 3: Analysis results of reliability and convergent validity.

Latent variable	Item no.	Factor loading coefficient	CR	AVE
PU	PU1	0.79		
	PU2	0.72	0.798	0.568
	PU3	0.75		
PEU	PEU1	0.71		
	PEU2	0.74	0.778	0.538
	PEU3	0.75		
PR	PR1	0.66		
	PR2	0.72	0.839	0.572
	PR3	0.92		
	PR4	0.70		
PT	PT1	0.69		
	PT2	0.64	0.771	0.530
	PT3	0.76		
SN	SN1	0.73		
	SN2	0.84	0.828	0.617
	SN3	0.81		
HM	HM1	0.94		
	HM2	0.70	0.813	0.597
	HM3	0.68		
AI	AI1	0.77		
	AI2	0.90	0.862	0.677
	AI3	0.84		

also the strongest influence relationship among all the variables in the model; each variable of the model has a good ability to explain acceptance intention and the amount of variation explained by each variable reaches 52%. The following is a comparative analysis of the influence of each variable in the research model on acceptance intention and previous studies, as well as the value of this research.

5.1. Comparative Analysis with Previous Studies

5.1.1. Impacts of PU and PEU on AI. According to the evaluation results of the structural model, PU showed the third greatest impact on AI. It indicates that travelers pay more attention to the usefulness of travel information on social networks, and only when the travelers believe that the travel information is helpful to make travel decisions will

TABLE 4: Analysis results of discriminant validity.

	PU	PEU	PR	PT	SN	HM	AI
PU	**0.893**						
PEU	0.587	**0.733**					
PR	0.415	0.488	**0.756**				
PT	0.507	0.452	−0.648	**0.728**			
SN	0.468	0.416	0.505	0.612	**0.785**		
HM	0.411	0.682	0.494	0.594	0.629	**0.773**	
AI	0.594	0.451	−0.529	0.525	0.581	0.587	**0.823**
CR	0.798	0.778	0.839	0.771	0.828	0.813	0.862
AVE	0.568	0.538	0.572	0.530	0.617	0.597	0.677

Note. Values on the diagonal of the coefficient matrix in the table are the square roots of the AVE values, and values in the lower triangular matrix are the PCC between variables.

TABLE 5: Test results of goodness-of-fit.

Evaluation index	$\chi^2/\mathrm{d}f$	CFI	TLI	RMSEA	SRMR
Criterion	(1, 3)	>0.90	>0.90	<0.08	<0.08
Test value	2.36	0.91	0.91	0.03	0.05

TABLE 6: Standardized path analysis results of the structural model and test result of hypothesis.

Hypotheses	Estimate	SE	CR	P	Hypothesis test result
H1: PU ⟶ AI	0.65	0.105	6.190	***	H1 supported
H2: PEU ⟶ AI	0.38	0.081	4.691	**	H2 supported
H3: PEU ⟶ PU	0.31	0.066	4.697	**	H3 supported
H4: PR ⟶ AI	−0.71	0.090	−7.889	***	H4 supported
H5: PT ⟶ AI	0.66	0.133	4.962	***	H5 supported
H6: PR ⟶ PT	−0.78	0.086	−9.070	***	H6 supported
H7: SN ⟶ AI	0.53	0.086	6.163	***	H7 supported
H8: SN ⟶ PU	0.46	0.096	4.792	**	H8 supported
H9: SN ⟶ PR	0.48	0.102	4.706	***	H9 supported
H10: HM ⟶ AI	0.42	0.092	4.565	***	H10 supported
H11: HM ⟶ PU	0.23	0.194	1.186	0.076	H11 rejected
H12: HM ⟶ PEU	0.29	0.104	2.788	*	H12 supported

Note. Estimate represents the estimated value of the standardized path coefficient; SE represents the standard error term of the path coefficient; CR is the critical ratio, CR = estimate/SE, and a CR value greater than 1.96 indicates that a significance level of 0.05 has been reached; ***denotes $P < 0.001$, **denotes $P < 0.01$, and *denotes $P < 0.05$.
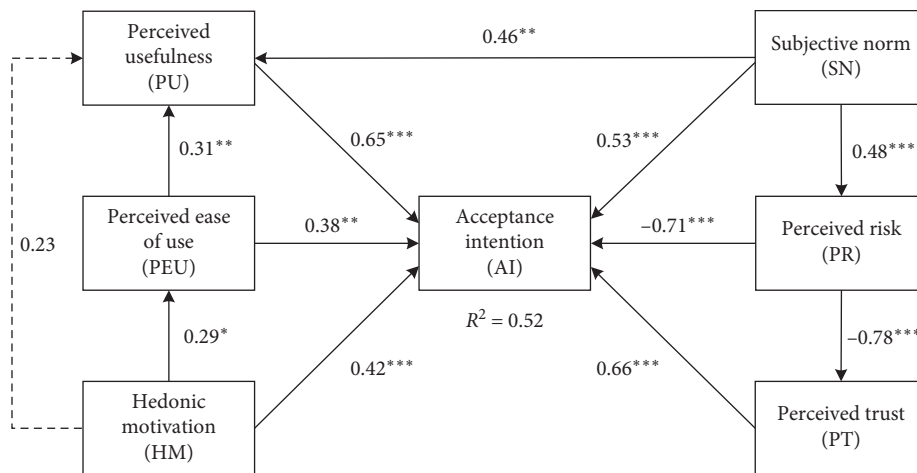
FIGURE 3: Analysis results of the structural model. *Note.* The solid arrows represent the paths with significant impacts, and the dotted arrow represents the path with insignificant impact; ***denotes $P < 0.001$, **denotes $P < 0.01$, and *denotes $P < 0.05$.

they be willing to accept the travel information, which is consistent with the findings of Chung et al. [14]. Although the impact of PEU on AI is found to be small, it cannot be ignored. It is worth noting that PEU here is the ease of process using the social network media platform to obtain travel information, which indicates that the easier the use of social network media platforms to obtain travel information is, the easier the travelers will accept it. In the study of Bilgihan et al., it was pointed out that PEU has a decisive influence on the sharing behavior of travel information on social networks [11].

5.1.2. Impacts of PR and PT on AI. In this study, PR showed the most significant impact on AI, and it is a negative impact. This observation indicates that travelers pay special attention to the security of travel information on social networks, which shows typical risk avoidance. The higher the risk perceived by travelers regarding this travel information is, the less willing the travelers will be to accept it. Tseng and Wang also pointed out the negative impact of PR on the use of travel information [46]. PT showed the second greatest impact on AI, only after PR. It indicates that the influence of trust factors on AI is also critical. The greater the trust travelers have in travel information on social networks, the stronger their AIs will be. This observation can be mutually supported by the research conclusions of Berhanu and Raj [15] and Li et al. [47].

5.1.3. Impacts of SN and HM on AI. SN represents the impacts of the surrounding social groups (friends, colleagues, etc.) on the traveler's own behavioral decisions, and it showed the fourth greatest impact on AI. It indicates that if the traveler's surrounding social groups have a very strong AI of travel information on social networks, it will be easier for him/her to accept the information. This is similar to the research findings of Bilgihan et al. on travel information sharing behavior on social networks [11]. There is no other research that explores the relationship between HM and AI of travel information on social networks. However, in the field of technology acceptance research of information systems, it has been confirmed by Venkatesh et al. that HM has a decisive impact on the intention to use information systems [28]. This study found that the impact of HM on AI of travel information on social networks exists but is not very strong, only higher than the impact of PEU on AI. It indicates that travelers do not pay much attention to the level of fun and pleasure involved in using the social network media platform to obtain travel information.

5.2. Research Value. This research explores the factors that influence travelers' acceptance intention of travel information on social networks. The data analysis results indicate the issues that traffic managers and travelers should pay attention to when publishing and sharing travel information on social media. PR and PT have the strongest influence on acceptance intentions. Therefore, when traffic managers and travelers publish travel information on social networks, they

should first ensure that the information is easy and quick to obtain, the information is updated in a timely and accurate manner, and the information cost is clear, to resolve the information recipient's concerns regarding the timeliness, accuracy, and cost. Simultaneously, attention should be paid to increasing the intensity of user and operator entry review and behavior supervision to ensure the security of obtaining and sharing travel information on social networks to increase the trust of travelers and make them more willing to accept travel information. PU and PEU also have a certain degree of influence on the acceptance intention, which shows that the practical value of travel information in social networks should be guaranteed and PEU should be improved from the aspects of information acquisition and software operation methods. Only useful and easy-to-use travel information will be readily accepted and used by travelers. The influence of HM and SN on acceptance intention is also significant, indicating that when travel information is released on social networks, attention should be paid to increase the interest in the information content and enhance its attraction to the user. In addition, relevant management departments should pay attention to publicity and promotion to increase the popularity of travel information on social networks and the scope of dissemination and focus on improving the usage efficiency of information through using the social influence power between users.

6. Conclusions

The purpose of this study is to explore the influencing factors that affect the travelers' AI of travel information on social networks and the relationships between the various factors. The research results can be expected to provide preliminary references for further research in this field. In the current study, it is found that PU, PEU, PR, PT, SN, and HM all have a significant impact on AI. Moreover, the impacts of the newly added variables—PR and PT—on the AI of travel information on social networks are especially prominent. It can be considered that the model constructed in this study demonstrates good applicability for the research of AI of travel information on social networks.

This study has two levels of significance in terms of theoretical enlightenment and practical application reference. At the theoretical level, this study extends the TAM and incorporates new variables that can characterize the features of travel information on social networks. An empirical basis has been provided for the effective combination of the PR, PT, HM, and TAM. The theoretical analysis methods have been enriched in the research field of technology acceptance. At the practical application level, this study explored the influencing factors that could affect the travelers' AI of travel information on social networks. It can provide reference opinions for social network media platform managers or users to publish travel information content and for the government to issue management policies for travel information on social networks. For example, it is found in this study that PR and PT are the most important influencing factors for AI. Therefore, the release, dissemination, and management of travel information on

social networks should strive to solve the traveler's risk concerns about the safety and personal privacy of travel information, enhance its accuracy and timeliness, and increase the trust of travelers to improve their AI.

However, this study inevitably has limitations that need to be resolved. First, the explanatory power of the current research model for the AI of travel information on social networks is 52%, which indicates that the model has not fully captured the influencing factors on the public AI. Whether other influencing factors in the research field of technology acceptance (such as perceived value and facilitating conditions) would significantly affect AI still requires further exploration. Second, the main purpose of this study is to explore the relationship between the degree of interpretation of the constructed extension model on the intention to accept travel information on social networks and the analysis of various influencing factors. Therefore, it does not consider the moderating effect of demographic attributes, such as traveler's gender, age, and education level, on the intention to accept travel information on social networks; this can be confirmed by subsequent studies in the future.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

References

[1] K. Simon, *Global Digital Report of 2020*, We Are Social and Hootsuite, London, UK, 2020.

[2] Z. Xiang and U. Gretzel, "Role of social media in online travel information search," *Tourism Management*, vol. 31, no. 2, pp. 179–188, 2010.

[3] K.-H. Yoo and U. Gretzel, "Influence of personality on travel-related consumer-generated media creationfluence of personality on travel-related consumer-generated media creation," *Computers in Human Behavior*, vol. 27, no. 2, pp. 609–621, 2011.

[4] N. Chung and C. Koo, "The use of social media in travel information search," *Telematics and Informatics*, vol. 32, no. 2, pp. 215–229, 2015.

[5] T. H. Rashidi, A. Abbasi, M. Maghrebi, S. Hasan, and T. S. Waller, "Exploring the capacity of social media data for modelling travel behaviour: opportunities and challenges," *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 197–211, 2017.

[6] E. Ampt and T. Ruiz, "Workshop synthesis: use of social media, social networks and qualitative approaches as innovative ways to collect and enrich travel data," *Transportation Research Procedia*, vol. 32, no. 1, pp. 93–98, 2018.

[7] A. M. Munar and J. K. S. Jacobsen, "Motivations for sharing tourism experiences through social media," *Tourism Management*, vol. 43, pp. 46–54, 2014.

[8] J. De Abreu E Silva, J. De Oña, and S. Gasparovic, "The relation between travel behaviour, ICT usage and social networks. the design of a web based survey," *Transportation Research Procedia*, vol. 24, pp. 515–522, 2017.

[9] N. A. Khan, M. A. Habib, and S. Jamal, "Investigation of weekend travel, social networking and transport-support application usage: a structural equation modeling approach," *Procedia Computer Science*, vol. 130, pp. 510–517, 2018.

[10] X. Chen and H. Deng, "A correlation analysis of information use, social networks and cooperation consciousness in travel behaviors," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 62, pp. 819–832, 2019.

[11] A. Bilgihan, A. Barreda, F. Okumus, and K. Nusair, "Consumer perception of knowledge-sharing in travel-related online social networks," *Tourism Management*, vol. 52, pp. 287–296, 2016.

[12] T. Oliveira, B. Araujo, and C. Tam, "Why do people share their travel experiences on social media?" *Tourism Management*, vol. 78, Article ID 104041, 2020.

[13] T. Oliveira, B. Araujo, and C. Tam, "Dataset for understanding why people share their travel experiences on social media: structural equation model analysis," *Data in Brief*, vol. 30, Article ID 105447, 2020.

[14] N. Chung, H. Han, and C. Koo, "Adoption of travel information in user-generated content on social media: the moderating effect of social presence," *Behaviour & Information Technology*, vol. 34, no. 9, pp. 902–919, 2015.

[15] K. Berhanu and S. Raj, "The trustworthiness of travel and tourism information sources of social media: perspectives of international tourists visiting Ethiopia," *Heliyon*, vol. 6, no. 3, Article ID e03439, 2020.

[16] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[17] J.-M. Tsai, M.-J. Cheng, H.-H. Tsai, S.-W. Hung, and Y.-L. Chen, "Acceptance and resistance of telehealth: the perspective of dual-factor concepts in technology adoption," *International Journal of Information Management*, vol. 49, pp. 34–44, 2019.

[18] R. Scherer, F. Siddiq, and J. Tondeur, "The technology acceptance model (TAM): a meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education," *Computers & Education*, vol. 128, pp. 13–35, 2019.

[19] T. Zhang, D. Tao, X. Qu et al., "Automated vehicle acceptance in China: social influence and initial trust are key determinants," *Transportation Research Part C: Emerging Technologies*, vol. 112, pp. 220–233, 2020.

[20] Y. Lu, S. Papagiannidis, and E. Alamanos, "Exploring the emotional antecedents and outcomes of technology acceptance," *Computers in Human Behavior*, vol. 90, pp. 153–169, 2019.

[21] R. A. Bauer, "Consumer behavior as risk taking," in *Dynamic Marketing for a Changing World, Proceedings of the 43rd*, R. S. Hancock, Ed., pp. 389–398, American Marketing Association: CHI, Chicago, IL, USA, 1960.

[22] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An integrative model of organizational trust," *The Academy of Management Review*, vol. 20, no. 3, pp. 709–734, 1995.

[23] D. Gefen, E. Karahanna, and D. W. Straub, "Trust and TAM in online shopping: an integrated model," *MIS Quarterly*, vol. 27, no. 1, pp. 51–90, 2003.

[24] M. Fishbein and I. Ajzen, "Belief, attitude, intention and behaviour: an introduction to theory and research," *Philosophy & Rhetoric*, vol. 41, no. 4, pp. 842–844, 1980.

[25] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: four longitudinal field studies," *Management Science*, vol. 46, no. 2, pp. 186–204, 2000.

[26] L. D. Kaczmarek, "Hedonic motivation," in *Encyclopedia of Personality and Individual Differences*, V. Zeigler-Hill and T. Shackelford, Eds., Springer, Cham, Switzerland, 2017.

[27] S. A. Brown and V. Venkatesh, "Model of adoption of technology in households: a baseline model test and extension incorporating household life cycle," *MIS Quarterly*, vol. 29, no. 3, pp. 399–426, 2005.

[28] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology," *MIS Quarterly*, vol. 36, no. 1, pp. 157–178, 2012.

[29] S. A. Nikou and A. A. Economides, "Mobile-based assessment: investigating the factors that influence behavioral intention to use," *Computers & Education*, vol. 109, pp. 56–73, 2017.

[30] J.-K. Liang, T. Eccarius, and C.-C. Lu, "Investigating factors that affect the intention to use shared parking: a case study of Taipei city," *Transportation Research Part A: Policy and Practice*, vol. 130, pp. 799–812, 2019.

[31] A. A. Abdallah, B. Abdullah, N. P. Rana, T. Kuttimani, and Y. K. Dwivedi, "Examining adoption of mobile internet in Saudi Arabia: extending tam with perceived enjoyment, innovativeness and trust," *Technology in Society*, vol. 55, pp. 100–110, 2018.

[32] J. R. Priester, "The use of structural equation models in consumer psychology: a methodological dialogue on its contributions, cautions, and concerns," *Journal of Consumer Psychology*, vol. 20, no. 2, pp. 205–207, 2010.

[33] H. Camgoz-Akdag and S. Zaim, "Education: a comparative structural equation modeling study," *Procedia—Social and Behavioral Sciences*, vol. 47, pp. 874–880, 2012.

[34] F. Boccia and P. Sarnacchiaro, "Structural equation model for the evaluation of social initiatives on customer behaviour," *Procedia Economics and Finance*, vol. 17, pp. 211–220, 2014.

[35] R. B. Kline, *Principles and Practice of Structural Equation Modeling*, Guilford Press, New York, NY, USA, 2nd edition, 2005.

[36] K. A. Bollen, *Structural Equations with Latent Variables*, Wiley, New York, NY, USA, 1989.

[37] C. Fornell and D. F. Larcker, "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research*, vol. 18, no. 1, pp. 39–50, 1981.

[38] R. P. Bagozzi and Y. Yi, "On the evaluation of structural equation models," *Journal of the Academy of Marketing Science*, vol. 16, no. 1, pp. 74–94, 1988.

[39] J. F. J. Hair, R. E. J. Anderson, R. L. Tatham, and W. C. Black, *Multivariate Data Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 5th edition, 1998.

[40] M. R. A. Hamid, W. Saml, and M. H. Mohmad-Sldek, "Discriminant validity assessment: use of Fornell & Larcker criterion versus HIMT criterion," *Journal of Physics Conference Series*, vol. 890, no. 1, Article ID 012163, 2017.

[41] D. Iacobucci, "Structural equations modeling: fit indices, sample size, and advanced topics," *Journal of Consumer Psychology*, vol. 20, no. 1, pp. 90–98, 2010.

[42] T. Niemand and R. Mai, "Flexible cutoff values for fit indices in the evaluation of structural equation models," *Journal of the Academy of Marketing Science*, vol. 46, no. 6, pp. 1148–1172, 2018.

[43] L. T. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 6, no. 1, pp. 1–55, 1999.

[44] J. J. Hoogland and A. Boomsma, "Robustness studies in covariance structure modeling," *Sociological Methods & Research*, vol. 26, no. 3, pp. 329–367, 1998.

[45] P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Psychology Press, London, UK, 2014.

[46] S.-Y. Tseng and C.-N. Wang, "Perceived risk influence on dual-route information adoption processes on travel websites," *Journal of Business Research*, vol. 69, no. 6, pp. 2289–2296, 2016.

[47] S. R. Li, Y. X. Ong, and N. Ito, "Credibility in question: travel information adoption among Chinese consumers in Canada and Singapore," in *Information and Communication Technologies in Tourism 2020*, J. Neidhardt and W. Wörndl, Eds., Springer, Cham, Switzerland, 2020.

Research Article

Evaluation Method of Wind Speed Time-Shifting Characteristics at Multiple Scales and Its Application in Wind Power System

Han Wang , **Shuang Han** , **Yongqian Liu, and Aimei Lin**

State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Renewable Energy, North China Electric Power University (NCEPU), Beijing 102206, China

Correspondence should be addressed to Shuang Han; hanshuang1008@sina.com

The wind speed sequences at different spatial positions have a certain spatiotemporal coupling relationship. It is of great significance to analyze the clustering effect of the wind farm(s) and reduce the adverse impact of large-scale wind power integration if we can grasp this relationship at multiple scales. At present, the physical method cannot optimize the time-shifting characteristics in real time, and the research scope is concentrated on the wind farm. The statistical method cannot quantitatively describe the temporal relationship and the speed variation among wind speed sequences at different spatial positions. To solve the above problems, a quantification method of wind speed time-shifting characteristics based on wind process is proposed in this paper. Two evaluation indexes, the delay time and the decay speed, are presented to quantify the time-shifting characteristics. The effectiveness of the proposed method is verified from the perspective of the correlation between wind speed sequences. The time-shifting characteristics of wind speed sequences under the wind farms scale and the wind turbines scale are studied, respectively. The results show that the proposed evaluation method can effectively achieve the quantitative analysis of time-shifting and could improve the results continuously according to the actual wind conditions. Besides, it is suitable for any spatial scale. The calculation results can be directly applied to the wind power system to help obtain the more accurate output of the wind farm.

1. Introduction

A high proportion of renewable energy is one of the basic characteristics of the future electric power system, and wind power is an important component of renewable energy [1]. With the rapid expansion of wind power integration, the volatility of wind power has a more adverse impact on the grid dispatching, safe and stable operation of the electric power system, and its power quality [2]. In fact, there is a certain spatiotemporal coupling relationship among the wind speed sequences at different spatial points [3]. If we can grasp this coupling relationship at multiple temporal-spatial scales, it is of great significance to analyze the clustering effect [4], deal with the ramp events and extreme events, obtain the real output of wind farm(s) [5], and so forth, which can effectively reduce the negative effect of the wind power integration [6].

At present, the research algorithms on the spatiotemporal coupling relationship among wind speed time series can be divided into two categories, the model-based physical method and the data-based statistical method, according to the driving mode.

The physical method takes the wind speed data at a certain point as input and considers factors such as terrain, roughness, altitude, and flow to establish the aerodynamic model to describe the coupling relationship among the wind speeds at different spatial locations. Zheng [7] assumed that the wind spreads at the speed of the starting point and established the motion equation of wind speed to obtain the average delay time between the wind speed time series at two spatial positions. She et al. [8] used the rotor diameter and geographic coordinate matrix of each wind turbine in the wind farm as the inherent characteristics and took the wind

speed and direction of the wind tower, rotor speed, and pitch angle of each wind turbine as inputs to establish the wake effect model and the time-lag effect model; then the dynamic spatial relationship among wind speed time series at each wind turbine location was obtained. Zeng et al. [9] first developed the continuous partial differential equation of each wind turbine by calculating the wake effects, according to the specific spatial locations and distributions of correlated wind turbines. Then, the finite volume method is used to differentiate the differential equation at each grid point and derive the spatial correlation matrix of wind speed. Finally, the wind speed at each wind turbine location is obtained by solving the above differential equation through the given boundary conditions. Ye et al. [10] proposed a novel spatial correlation matrix by using computational fluid dynamics to mathematically describe the correlations among wind speeds at different wind turbine locations.

The physical method studies the temporal and spatial distribution and the coupling relationship of wind speeds by establishing the aerodynamic model among the wind speeds at different spatial positions. However, this aerodynamic model cannot be optimized in real time according to the actual wind conditions. Besides, when the spatial scale increases, the complexity of the constructed model increases sharply, which leads to the increasing difficulty of modeling and solving the model simultaneously. But it is difficult to ensure the accuracy of the aerodynamic model if the factors affecting wind speed are simplified. Therefore, the application of this method is mainly concentrated in the interior of the wind farm, which is difficult to expand to a larger spatial scale.

The statistical method uses the actual wind data at different spatial locations to analyze the spatiotemporal coupling relationship among the wind speeds. The research methods can be mainly divided into the following three types:

(1) The spatiotemporal coupling relationship is studied by calculating the distribution of wind speeds at different spatial locations, from the perspectives of average wind speed, wind speed frequency distribution, wind shear exponent, and so forth [11]. Liu et al. [12] analyzed the spatiotemporal distributions of the wind resource at the hub height in China from the aspects of average wind speed and average wind power density, based on the large network of wind speed observations, including 2430 meteorological stations from 2006 to 2015. The results show that northeast China has the highest wind speed, while South China has the lowest wind speed. Li and Yu [13] conducted the statistical analysis of the onshore, nearshore, and offshore wind energy conditions from the aspects of Weibull shape and scale factors, average turbulence intensity, and average wind power density. The results indicate that the average wind speed at the offshore location is 1-2 times higher than that nearby onshore area at the height of 50 m above ground level. Bosch et al. [14] estimated the distributions of global offshore wind energy potential from the aspects of average wind speed,

mean capacity factor which could be built across the viable offshore area, and generation potential per year. The results show that the wind speed increases with the distance to the coast.

(2) The spatiotemporal coupling relationship is studied by establishing the mapping relationship among wind speed time series at different spatial positions through interpolation, neural network, principal component analysis, and so forth [15]. Philippopoulos and Deligiorgi [16] used two feedforward neural network architectures to establish the relationship among hourly wind speeds at six meteorological sites. An insight into the underlying input-output function approximation of the neural networks is obtained by examining their ability to incorporate the mean wind variability characteristics. Miranda and Dunn [17] proposed a spatially correlated wind speed model to characterize the wind resources between various sites by using the multivariate time-series approach with actual wind speed data in different zones. Park and Jang [18] presented a modified Korean parameter-elevation regression on independent slopes model for spatial interpolation of monthly wind speeds. The results show that the proposed model can perform more appropriately to represent the phenomena where similar wind speeds appear continuously along ridges and coastlines, compared with the cokriging model.

(3) The correlation coefficient is used to study the correlation characteristics of wind speed time series at different spatial positions, which is an important part of the spatiotemporal coupling relationship. The calculation method can be divided into the following categories. One is to directly calculate the correlation coefficient based on the wind speed time series [19]. Another is to use the Copula function to fit the joint probability distribution of the wind speed sequences and then calculate the correlation coefficient based on the fitted Copula function [20]. The commonly used correlation coefficients include the Pearson linear correlation coefficient [21], Kendall rank correlation coefficient [22], and cross-correlation coefficient [23].

The statistical method studies the spatiotemporal coupling relationship of wind speed sequences at different spatial positions based on the actual wind data. The obtained results of this method can be continuously improved according to actual conditions, and the research method can be applied to any spatial scale. However, only the correlation relationship between wind speed sequences is quantified in this method. Compared with the physical method, the temporal relationship and the law of speed variation are both implicit in the spatial distribution, and the mapping relationship is not quantified yet.

In conclusion, when the physical method is used to study the spatiotemporal coupling relationship among wind speeds at different spatial positions, the established aerodynamic model cannot be optimized in real time according

to the actual wind conditions, and the application of this method mainly concentrated in the interior of the wind farm; it is difficult to expand to a larger spatial scale. When the statistical method is applied, only the correlation relationship of wind speed sequences is quantified, and the temporal relationship and the law of speed variation among wind speed sequences at different spatial positions have not been quantified yet. In order to solve the above problems, a quantification method of time-shifting characteristics based on wind process is proposed in this paper. The results of the proposed method can be continuously improved according to the actual wind conditions and could be applied to any spatial scale. The main contributions of this paper are as follows:

(1) The concept of wind process is introduced to study the time-shifting characteristics. In order to cope with the complexity and variability of wind conditions, the wind sequences at different spatial positions are divided according to the certain time window at first, and then the time-shifting characteristics are studied based on the wind processes.

(2) Two evaluation indexes, the delay time and the decay speed, are proposed to quantify the time-shifting characteristics of the temporal relationship and the speed variation in each wind process.

(3) The time-shifting characteristics are studied at two spatial scales, the wind farms scale and the wind turbines scale.

(4) The improved clustering method of k-means, x-means, is employed to extract the typical wind processes to explore the time-shifting characteristics among wind speeds under various typical wind processes. The obtained results can be directly used in the wind power system.

(5) The validity and practicability of the proposed method are verified by applying the calculation results of time-shifting among wind speed sequences at different spatial positions to the wind power system.

The remainder of this paper is organized as follows. Section 2 describes the research process of wind speed time-shifting. Section 3 studies the wind speed time-shifting characteristics under two spatial scales, the wind farms scale and the wind turbines scale. Section 4 verifies the effectiveness and practicability of the proposed time-shifting evaluation method. Section 5 concludes this paper.

2. Research Process of Wind Speed Time-Shifting

This section proposes an evaluation method of the time-shifting characteristics among wind speed sequences at different spatial positions. The research process is as follows.

2.1. Outline of the Research Process on Wind Speed Time-Shifting. First, the wind sequence at each point is pre-processed. The wind sequence is divided according to a certain time window to obtain the wind processes at each point. Then some wind processes are screened out to compose a new wind sequence that can be used to study the time-shifting characteristics by judging whether the main wind direction in the wind process is consistent with the relative position of points. For example, the wind sequence at point p can be expressed as a composed sequence of wind process α_p, wind process β_p, wind process m_p, and so forth.

Then, two evaluation indexes, the delay time (Δt) and the decay speed (Δv), are proposed to study the time-shifting characteristics between wind processes at two points. The firefly optimization algorithm is used to calculate the proposed indexes, in which the constraint conditions are the distance between two points and the actual wind conditions; the optimal objective is the minimum Euclidean distance between the corresponding wind processes. Taking the wind process α_p and the wind process α_q as an example, the wind process α_p at point p and the wind process α_q at point q can be expressed as $v_p^\alpha(t - x), \ldots, v_p^\alpha(t), \ldots, v_p^\alpha(t + y)$ and $v_q^\alpha(t - x), \ldots, v_q^\alpha(t), \ldots, v_q^\alpha(t + y)$, respectively. The distance between two points and the wind conditions of the wind process α_p at point p are set as the constraint conditions, and the wind process α_q is offset through the time axis and the speed axis to achieve the objective of the minimum Euclidean distance between the wind process α_p at point p and the new wind process α_q' at point q. Then the corresponding delay time Δt_α, decay speed Δv_α, and the new wind process α_q' ($v_q^\alpha(t - x + \Delta t_\alpha) + \Delta v_\alpha, \ldots, v_q^\alpha(t + \Delta t_\alpha) + \Delta v_\alpha, \ldots, v_q^\alpha(t + y + \Delta t_\alpha) + \Delta v_\alpha$) could be obtained at the same time.

Finally, the time-shifting characteristics of wind speed sequences at two points are studied based on the proposed indexes. The delay time and the decay speed of each wind process at point p and point q are calculated, and then the statistical analysis method is used to analyze the time-shifting. Furthermore, the improved clustering method of k-means, x-means, is employed to extract the typical wind processes, and the time-shifting characteristics of different typical wind processes are studied so that the research results can be applied to the actual planning and operation of the electric power system.

In order to verify the effectiveness of the proposed time-shifting analysis method, the wind speed sequence at point q is reconstructed after considering the time-shifting characteristics. t-Copula function is used to fit the joint probability distribution between wind speed sequences under two conditions, that is, whether the time-shifting is considered. Then the Kendall rank correlation coefficient τ, the Spearman rank correlation coefficient ρ_s, and the tail correlation coefficient λ are calculated based on the t-Copula function. The correlation coefficients increase with the increasing similarity between two wind speed sequences.

The outline of the research process on wind speed time-shifting is shown in Figure 1.

It should be noted that, due to the fact that the wind speed can gradually recover within a certain distance after passing through the wind turbine(s) or the wind farm(s), the evaluation indexes for time-shifting analysis should be different according to the distance and the installed capacity of wind power, that is, whether the decay speed should be
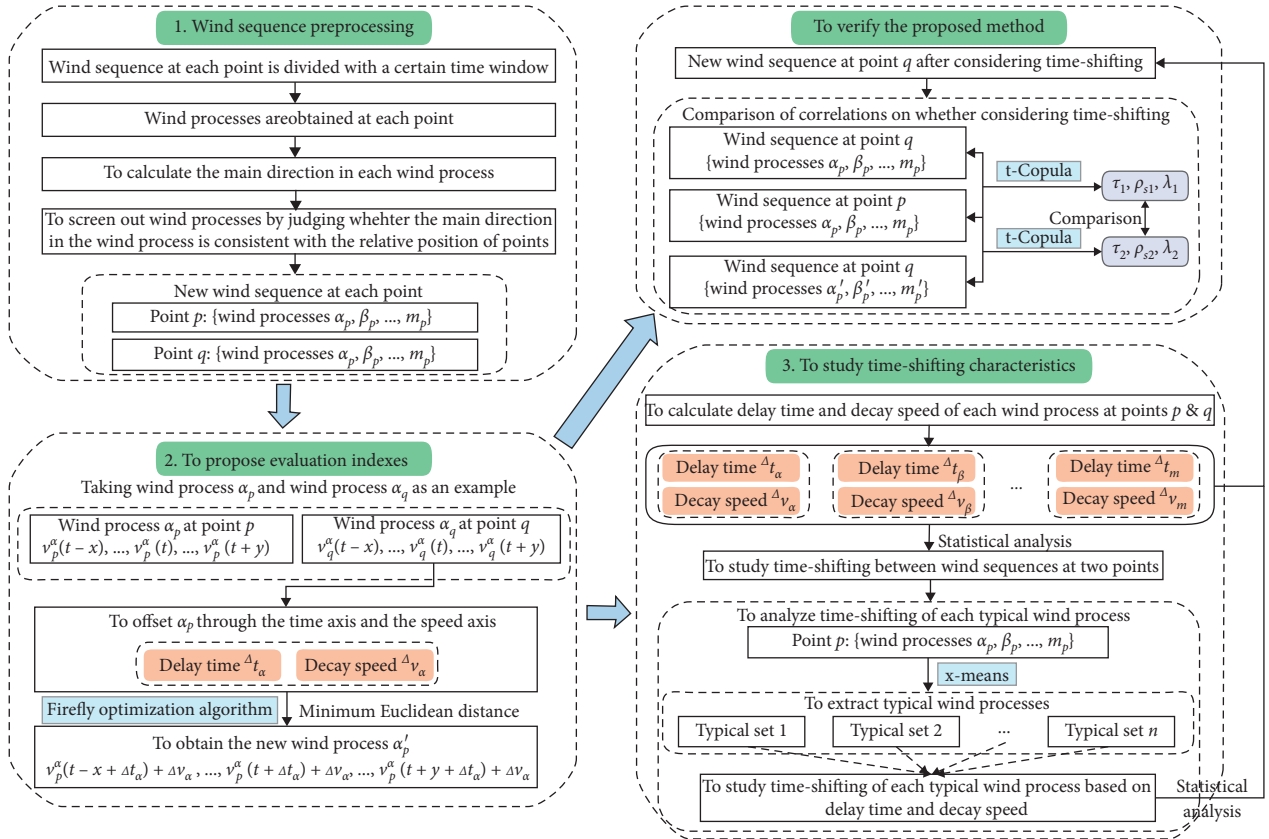
Figure 1: Outline of the research process on wind speed time-shifting.

contained when studying the time-shifting characteristics of wind speed sequences at different spatial positions. According to the research results of Risø Laboratory at the Technical University of Denmark, even for a large-scale wind farm of $30 \, \text{km} \times 30 \, \text{km}$ at the upwind direction, the weakening effect on the wind speed at the downwind direction could vanish after a distance of 30 km to 60 km [24]. Therefore, it can be considered that there is no influence of the decay speed among wind speed sequences at different locations when the distance between the wind farms exceeds 60 km.

2.2. Calculation Method of the Proposed Indexes Based on Firefly Optimization Algorithm.

The firefly optimization algorithm [25] is applied to calculate the proposed indexes, the delay time and decay speed. Each firefly is regarded as a point in the algorithm, and the fireflies are distributed randomly in the search space at first. There are two parameters in the firefly optimization algorithm, the brightness and the attractiveness. The brightness of the firefly is decided by the objective function, which indicates the position and decides the next moving direction. The attractiveness between two fireflies is inversely proportional to the distance, which decides the next moving distance. The firefly is attracted and moves toward the relatively brighter one according to the brightness and attractiveness; then the fireflies gather through the movement. In the optimization process, the brightness and attractiveness are updated

iteratively until the optimization objective is achieved. The specific steps of the firefly algorithm are as follows.

Step (a): to set the initial parameters of the algorithm, including the maximum attraction $\beta 0$, the light attraction coefficient γ, the step factor α, the maximum number of iterations, and the number of fireflies.

Step (b): to initialize the position of fireflies.

Step (c): to calculate the objective function of each firefly and regard it as the maximum fluorescent brightness I_0.

Step (d): to define the relative brightness between fireflies, as shown in the following equation:

$$I = I_0 e^{-\gamma r_{ij}}, \tag{1}$$

where r_{ij} is the relative distance between firefly i and firefly j and the light attraction coefficient γ reflects the weakening characteristics with the increase of the distance and the absorption of the propagation medium.

to define the attractiveness between fireflies, as shown in the following equation:

$$\beta = \beta_0 e^{-\gamma r_{ij}^2}. \tag{2}$$

to calculate the relative brightness and attractiveness between fireflies according to equations (1) and (2) and sort them by the relative brightness.

Step (e): to update the position of fireflies, as shown in the following equation:

$$x_i = x_i + \beta\left(x_j - x_i\right) + \alpha\varepsilon_i, \tag{3}$$

where x_i and x_j are the positions of firefly i and firefly j, respectively. The step factor α is a disturbance term, which is added to enlarge the search area and avoid the optimization algorithm falling into partial convergence prematurely. In general, $\alpha \in [0, 1]$ and $\varepsilon_i \in [-0.5, 0.5]$.

Step (f): to repeat Step (c) to Step (e) until the maximum iterations number is reached, and then the optimal results can be obtained.

2.3. Correlation Analysis Based on t-Copula Function.

The correlation analysis between wind speed time series at different spatial points is an important method to test the validity of the proposed time-shifting indexes and evaluate the effect of the time-shifting characteristics. At present, most researchers use the linear correlation coefficient to measure the correlation relationship between wind speed sequences. However, when the distribution of the variable follows the nonnormal distribution or nonelliptic distribution, it may get wrong conclusions [26]. The distribution of wind speed is more consistent with Weibull distribution, so it is more reasonable to use the rank correlation coefficient to measure the correlation relationship between wind speed sequences. Furthermore, compared with the normal weather conditions, the wind power output in the extreme wind conditions has a greater impact on the electric power system, and the output correlation among wind turbines or wind farms should be paid more attention in these weather conditions.

t-Copula [27] function is one of the commonly used Copula functions. It has a thicker tail and is more sensitive to the changes of the tail correlation between random variables, which can capture the symmetric tail correlation. Therefore, the t-Copula function is used to fit the joint probability density of wind speed sequences under two conditions in this paper, including the original sequences and the sequences considering time-shifting characteristics. Then, the Kendall rank correlation coefficient, Spearman rank correlation coefficient, and tail correction coefficient can be calculated to verify the validity of the proposed time-shifting indexes and evaluate the effect of the time-shifting characteristics. The specific steps of the correlation analysis based on t-Copula function are as follows.

The binary t-Copula function is shown in equation (4), where the linear correlation coefficient between two variables is ρ and the degree of freedom is k. The density function and the distribution function of t-Copula are depicted in Figure 2:

$$C^t\left(\mu, \nu; \rho, k\right) = \int_{-\infty}^{t_k^{-1}(\mu)} \int_{-\infty}^{t_k^{-1}(\nu)} \frac{1}{2\pi\sqrt{1 - \rho^2}}$$
$$\cdot \left[1 + \frac{s^2 - 2\rho st + t^2}{k\left(1 - \rho^2\right)}\right]^{-(k+2)/2} ds dt, \tag{4}$$

where t_k^{-1} is the inverse function of the distribution function of unary t distribution with freedom degree of k.

The calculation equations of Kendall rank correlation coefficient τ, Spearman rank correlation coefficient ρ_s, and tail correlation coefficient λ based on t-Copula function are shown in the following equations:

$$\tau = 4 \int_0^1 \int_0^1 C^t\left(\mu, \nu\right) dC^t\left(\mu, \nu\right) - 1$$
$$= \frac{2\arcsin\rho}{\pi}, \tag{5}$$

$$\rho_s = 12 \int_0^1 \int_0^1 C^t\left(\mu, \nu\right) d\mu d\nu - 3$$
$$= \frac{6\arcsin\left(\rho/2\right)}{\pi}, \tag{6}$$

$$\lambda = \lim_{\mu \longrightarrow 1^-} \frac{1 - 2\mu + C^t\left(\mu, \mu\right)}{1 - \mu}$$
$$= \lim_{\mu \longrightarrow 0^+} \frac{C^t\left(\mu, \mu\right)}{\mu}$$
$$= 2 - 2t_{k+1}\left(\frac{\sqrt{k+1}\sqrt{1 - \rho}}{\sqrt{1 + \rho}}\right), \tag{7}$$

where $t_{k+1}\left(\left(\sqrt{k+1}\sqrt{1 - \rho}\right)/\sqrt{1 + \rho}\right)$ is the value of the distribution function of unary t with freedom degree of $k + 1$ at $\left(\sqrt{k+1}\sqrt{1 - \rho}\right)/\sqrt{1 + \rho}$.

2.4. Extraction of Typical Wind Processes Based on the x-Means Clustering Method.

The improved algorithm of k-means algorithm, x-means [28], is adopted to extract the typical wind processes when analyzing the time-shifting characteristics of each typical wind process. Compared with k-means algorithm, x-means algorithm is not required to give the specific value of the clustering number n in advance but only needs to give a range within which the true n reasonably lies. The clustering result is not only the set of centroids but also the value for n which scores best in this range. The specific steps of x-means are as follows.

Step (a): to input m wind processes to be clustered and set the range of clustering number as $[n_{\min}, n_{\max}]$, where $n_{\max} \leq m$.

Step (b): to select $n_{\min}$ initial clustering centers from original wind processes randomly.

Step (c): to calculate the Euclidean distance between other wind processes and each clustering center and assign the corresponding wind process to the nearest category.

Step (d): to calculate the average value of wind processes in each clustering set and use the results as the new clustering centers.

Step (e): to repeat Step (c) and Step (d) until the clustering centers no longer change, and the optimal clustering results can be obtained with the clustering number of $n_{\min}$.
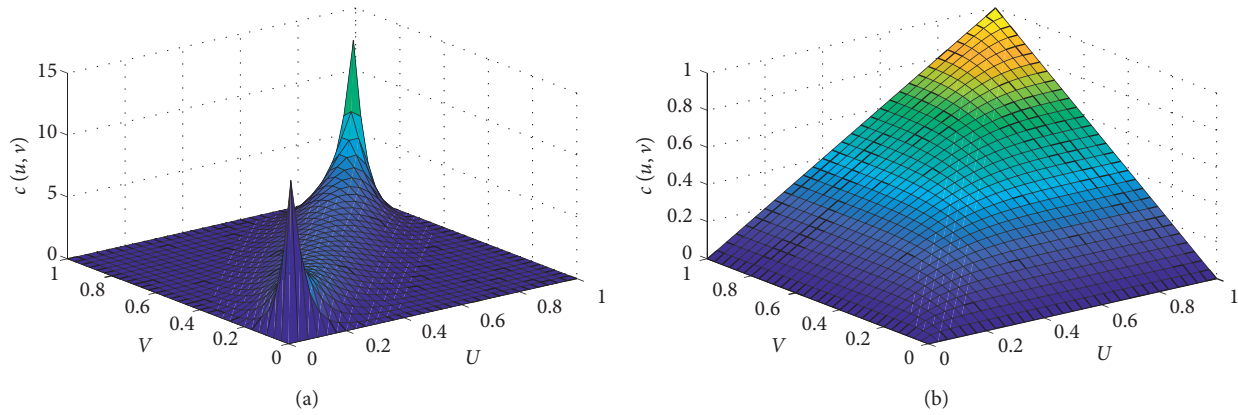
Figure 2: The density function and distribution function of binary t-Copula. (a) Density function. (b) Distribution function.

Step (b) to Step (e) are the main steps of k-means algorithm when the clustering number is $n_{\min}$.

Step (f): to apply the 2-means algorithm in each clustering set, and the clustering results are evaluated according to the Bayesian Information Criterion (BIC). The calculation method of BIC can be seen in [28]. If the BIC score is larger when 2-means algorithm is used in the clustering set, the set will be divided into two sets; otherwise, the clustering set will be retained.

Step (g): to repeat Step (f) in each new clustering set until the clustering number n is no longer increased or $n \geq n_{\max}$.

3. Case Study and Analysis

This section analyzes the time-shifting characteristics of wind speed time series in different points under the wind farms scale and the wind turbines scale, respectively. The data includes wind speed and wind direction. In the wind farms scale, the data is the mesoscale data located in South Gobi of Mongolian; the time range is from January 1, 2004, to October 31, 2018, and the temporal resolution is 1 h. All points are located in the flat terrain in this case. In the wind turbines scale, the data is the real operation data of one wind farm in Northern China; the time range is from January 1, 2014, to December 30, 2014, and the temporal resolution is 1 min. All points are located in the hilly terrain in this case.

3.1. Time-Shifting under Wind Farms Scale. The wind sequence at each point is preprocessed at first. The wind sequence is divided into different wind processes with a time window of 1 day in this case, and then the wind process sets are obtained at each point. The relative distribution of each point is shown in Figure 3, so the wind processes with the main wind direction between 0265° and 275° at the corresponding point should be screened out for the time-shifting study. For example, when studying the time-shifting characteristics of wind speed sequences from point 1 to other points (2–5), the wind processes with the main direction between 265° and 275° at point 1 and the contemporaneous wind processes at other points should be screened out
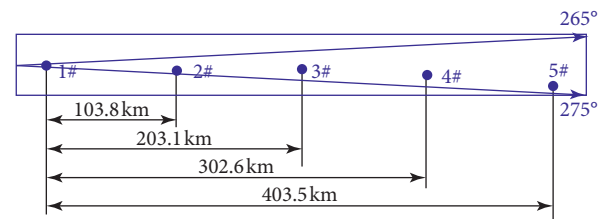
Figure 3: The relative distribution of points (wind farms scale).

simultaneously. Then, the new wind speed sequence that can be used to study the time-shifting characteristics at each point is reconstructed. The average speeds of new wind speed sequences at points 1 to 4 are 8.32 m/s, 8.44 m/s, 6.68 m/s, and 6.06 m/s, respectively.

Then, the time-shifting characteristics of new wind speed sequences between point 1 and points 2 to 5, between point 2 and points 3 to 5, between point 3 and points 4 to 5, and between point 4 and point 5 are studied, respectively. Since the distances between two points are all more than 100 km, only the delay time is used as the time-shifting evaluation index in this case. The delay time distributions of wind speed time series between different points are shown in Figure 4. It can be seen that, with the increase of distance, the median delay time between wind speed sequences at two points increases, and the frequency distribution becomes more dispersed. The delay time of the wind speed sequence from point 1 to other points (2–5) varies in 0 h–26 h, 0 h–31 h, 0 h–49 h, and 0 h–78 h, respectively. The average delay time between wind speed sequences at two points has a correlation relationship with the average wind speed at the starting point and the distance between two points, but it is not completely equal to the quotient between these two variables. Take the wind speed sequence at point 1 as an example; if the distances between point 1 and points 2 to 5 are divided by the average wind speed at point 1, the corresponding results are 3 h, 7 h, 10 h, and 13 h, respectively. When the proposed time-shifting analysis method is applied, the average delay time of wind speed sequence from point 1 to other points (2–5) is 2 h, 5 h, 8 h, and 14 h, respectively. This is because even if the average speed of two wind processes is the same, the wind conditions can also be quite
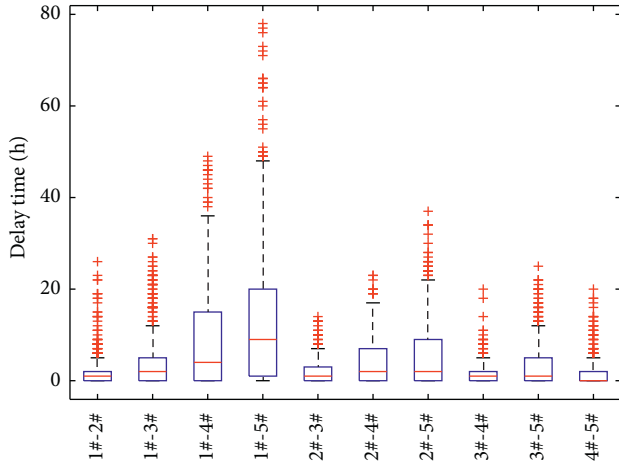
FIGURE 4: Delay time distributions of wind speed sequences between different points (wind farms scale).

FIGURE 5: Average delay time between different points in each wind speed bin (wind farms scale).

FIGURE 6: Average delay time in each wind process set between point 1 and other points (wind farms scale).

different, and the time-shifting characteristics are various under different wind processes. Therefore, it is more reasonable to calculate the delay time between wind speed sequences at two points based on the similarity of wind processes.

In order to preliminarily analyze the mapping relationship among the time-shifting characteristics of wind speed sequences at different points, the distances between points, and the actual wind conditions, the wind processes at point 1 are divided according to the average wind speed of each process, with the interval of 1 m/s. The average delay time of point 1 to points 2–5 is calculated in each wind speed bin, respectively, and the results are depicted in Figure 5. As can be seen, the delay time increases with the distance when the wind speed is almost the same and decreases with the wind speed when the distance is basically the same. However, since the wind speed is inevitably affected by various factors such as terrain and obstacles when it propagates from one point to another point in space, the propagation time is not equal to the quotient of distance and its value. Furthermore, the wind process is the time series composed of different wind speeds, so the propagation law of the wind process is more complicated. This is why some dots in Figure 5 look like the "abnormal dots," such as the average delay time from point 1 to point 2 when the average speed of the wind process is between 8 m/s and 9 m/s and the average delay time from point 1 to point 3 when the average speed of the wind process is between 11 m/s and 12 m/s. Therefore, it is necessary to further analyze the time-shifting characteristics of different wind processes at different spatial scales.

x-Means algorithm is applied to cluster the wind processes at point 1 to extract typical wind processes, and then the typical wind process sets are constructed. The range of clustering number is set as in [3, 15] when using x-means algorithm. The optimal clustering number is 24 in this case. The average delay time of each wind process set from point 1 to other points (2–5) is calculated, respectively, as shown in Figure 6. Six typical wind process sets are selected randomly, as depicted in Figure 7. The red solid line is the clustering center and the black-dashed lines are the actual wind
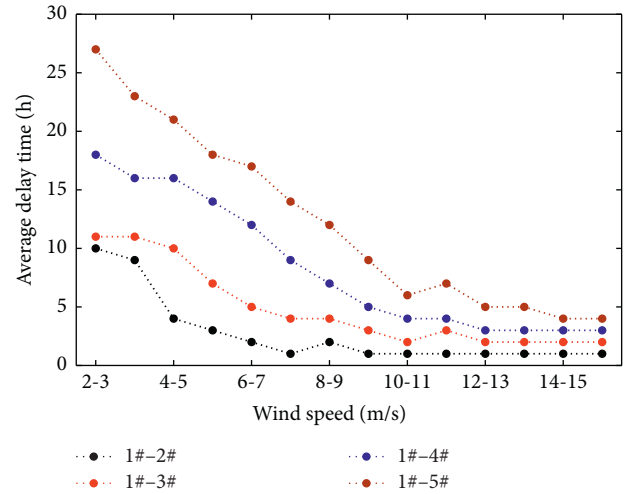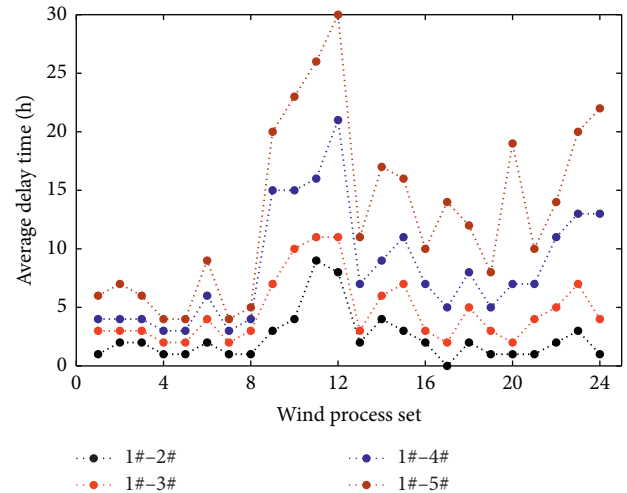
processes of each set. The average delay time from point 1 to other points and the average wind speed of these six typical wind process sets are listed in Table 1.

The new wind speed sequence at each point considering the time-shifting characteristics is reconstructed, respectively, according to the calculation results of the delay time. The wind speed time series of whether considering the time-shifting is shown in Figure 8. As can be seen, the wind speed sequences at different points could be more similar if the time-shifting characteristics are considered.

In order to further verify the validity of the proposed method, the correlation analysis is used to test the effectiveness of the time-shifting characteristics. t-Copula function is used to fit the original wind speed sequences of two points and the new wind speed sequences considering the time-shifting characteristics of two points, respectively. Then the Kendall correlation coefficient τ, Spearman correlation coefficient ρ_s, and tail correlation coefficient λ are
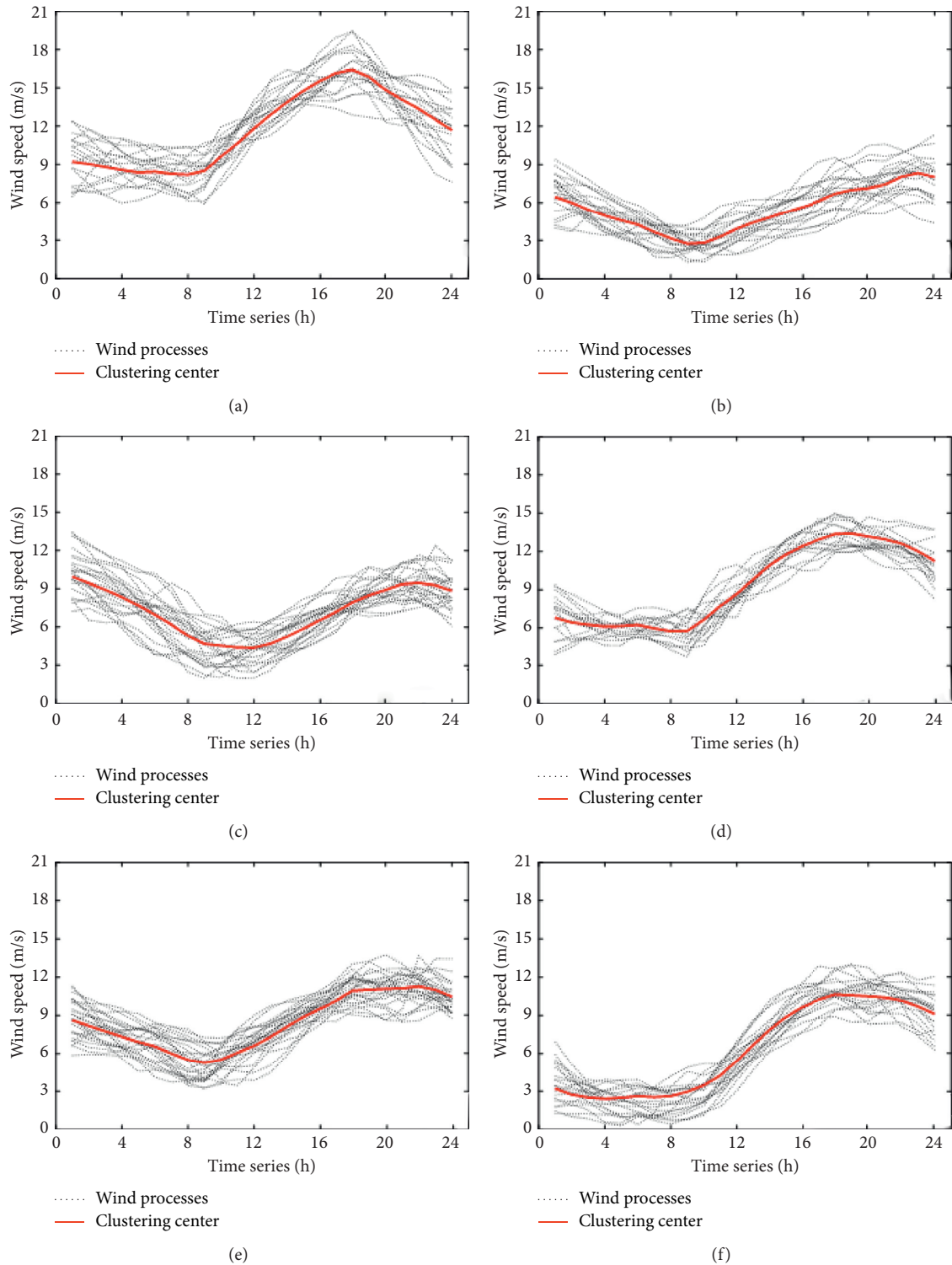
FIGURE 7: Six typical wind process sets at point 1 (wind farms scale). (a) Typical wind process set 6. (b) Typical wind process set 10. (c) Typical wind process set 15. (d) Typical wind process set 18. (e) Typical wind process set 22. (f) Typical wind process set 23.

calculated based on the fitted function, and the results are listed in Table 2. It can be seen that the correlation of wind speed sequences at different points is enhanced significantly when the time-shifting characteristics are considered. Since all points are located in the flat terrain in this case, the effect

of time-shifting becomes more obvious as the distance increases. For example, taking point 1 as the starting point, if the time-shifting characteristics are considered, the correlations of wind speed sequences between point 1 and points 2 to 5 are increased by 0.048, 0.097, 0.123, and 0.227,

TABLE 1: Average delay time in six wind process sets between point 1 and other points (wind farms scale).

Typical wind process set	Average delay time (h)				Average wind speed (m/s)
	1#-2#	1#−3#	1#−4#	1#−5#	
6	2	4	6	9	11.84
10	4	10	15	23	5.54
15	3	7	11	16	7.27
18	2	5	8	12	9.49
22	2	5	11	14	8.37
23	3	7	13	20	6.26
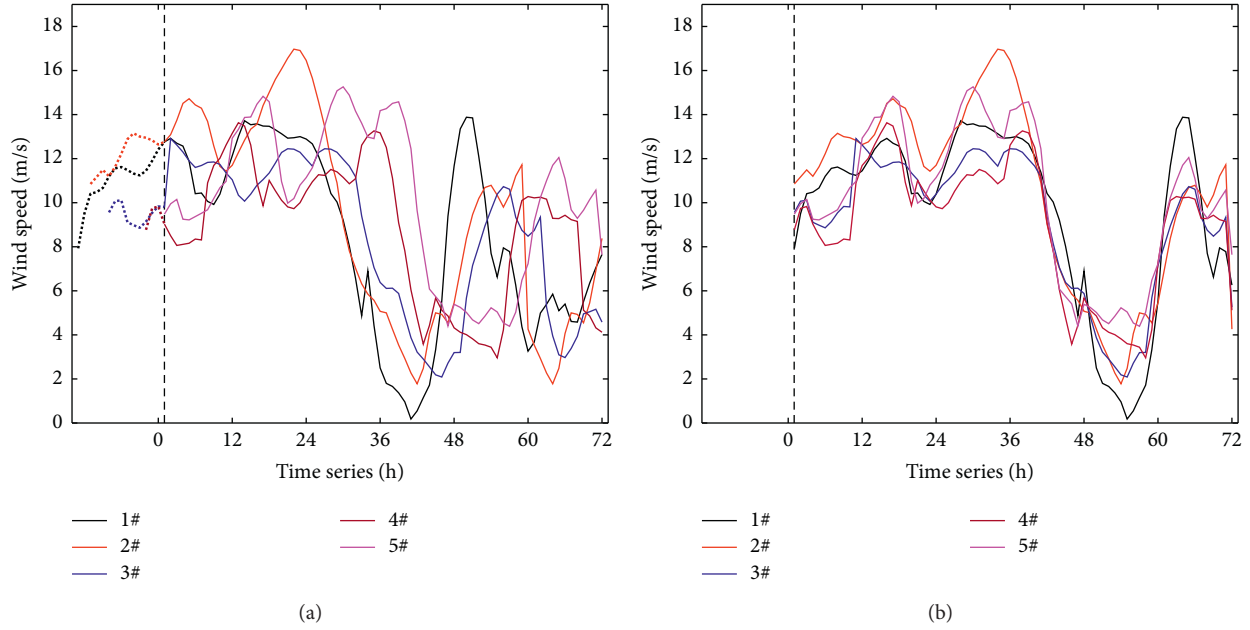
FIGURE 8: Wind speed sequences of whether considering time-shifting characteristics (wind farms scale). (a) Original. (b) Considering time-shifting.

TABLE 2: Correlation of wind speed sequences between points under two conditions (wind farms scale).

τ	1#-2#	1#−3#	1#−4#	1#−5#	2#-3#	2#−4#	2#−5#	3#-4#	3#-5#	4#-5#
Original	0.569	0.479	0.399	0.283	0.641	0.485	0.464	0.681	0.555	0.615
Considering time-shifting	0.616	0.577	0.522	0.510	0.655	0.543	0.567	0.712	0.611	0.648
Difference	**0.048**	**0.097**	**0.123**	**0.227**	**0.014**	**0.058**	**0.103**	**0.030**	**0.056**	**0.033**
ρ_s	1#-2#	1#−3#	1#−4#	1#−5#	2#-3#	2#−4#	2#−5#	3#-4#	3#-5#	4#-5#
Original	0.756	0.660	0.563	0.411	0.831	0.670	0.640	0.862	0.743	0.805
Considering time-shifting	0.802	0.767	0.712	0.699	0.845	0.735	0.758	0.886	0.801	0.834
Difference	**0.046**	**0.107**	**0.149**	**0.288**	**0.014**	**0.065**	**0.118**	**0.024**	**0.058**	**0.029**
λ	1#-2#	1#−3#	1#−4#	1#−5#	2#-3#	2#−4#	2#−5#	3#-4#	3#-5#	4#-5#
Original	0.348	0.226	0.093	0.040	0.095	0.106	0.241	0.463	0.318	0.301
Considering time-shifting	0.438	0.276	0.118	0.085	0.138	0.119	0.244	0.507	0.333	0.390
Difference	**0.090**	**0.051**	**0.025**	**0.045**	**0.043**	**0.013**	**0.003**	**0.044**	**0.015**	**0.089**

respectively, when the Kendall correlation coefficient τ is used as the evaluation index; the correlations are increased by 0.046, 0.107, 0.149, and 0.288, respectively, when the Spearman correlation coefficient ρ_s is used as the evaluation index. Furthermore, the tail correlation of wind speed sequences is also improved when the time-shifting characteristics are taken into account, which means the appearance times under high wind conditions and small wind conditions are more consistent. The tail correlation coefficients of wind speed sequences between point 1 and points 2 to 5 are increased by 0.090, 0.050, 0.025, and 0.045, respectively.

3.2. Time-Shifting under Wind Turbines Scale Containing Wake Effect. The wind sequence at each point is preprocessed at first. The wind sequence is divided into different wind processes with a time window of 1 hour in this case, and then the wind process sets are obtained at each point. The relative distribution of each point is shown in Figure 9, so the wind processes with the main wind direction between 130° and 140° at the corresponding point should be screened out for the time-shifting study. The average speeds of new wind speed sequences at points 1 to 3 are 11.45 m/s, 10.17 m/s, and 10.49 m/s, respectively.

Then, the time-shifting characteristics of new wind speed sequences between point 1 and points 2 to 4, between point 2 and points 3 to 4, and between point 3 and point 4 are studied, respectively. Since the data used in this case is the actual operation data of the wind farm, the factors that affect wind speed sequences at different locations should include both the delay time and the decay speed. The distributions of delay time and decay speed of wind speed time series between different points are shown in Figure 10. It can be seen that, with the increase of distance, the median delay time between wind speed sequences at two points increases, and the frequency distribution becomes more dispersed. The wind speed at the downstream location is not always lower than that at the upstream location. The average decay speeds of the wind speed sequence from point 1 to other points (2–4) are 0.62 m/s, −0.16 m/s, and 0.87 m/s, respectively. This is because the points are all located in the hilly terrain in this case, and the propagation of wind speed between different points is greatly affected by the terrain. Furthermore, the wake effect of wind turbines is weakened with the increase of the distance and the relief of the terrain.

In order to preliminarily analyze the mapping relationship among the time-shifting characteristics of wind speed sequences at different points, the distances between points, and the actual wind conditions, the wind processes at point 1 are divided according to the average wind speed of each process, with the interval of 1 m/s. The average delay time and average decay speed of wind processes from point 1 to points 2–4 are calculated in each wind speed bin, respectively, and the results are depicted in Figure 11. As can be seen, the delay time increases with the distance when the wind speed is almost the same and decreases with the wind speed when the distance is basically the same. However, due to the influence of the small distance between two points and the low time resolution of the data, the delay time of wind speed sequence between two points does not decrease with the increase of wind speed in some situations. This situation happens more frequently with the shortening of the distance between two points. The decay speed of wind speed sequences between different points varies with the value of wind speed.

The improved algorithm of *k*-means, *x*-means, is employed to cluster the wind processes at point 1 to extract typical wind processes, and then the typical wind process sets are constructed. The range of clustering number is set as in [4, 20], and the optimal clustering number is 32 in this case. The average delay time and the average decay speed of each wind process set from point 1 to other points (2–4) are calculated, respectively, and the results are shown in Figure 12. Six typical wind process
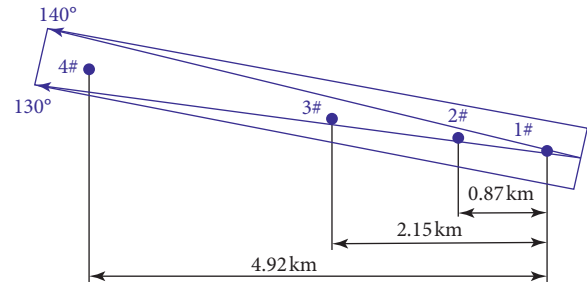
Figure 9: The relative distribution of points (wind turbines scale).

sets are selected randomly, as depicted in Figure 13. The average delay time and the average decay speed from point 1 to other points and the average wind speeds of these six typical wind process sets are listed in Table 3.

The new wind speed sequence at each point considering the time-shifting characteristics is reconstructed, respectively, according to the calculation results of the delay time and decay speed. The wind speed time series of whether considering the time-shifting is shown in Figure 14. As can be seen, the wind speed sequences at different points could be more similar if the time-shifting characteristics are considered.

In order to further verify the validity of the proposed method, the correlation analysis is used to test the effectiveness of the time-shifting characteristics. t-Copula function is used to fit the original wind speed sequences of two points and the new wind speed sequences considering the time-shifting characteristics of two points, respectively. Then the Kendall correlation coefficient τ, Spearman correlation coefficient ρ_s, and tail correlation coefficient λ are calculated based on the fitted function, and the results are listed in Table 4. It can be seen that the correlation of wind speed sequences at different points is enhanced significantly when the time-shifting characteristics are considered. For example, taking point 1 as the starting point, if the time-shifting characteristics are taken into account, the correlations of wind speed time series between point 1 and points 2 to 4 are increased by 0.142, 0.103, and 0.202, respectively, when the Kendall correlation coefficient τ is used as the evaluation index; the correlations are increased by 0.065, 0.046, and 0.124, respectively, when the Spearman correlation coefficient ρ_s is used as the evaluation index. Furthermore, the tail correlations of wind speed sequences between point 1 and points 2 to 4 are increased by 0.374, 0.147, and 0.802, respectively. It should be noted that, different from the case in wind farms scale, the points are all located in the hilly terrain in this case, the mapping relationship among the wind speed sequences at different points is more complicated, and the time-shifting effect does not show a more obvious trend with the increase of distance.

4. Verification of the Proposed Time-Shifting Evaluation Method

This section takes the data on December 31, 2014, under the wind turbines scale as an example to verify the calculation results of time-shifting and illustrate the effectiveness of the
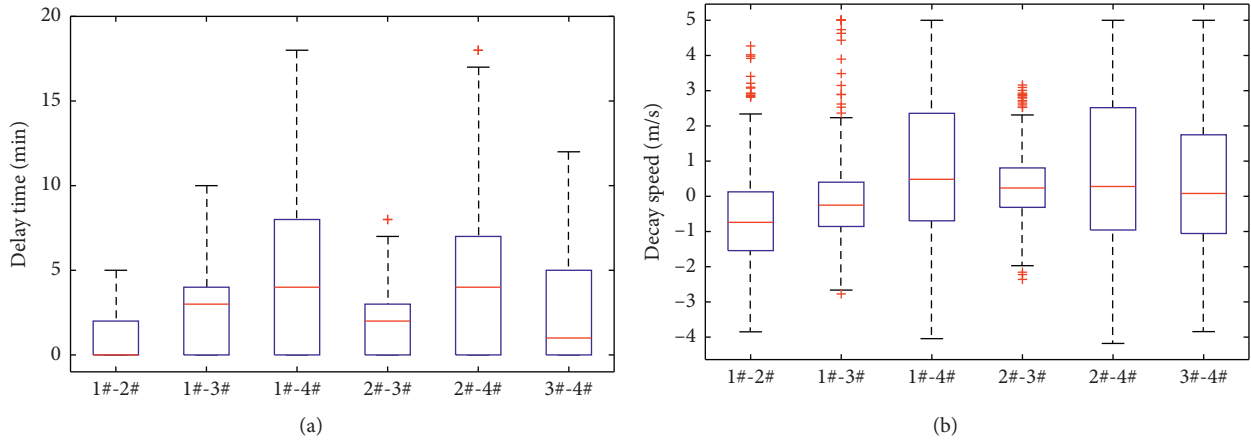
FIGURE 10: The time-shifting distributions of wind speed sequences between different points (wind turbines scale). (a) Delay time. (b) Decay speed.
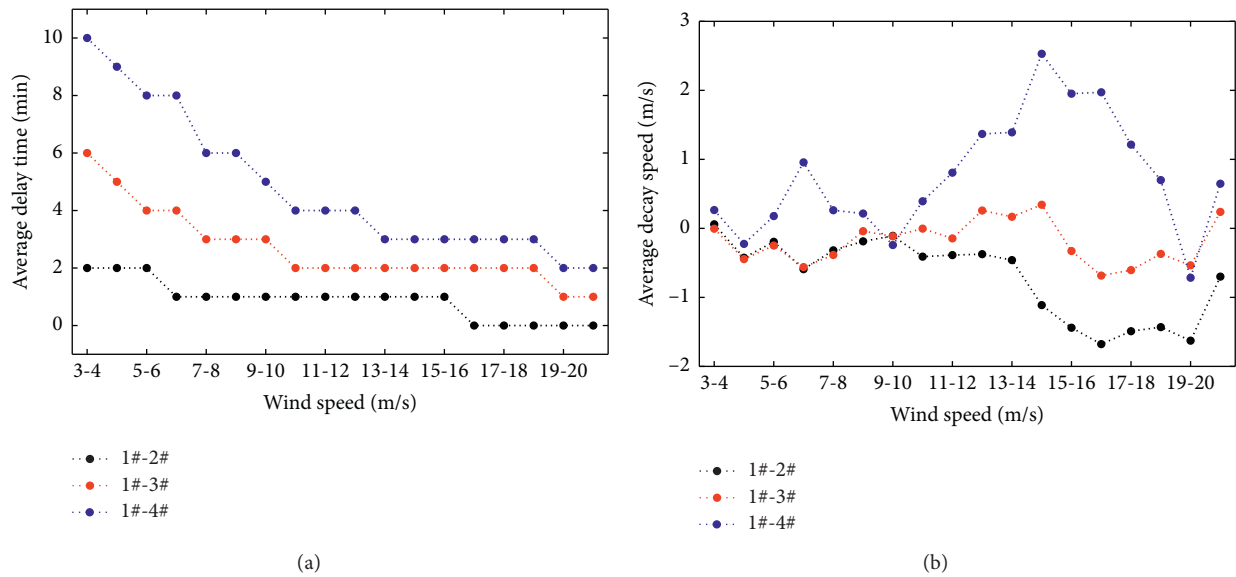
FIGURE 11: The time-shifting characteristics between different points in each wind speed bin (wind turbines scale). (a) Average delay time. (b) Average decay speed.
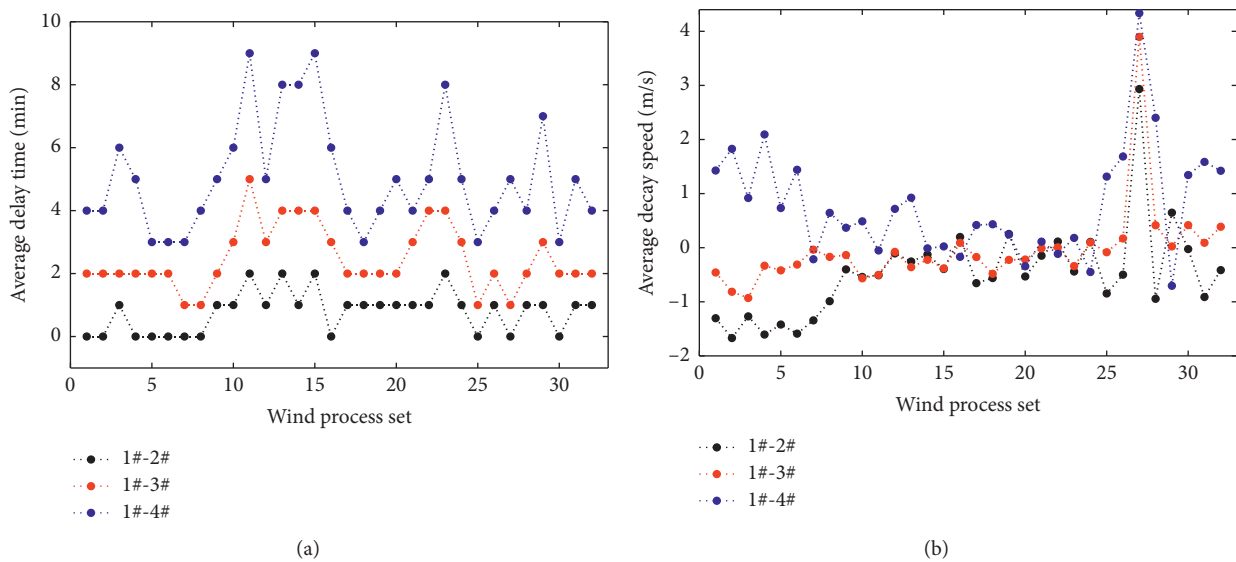
FIGURE 12: The time-shifting characteristics in each wind process set between point 1 and other points (wind turbines scale). (a) Average delay time. (b) Average decay speed.
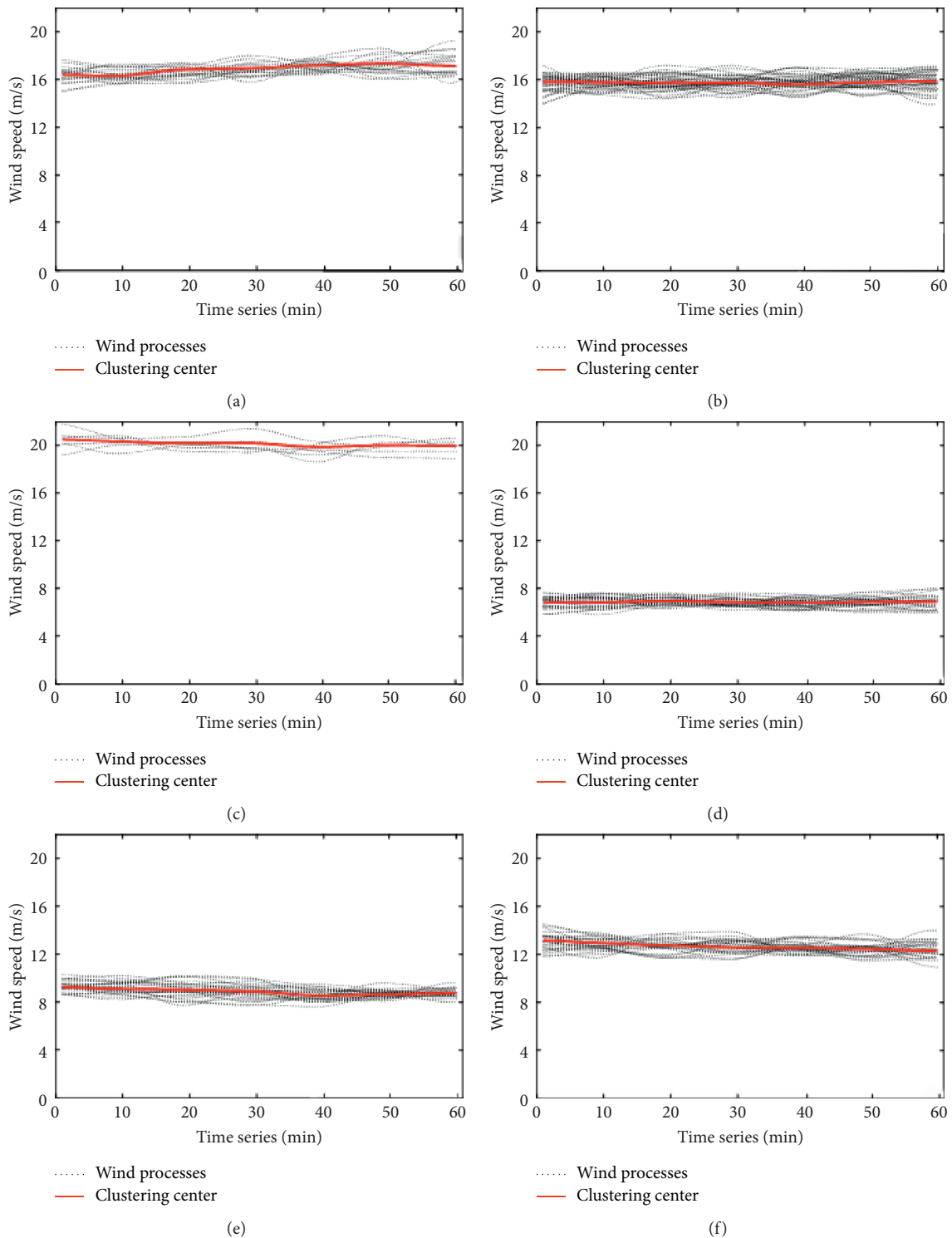
FIGURE 13: Six typical wind process sets at point 1 (wind turbines scale). (a) Typical wind process set 2. (b) Typical wind process set 4. (c) Typical wind process set 7. (d) Typical wind process set 11. (e) Typical wind process set 23. (f) Typical wind process set 31.

proposed evaluation method in this paper. The wind speed sequences at points 2 to 4 can be considered the same as that at point 1 if the time-shifting characteristics are not taken into account. When the time-shifting characteristics are considered, the wind speed sequences at points 2 to 4 should be calculated, respectively, based on the research results of

time-shifting. The specific steps are as follows: first, to divide the wind speed sequence and the wind direction sequence at point 1 with a time window of 1 hour and, second, to judge whether the main wind direction at point 1 is consistent with the relative position of wind turbines in each wind process, that is, whether the main direction of the wind process at

TABLE 3: Average delay time and average decay speed in six wind process sets between point 1 and other points (wind turbines scale).

| Typical wind process set	Average delay time (min)			Average decay speed (m/s)			Average wind speed (m/s)
	1#-2#	1#−3#	1#−4#	1#-2#	1#−3#	1#−4#	
2	0	2	4	−1.67	−0.81	1.83	16.89
4	0	2	5	−1.61	−0.33	2.09	15.75
7	0	1	3	−1.34	−0.04	−0.21	20.12
11	2	5	9	−0.51	−0.50	−0.05	6.83
23	2	4	8	−0.44	−0.34	0.18	8.81
31	1	2	5	−0.91	0.09	1.59	12.67
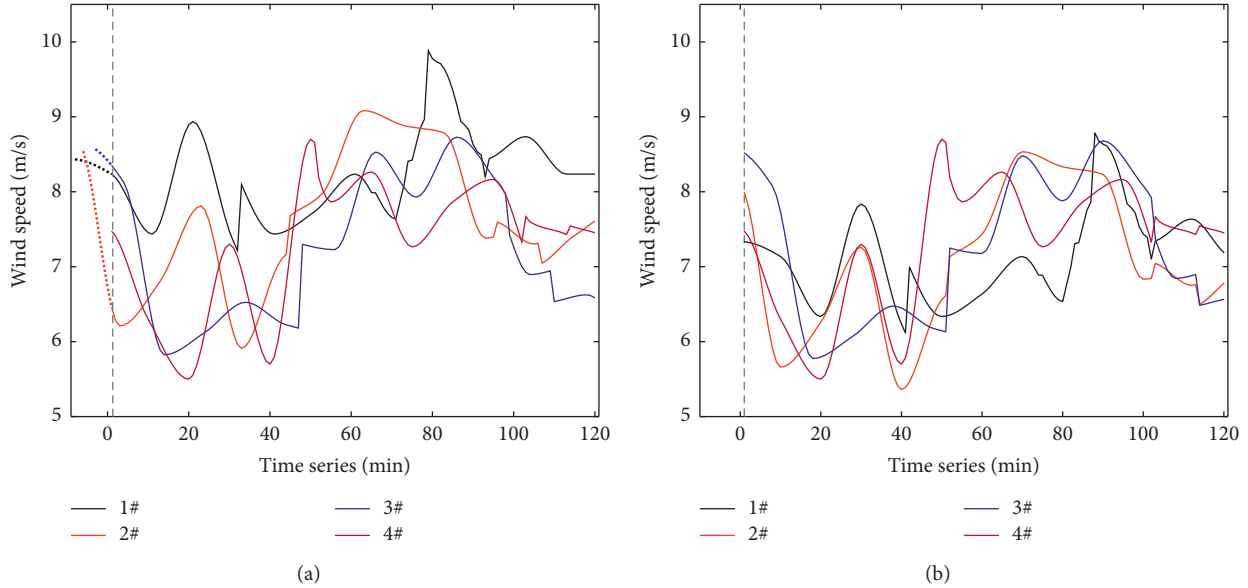
FIGURE 14: Wind speed sequences of whether considering time-shifting characteristics (wind turbines scale). (a) Original. (b) Considering time-shifting.

point 1 is between 130° and 140°. If so, it should be judged at first which wind process set the wind process belongs to, and then the wind process at points 2 to 4 should be constructed according to the above calculation results of time-shifting. If not, the corresponding wind speed at points 2 to 4 can be considered the same as that at point 1. The actual wind speed sequence and wind speed sequences of whether considering time-shifting at points 2 to 4 are shown in Figure 15. It can be seen that the wind speed sequence is more similar to the actual sequence at each point if the time-shifting characteristics are considered.

To further verify the calculation results of time-shifting, the average absolute deviations between the calculated wind speed sequences of whether considering the time-shifting and the actual wind speed sequence are used as the evaluation index to quantify the effectiveness of time-shifting. The average absolute deviations between the calculated wind speed sequence and the actual wind speed sequence at points 2 to 4 are 1.27 m/s, 1.18 m/s, and 1.26 m/s, respectively, if the time-shifting is not taken into account. If the time-shifting is considered, the average absolute deviations are 1.06 m/s, 1.02 m/s, and 1.04 m/s, respectively. The average absolute deviations are reduced by 16.5%, 13.6%, and 17.5%, respectively, when

TABLE 4: Correlation of wind speed sequences between points under two conditions (wind turbines scale).

τ	1#-2#	1#−3#	1#−4#	2#−3#	2#−4#	3#−4#
Original	0.758	0.794	0.651	0.827	0.587	0.649
Considering time-shifting	0.900	0.897	0.853	0.912	0.853	0.872
Difference	**0.142**	**0.103**	**0.202**	**0.086**	**0.266**	**0.223**
ρ_s	1#-2#	1#−3#	1#−4#	2#−3#	2#−4#	3#-4#
Original	0.920	0.938	0.842	0.959	0.783	0.839
Considering time-shifting	0.985	0.983	0.966	0.988	0.965	0.974
Difference	**0.065**	**0.046**	**0.124**	**0.029**	**0.182**	**0.135**
λ	1#-2#	1#−3#	1#−4#	2#-3#	2#-4#	3#-4#
Original	0.467	0.703	0.004	0.517	0.000	0.058
Considering time-shifting	0.841	0.850	0.806	0.872	0.828	0.813
Difference	**0.374**	**0.147**	**0.802**	**0.354**	**0.828**	**0.754**

the time-shifting is considered. The average absolute deviations between wind speed sequences of whether considering the time-shifting and the actual wind speed sequence in each time window at points 2 to 4 are depicted in Figure 16. It can be seen that the wind speed

(a)

(b)

(c)
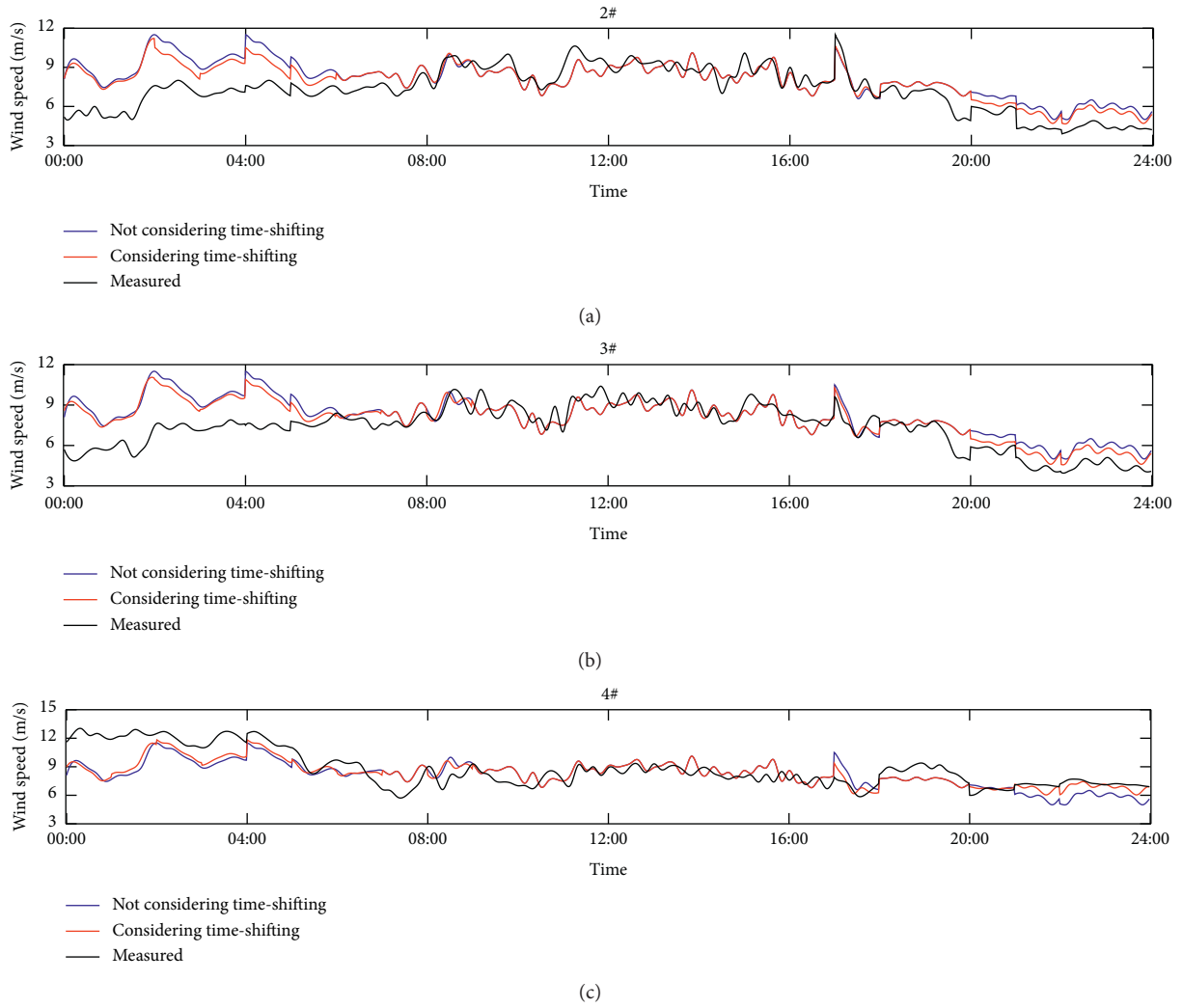
FIGURE 15: Wind speed sequences of whether considering the time-shifting characteristics at each point.

(a)

(b)
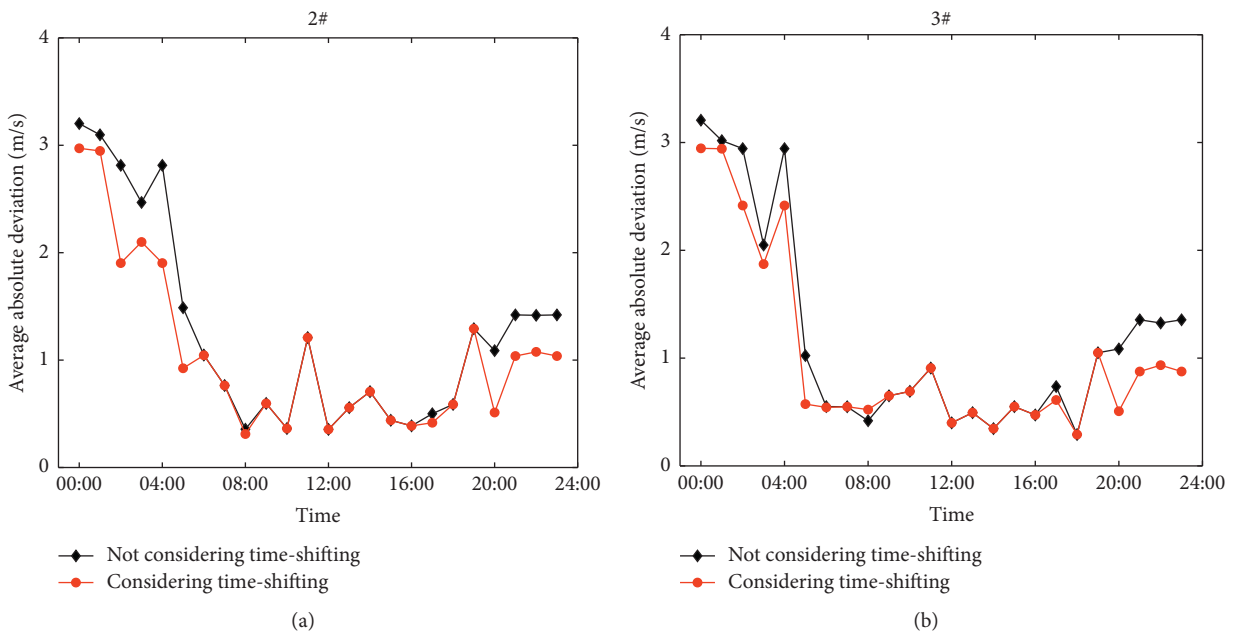
FIGURE 16: Continued.

(c)
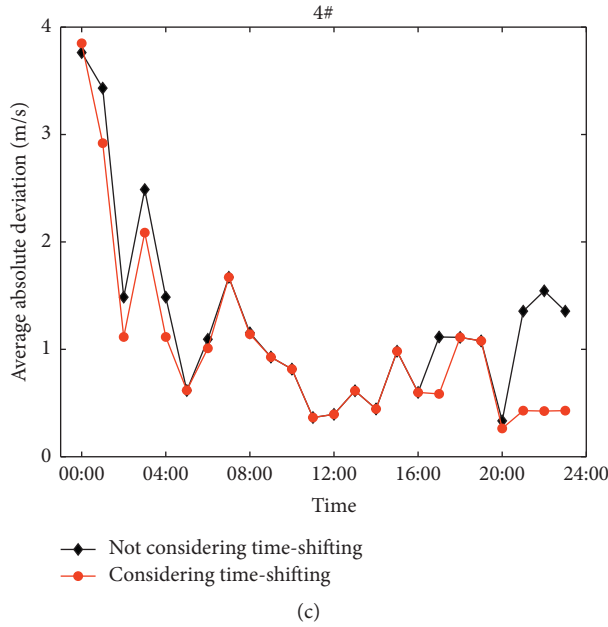
FIGURE 16: Average absolute deviations of wind speed sequences under two conditions at each point.

sequences are more similar to the actual wind speed sequences in all time windows if the time-shifting is considered.

In order to study the influence of time-shifting on the wind power generation system, the measured wind speed sequence and the calculated wind speed sequences under two conditions (considering time-shifting or not) at each point are converted to the power sequences based on the theoretical power curve of the wind turbine, respectively; then the total output of wind turbines at points 1 to 4 is obtained by adding the output of wind turbine at each point. The total power sequences of four wind turbines calculated by the measured wind speed sequence and the wind speed sequences of whether considering time-shifting are shown in Figure 17. It can be seen that the calculated total power is more similar to the actual power when the time-shifting is considered. To further illustrate the necessity of time-shifting, the influence of time-shifting is quantified from two perspectives: one is the volatility of the total power sequence and the other is the generation of four wind turbines. Fluctuation ratio (FR) [29] is used to evaluate the volatility of the power sequence in this paper. The calculation method of FR is shown in the following equation:

$$FR = \frac{\sum_{t=1}^{N-1} \left| (P(t+1) - P(t))/P_{cap} \right|}{N-1}, \qquad (8)$$

where $P(t)$ is the power of wind turbines at time t and P_{cap} is the rated power of wind turbines, 8.02 MW in this study.

The total power generation and the volatility of the power sequence of four wind turbines are calculated through the wind speed sequences under two conditions, and the results are compared with the results when using the measured wind speed sequences, as shown in Table 5. As can be seen from the table, if the time-shifting is taken
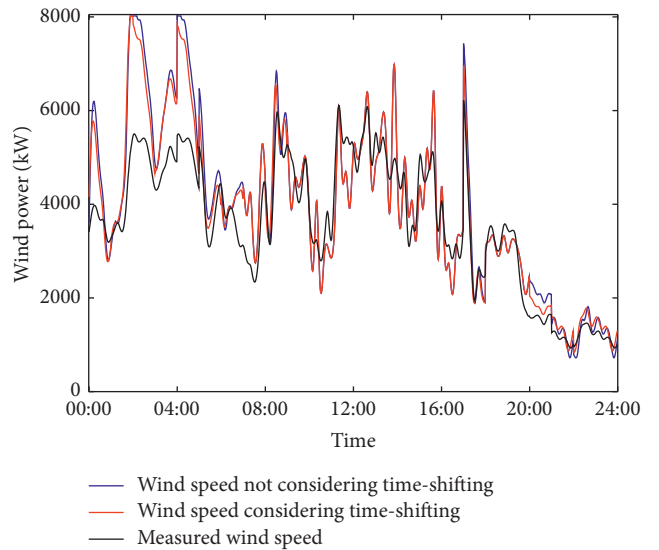
FIGURE 17: The power sequences of four wind turbines of whether considering time-shifting.

into account, the calculated daily power generation is closer to the actual generation, and the volatility of the calculated power sequence is more similar to that of the actual power sequence. The aggregation characteristics of wind turbines output are reflected. Compared with the actual daily power generation, the deviation of the calculated generation is 8.6% if the time-shifting is considered and is 10.6% if the time-shifting is not considered. Compared with the volatility of the actual power sequence, the deviation of the volatility of the calculated power sequence is 70.2% if the time-shifting is considered and is 77.2% if the time-shifting is not considered.

TABLE 5: Comparison of wind power output under two conditions.

Power generation (MW·h)			Power volatility		
Not considering time-shifting	Considering time-shifting	Measured	Not considering time-shifting	Considering time-shifting	Measured
96.51	94.77	87.28	0.0101	0.0097	0.0057

5. Conclusions

A quantification method of wind speed time-shifting at different spatial positions based on wind process is proposed in this paper. As a result of this study, the following conclusions can be drawn:

(1) The proposed evaluation method can effectively achieve the quantitative analysis of the time-shifting characteristics of wind speed sequences at different spatial positions. Furthermore, the results of the proposed method can be continuously improved according to the actual wind conditions and could be applied to any spatial scale.

(2) The time-shifting effect between wind speed sequences increases with the distance when points are located in the flat terrain. But when points are located in the hilly terrain, the time-shifting effect does not show a more obvious trend with the increase of distance.

(3) The median delay time of wind speed sequences increases and the frequency distribution of the delay time becomes more dispersed with the increase of distance, no matter what terrain the points are located in.

(4) The wind speed at the downstream location could be larger than that at the upstream location in the wind farm when points are located in the hilly terrain.

(5) x-Means clustering method can achieve the effective division of wind processes, and the time-shifting characteristics of different wind processes are quite various.

(6) The calculated output of wind turbines is more similar to the actual output in the wind power system if the time-shifting characteristics are taken into account.

The time-shifting characteristics of wind speeds are studied when the wind direction is consistent with the relative position of points in this paper. In future research, the proposed evaluation method will be applied to analyze the time-shifting characteristics of wind speeds in full wind directions to better grasp the spatiotemporal coupling relationship of wind speeds at multiple scales.

Data Availability

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

Conflicts of Interest

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

References

[1] Y. Liu, Y. Ye, X. Chen, H. Li, and Y. Huang, "Robust day-ahead dispatch for integrated power-heat-gas microgrid considering wind power uncertainty," *Mathematical Problems in Engineering*, vol. 2020, no. 9, Article ID 4215906, 12 pages, 2020.

[2] Y. Zhu, X. Liu, Y. Zhai, and R. Deng, "Monthly unit commitment model and algorithm with renewable energy generation considering system reliability," *Mathematical Problems in Engineering*, vol. 2019, Article ID 3835296, 9 pages, 2019.

[3] C. Lucheroni, J. Boland, and C. Ragno, "Scenario generation and probabilistic forecasting analysis of spatio-temporal wind speed series with multivariate autoregressive volatility models," *Applied Energy*, vol. 239, pp. 1226–1241, 2019.

[4] M. Cellura, G. Cirrincione, A. Marvuglia, and A. Miraoui, "Wind speed spatial estimation for energy planning in sicily: a neural kriging application," *Renewable Energy*, vol. 33, no. 6, pp. 1251–1266, 2008.

[5] S. Z. Moghaddam, "Generation and transmission expansion planning with high penetration of wind farms considering spatial distribution of wind speed," *Electrical Power and Energy Systems*, vol. 106, pp. 232–241, 2019.

[6] C. M. Grams, R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, "Balancing Europe's wind-power output through spatial deployment informed by weather regimes," *Nature Climate Change*, vol. 7, no. 8, pp. 557–562, 2017.

[7] W. Zheng, *Research on Wind Speed Forecasting of Regional Wind Farm Group Based on Spatio-Temporal Correlation*, North China Electric Power University, Beijing, China, 2014.

[8] S. She, Z. Li, and X. Cai, "Research on wind speed distribution model of wind farm based on its dynamic space-time relation," *Power System Technology*, vol. 38, no. 6, pp. 1432–1438, 2014.

[9] C. Zeng, L. Ye, and Y. Zhao, "Spatial model for short term wind power prediction considering wake effects," *Power System Protection and Control*, vol. 40, no. 24, pp. 59–142, 2012.

[10] L. Ye, Y. Zhao, C. Zeng, and C. Zhang, "Short-term wind power prediction based on spatial model," *Renewable Energy*, vol. 101, pp. 1067–1074, 2017.

[11] Y. L. Pichugina, R. M. Banta, T. Bonin et al., "Spatial variability of winds and HRRR-NCEP model error statistics at three Doppler-lidar sites in the wind-energy generation region of the Columbia river basin," *Journal of Applied Meteorology and Climatology*, vol. 58, no. 8, pp. 1633–1656, 2019.

[12] F. Liu, F. Sun, W. Liu et al., "On wind speed pattern and energy potential in China," *Applied Energy*, vol. 236, pp. 867–876, 2019.

[13] J. Li and X. Yu, "Onshore and offshore wind energy potential assessment near Lake Erie shoreline: a spatial and temporal analysis," *Energy*, vol. 147, pp. 1092–1107, 2018.

[14] J. Bosch, I. Staffell, and A. D. Hawkes, "Temporally explicit and spatially resolved global offshore wind energy potentials," *Energy*, vol. 163, pp. 766–781, 2018.

[15] Y. Chen, S. Zhang, W. Zhang, J. Peng, and Y. Cai, "Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting," *Energy Conversion and Management*, vol. 185, pp. 783–799, 2019.

[16] K. Philippopoulos and D. Deligiorgi, "Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography," *Renewable Energy*, vol. 38, no. 1, pp. 75–82, 2012.

[17] M. S. Miranda and R. W. Dunn, "Spatially correlated wind speed modelling for generation adequacy studies in the UK," in *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, Tampa, FL, USA, June 2007.

[18] J. Park and D.-H. Jang, "Application of MK-PRISM for interpolation of wind speed and comparison with co-kriging in South Korea," *GIScience & Remote Sensing*, vol. 53, no. 4, pp. 421–443, 2016.

[19] G. Ren, J. Wan, J. Liu, and D. Yu, "Assessing temporal variability of wind resources in China and the spatial correlation of wind power in the selected regions," *Journal of Renewable and Sustainable Energy*, vol. 12, Article ID 13302, 2020.

[20] Z. Q. Xie, T. Y. Ji, M. S. Li, and Q. H. Wu, "Quasi-Monte Carlo based probabilistic optimal power flow considering the correlation of wind speeds using copula function," *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 2239–2247, 2018.

[21] X. Hui, *Calculation of Available Transfer Capability Considering Temporal and Spatial Correlation in Wind Power Integrated System*, Northeast Electric Power University, Jilin, China, 2018.

[22] G. Ren, J. Wan, J. Liu, and D. Yu, "Spatial and temporal assessments of complementarity for renewable energy resources in China," *Energy*, vol. 177, pp. 262–275, 2019.

[23] M. Zeng, J.-h. Li, Q.-h. Meng, and X.-n. Zhang, "Temporal-spatial cross-correlation analysis of non-stationary near-surface wind speed time series," *Journal of Central South University*, vol. 24, no. 3, pp. 692–698, 2017.

[24] S. T. Frandsen, H. E. Jørgensen, R. Barthelmie et al., "The making of a second-generation wind farm efficiency model complex," *Wind Energy*, vol. 12, no. 5, pp. 445–458, 2009.

[25] M. B. Imamoglu, M. Ulutas, and G. Ulutas, "A new reversible database watermarking approach with firefly optimization algorithm," *Mathematical Problems in Engineering*, vol. 2017, Article ID 1387375, 14 pages, 2017.

[26] Y. Kang, Z. Zhang, and X. Guo, "Portfolio risk analysis in electricity market based on copula approach," *Power System Protection and Control*, vol. 40, no. 6, pp. 50–56, 2012.

[27] Z. Jiang and J. He, "Method of fusion diagnosis for dam service status based on joint distribution function of multiple points," *Mathematical Problems in Engineering*, vol. 2016, Article ID 9049260, 10 pages, 2016.

[28] D. Pelleg and A. Moore, "*X*-means: extending *K*-means with efficient estimation of the number of clusters," in *Proceedings of the Seventeenth International Conference on Machine Learning*, Stanford, CA, USA, July 2000.

[29] S. Han, L.-n. Zhang, Y.-q. Liu et al., "Quantitative evaluation method for the complementarity of wind-solar-hydro power and optimization of wind-solar ratio," *Applied Energy*, vol. 236, pp. 973–984, 2019.

Review Article

Research Hotspots and Evolution of Energy Prosumer: A Literature Review and Bibliometric Analysis

Chunyan Dai ![ORCID],[1,2] Kai Cheng,[2] Yiting Lei,[2] and Ying Yang[2]

[1]*Research Center for Economy of Upper Reaches of the Yangtze River, Chongqing Technology and Business University, Chongqing 400067, China*
[2]*School of Management, Chongqing Technology and Business University, Chongqing 400067, China*

Correspondence should be addressed to Chunyan Dai; daichunyan@ctbu.edu.cn

Energy prosumer participates in the balance of the energy system actively through their production and consumption of energy and will become a major participant in the open energy market increasingly in the future. This paper used the bibliometric method to sort out 1251 literature data related to energy prosumer in the "Web of Science Core Collection" and used the software named CiteSpace to perform statistics and cooccurrence analysis. The research found the following: (a) journals in the field of prosumer have their emphasis on theory, application, policy, and modeling; (b) the United States and European institutions have published the articles most, and the academic influence of Chinese scholars needs to be further improved; (c) the focus of the prosumers' research in the past decade has turned the energy system designing to the market transaction research; (d) P2P, P2G, and prosumer community are three market models that have attracted attention for scholars; besides, self-consumption, market matching, district heating, aggregators, and household energy management systems are also the current research hotspots; (e) digital technologies such as big data, the Internet of Things, and blockchain have a significant impact on the prosumer participating in energy market transactions, which will change the existing energy supply chain model profoundly. In the context of the energy Internet, energy trading platforms, digital ecosystems, cross-border innovation, and other fields will become the hotspots in future research on prosumer.

1. Introduction

The term "prosumer" was coined by American's futurist named Alvin Toffler in 1980. It refers to those consumers who participate in production activities. They are both consumers and producers (Figure 1) [1]. In the energy sector, distributed energy sources (DERs) such as rooftop solar panels, battery storage, and control equipment continue to integrate and their prices fall, as well as the advancement of information and communication technology (ICT) equipment, some energy consumers are turning into prosumers, whom produce and consume energy, and they can share the excess energy with other users or feed it back into the energy network [2]. Most of the current prosumers refer to the electricity prosumer in the power grid [3], and there are also heat prosumers in the district heating network

[4] and multienergy prosumers with cooling, heating, and electricity [5]. The emergence of prosumers will change the top-down value chain in traditional power grids or heating networks. Prosumers will become active participants in the smart grid/heating grid, promoting energy sharing and contributing to the digital energy ecosystem [6].

After nearly 10 years of development, some scholars have summarized the research on prosumers in different perspectives. Zafar et al. [7] outlined the energy management and sharing of the prosumer in the smart grid environment, analyzed related ICT technologies and optimization technologies, and detailed the mechanisms of energy sharing, management, and trading based on prosumer. Espe et al. [3] analyzed the latest research results of the prosumer community in the smart grid from the two perspectives of "prosumer community" and "prosumer relationship" and
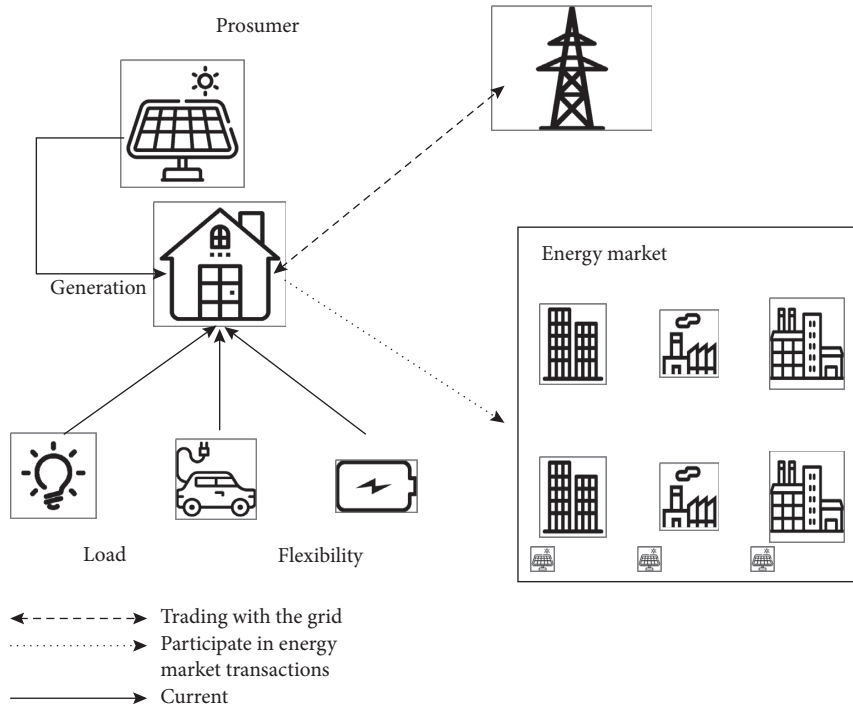
FIGURE 1: The concept of prosumer.

outlined the future research directions in these two aspects. Sousa et al. [8] summarized the three structures of the P2P market: a complete P2P market, a community-based market, and a mixed P2P market. They analyze the opportunities and challenges of the future development of the P2P market. Huang et al. [9] regarded the data center as a large energy prosumer in the regional energy system, considering the integration of upstream green energy supply and the reuse of downstream waste heat. Also, the related researches include energy blockchain [10], electric vehicles in the smart grid [11], etc.

Up to now, most of researches focused on a certain research branch (such as the community prosumers community) only. Gough et al. [12] used a bibliometric method to analyze the flexibility of prosumers and study the relationship between the five high-frequency keywords (prosumers, smart grids, microgrids, renewable energy, and demand response) and the flexibility of prosumers. Different from the former, this article will make a systematic evaluation of the internal relevance of the research content and the evolutionary trend of the period from the entire field of prosumers. Given this, this article sorts out the development history of energy prosumer based on the perspective of bibliometrics. Using CiteSpace bibliometric analysis software [13], this article systematically analyzed 1251 literatures data related to energy prosumer in the "Web of Science Core Collection." Through the analysis of core journals, key authors, and key regions, we have sorted out the literature characteristics of the research on prosumers. Through keyword cooccurrence analysis, research hotspot analysis in

the past 180 days, and literature cocitation knowledge graph analysis, we have unearthed the research foundation of the prosumer and the research hotspots in different periods. From a broader longitudinal perspective, it systematically expounds the knowledge evolution and development frontiers of the prosumer in the international field, hoping to provide reference for future-related studies of prosumer.

2. Research Design

2.1. Research Methods and Tools. This article adopts the research method that combines bibliometrics analysis and literature review. First of all, based on CiteSpace software and Web of Science's literature analysis tools, this paper realizes the basic statistical analysis of the major journals, core authors, and key countries and regions in the field of prosumer research and presents the basic situation of international research on prosumers. After that, this article conducts a keywords cooccurrence network analysis on the research literature to study the research hotspots and evolution process of the prosumers in the field. Finally, based on the literature cocitation knowledge graph, this article shows the mainstream subdivision areas of the field of the prosumer. Cocitation analysis means that if two documents appear together in the reference list of the third citing document, the two documents will form a cocitation relationship [14]. Mapping knowledge domains is an important bibliometric method. It visualizes the structure, rules, and distribution of scientific knowledge in the research field of the pros and cons and explores the development of the

subject knowledge field and its research hotspots, frontiers, and trend [14].

2.2. Data Source and Processing.

2.2. Data Source and Processing. This article focuses on the evolution and structure of research on prosumers, and all the research literatures come from the Web of Science Core Collection. At first, search according to the subject search formula "TS = ("prosumer*") AND TS = ("electricity" OR "energy" OR "electric" OR "power")". Through screening, irrelevant documents were eliminated, and 1251 English literatures were retrieved. The actual literature data year is 2009–2020, and the data search time is May 26, 2020. This paper uses these data to establish the original database for the study and then import the original data into the software of CiteSpace 5.6.R5, the time parameter is set to 2009–2020, year per slice is set to 1, and G-index $K = 25$.

3. Basic Statistical Analysis of Research on Energy Prosumer

3.1. Posting Trend. As shown in Figure 2, from 2009 to 2020, the number of publications in the field of prosumers has increased year by year, and the upward trend is obvious. This shows that the field of prosumer has received great attention by scholars in the world. According to data retrieved by the Web of Science, the first paper referring to energy prosumer by Mauri is (2009) "Energy conservation and smart grids: new challenge for multimetering infrastructures." In this paper, the author believes that the future power system will be a combination of centralized and distributed, and power prosumer will become a part of the grid [15]. Besides, scholars have not yet formed a unified academic concept for the term "prosumer," and some articles may still use consumer to express it. With the popularization of the term prosumer, relevant literature will become more and more concentrated.

3.2. Distribution of Journals. Table 1 lists the major journals published in the field of prosumer from 2009 to 2020. The topics and contents of different journals have their focuses. Among them, "IEEE Transactions on Smart Grid" focuses on research related to energy generation, transmission, distribution, and delivery in the field of smart grids. "Applied Energy" focuses on energy conversion and conservation, optimal utilization of energy resources, and energy process analysis and optimization. "Renewable and Sustainable Energy Reviews" publishes review papers, original research, case studies, and new technology analysis related to renewable and sustainable energy. "Energy Policy" focuses on energy politics, economics, planning, environmental, and social research. "Energy" focuses on energy analysis, energy modeling and forecasting, integrated energy systems, energy planning, and energy management. The scope of transactions of "IEEE Transactions on Power Systems" includes the planning, analysis, reliability, operation, and economics of power generation, transmission, and distribution systems.

3.3. Author. Table 2 lists the top ten high-published authors and highly cited authors in the field of prosumers. Among them, the high number of publications indicates that there is more output in the field of research on the prosumer, and high citation is an important indicator to measure the academic level and academic influence. According to the ranking of high-published authors in Table 2, the four scholars named Menniti Daniele, Pinnarelli Anna, Sorrentino Nicola, and Burgio Alessandro from the University of Calabria in Italy published the most, followed by Liu from North China Electric Power University with 13 articles and Morstyn from Oxford University with 11 articles. The highly cited authors Zhang and Long in Table 2 are both from Wu's team at Cardiff University, UK. It can be found that the number of articles published is not necessarily related to the frequency of citations. The top ten scholars in the number of publications did not stand out in the evaluation of the citation frequency index.

Highly cited authors are generally one of the important indicators to measure the important influence of a research field. According to the ranking of the highly cited authors in Table 2, the highly cited authors have put forward important opinions on the research of the various subdivisions of the prosumer. Mengelkamp et al. [16] proposed the concept of a microgrid energy market based on blockchain and summarized the seven components of an efficient microgrid energy market. Parag et al. [17] designed three power markets in the era of prosumers: P2P mode, P2G mode, and prosumer community mode. Zhang et al. [18] established a four-layer P2P energy trading system architecture model and, for the first time, proposed a P2P energy trading bidding system between consumers and prosumers in the grid-connected microgrid. The main contribution of Liu et al. lies in the management of energy sharing within the microgrid, including the Stackelberg game between microgrid operators (MGO) and prosumer [19], P2P energy sharing based on dynamic pricing of supply-demand ratio (SDR) [20], and energy sharing providers with energy storage coordinate multiple producers and consumers to form an energy sharing network [21]. Tushar et al. [22] studied P2P energy trading from the perspective of the prosumer alliance.

3.4. Country/Region and Organization. Table 3 lists the top ten countries or regions and top ten organizations with the largest number of postings. The number of papers published can reflect the research level and contribution degree of different countries or regions and scientific research institutions to a certain extent. First of all, except for the United States, China, and Australia, the top ten countries are concentrated in Europe, such as Italy, Germany, the United Kingdom, Spain, and the Netherlands. Secondly, judging from the top ten scientific research institutions, the amount of publications is not much different. Except for a Chinese institution, the others are from Europe. Also, countries such as the United States and the United Kingdom with a large number of publications have not appeared in the top ten high-publishing institutions. It can be seen that the current
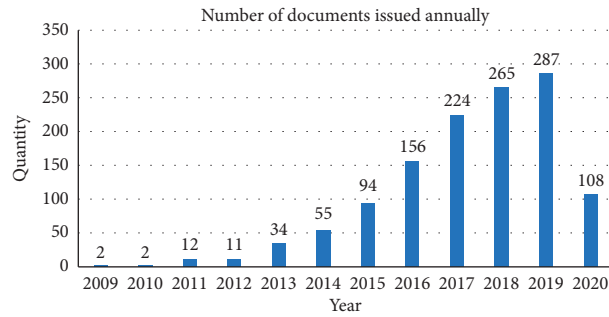
Figure 2: Annual number of papers published.

Table 1: Distribution of the top ten journals published in the field of prosumer.

Number	Number of articles	Source journal	2018 impact factor
1	77	Energies	2.707
2	53	Applied Energy	8.426
3	32	IEEE Transactions on Smart Grid	10.486
4	29	Renewable & Sustainable Energy Reviews	10.556
5	26	Energy	5.537
6	25	IEEE Access	4.098
7	22	Energy Policy	4.88
8	22	IEEE Transactions on Power Systems	6.807
9	17	Sustainability	2.592
10	12	Energy Research & Social Science	5.525

Table 2: The top 20 high-volume authors and highly cited authors in the field of prosumers.

		High-published authors			Highly cited author	
Number	Number of articles	Author	Organization	Highly cited	Cited author	Organization
1	16	D. Menniti		81	E. Mengelkamp	Karlsruhe Institute of Technology
2	16	A. Pinnarelli	University of Calabria	80	Y. Parag	Interdisciplinary Center Herzliya
3	16	N. Sorrentino		78	C. H. Zhang	Cardiff University
4	14	A. Burgio		77	N. Liu	North China Electric Power University
5	13	N. Liu	North China Electric Power University	57	W. Tushar	The University of Queensland
6	12	T. Morstyn	Oxford University	54	C. Long	Cardiff University
7	11	S. Grijalva	Georgia Institute of Technology	53	T. Morstyn	Oxford University
8	11	G. C. Christoforidis	University of Western Macedonia	52	A. J. D. Rathnayaka	Curtin University of Technology
9	11	M. Sanduleac	Politehnica University of Bucharest	49	H. Lund	Aalborg University
10	11	B. Adebisi	Manchester Metropolitan University	48	W. Saad	Virginia Tech

research is still at a preliminary stage and research cooperation is relatively scattered. Based on the high-published authors and highly cited authors mentioned above, the research results of Liu's team at North China Electric Power University are relatively outstanding [19–21]. Also worthy of attention are Wu's team from Cardiff University [23, 24] and Morsty's team from Oxford University [25].

To analyze the differences in the research directions and focus of different institutions, we further analyzed and compared the 5 subject terms most concerned by the top 8 institutions. The results are shown in Table 4.

In addition to the topics of common concern of various institutions, such as demand-side management and systems, INESC TEC focuses on flexibility [26] and aggregators

TABLE 3: High-published countries or regions and organizations.

Number	High-published countries/regions		High-published organizations	
	Number of articles	Country/region	Number of articles	Organizations
1	142	USA	26	INESC TEC
2	138	Italy	24	North China Electric Power University
3	99	Germany	23	Aalto University
4	93	Peoples R China	23	Polytechnic University of Bucharest
5	81	England	19	University of Calabria
6	79	Spain	18	Aalborg University
7	75	Netherlands	17	Polytechnic University of Turin
8	75	Australia	17	The University of Lisbon
9	67	Poland	16	Eindhoven University of Technology
10	65	Portugal	16	Helmholtz Association of German Research Centres

TABLE 4: Comparative analysis of research topics of the top eight institutions.

Organization	Top 5 topics that have received the most attention (in order of frequency)
INESC	Prosumer (14), flexibility (8), aggregators (8), electricity market (8), energy storage (5)
North China Electric Power University	Demand response (10), prosumer (10), optimization (9), system (6), microgrid (6)
Aalto University	Prosumer (12), systems (7), district heating (5), impact (4), demand-side management (3)
Polytechnic University of Bucharest	Prosumer (8), smart grid (7), renewable energy (4), power distribution system (4), microgrid (3)
University of Calabria	Prosumer (10), demand response (7), system (6), home automation system (4), management (3)
Aalborg University	System (6), prosumer (5), microgrid (5), demand response (3), blockchain (2)
Polytechnic University of Turin	Smart grid (5), energy management (4), prosumer (4), system (4), demand-side management (2)
Universidade de Lisboa	Self-consumption (5), prosumer (5), demand response (4), demand (3), influence (3)

[27–30], North China Electric Power University mainly studies issues such as microgrid optimization [31, 32], Aalto University focuses on district heating [33, 34], the Polytechnic University of Bucharest noticed the power distribution system [35, 36], the University of Calabria focuses on home automation system [37, 38], Aalborg University researches blockchain [39], the Polytechnic University of Turin considers energy management [40, 41], and the University of Lisbon focuses more on self-consumption [42, 43].

4. Analysis of the Evolution of Research Hotspots for Energy Prosumer

4.1. Identification and Evolution of Research Hotspots Based on Keyword Cooccurrence. Table 5 lists the top 20 keywords in the field of prosumers according to the frequency of keyword cooccurrence. Keywords represent knowledge points and research hotspots in a certain research field. The year in the table represents the time when the keyword-containing article first cooccurs with high frequency and not necessarily the time when the keyword first appeared mainly including prosumer, smart grid, demand response, microgrid, optimization, energy storage, distributed power generation, and electric vehicles.

In Table 5, keyword frequency ranking shows the hotspots in the research field of prosumer from 2009 to 2020 at a certain level. The "prosumer" is the keyword with the highest frequency of cooccurrence in the field of energy prosumer, reflecting the research theme in this field. The "smart grid" and "microgrid" indicate the network background studied by prosumer. "System," "optimization," "management,"

"energy management," and "model" reflect different levels and scales of energy system management and optimization issues based on prosumer. "Distributed generation," "storage," and "electric vehicle" reflects the flexibility of prosumer fully. Also worthy of attention are "demand-side management" to improve energy efficiency, and "blockchain" promotes decentralized P2P energy transactions.

To better understand the evolution law of research hotspots in the field of the prosumer and identify research frontiers, this article lists keywords with relatively high frequency or high centrality in each year, aiming to analyze the development trend of research hotspots in the field of prosumer. "Centrality" means "betweenness centrality," which is one of the important indicators for measuring nodes in the network. It is usually a hub connecting two different fields. The higher the centrality value, the more important the node. As shown in Table 6, in 2014, hot keywords were more numerous and scattered, reflecting that the concept of "prosumer" has attracted more scholars' attention this year. The research in these subdivisions has laid the foundation for the subsequent research and development of prosumers. Besides, it can be seen in the research roughly that it were focused more on the scientific and technological research of various energy managements and energy system elements before 2015; and in 2015 and later, it were focused more on the soft science research of market transactions.

4.2. Analysis of Research Hotspots of Prosumer in the past 180 Days. To further analyze the research hotspots of the prosumer in the past 180 days, that is, to identify cutting-edge

TABLE 5: The ranking of the cooccurrence frequency of the top 20 keywords.

Number	Frequency	year	Keyword
1	354	2010	Prosumer
2	226	2010	Smart grid
3	184	2011	System
4	142	2013	Demand response
5	135	2011	Microgrid
6	115	2014	Optimization
7	114	2014	Renewable energy
8	106	2014	Management
9	93	2014	Storage
10	78	2016	Power
11	77	2014	Energy
12	75	2010	Distributed generation
13	72	2016	Model
14	69	2014	Generation
15	68	2012	Energy management
16	64	2009	Demand-side management
17	58	2016	Network
18	54	2010	Electric vehicle
19	54	2017	Blockchain
20	54	2014	Electricity

TABLE 6: Hot keywords in the field of prosumer.

Number	Keyword
2009	Demand-side management (64)
2010	Prosumer (354), smart grid (226), distributed generation (75), electric vehicle (54), distributed energy resource (39), load (23)
2011	System (184), microgrid (135), renewable energy source (34), virtual power plant (21), ancillary service (12)
2012	Energy management (68), operation (47), energy management system (30), multiagent system (20), energy efficiency (18)
2013	Demand response (142), energy storage (52), electricity market (50), aggregator (28)
2014	Optimization (115), renewable energy (114), management (106), storage (93), energy (77), generation (69), electricity (54), battery (42), algorithm (32), building (27), distribution system (26), district heating (23), business model (18), battery storage (17)
2015	Game theory (34), photovoltaics (21), energy market (16), price (14), net metering (14)
2016	Power (78), model (72), network (58), self-consumption (37), technology (35), integration (33), flexibility (32), policy (30)
2017	Blockchain (54), market (38), energy trading (32), transactive energy (22)
2018	Challenge (17), household (14), sustainability (13), IoT (8)
2019	Peer-to-peer energy trading (10), optimal operation (7), peer to peer (7), local energy market (6), big data analytics (2)
2020	Game (6), home energy management system (3), energy community (2)

research within the 180 days before May 26, 2020, scholars focus on this article. This article uses the function of "Usage180" in CiteSpace, according to the number of visits to the full text in the past 180 days, or the number of times the record is saved, which counts recent high-frequency keywords in the field of prosumer [14]. Table 7 reflects the high-frequency keywords that appeared in the field of prosumer within the 180 days before May 26, 2020, which can identify the latest research hotspots. Among them, battery energy storage (battery and battery storage) has the highest frequency of occurrence, with a total frequency of 43 times. Battery energy storage is a recent focus of research by prosumer, which can improve the self-consumption of prosumers and the flexibility of energy exchange. Besides, it also includes aggregator, algorithm, ancillary services, auction, and architecture.

4.3. Research on Prosumer Based on the Knowledge Graph of Article Cocitation.
Article cocitation represents the knowledge base in the field. Through citation cluster analysis of document cocitation relationships, it is possible to understand the migration of research hotspots and research trends. This article uses the timeline diagram in CiteSpace, which describes clustering along the horizontal timeline. As shown in Figure 2, colored lines indicate the correlation between different clusters. Under each timeline, the three most cited references in a particular year will be displayed. According to the number of articles contained in the clusters, numbering starts from 0, and the 10 largest clusters are obtained from 73 clusters. Large nodes or nodes with red tree lines are particularly interesting because they are either highly cited, or have citation bursts, or both [44]. Table 8 shows the specific characteristics of each cluster in Figure 3,

TABLE 7: Keyword ranking in the past 180 days.

Number	Frequency	Centrality	Year	Keywords
1	33	0.04	2014	Battery
2	32	0.06	2014	Algorithm
3	29	0.18	2013	Aggregator
4	14	0.09	2017	Barrier
5	14	0.08	2009	Ancillary service
6	13	0.1	2017	Adoption
7	12	0.01	2015	Auction
8	10	0.02	2014	Battery storage
9	10	0.1	2010	Smart grid
10	8	0	2013	Architecture

TABLE 8: Overview of clustering of cocited documents.

Cluster	Quantity	Average year	Label words	Research focus
0	84	2017	Energy trading	Energy trading, peer-to-peer, automated negotiation, energy prosumer, local electricity markets, blockchain
1	68	2013	Self-consumption	Self-consumption, net metering, battery storage, distributed generation, photovoltaic (PV), vehicle-to-grid
2	63	2012	Smart grid	Smart grid, energy efficiency, virtual power plant (VPP), genetic algorithm (GA)
3	57	2014	Matching markets	Matching markets, game theory, microgrid, transactive energy, demand response
4	35	2015	District heating	District heating, mathematical modeling, district heating system
5	30	2010	Variable grid prices	Variable grid prices, complex network, network evolution, automated demand-side management, smart household appliance
6	26	2009	Information management	Information management, web services, power system economics
7	25	2012	Local electricity market	Local electricity market, multiagent system, intelligent agents
8	24	2015	Aggregators	Aggregators, flexibility, electricity markets, distribution network
9	19	2011	Profiling	Profiling, energy demand estimation, prosumers behavior, peer-to-peer communication
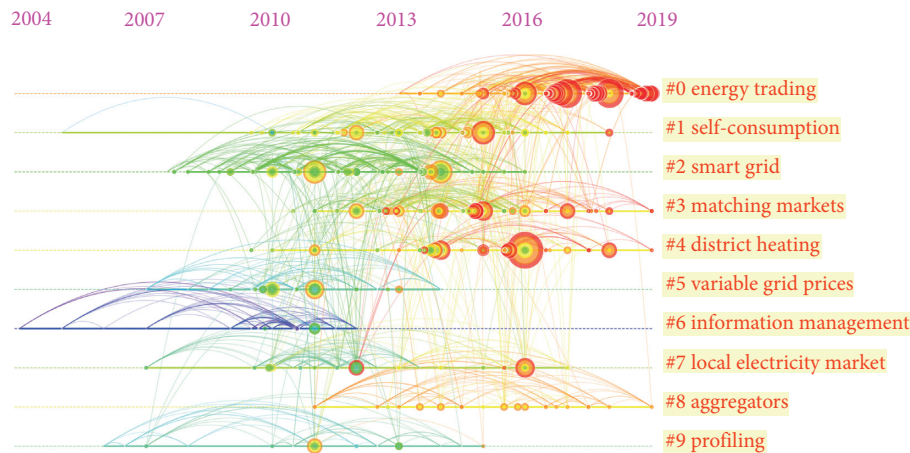
FIGURE 3: Time line diagram of the largest 10 clusters.

including the cluster number, the number of cluster members, the average publication years of the cluster literature, the main tags of the cluster, and the research focus. In this paper, the clustering tags are extracted from keywords, and the LLR (log-likelihood ratio) algorithm is used to generate the clustering tag words.

According to the research focus of the categories in Table 8, we can summarize several main research paths in the field of energy prosumer at the present stage. (1) Types of the prosumer, including electricity prosumer and thermal prosumer. Cluster 4 is concentrated in the district heating system, and the remaining clusters are concentrated in the

field of smart grid and microgrid. (2) Prosumer application scenarios: the research of clusters 1, 5, 6, and 9 is closely related to the distributed energy system. It can be seen that the prosumer model is beginning to be applied in the distributed energy system, which improves the flexibility of end-user's energy demand. (3) Prosumer transactions: clusters 0, 1, 2, 3, 5, 7, and 8 are all concentrated on prosumer transactions. Prosumer use distributed photovoltaics, batteries, and electric vehicles to meet and regulate loads, increase self-sufficiency, and reduce power purchase costs. Excess or insufficient energy is traded in local power markets such as communities through P2P mode, or through the P2G mode, transactions are based on grid electricity prices such as grid-connected electricity prices.

Among the above clusters, cluster 0 "energy trading" is particularly noteworthy. It has the largest number of documents, with an average age of 2017, and is closely related to the research content of other clusters. It is the cutting-edge content of research on prosumers. In recent years, many scholars have devoted themselves to research in this field, including trading market and trading platform design, operation models of different trading entities, and multi-agent trading pricing mechanism. This article will conduct a further cocitation analysis on the cross-references of the three market models (P2P mode, P2G mode, and community model of the prosumer) to find the cutting-edge research results and trends of prosumer.

This article retrieved articles related to the three market models of prosumers and consumers in the "Web of Science Core Collection." As of July 19, 2020, the search results are shown in Table 9. It can be seen that some literature studies involve two or more themes. Using highly cited literature on three themes as the research samples, the following research trends can be summarized.

The P2P model refers to direct negotiation and transaction between prosumers, mainly in complete P2P energy transaction. Mengelkamp et al. [16] proposed the concept of a blockchain-based microgrid energy market; besides, they designed and simulated the application of blockchain in the local energy market [45]. Zhang et al. [18] designed a P2P energy trading platform and bidding mechanism. Liu et al. [20] designed a shared pricing model between prosumers within the microgrid. Zhang et al. [46] analyzed the characteristics of existing P2P projects comparatively. Morstyn et al. suggested that prosumer jointly forms a virtual power plant [47]; also, to coordinate prosumer of different sizes, they proposed bilateral contracts for real-time and forward markets [48]. The P2P market may also involve prosumer agents, and P2P market transactions with smart agents seem to be more able to get the lowest electricity price [49]. Luo et al. [50] proposed to promote electricity trading by forming alliances between prosumer through agents. Etukudor et al. [51] used an autonomous agency model to establish an automatic P2P power transaction negotiation framework between consumers and prosumers. P2P energy trading can produce obvious benefits, but there are also a series of severe challenges at the technical, institutional, economic, and social levels. How to achieve key technological breakthroughs at the physical level of trading? How to reform the existing power market system and mechanism to adapt to distributed power trading? How to improve the economics of participating in transactions by prosumers? How to increase the willingness of the prosumer to participate and ensure fairness and privacy? These issues are still worthy of further exploration.

In the P2G mode, the prosumer can connect to the microgrid, and the MGO determines the transaction price and balances the power of the microgrid. Through the Stackelberg game, MGO maximizes profits, and prosumer maximizes utility [19, 52]. In addition to PV inside the microgrid, cogeneration [53] and battery energy storage [32] are also considered. Prosumers can also directly trade with the main grid. Generally, the state adopts preferential electricity tariff policies when encouraging PV popularization, such as feed-in tariffs and net metering. In recent years, compared with the FIT, net metering has received more attention in coordinating the power generation and consumption of prosumer [54]. Yamamoto et al. [55] compared the possible differences in social welfare and retail electricity prices among the three mechanisms of the feed-in tariff, net metering, and net purchase and sale. Gautier et al. [56] pointed out two drawbacks of the net metering policy: while reducing the cost of the prosumer, it will increase the cost of traditional consumer; it is not conducive to promoting local consumption of electricity. Eid et al. [57] designed cross-subsidies to coordinate policy objectives and the cost recovery balance of utility companies to change the unfairness of net measurement to non-PV owners. It is worth noting that the economic motivation for prosumer to install photovoltaic battery energy storage systems is more dependent on the self-consumption of photovoltaic power generation rather than trading with the grid [58]. Luthander et al. [59] reviewed two technologies for improving self-consumption: energy storage and demand-side management. According to the existing literature, they believe that the combination of the two can maximize the self-consumption rate. Weniger et al. [60] analyzed the proper configuration of residential photovoltaics and batteries and selected the most cost-optimized configuration for various cost scenarios. Noor et al. [61] introduced blockchain into demand-side management and implemented decentralized energy supply and demand management in practice. The P2G model is relatively easier to implement, but the two-way power flow will involve the issue of electricity prices and grid management pressure. Also, aggregators' participation in demand-side management and power market transactions can play an intermediary coordination role.

The prosumer community model, which is somewhere in between the former two, aims to promote local energy trading. Compared with the traditional P2G model, community P2P energy sharing can save energy costs. Rathnayaka et al. [62] and Yan [63] put forward the issues of community formation, common goals, and distribution of prosumers; Koirala et al. [64] reviewed the technical, socioeconomic, environmental, and institutional issues of the development of integrated community energy systems. Moret et al. [65] introduced a community-based collective energy market structure and evaluated it through fairness

TABLE 9: Number of documents in the three topic search formulas.

Number	Search	Quantity
1	TS = "prosumer*" AND TS = ("P2P" OR "peer-to-peer")	168
2	TS = "prosumer*" AND TS = ("P2G" OR "peer to grid" OR "feed-in" OR "sold to the grid" OR "prosumer* to grid")	75
3	TS = "prosumer*" AND TS = "community*"	292
4	#1 AND #2 AND #3	440

indicators. In terms of community energy transactions, Paudel et al. [66] pointed out that, in P2P transactions, in addition to the game between the buyer and the seller, there is also price competition within the buyer and the seller. Wang et al. [67] proposed a P2P energy transaction method using double auctions, which can improve system stability and protect the privacy of prosumer. The community also provides new ideas for energy storage sharing. Luth et al. [68] studied the benefits of decentralized and centralized energy storage in P2P energy transactions in the community. Van et al. [69] compared household energy storage and community energy storage methods and believed that the current two types of energy storage are not economical. Rodrigues et al. [70] considering that, in the case of producers and consumers participating in community P2P, community energy storage can reduce the investment costs of the prosumer. The prosumer community is a regulated P2P model, and future research hotspots are mainly the fairness of internal P2P transactions and the greatest overall benefits.

5. Research Conclusions and Prospects

5.1. Main Research Conclusions. This article uses qualitative literature review and quantitative bibliometric methods to systematically review the research in the field of prosumers. In this paper, CiteSpace software is used to analyze 1251 articles in the field of consumer and producer from 2009 to 2020 as follows: annual publishing trend, core journals, important authors, key countries/regions and organizations, keyword cooccurrence, hot keywords in recent 180 days, and the literatures cooccurrence. The main research conclusions of this paper are as follows.

Firstly, the trend of publishing articles in the field of the prosumer is increasing year by year. With the popularization of distributed photovoltaic, battery energy storage, and electric vehicles, this field is bound to gain more attention.

Secondly, authoritative journals in the field of prosumer focus on theoretical foundations, energy applications, energy policies, and energy modeling. Among them, "Renewable & Sustainable Energy Reviews" focuses on analysis and review of academic viewpoints, "Applied Energy" focuses on energy conversion applications, "Energy Policy" emphasizes energy policy, and "Energy" focuses on energy modeling and forecasting.

Thirdly, the main countries and regions in the field of prosumers are distributed in Europe, North America, Asia, and Australia. Among them, the United States ranked first with 138 articles. After that, except for China and Australia, the remaining seven were all European countries. Among the high-published organizations in the field of the prosumers, INESC TEC, located at the University of Porto in Portugal, topped the list with 26 articles, followed by North China Electric Power University, and the remaining eight institutions are also in Europe. It is worth noting that institutions in the United Kingdom, Spain, Australia, and Poland are not among the top ten high-published institutions, but there are institutions from Finland, Romania, and Denmark. It can be seen that Europe is still an important research area for prosumers, which benefits from Europe's pursuit of carbon neutrality in recent years and the promotion of the popularization of distributed power generation and energy storage by policies. Also, this shows that research on prosumer is still in its infancy, and various research institutions are exploring from different angles, and no institution has formed a strong academic influence.

Fourth, the results of the cooccurrence of keywords in the research on prosumer show that, in the context of smart grids, research on prosumer focuses on energy system management on the microgrid side, flexible resources on the prosumer side, and flexible response on the demand side and energy trading between the three. Besides, the research found that since 2015, the research focus has shifted from scientific and technological research on energy systems to soft scientific research on market transactions.

Fifth, the results of high-frequency keyword analysis in the research field of the prosumer in the past 180 days show that battery energy storage has become the biggest recent hotspot in the research of prosumers; the emergence of prosumer provides an opportunity for the development of various aggregators. It has also become a recent research hotspot.

Sixth, the analysis results of the cocited knowledge map of the literatures show that the mainstream research on energy prosumer is summarized into three parts, namely, the types of the prosumer, the application scenarios of the prosumer, and the trade of prosumer. The various subfields are still relatively scattered, but their rise and development explain the new market entity "energy prosumer" from different perspectives. Among them, energy trading is the most important hot area, and further research frontier analysis has been done from the three market models of P2P, P2G, and the prosumer community. Also, self-consumption, market matching, district heating, aggregators, and household energy management systems have been researched hotspots since 2015, indicating that research in these six fields is in a rapid development stage and there is still a lot of research space.

5.2. Outlook. Based on the analysis of the cocited literature of research on prosumer, this article finds that the main research hotspots in recent years are energy trading, self-

consumption, and distributed energy system management. Keyword analysis found that energy storage, aggregators, buildings, and home energy management systems are currently hotspots of high attention; social issues such as policies and motivations are hotspots of continuous attention, which need to be further studied and expanded. Also, the use of digital technologies such as big data, the Internet of Things, and blockchain in this field has attracted more and more attention. In the future, digital technology will have a significant impact on the participation of prosumer in energy market transactions, including upstream distribution network operations, microgrid operations, peer interaction between prosumer, and downstream demand response. This will profoundly change the existing energy supply chain model.

With the rise of the concept of the energy Internet, combining multienergy complementary systems and information and communication technologies to achieve the integration and development of multiple energy supply networks, research on prosumer and related fields has also attracted more attention. Based on the bibliometric analysis and the actual needs of the development of prosumer, this article summarizes some related research hotspots: ① in the context of the energy Internet, prosumers as transaction entities will not be limited to electricity prosumers, and the transaction field will be the whole energy Internet not limited to the grid, and the transaction objects will be a variety of energy categories; based on this, future research on prosumer can be placed in the context of the energy Internet to explore the key technologies, behavioral characteristics, transaction decisions, and path trends of the development of prosumer groups. ② Current research mostly focuses on engineering technology and economic analysis, and the research methods adopted are mainly quantitative mathematical models and relatively few research studies involving the perspective of social science. Research perspectives are mostly system optimization and transaction design, and there is less research on various corporate roles and the adjustment of business models in the market. Based on this, in future research, the management of prosumer can be discussed from the perspective of social science, on the one hand, in combination with management, sociology, and psychology and, on the other hand, the business model innovation of various enterprises in the existing market and the design of business models for new entities such as aggregators, virtual power plants, and integrated energy services. ③ Consumers and prosumers are more involved in energy transactions. Value cocreation blurs the boundaries between users and energy suppliers. The value cocreation and effects of the two in the interactive process need to be studied in depth. Based on this, from the value chain of the activities of the prosumer, consider the cooperative relationship between various enterprises and the prosumer and ordinary prosumers and improve the value cocreation ability between elements. ④ The development of prosumer still needs policy guidance and energy market reforms to adapt. Therefore, supporting policies and market reforms have become a common concern of enterprises and scholars.

In summary, domestic research can be further expanded in combination with the above current research hotspots and possible future hotspot issues. In the context of the energy Internet, energy trading platforms, digital ecosystems, cross-border innovation, and other fields will also become the focus of future research on the production and prosumer groups.

There are still deficiencies in the research in this paper. First of all, this research only selects data from the Web of Science Core Collection, and some research results may be omitted. Secondly, since the concept of energy prosumer has just emerged, the current research hotspots are relatively scattered. Although the overall picture of the research on energy prosumer can be presented, it is difficult to conduct an in-depth analysis of each cluster. Finally, this article only focuses on the international research results of energy prosumer, which can be compared with domestic research in the future, and provides reference suggestions for the development direction of domestic prosumer research.

Data Availability

The data used to support the findings of the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. Toffler, *The Third Wave*, B. Williams and T. Tapscott, Eds., Bantam Books, New York, NY, USA, 1981.

[2] R. Ford, J. Whitaker, and J. Stephenson, *Prosumer Collectives: A Review A Report for the Smart Grid Forum*, Centre for Sustainability, University of Otago, Dunedin, New Zealand, 2016.

[3] E. Espe, V. Potdar, and E. Chang, "Prosumer communities and relationships in smart grids: a literature review, evolution and future directions," *Energies*, vol. 11, no. 10, p. 2528, 2018.

[4] L. Brange, J. Englund, and P. Lauenburg, "Prosumers in district heating networks - a Swedish case study," *Applied Energy*, vol. 164, pp. 492–500, 2016.

[5] Y. Hongming, X. Tonglin, Q. Jing, Q. Duo, and Y. D. Zhao, "Optimal operation of DES/CCHP based regional multi-energy prosumer with demand response," *Applied Energy*, vol. 167.

[6] K. Kotilainen, M. Sommarberg, P. Järventausta et al., "Prosumer centric digital energy ecosystem framework," in *Proceedings of the 8th International Conference on Management of*

Digital EcoSystems, pp. 47–51, Biarritz, France, November 2016.

[7] R. Zafar, A. Mahmood, S. Razzaq, W. Ali, U. Naeem, and K. Shehzad, "Prosumer based energy management and sharing in smart grid," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 1675–1684, 2018.

[8] T. Sousa, T. Soares, P. Pinson, F. Moret, T. Baroche, and E. Sorin, "Peer-to-peer and community-based markets: a comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 104, pp. 367–378, 2019.

[9] P. Huang, B. Copertaro, X. Zhang et al., "A review of data centers as prosumers in district energy systems: renewable energy integration and waste heat reuse for district heating," *Applied Energy*, vol. 258, Article ID 114109, 2020.

[10] A. Ahl, M. Yarime, K. Tanaka, and D. Sagawa, "Review of blockchain-based distributed energy: implications for institutional development," *Renewable and Sustainable Energy Reviews*, vol. 107, pp. 200–211, 2019.

[11] K. Mahmud, G. E. Town, S. Morsalin, and M. J. Hossain, "Integration of electric vehicles and management in the internet of energy," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 4179–4203, 2018.

[12] M. Gough, F. S. Santos, M. Javadi, R. Castro, and P. S. Catalão, "Prosumer flexibility: a comprehensive state-of-the-art review and scientometric analysis," *Energies*, vol. 13, no. 11, p. 2710, 2020.

[13] C. Yue, C. Chen, Z. Liu et al., "Methodological function of CiteSpace knowledge graph," *Studies in Science of Science*, no. 2, pp. 242–253, 2015.

[14] L. Jie, *Citespace: Science and Technology Text Mining and Visualization*, Capital University of Economics and Business Press, Beijing, China, 2017.

[15] G. Mauri, D. Moneta, and C. Bettoni, "Energy conservation and smartgrids: new challenge for multimetering infrastructures," in *Proceedings of the 2009 IEEE Bucharest PowerTech*, pp. 1–7, IEEE, Bucharest, Romania, October 2009.

[16] E. Mengelkamp, J. Gärttner, K. Rock, S. Kessler, L. Orsini, and C. Weinhardt, "Designing microgrid energy markets," *Applied Energy*, vol. 210, pp. 870–880, 2018.

[17] Y. Parag and B. K. Sovacool, "Electricity market design for the prosumer era," *Nature Energy*, vol. 1, no. 4, pp. 1–6, 2016.

[18] C. Zhang, J. Wu, Y. Zhou, M. Cheng, and C. Long, "Peer-to-Peer energy trading in a Microgrid," *Applied Energy*, vol. 220, pp. 1–12, 2018.

[19] N. Liu, X. Yu, C. Wang, and J. Wang, "Energy sharing management for microgrids with PV prosumers: a Stackelberg game approach," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 3, pp. 1088–1098, 2017.

[20] N. Liu, X. Yu, C. Wang, C. Li, L. Ma, and J. Lei, "Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3569–3583, 2017.

[21] N. Liu, M. Cheng, X. Yu, J. Zhong, and J. Lei, "Energy-sharing provider for PV prosumer clusters: a hybrid approach using stochastic programming and Stackelberg game," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 8, pp. 6740–6750, 2018.

[22] W. Tushar, T. K. Saha, C. Yuen, P. Liddell, R. Bean, and H. V. Poor, "Peer-to-Peer energy trading with sustainable user participation: a game theoretic approach," *IEEE Access*, vol. 6, pp. 62932–62943, 2018.

[23] C. Long, J. Wu, Y. Zhou, and N. Jenkins, "Peer-to-peer energy sharing through a two-stage aggregated battery control in a community Microgrid," *Applied Energy*, vol. 226, pp. 261–276, 2018.

[24] Y. Zhou, J. Wu, and C. Long, "Evaluation of peer-to-peer energy sharing mechanisms based on a multiagent simulation framework," *Applied Energy*, vol. 222, pp. 993–1022, 2018.

[25] T. Morstyn and M. D. McCulloch, "Multiclass energy management for peer-to-peer energy trading driven by prosumer preferences," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 4005–4014, 2018.

[26] J. P. Iria, F. J. Soares, and M. A. Matos, "Trading small prosumers flexibility in the energy and tertiary reserve markets," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2371–2382, 2018.

[27] J. Iria, F. Soares, and M. Matos, "Optimal supply and demand bidding strategy for an aggregator of small prosumers," *Applied Energy*, vol. 213, pp. 658–669, 2018.

[28] M. Yazdani-Damavandi, N. Neyestani, M. Shafie-khah et al., "Strategic behavior of multi-energy players in electricity markets as aggregators of demand side resources using a bi-level approach," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 397–411, 2017.

[29] J. Iria, F. Soares, and M. Matos, "Optimal bidding strategy for an aggregator of prosumers in energy and secondary reserve markets," *Applied Energy*, vol. 238, pp. 1361–1372, 2019.

[30] J. Iria and F. Soares, "Real-time provision of multiple electricity market products by an aggregator of prosumers," *Applied Energy*, vol. 255, Article ID 113792, 2019.

[31] S. Cui, Y. W. Wang, J. W. Xiao et al., "A two-stage robust energy sharing management for prosumer microgrid," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2741–2752, 2018.

[32] H. Huang, S. Nie, J. Lin, Y. Wang, and J. Dong, "Optimization of peer-to-peer power trading in a microgrid with distributed PV and battery energy storage systems," *Sustainability*, vol. 12, no. 3, p. 923, 2020.

[33] K. Kontu, S. Rinne, and S. Junnila, "Introducing modern heat pumps to existing district heating systems-global lessons from viable decarbonizing of district heating in Finland," *Energy*, vol. 166, pp. 862–870, 2019.

[34] B. M. Delgado, S. Cao, A. Hasan et al., "Thermoeconomic analysis of heat and electricity prosumers in residential zero-energy buildings in Finland," *Energy*, vol. 130, pp. 544–559, 2017.

[35] M. Sanduleac, G. Lipari, A. Monti et al., "Next generation real-time smart meters for ICT based assessment of grid data inconsistencies," *Energies*, vol. 10, no. 7, p. 857, 2017.

[36] M. Sanduleac, J. Martins, I. Ciornei et al., "Resilient and immune by design microgrids using solid state transformers," *Energies*, vol. 11, no. 12, p. 3377, 2018.

[37] G. Brusco, A. Burgio, D. Menniti, A. Pinnarelli, N. Sorrentino, and L. Scarcello, "An energy box in a cloud-based architecture for autonomous demand response of prosumers and prosumages," *Electronics*, vol. 6, no. 4, p. 98, 2017.

[38] G. Belli, A. Giordano, C. Mastroianni et al., "A unified model for the optimal management of electrical and thermal equipment of a prosumer in a DR environment," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1791–1800, 2017.

[39] M. F. Zia, M. Benbouzid, E. Elbouchikhi, S. M. Muyeen, K. Techato, and J. M. Guerrero, "Microgrid transactive energy: review, architectures, distributed ledger technologies, and market analysis," *IEEE Access*, vol. 8, pp. 19410–19432, 2020.

[40] Y. Cai, T. Huang, E. Bompard et al., "Self-sustainable community of electricity prosumers in the emerging distribution

system," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2207–2216, 2016.

[41] X. Hou, J. Wang, T. Huang, T. Wang, and P. Wang, "Smart home energy management optimization method considering energy storage and electric vehicle," *IEEE Access*, vol. 7, pp. 144010–144020, 2019.

[42] C. H. Villar, D. Neves, and C. A. Silva, "Solar PV self-consumption: an analysis of influencing indicators in the Portuguese context," *Energy Strategy Reviews*, vol. 18, pp. 224–234, 2017.

[43] V. Reis, R. H. Almeida, J. A. Silva, and M. C. Brito, "Demand aggregation for photovoltaic self-consumption," *Energy Reports*, vol. 5, pp. 54–61, 2019.

[44] C. Chen, "Science mapping: a systematic review of the literature," *Journal of Data and Information Science*, vol. 2, no. 2, pp. 1–40, 2017.

[45] E. Mengelkamp, B. Notheisen, C. Beer et al., "A blockchain-based smart grid: towards sustainable local energy markets," *Computer Science-Research and Development*, vol. 33, no. 1-2, pp. 207–214, 2018.

[46] C. Zhang, J. Wu, C. Long, and M. Cheng, "Review of existing peer-to-peer energy trading projects," *Energy Procedia*, vol. 105, pp. 2563–2568, 2017.

[47] T. Morstyn, N. Farrell, S. J. Darby, and M. D. McCulloch, "Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants," *Nature Energy*, vol. 3, no. 2, pp. 94–101, 2018.

[48] T. Morstyn, A. Teytelboym, and M. D. McCulloch, "Bilateral contract networks for peer-to-peer energy trading," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2026–2035, 2018.

[49] E. Mengelkamp, P. Staudt, J. Garttner et al., "Trading on local energy markets: a comparison of market designs and bidding strategies," in *Proceedings of the 2017 14th International Conference on the European Energy Market (EEM)*, pp. 1–6, IEEE, Dresden, Germany, July 2017.

[50] F. Luo, Z. Y. Dong, G. Liang et al., "A distributed electricity trading system in active distribution networks based on multi-agent coalition and blockchain," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 4097–4108, 2018.

[51] C. Etukudor, B. Couraud, V. Robu, W.-G. Früh, D. Flynn, and C. Okereke, "Automated negotiation for peer-to-peer electricity trading in local energy markets," *Energies*, vol. 13, no. 4, p. 920, 2020.

[52] S. Cui, Y. W. Wang, and N. Liu, "Distributed game-based pricing strategy for energy sharing in microgrid with PV prosumers," *IET Renewable Power Generation*, vol. 12, no. 3, pp. 380–388, 2017.

[53] L. Ma, N. Liu, J. Zhang, W. Tushar, and C. Yuen, "Energy management for joint operation of CHP and PV prosumers inside a grid-connected microgrid: a game theoretic approach," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 5, pp. 1930–1942, 2016.

[54] G. Christoforidis, I. Panapakidis, T. Papadopoulos et al., "A model for the assessment of different net-metering policies," *Energies*, vol. 9, no. 4, p. 262, 2016.

[55] Y. Yamamoto, "Pricing electricity from residential photovoltaic systems: a comparison of feed-in tariffs, net metering, and net purchase and sale," *Solar Energy*, vol. 86, no. 9, pp. 2678–2685, 2012.

[56] A. Gautier, J. Jacqmin, and J.-C. Poudou, "The prosumers and the grid," *Journal of Regulatory Economics*, vol. 53, no. 1, pp. 100–126, 2018.

[57] C. Eid, J. Reneses Guillén, P. Frías Marín, and R. Hakvoort, "The economic effect of electricity net-metering with solar PV: consequences for network cost recovery, cross subsidies and policy objectives," *Energy Policy*, vol. 75, pp. 244–254, 2014.

[58] E. McKenna, J. Pless, and S. J. Darby, "Solar photovoltaic self-consumption in the UK residential sector: new estimates from a smart grid demonstration project," *Energy Policy*, vol. 118, pp. 482–491, 2018.

[59] R. Luthander, J. Widén, D. Nilsson, and J. Palm, "Photovoltaic self-consumption in buildings: a review," *Applied Energy*, vol. 142, pp. 80–94, 2015.

[60] J. Weniger, T. Tjaden, and V. Quaschning, "Sizing and grid integration of residential PV battery systems," in *Proceedings of the 8th International Renewable Energy Storage Conference and Exhibition (IRES 2013)*, Berlin, Germany, November 2013.

[61] S. Noor, W. Yang, M. Guo, K. H. van Dam, and X. Wang, "Energy Demand Side Management within micro-grid networks enhanced by blockchain," *Applied Energy*, vol. 228, pp. 1385–1398, 2018.

[62] A. J. D. Rathnayaka, V. M. Potdar, T. Dillon, O. Hussain, and S. Kuruppu, "Goal-oriented prosumer community groups for the smart grid," *IEEE Technology and Society Magazine*, vol. 33, no. 1, pp. 41–48, 2014.

[63] Y. Yan, Z. Haoran, L. Yin et al., "A factor-based bottom-up approach for the long-term electricity consumption estimation in the Japanese residential sector," *Journal of Environmental Management*, vol. 270, pp. 260–273, 2020.

[64] B. P. Koirala, E. Koliou, J. Friege, R. A. Hakvoort, and P. M. Herder, "Energetic communities for community energy: a review of key issues and trends shaping integrated community energy systems," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 722–744, 2016.

[65] F. Moret and P. Pinson, "Energy collectives: a community and fairness based approach to future electricity markets," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3994–4004, 2018.

[66] A. Paudel, K. Chaudhari, C. Long et al., "Peer-to-peer energy trading in a prosumer-based community microgrid: a game-theoretic model," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 8, pp. 6087–6097, 2018.

[67] Z. Wang, X. Yu, Y. Mu, and H. Jia, "A distributed Peer-to-Peer energy transaction method for diversified prosumers in Urban Community Microgrid System," *Applied Energy*, vol. 260, Article ID 114327, 2020.

[68] A. Lüth, J. M. Zepter, P. Crespo del Granado, and R. Egging, "Local electricity market designs for peer-to-peer trading: the role of battery flexibility," *Applied Energy*, vol. 229, pp. 1233–1243, 2018.

[69] S. Van Der Stelt, T. AlSkaif, and W. van Sark, "Techno-economic analysis of household and community energy storage for residential prosumers with smart appliances," *Applied Energy*, vol. 209, pp. 266–276, 2018.

[70] D. L. Rodrigues, X. Ye, X. Xia, and B. Zhu, "Battery energy storage sizing optimisation for different ownership structures in a peer-to-peer energy sharing community," *Applied Energy*, vol. 262, Article ID 114498, 2020.

Review Article

Optimization and Simulation of Carsharing under the Internet of Things

Yuxuan Wang[1] and Huixia Feng [ID][2]

[1]*Research Department of Shanghai Maicon Technology Co., Ltd., 555 Dongchuan Road, Minhang District, Shanghai 200241, China*
[2]*School of Mechanical and Transportation Engineering, China University of Petroleum Beijing, Fuxue Road No. 18, Changping District, Beijing 102249, China*

Correspondence should be addressed to Huixia Feng; 18434390388@163.com

Internet of Things devices are popular in civilian and military applications, including smart device cities, smart grids, smart pipelines, and medical Internet of Things. Among them, carsharing supported by the Internet of Things is developing rapidly due to their advantages in environmental protection and reducing traffic congestion. The optimization of the carsharing system needs to consider the uncertainty of demand and the coupling relationship of multiple decision variables, which brings difficulties to the establishment of mathematical models and the design of efficient algorithms. Existing studies about carsharing optimization are mainly divided into four subproblems: the operation mode selection, vehicle type selection, demand analysis, or decision-making, rather than comprehensive consideration. This paper summarizes the four subproblems from the perspective of mathematical models, solving algorithms, and statistical methods and provides references for more comprehensive research in the future.

1. Introduction

In 1999, the Massachusetts Institute of Technology defined the Internet of Things: connecting all items to the Internet through information sensing devices such as radio frequency identification. The Internet of Things was widely used in smart device cities, smart grids, smart pipelines, and medical Internet of Things. The Internet of Things used the Internet as a cornerstone of further expansion and development. With the help of GPS, infrared sensor, and other sensing devices, it transmitted and exchanged information between different mobile digital devices, namely, different entities. It had three characteristics: (1) intelligent sensing, (2) two-way transmission, and (3) intelligent control. There had been several proposals for unique object identifiers that uniquely identified objects and locations in the real world. Information could be associated with objects and places, and decoding could be used to retrieve relevant information. Karakostas [1] proposed a Domain Name Service (DNS)

infrastructure of the Internet of Things that could translate the unique identifier of a physical object into a specific network address and then extract information such as status and location. Due to the advantages of large capacity and high reliability of the Internet of Things, it provided an opportunity for the development of a new transportation mode named carsharing. Users could download the APP and register online to become a customer. Although the Internet of Things provided technical support for carsharing, there were still various problems in terms of its application and promotion. Figure 1 shows the operation process of the carsharing system. The user placed an order to reserve a car that has been charged or refueled. After the order was completed, the operator needed to relocate the car to meet the balance of supply and demand. The Internet of Things realized data feedback and instruction issuance.

Carsharing emerged informally as a consequence of gasoline prices rising in the 1940s [2] and had become popular in Europe and the United States since the 1980s and
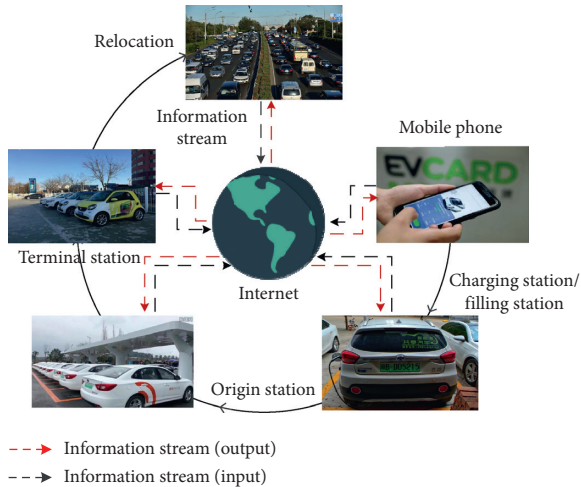
FIGURE 1: Structure of carsharing system.

then was introduced to many other countries. The growing number of private cars in developed cities brought about problems such as traffic congestion, environmental pollution, and insufficient social resources. For example, the average daily number of trips increased to more than 28 million, and the number of private vehicles covered more than 31% in Beijing which was affected by these problems deeply [3]. Studies showed that the introduction of carsharing had alleviated the seriousness of the congestion and air pollution problems to a certain extent and made the social resources more fully utilized [4–6]. It also increased the mobility of the city and provided a new transportation option.

The research on optimal design and operation of carsharing was divided into four subproblems, as shown in Figure 2. The characteristics of each part of the main decision-making contents are listed in the rectangle in Figure 2. Operation modes' selection should consider operating costs and demand. Vehicle type selection should consider the feasibility and environmental pollution. The total customer demand needed to be predicted by function fitting or neural network, and changes in uncertain demand over time and space should be considered. Decision-making was the most difficult research content because there were many decision variables and the variable coupling relationship was complicated.

2. Operation Modes

The operation modes of the carsharing were mainly divided into three types according to the terminal location: round trip, one-way trip, and free floating. Round trip and one-way trip were based on stations that differed from free floating [3]. The round trip required customers to return a car to the station where they picked up. Free floating was the most convenient for customers that they could return cars anywhere just within the operation areas [7]. While the one-way trip was a compromise between the two modes, in which customers should return cars at any station. The driving routes of the three operation modes are shown in Figure 3.

Apparently, the round trip cannot satisfy a lot of demands due to the rigorous stop condition [8], but the cost of relocation was decreased simultaneously. For the one-way trip and free-floating trip, the relocation aiming to rebalance the demand and supply was a critical factor that contributed to the cost [9, 10].

Brendel et al. [11] summarized 22 influential pieces of literature on carsharing. The results showed that most of the pieces of literature (about 82% of the total literature) studied station-based trip, while free floating is less studied due to the difficulty of rescheduling. Although a station-based one-way trip was suitable for more travel demands compared to a round trip, the round-trip demand would not be significantly reduced under the introduction of a one-way trip because that round trip was appropriate for purposes such as shopping, entertainment, and sightseeing [12, 13]. In addition, there were other factors affecting the choice of operation modes, such as age, gender, income, car ownership, and weather [14]. Some companies such as Zipcar have two modes named round trip for short trips and a one-way trip for long-distance travel. Customers who chose round trip had priority to reserve. This kind of service had already been studied. On one hand, some scholars started to add new services to the present modes to maximize the profit. Jorge et al. [15] developed an integer programming model to determine the station location that should be open to a one-way trip when integrating both one-way trip mode and the original round-trip mode. It was proved that there was a potential market for this model and it would greatly satisfy demands with the case study of Boston. But the disadvantage of this model was that it could only show its efficiency in some specified cases. On the other hand, some other scholars also started to add new services to the present modes to improve the service. Molnar and Correia [16] proposed a kind of long-term reservation service for a one-way and free-floating trip to improve user's satisfaction. And he developed the relocation-based reservation enforcement method (RB-R) to provide customers the ideal car. What is more, a quality of service model was developed and used to estimate and guarantee the satisfaction of customers. Finally, this model was tested in both small town and large major city and a case of Lisbon Municipality was studied to prove the utility of this model.

3. Vehicle Types

There were two main types of vehicles according to the power system: green energy vehicles (GEVs) and gasoline vehicles (GVs). GEVs could better reflect the main characteristic of carsharing, which was environmental protection. The most studied and introduced type of GEVs among the studies was the group of electric vehicles (EVs). Further discussion about the GEVs would mainly focus on EVs. The charging time of EVs was long, the travel distance was limited, and the investment could be enormous (the cost of charging station and charging facility). Normally, GEVs were more suitable for a round trip and a one-way trip. GVs were just the opposite, so they were more suitable for free-floating mode. It was the main research direction for
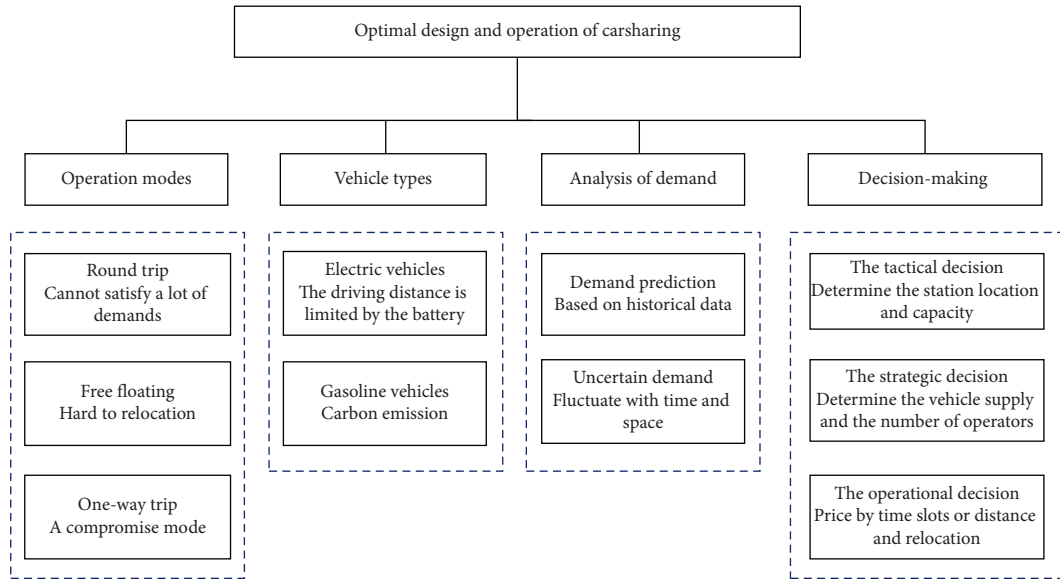
FIGURE 2: Classification of the research on optimal design and operation of carsharing.
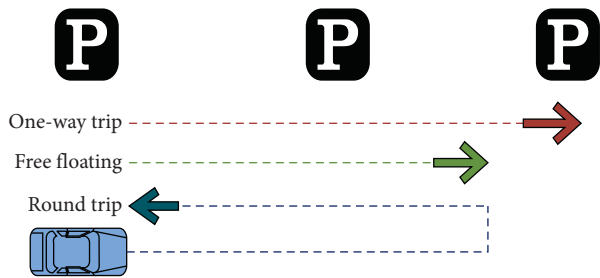
FIGURE 3: The driving routes of operation modes.

operators to choose which type of vehicles was more economical or whether operators that had adopted GVs in the early stage should introduce GEVs.

The social costs of electric vehicles and conventional vehicles could be a standard to determine which way was more profitable, and we should also consider the air pollution costs and the noise costs of conventional vehicles [17]. Based on a large number of GPS data, Kihm and Troomer [18] analyzed the prospects for the use of EVs considering different consumption attitudes. The results showed that it was more effective to reduce the cost of electric cars, improving public charging facilities and increasing government subsidies rather than increasing battery capacity in increasing the potential of electric cars. In addition, a powertrain selection model could be developed considering economic, technical, and social condition constraints on vehicle registration and inventory to analyze the market development and estimate the demand for vehicles [19]. For the problem of introducing electric vehicles to the original gasoline vehicles market, Yoon et al. [20] built a simulation model considering that the car might not be fully recharged before reentering the market which was always simplified in other papers and studied whether the electric car could compete with the gasoline fuel car under different market conditions. Actually, the battery capacity of electric vehicles also needed to be selected economically. According to the battery capacity, it can be divided into level 2 charging and level 3 charging. The level 3 charging was more expensive but drives a longer distance than level 2 charging. Comparing the type of vehicles and the charging level of electric vehicles, the most economical scheme could be more forceful [21].

4. Analysis of Demand

The demand in a certain area was partly affected by many territorial factors, such as population density, education levels, age, and private car park rate [22]. Also, the operation mode (round trip, one-way trip, and free floating) was another aspect that greatly accounted for a large proportion since there were different demands for different travel modes. Particularly, the free-floating mode could lead to many uncertainties. The determination of demand could be predicted by mathematical models based on the historical travel data of the region. To ensure the quality of the model, the data collection and analysis tools must be comprehensive and precisely suitable. In addition, there had been many pieces of literature accounting for the asymmetry and elasticity of the demand.

4.1. Demand Prediction. This subsection was to state how to predict the demand for a new region to provide advice for operators. When we make the demand prediction, various scenarios needed to be considered, including the travel distance, the number of travelers, the ratio of public transportation users, and the ratio of households without cars [21]. Because there were many factors affecting the demand, the differences in different regions lead to large errors in the prediction model [23], so the prediction model might only be used in specific regions. In order to predict the demand of a new region based on historical data of another region, it was necessary to clarify the differences between the

two regions and which differences were the main factors affecting the demand; then we used regression models to solve the functional relationship between the demand and these factors.

In the early years, scholars studied key indicators of whether or not the carsharing system could be successfully introduced [24]. Later, scholars explored agent-based simulation software to estimate the demand in different scenarios [25]. The GPS data of mobile phones can also be used to count the travel state of users during each period, and then the potential demand for EVs was analyzed [26–29]. After the carsharing was put into operation, the relationship between urban structure and high demand areas as well as the spatial and temporal distribution of supply and demand imbalance could be analyzed based on the historical data [7]. To predict the demand of vehicles, papers should consider multiple objective functions (such as minimum customer wait time and the minimum number of scheduling) and characterize the system performances according to the proposed evaluation indicators, such as the average wait time of the users, the total wait time, the number of waiting users, and the ratio of the number of available vehicles to the total demand of the trip [30]. The spatial decision support system was widely used to identify areas with high demand and exclude areas with low demand and Point of Interests (POIs) (such as shopping malls and schools) which could interpret the spatiotemporal dynamics and helped managers understand customer's behavior better [31]. In addition, some scholars summarized the factors affecting the use of carsharing in small- and medium-sized cities and determined the possibility of introduction [32, 33].

4.2. Uncertain Demand. This subsection was to state how to consider the uncertain demand in space and time.

The uncertain demand was a hot spot. Scholars before generally studied the linear elastic demand function [34]. In addition, the elastic demand could be considered as a random variable, and the Scenario Tree Approach could be used to solve the stochastic programming model, the first step of which was obtained by SAS macrocode. Others were assumed to follow a discrete distribution (such as low, medium, and high) and were predicted in turn according to the stochastic programming approach [35]. Xu et al. [36] firstly described the elastic demand more accurately as a nonlinear model. When the price was less than a threshold value, the demand satisfied the logistic regression function related to the price; otherwise, the demand became zero because the price was too high. Particularly, the mixed-integer nonlinear and nonconvex programming models were forged to be mixed-integer convex programming models. It could be solved by an efficient outer-approximation method. In addition, some scholars obtained specific functional relationships to describe the elastic demand through the logistic regression model considering the utility of carsharing and its competitors [37]. Developing an efficient algorithm could greatly reduce the time of solving and find the optimal or near-optimal solution. An et al. [38] adopt adaptive a heuristic algorithm of large-scale

neighborhood search to solve the complicated model. Considering the stochasticity of demand to establish a mixed-integer programming model with a nonconvex feasible region, a partial redistribution plan would be generated if the demand exceeded the supply [39]. Li et al. [40] proposed a continuum approximation model to determine the optimal station location and the corresponding fleet size of EVs considering stochastic and dynamic travel demand. By dividing the study area into multiple small neighborhoods, each small neighborhood approximates as an Infinite Homogeneous Plane and is finally solved by the bisection algorithm. Zhang et al. [41] considered the uncertain demand and established a multiscenario integer linear programming model to optimize the rebalancing procedures.

Due to so many factors affecting the demand, the model might not be realistic considering the limited factors. Therefore, some scholars used the neural network and support vector machine based on historical data to predict demand. The support vector machine could accurately predict the demand by selecting the appropriate kernel function. Cheu et al. [42] chose the radial basis kernel function and the average error was 0.42–0.83 vehicles per three hours. The calculation result was accurate but slightly worse than the multilayer perceptron. A mixed approach of genetic algorithm and backpropagation was proved to be efficient to train the neural network. The genetic algorithm was used to avoid that the neural network falls into local optimum and the backpropagation was used to accelerate convergence. Each chromosome represented a neural network model and had different weighs from each other (the number of nodes in the neural network model) but the same structure (the number of layers in the neural network model) and thus limited the diversity of the offspring [43]. Xu and Lim [44] improved the model by keeping the structure and weighs different between chromosomes, and for each generation, backpropagation only works on one chromosome, which accelerates the calculation. It should be noted that the neural network required a large amount of data for training.

5. Decision-Making

The business strategy could be classified into the following three decision level: strategic decision, tactical decision, and operational decision. The strategic decision determined the station location and capacity (the number of parking spaces). The tactical decision determined the vehicle supply and the number of operators, while operational decision determined the relocation scheme and how to price by time slots or distance. Actually, the established model should consider the three levels simultaneously for the strong interaction among them, but the model would be too large and cannot be used in cities with large demand. To address this problem, Boyacı et al. [45] introduced the aggregation model using the concept of a virtual hub which made a branch-and-bound approach available. For the optimization of one-way electric carsharing systems, Huang et al. [46] tracked the stage of charge over time and optimized the fleet size, station capacity, demand satisfaction, and vehicle relocation. A hierarchical strategy was adopted to solve the two

subproblems of strategy and operation. However, other existing pieces of literature often studied the three aspects separately or considered two aspects to avoid the surplus size of the model.

5.1. The Strategic Decision.

The places with more parking spaces, longer business hours, and higher population density had more booking demands and higher turnover rates, and they provided a basis for station selection [47]. The choice of station location should consider not only the demand but also the least relocation operations which minimize the imbalance between supply and demand. Actually, relocation during the day could lead to a tremendous cost, so that the model could consider the maintenance costs and the relocation costs only at the end of the day which provided a new strategy for companies [2]. Heuristic algorithms were popular these years, but they could only handle medium-sized instances [45]. Huang et al. [37] established an MINLP model to solve the station positioning and capacity considering the relocation operations and relocation costs except the allocation of staff and solved it with a customized gradient algorithm.

5.2. The Tactical Decision.

The carsharing system could be expressed as a hybrid queuing network model which took the road congestion into account in the optimization model to solve parking capacities and fleet size [48]. There was a function between the size of the city and demand density which was always used to explore the balance between fleet size and vehicle relocation [49]. Scholars mostly optimized the fleet size as well as the operational decision or the strategic decision rather than the fleet size individually. Cepolina and Farina [50] used the position, quantity, and capacity of the station as the input of the model and used the simulated annealing algorithm to solve the fleet size and vehicle distribution. For the optimization problem of the University of Tennessee (UT) motor pool, Yoon and Cherry [51] proposed a queuing model with the constraints of the limited distance and the limited charging time of electric vehicles to solve the fleet size of different types of the vehicle when the customer waiting time was close to zero.

5.3. The Operational Decision.

Relocation of vehicles which belonged to the operational stage could be operated in two ways, including operator-based location and user-based location [52]. The operator-based relocation was that the company hires employees for vehicle scheduling [53]. The user-based relocation was to provide customers with a reward and punishment mechanism to encourage customers to return the cars to places with larger demand or pick up a car from the lesser one [34].

For the operator-based location problems, scholars might adopt two-stage optimization or three-stage optimization to reduce the size of the model which could reduce the scheduling cost and shorten the time of solving simultaneously [54].

Establishing both the optimization model and the simulation model was adopted in some papers which could find an optimal scheme and study different real-time relocation strategies [55, 56]. The model of relocation was generally large and difficult to solve due to the large driving demand, so developing an efficient algorithm was a hotspot. Bruglieri et al. [3] developed an Adaptive Large Neighborhood Search metaheuristic solution for a relocation model and compared it with the Tabu Search. A previous Ruin and Recreate metaheuristic and the optimal results are obtained via Mixed-Integer Linear Programming to verify the superiority of the proposed algorithm on time of solving and optimization results. In addition to algorithm research, there are scholars who simplified models through mathematical reasoning to accelerate the solution and improve the quality of the solution. Zhao et al. [57] proposed a Lagrangian relaxation-based solution approach to divide an MINLP model into two MILP models and finally solved it with a three-phase implementing algorithm.

In addition, the regions were divided into blocks according to the demand or peak and nonpeak periods according to time, and different pricing strategies were adopted for different blocks to minimize the relocation operation [34, 58]. The most profitable pricing mechanism could be formulated by increasing the travel costs that would cause an unfavorable imbalance and reduce the travel costs that would help to slow down the imbalance [34]. Combining the two rescheduling mechanisms (operator-based relocation and user-based relocation) was sometimes more profitable. It rewarded the customer dynamically and periodically planned the routing for the dispatcher [59].

6. Conclusions

In the past fifteen years, research on carsharing has become more and more plentiful. Scholars have optimized and simulated the carsharing system through mathematical models and advanced algorithms, which provide a theoretical basis for future research. Future research can be based on the following aspects:

(1) In the modeling solution, literature introduces the assumptions for the simplified model, and the solution results deviate from the actual situation. For example, (1) there are few documents that consider the situation of delaying or canceling travel after the customer has made a reservation and the flexible choice of destination choice for customers; (2) charging time and level are ignored mostly.

(2) The model size of relocation in the carsharing system increases greatly with the increase of the number of stations and the demand for carsharing services. Therefore, it is urgent to develop an efficient algorithm or learn from mature fields such as refined oil scheduling and shared bikes scheduling.

(3) The model should be closer to the real-world instance and integrate the three decisions under uncertain demands, which is rarely considered in the existing research.

(4) Most researches lead providers to choose large cities with high demand. However, from a social perspective, the underdeveloped areas with lower demand should also enjoy the same convenience of sharing cars. Future research can focus on the operation of low-demand areas under the support of the government.

Disclosure

Yuxuan Wang and Huixia Feng are co-first authors of the paper.

Conflicts of Interest

The authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Authors' Contributions

Yuxuan Wang wrote the initial paper. Huixia Feng provided the overall idea of the paper and revised it. All authors contributed to the final paper.

References

[1] B. Karakostas, "A DNS architecture for the internet of things: a case study in transport logistics," *Procedia Computer Science*, vol. 19, pp. 594–601, 2013.

[2] G. H. D. A. Correia and A. P. Antunes, "Optimization approach to depot location and trip selection in one-way carsharing systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 1, pp. 233–247, 2012.

[3] M. Bruglieri, F. Pezzella, and O. Pisacane, "An adaptive large neighborhood search for relocating vehicles in electric carsharing services," *Discrete Applied Mathematics*, vol. 253, pp. 185–200, 2018.

[4] E. W. Martin and S. A. Shaheen, "Greenhouse gas emission impacts of carsharing in north America," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1074–1086, 2011.

[5] T. H. Stasko, A. B. Buck, and H. Oliver Gao, "Carsharing in a university setting: impacts on vehicle ownership, parking demand, and mobility in Ithaca, NY," *Transport Policy*, vol. 30, pp. 262–268, 2013.

[6] E. Martin, S. Shaheen, and J. Lidicker, "Impact of carsharing on household vehicle holdings: results from north american shared-use vehicle survey," *Transportation Research Record*, vol. 2143, no. 1, pp. 150–158, 2010.

[7] S. Schmöller, S. Weikl, J. Müller, and K. Bogenberger, "Empirical analysis of free-floating carsharing usage: the Munich and Berlin case," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 34–51, 2015.

[8] S. Le Vine, M. Lee-Gosselin, A. Sivakumar, and J. Polak, "A new approach to predict the market and impacts of round-trip and point-to-point carsharing systems: case study of London," *Transportation Research Part D: Transport and Environment*, vol. 32, pp. 218–229, 2014.

[9] C. A. Folkestad, N. Hansen, K. Fagerholt, H. Andersson, and G. Pantuso, "Optimal charging and repositioning of electric vehicles in a free-floating carsharing system," *Computers and Operations Research*, vol. 113, p. 104771, 2020.

[10] L. Shen and F. Tao, "Optimization of vehicle management for one-way carsharing systems based on time windows of users," in *Proceedings of the 8th International Conference on Industrial Technology and Management, ICITM*, Cambridge, UK, March 2019.

[11] A. B. Brendel, S. Lichtenberg, B. Brauer, I. Nastjuk, and L. M. Kolbe, "Improving electric vehicle utilization in carsharing: a framework and simulation of an e-carsharing vehicle utilization management system," *Transportation Research Part D: Transport and Environment*, vol. 64, pp. 230–245, 2018.

[12] M. Barth, S. A. Shaheen, and T. R. B. Trb, "Shared-use vehicle systems-framework for classifying carsharing, station cars, and combined approaches," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1791, no. 1, pp. 105–112, 2002.

[13] C. Costain, C. Ardron, and K. N. Habib, "Synopsis of users' behaviour of a carsharing program: a case study in Toronto," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 3, pp. 421–434, 2012.

[14] T. Yoon, C. R. Cherry, and L. R. Jones, "One-way and round-trip carsharing: a stated preference experiment in Beijing," *Transportation Research Part D: Transport and Environment*, vol. 53, pp. 102–114, 2017.

[15] D. Jorge, C. Barnhart, and G. H. De Almeida Correia, "Assessing the viability of enabling a round-trip carsharing system to accept one-way trips: application to Logan Airport in Boston," *Transportation Research Part C: Emerging Technologies*, vol. 56, pp. 359–372, 2015.

[16] G. Molnar and G. H. D. A. Correia, "Long-term vehicle reservations in one-way free-floating carsharing systems: a variable quality of service model," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 298–322, 2019.

[17] K. Funk and A. Rabl, "Electric versus conventional vehicles: social costs and benefits in France," *Transportation Research Part D: Transport and Environment*, vol. 4, no. 6, pp. 397–411, 1999.

[18] A. Kihm and S. Trommer, "The new car market for electric vehicles and the potential for fuel substitution," *Energy Policy*, vol. 73, pp. 147–157, 2014.

[19] S. Illgen and M. Höck, "Electric vehicles in car sharing networks-challenges and simulation model analysis," *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 377–387, 2018.

[20] T. Yoon, C. R. Cherry, M. S. Ryerson, and J. E. Bell, "Carsharing demand estimation and fleet simulation with EV adoption," *Journal of Cleaner Production*, vol. 206, pp. 1051–1058, 2019.

[21] H. Zhang, X. Song, T. Xia et al., "Battery electric vehicles in Japan: human mobile behavior based adoption potential analysis and policy target response," *Applied Energy*, vol. 220, pp. 527–535, 2018.

[22] F. Jin, E. Yao, and K. An, "Understanding customers' battery electric vehicle sharing adoption based on hybrid choice model," *Journal of Cleaner Production*, vol. 258, Article ID 120764, 2020.

[23] J. E. An and A. Millard-Ball, "Who is attracted to carsharing?" *Transportation Research Record*, vol. 1986, no. 1, pp. 98–105, 2006.

[24] C. Celsor and A. Millard-Ball, "Where does carsharing work?" *Transportation Research Record*, vol. 1992, no. 1, pp. 61–69, 2007.

[25] F. Ciari, B. Bock, and M. Balmer, "Modeling station-based and free-floating carsharing demand," *Transportation Research*

Record: Journal of the Transportation Research Board, vol. 2416, no. 1, pp. 37–47, 2014.

[26] Q. Yu, H. Zhang, W. Li, X. Song, D. Yang, and R. Shibasaki, "Mobile phone GPS data in urban customized bus: dynamic line design and emission reduction potentials analysis," *Journal of Cleaner Production*, vol. 272, Article ID 122471, 2020.

[27] H. Zhang, J. Chen, W. Li, X. Song, and R. Shibasaki, "Mobile phone GPS data in urban ride-sharing: an assessment method for emission reduction potential," *Applied Energy*, vol. 269, Article ID 115038, 2020.

[28] Q. Yu, H. Zhang, W. Li et al., "Mobile phone data in urban bicycle-sharing: market-oriented sub-area division and spatial analysis on emission reduction potentials," *Journal of Cleaner Production*, vol. 254, Article ID 119974, 2020.

[29] Y. Sui, H. Zhang, X. Song et al., "GPS data in urban online ride-hailing: a comparative analysis on fuel consumption and emissions," *Journal of Cleaner Production*, vol. 227, pp. 495–505, 2019.

[30] M. Barth and M. Todd, "Simulation model performance analysis of a multiple station shared vehicle system," *Transportation Research Part C: Emerging Technologies*, vol. 7, no. 4, pp. 237–259, 1999.

[31] C. Willing, K. Klemmer, T. Brandt, and D. Neumann, "Moving in time and space-location intelligence for car-sharing decision support," *Decision Support Systems*, vol. 99, pp. 75–85, 2017.

[32] K. W. Steininger and G. Bachner, "Extending car-sharing to serve commuters: an implementation in Austria," *Ecological Economics*, vol. 101, pp. 64–66, 2014.

[33] L. Rotaris and R. Danielis, "The role for carsharing in medium to small-sized towns and in less-densely populated rural areas," *Transportation Research Part A: Policy and Practice*, vol. 115, pp. 49–62, 2018.

[34] D. Jorge, G. Molnar, and G. H. De Almeida Correia, "Trip pricing of one-way station-based carsharing networks with zone and time of day price variations," *Transportation Research Part B: Methodological*, vol. 81, pp. 461–482, 2015.

[35] W. D. Fan, "Optimizing strategic allocation of vehicles for one-way car-sharing systems under demand uncertainty," *Journal of the Transportation Research Forum*, vol. 53, 2014.

[36] M. Xu, Q. Meng, and Z. Liu, "Electric vehicle fleet size and trip pricing for one-way carsharing services considering vehicle relocation and personnel assignment," *Transportation Research Part B: Methodological*, vol. 111, pp. 60–82, 2018.

[37] K. Huang, G. H. D. A. Correia, and K. An, "Solving the station-based one-way carsharing network planning problem with relocations and non-linear demand," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 1–17, 2018.

[38] G. An, R. Musmanno, and F. Vocaturo, "An adaptive large neighbourhood search heuristic for the capacitated arc-routing problem with stochastic demands," *Transportation Science*, vol. 44, no. 1, pp. 125–135, 2010.

[39] R. Nair and E. Miller-Hooks, "Fleet management for vehicle sharing operations," *Transportation Science*, vol. 45, no. 4, pp. 524–540, 2011.

[40] X. Li, J. Ma, J. Cui, A. Ghiasi, and F. Zhou, "Design framework of large-scale one-way electric vehicle sharing systems: a continuum approximation model," *Transportation Research Part B: Methodological*, vol. 88, pp. 21–45, 2016.

[41] H. Zhang, X. Song, Y. Long et al., "Mobile phone GPS data in urban bicycle-sharing: layout optimization and emissions

[42] R. L. Cheu, J. Xu, A. G. H. Kek, W. P. Lim, and W. L. Chen, "Forecasting shared-use vehicle trips with neural networks and support vector machines," *Transportation Research Record*, vol. 1968, no. 1, pp. 40–46, 2006.

[43] M. Mcinerney and A. P. Dhawan, "Use of genetic algorithms with backpropagation in training of feedforward neural networks," in *Proceedings of the IEEE International Conference on Neural Networks*, Nagoya, Japan, October 1993.

[44] J. X. Xu and J. S. Lim, "A new evolutionary neural network for forecasting net flow of a. car sharing system," in *Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, Singapore, September 2007.

[45] B. Boyacı, K. G. Zografos, and N. Geroliminis, "An optimization framework for the development of efficient one-way car-sharing systems," *European Journal of Operational Research*, vol. 240, no. 3, pp. 718–733, 2015.

[46] K. Huang, K. An, and G. H. D.A. Correia, "Planning station capacity and fleet size of one-way electric carsharing systems with continuous state of charge functions," *European Journal of Operational Research*, vol. 287, no. 3, pp. 1075–1091, 2020.

[47] S. Hu, P. Chen, H. Lin, C. Xie, and X. Chen, "Promoting carsharing attractiveness and efficiency: an exploratory analysis," *Transportation Research Part D: Transport and Environment*, vol. 65, pp. 229–243, 2018.

[48] L. Hu and Y. Liu, "Joint design of parking capacities and fleet size for one-way station-based carsharing systems with road congestion constraints," *Transportation Research Part B: Methodological*, vol. 93, pp. 268–299, 2016.

[49] J. A. Barrios and J. D. Godier, "Fleet sizing for flexible carsharing systems simulation-based approach," *Transportation Research Record*, vol. 2416, no. 1, pp. 1–9, 2014.

[50] E. M. Cepolina and A. Farina, "A new shared vehicle system for urban areas," *Transportation Research Part C: Emerging Technologies*, vol. 21, no. 1, pp. 230–243, 2012.

[51] T. Yoon and C. R. Cherry, "Migrating towards using electric vehicles in campus-proposed methods for fleet optimization," *Sustainability*, vol. 10, no. 2, 285 pages, 2018.

[52] M. Barth, W. X. Li, and M. Todd, "Interoperability options for shared-use vehicle systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1887, no. 1, pp. 137–144, 2004.

[53] A. Ait-Ouahmed, F. Zhou, and D. Josselin, "Optimization of electric carsharing service optimisation d'un service d'auto-partage de vehicules electriques," in *Proceedings of the 11th Annual International Conference on Spatial Analysis and GEOmatics SAGEO*, Hammamet, Tunisia, November 2015.

[54] A. G. H. Kek, R. L. Cheu, Q. Meng, and C. H. Fung, "A decision support system for vehicle relocation operations in carsharing systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 45, no. 1, pp. 149–158, 2009.

[55] D. Jorge, G. H. A. Correia, and C. Barnhart, "Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1667–1675, 2014.

[56] G. Alfian, J. Rhee, and B. Yoon, "A relocation simulation model for one-way carsharing service," in *IEEE International Conference on Industrial Technology, ICIT 2014*, Busan, South Korea, March 2014.

[57] M. Zhao, X. Li, J. Yin, J. Cui, L. Yang, and S. An, "An integrated framework for electric vehicle rebalancing and staff relocation in one-way carsharing systems: model formulation

and Lagrangian relaxation-based solution approach," *Transportation Research Part B: Methodological*, vol. 117, pp. 542–572, 2018.

[58] A. Angelopoulos, D. Gavalas, C. Konstantopoulos, D. Kypriadis, and G. Pantziou, "Incentivized vehicle relocation in vehicle sharing systems," *Transportation Research Part C: Emerging Technologies*, vol. 97, pp. 175–193, 2018.

[59] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari, "Dynamic vehicle redistribution and online price incentives in shared mobility systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1567–1578, 2014.

Research Article

Does Economic Growth Driving Force Convert? Evidence from China

Zejun Li,[1] **Zimei Huang,**[2] **and Tinghui Li** [2]

[1]*School of Computer and Information Science, Hunan Institute of Technology, Hengyang 421002, China*
[2]*School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China*

Correspondence should be addressed to Tinghui Li; lith@gzhu.edu.cn

The conversion from conventional to new driving forces and their time-varying characteristics are of great importance to China's economic transformation. In this paper, we attempt to investigate the economic growth driving force conversion in China and the time-varying characteristics of driving forces by constructing a time-varying coefficient panel data model during the period of 1998–2015. The empirical results indicate that China's economy has undergone driving force conversion. Specifically, China's economic growth driving forces have been transformed from traditional ones (human capital and gross fixed capital formation) to new ones (innovation capacity and structural transformation). Furthermore, we find that the characteristics of the driving forces are time-varying and heterogeneous. Innovation capacity and structural transformation have a more crucial impact on economic growth. Finally, based on the conclusions of the quantitative analysis, some important policy implications can be pursued to foster economic growth. Chinese government ought to enact various policies that are conducive to enhancing innovation capacity and accelerating structural transformation.

1. Introduction

The driving force of economic growth is an ongoing topic that has lasted for a long time. The evolution of the world economy is a leap from relying mainly on simple labour to relying mainly on mental labour. With the improvement of the level of economic development, more and more simple labour will be replaced by intelligence. Knowledge, innovation [1, 2], technology [3], information, and other intangible elements are playing a crucial role in economic growth. Through economic transformation, stimulate the vitality of new production factors and promote the rational flow and effective agglomeration of new production factors such as knowledge, technology, information, and data. Therefore, sustainable economic growth requires economic growth driving force conversion. Economic growth driving force conversion is through the new model, forms, technology, materials, and energy to replace the old ones, to realize the industrial upgrading, and to transfer from quantity, epitaxial, and labour-intensive economic growth to quality, connotation, and knowledge-intensive economic growth.

As a developing country, China has paid much attention to economic growth driving force conversion. China's economy has experienced years of sustained and rapid growth since the reform and opening up [4]. It is well known that China's economic growth is mainly driven by huge investment and energy-intensive industry [5]. At present, China's economy is transitioning from a phase of rapid growth to a stage of high-quality development. However, the labour- and investment-driven and energy-intensive economic growth mode of China has increased energy consumption and entailed serious environmental, resource, ecological, and social problems, thereby undermining green productivity [6] and resulting in less sustainable economic growth. To maintain sustainable economic development, China is required to undergo an economic transformation. Hence, this is a pivotal stage for transforming the growth model, improving economic structure, and fostering new driving forces of economic growth. Therefore, to foster economic growth, an accurate understanding of economic growth driving force conversion and its time-varying characteristics is extremely crucial.

In the process of transforming China's economic growth driving forces, green total factor productivity (GTFP) can effectively reflect the sustainability of economic growth [7]. Broadly, GTFP, also called environmental total factor productivity (TFP), is most commonly defined as the TFP that takes environmental factors and energy consumption into consideration [8, 9]. GTFP is an important index for measuring economic vitality because it can reflect the engine and quality of economic development [10]. A review of the available literature manifests that several advancements have been made in the methodological development of the typical productivity index. The measurement method of GTFP mainly includes the Solow residual method [11], stochastic frontier analysis (SFA), and DEA. The Solow residual method and SFA are based on the production function and required large sample data, which are subject to the conditions of the functional form itself and are unsuitable not only for a situation involving multiple inputs and outputs but also for small sample problems, thus, resulting in a certain gap between assumptions and reality [12]. However, DEA is a linear programming-based technique for measuring the relative performance of organizational units. Compared with the Solow residual method and SFA, DEA is unnecessary for fitting the production function and estimating parameters and suitable for a situation involving multiple inputs and outputs. Moreover, our problem is a small sample problem and we employed three inputs and three outputs to measure GTFP. Therefore, in this paper, DEA is a more powerful tool to measure GTFP.

Since the seminal studies of Caves et al. [13], there has been an explosion of work exploring the productivity index by employing the DEA method. The methods include Malmquist (M) productivity index [13], Luenberger (L) productivity index [14], Malmquist–Luenberger (ML) productivity index [15], global Malmquist (GM) productivity index [16], GML index [17], and GML index based on the SBM-DDF [10]. Specifically, M productivity index is not circular and does not account for the effect of environmentally harmful byproducts. GML productivity index is not only circular and immune to linear programming infeasibility, but also it considers undesirable output. To account for the potential of slack in technological constraints, Fukuyama and Weber [18] propose a directional SBM of technical inefficiency. To more accurately measure the GTFP, combining the advantages of the above method, Liu and Xin [10] construct a GML index based on the SBM-DDF. Compared with the above methods, this method can not only effectively deal with radial and oriented problems and achieve global comparability in the production frontier, but also take undesirable output into consideration. The GML index based on the SBM-DDF is extensively adopted to calculate the GTFP [19]. Therefore, in this paper, a GML index based on the SBM-DDF is employed to measure the GTFP.

According to the above, determinants of GTFP can be regarded as driving forces of economic growth. Various mainstream research studies have examined the influential factors of GTFP and have been studied from different perspectives [20]. For instance, Chen et al. [21] revealed that environmental regulation, technological innovation, independent research and development, and endowment and property right structure are the determinants of GTFP. Besides that, some scholars conducted empirical analyses to find that technological change [22], environmental governance [23], degree of openness, financial development [24–26], and monetary policy [27] are the main influencing factors of GTFP and economic growth.

The aim of this paper is to shed light on the economic growth driving force conversion in China. In this paper, we expand and supplement the existing literature in the following three regards. The first of this article's contributions is that we provide evidence on economic growth from the GTFP perspective. The main contribution of this perspective is that we can gain insight into the sustainability of economic growth by observing GTFP.

The second important contribution of our studies is that we attempt to further explore the economic growth driving force conversation by dividing the driving forces into conventional and new driving forces. In the past few years, China's rapid economic growth has been through labour and investment-driven modes, which we regarded as the main traditional economic growth driving forces. However, this mode promotes growth but entails serious environmental, resource, ecological, and social problems, resulting in less sustainable economic growth. Thus, it is urgent to foster new economic growth driving forces. Improving innovation capacity can accelerate technological progress, thereby promoting economic development. At the same time, optimizing and updating industrial structure will reduce environmental equality as well as improve resource allocation efficiency, thus enhancing GTFP. Hence, we consider innovation capacity and industrial structural transformation as new economic growth driving forces.

The third contribution of our work is to find empirical evidence for the time-varying characteristics of the conventional and new driving forces that contribute to GTFP. These time-varying characteristics are crucial for China to enact corresponding policies to foster economic growth. However, the vast majority of literature has focused on the theoretical research of economic growth driving force conversion. There is little empirical evidence on the time-varying characteristics of economic growth driving force conversion. In this article, we construct such an empirical analysis. Employing a sample of 30 Chinese provinces for the 1998–2015 periods, we construct a time-varying coefficient panel data model to exploit the characteristics of economic growth driving force conversion.

The paper proceeds as follows: Section 2 introduces the methods employed in this paper. In Section 3, we introduce the variables selection and data source. In Section 4, we perform a case study. Section 5 includes the conclusion.

2. Methods

In this section, we mainly focus on the methods that we employ in this paper. First, in Section 2.1, we introduce the global Malmquist–Luenberger (GML) index based on the slacks-based measure (SBM) directional distance function (DDF) to measure the GTFP. Section 2.2 introduces the panel regression model, which is adapted to investigate economic growth driving forces conversion. In Section 2.3, we briefly introduce the time-varying coefficient panel data model, which is employed to further investigate the time-

varying characteristics of the conventional and new driving forces' contributions to the GTFP.

2.1. GML Based on SBM.

Each province in China is regarded as a decision-making unit (DMU). First, following the contribution of Oh [17], we produce a global production possibility set (PPS), including M desirable outputs (real GDP), $y = (y_1, \cdots, y_n) \in R_M^+$, and I undesirable outputs (waste water and waste gas), $b = (b_1, \cdots, b_n) \in R_I^+$, and N inputs (i.e., labour input, capital input, energy input), $x = (x_1, \cdots, x_n) \in R_N^+$. The global PPS is defined as follows:

$$P^G(x) = \left\{ (y^t, b^t) : \sum_{t=1}^{T}\sum_{k=1}^{K} z_k^t y_{km}^t \geq y_{km}^t, \quad \forall m, \sum_{t=1}^{T}\sum_{k=1}^{K} z_k^t b_{ki}^t \geq b_{ki}^t, i, \sum_{t=1}^{T}\sum_{k=1}^{K} z_k^t x_{kn}^t \leq x_{kn}^t, \forall n, \sum_{k=1}^{K} z_k^t = 1, z_k^t \geq 0, \forall k \right\}, \quad (1)$$

where z_k^t denotes the weight of each province. If $z_k^t \geq 0$, it indicates the constant returns to scale (CRS). If $\sum_{k=1}^{K} z_k^t = 1$, $z_k^t \geq 0$, it indicates the variable returns to scale (VRS). In this paper, CRS are assumed.

Second, drawing on the research of Fukuyama and Weber [10, 18], we construct a global SBM-DDF that covers undesired output, as follows:

$$\vec{S}^G\left(x^{t,k'},\, y^{t,k'},\, b^{t,k'},\, g^x, g^y, g^b\right) = \max_{s^x, s^y, s^b} \frac{(1/N)\sum_{n=1}^{N}\left((s_n^x/g_n^x) + (1/(M+I))\right)\left(\sum_{m=1}^{M} s(s_m^y/g_m^y) + \sum_{i=1}^{I}\left(s_i^b/g_i^b\right)\right)}{2},$$

$$\text{s.t.} \begin{cases} \sum_{t=1}^{T}\sum_{k=1}^{K} z_k^t x_{kn}^t + s_n^x = x_{k'n}^t, \quad \forall n, \\[2mm] \sum_{t=1}^{T}\sum_{k=1}^{K} z_k^t y_{km}^t + s_m^y = y_{k'm}^t, \quad \forall m, \\[2mm] \sum_{t=1}^{T}\sum_{k=1}^{K} z_k^t b_{ki}^t + s_i^b \\[2mm] = b_{k'i}^t, \quad \forall i, \\ z_k^t \geq 0, \quad \forall k, \\ s_m^y \geq 0, \quad \forall m, \\ s_i^b \geq 0, \quad \forall i, \end{cases} \quad (2)$$

where (g^x, g^y, g^b) represents the direction vectors for decreasing labour, capital, and energy inputs, increasing real GDP, and decreasing waste water and gas, respectively, and (s_n^x, s_m^y, s_i^b) denotes the slack vectors for redundant labour, capital and energy inputs, inadequate real GDP, and redundant waste water and gas, respectively. If the value is greater than 0, the actual labour, capital, and energy inputs and the waste water and gas are greater than the boundary labour, capital, and energy inputs and waste water and gas, while the actual real GDP are less than the boundary real GDP.

Finally, we construct the GML index as follows:

$$\text{GML}^{t,t+1}\left(x^t,\, y^t,\, b^t,\, x^{t+1},\, y^{t+1},\, b^{t+1}\right)$$

$$= \frac{1 + \vec{S}^G\left(x^t, y^t, b^t; g^x, g^y, g^b\right)}{1 + \vec{S}^G\left(x^{t+1}, y^{t+1}, b^{t+1}, g^x, g^y, g^b\right)}, \quad (3)$$

where $\vec{S}^G(x^t, y^t, b^t; g^x, g^y, g^b)$ represents the global SBM-DDFs based on nonradial and nonoriented measurements.

The GML index denotes the change from period t to period $t+1$. If the GML is greater than 1, it indicates that GTFP increases. If the GML is equal to 1, the GTFP is in a stable state. Otherwise, the GTFP decreases.

However, the GML index is not the GTFP, but rather its change rate. Thus, regression treatment is needed. Following the work of Liu and Xin [10], we suppose that the GTFP of each province in 1997 is 1. Then, the GTFP in 1998 would be the GTFP in 1997 multiplied by the GML index: $\text{GTFP}_{1998} = \text{GTFP}_{1997} * \text{GML}_{1997-1998}$. The GTFP of other years can be calculated similarly.

2.2. Panel Regression Model.

Economic growth driving forces convert in China. Since the reform and opening up, China has made full use of the advantages of resource endowment and demographic dividend to form comparative advantages in product production, processing, and manufacturing, to promote economic growth. With China joining the World Trade Organization, the investment-driven development mode has become more important. However, since 2010, the

contradictions in economic development gradually appeared: the economic structure was unbalanced, the investment efficiency was reduced, the environmental carrying capacity continues to decline, and the economic constraints of resources and environment are gradually strengthened. China's economic growth rate has decreased to some extent [5]. With the rising labour cost, the advantages of conventional driving force are gradually weakened. The traditional growth pattern is no longer adapted to the new situation, and conventional driving forces of economic growth are being replaced by new ones. In other words, China converts economic growth driving forces.

To investigate the economic growth driving forces conversion, we specify the panel regression model. Over the past decade, the use of panel data in the econometric investigation has become popular [28]. The panel data approach encompasses data across cross sections and over time series, thus providing a comprehensive analysis to examine the influencing factors of economic growth driving force conversion [14]. The panel regression model typically assumes heterogeneity in the data, which allows us to completely capture by means of individual and time effects. The panel data model should be tested before estimating the model parameters, in order to facilitate the correct setting of the model in [29]. Panel regression models are estimated using recently developed techniques such as mean group estimators, which allow for heterogeneity in the estimation of the slope coefficients. Therefore, it can provide a more efficient estimation and information to discover the influencing factors of economic growth driving force conversion. To explore the conventional and new driving forces that affect economic growth, we perform the following panel regression model:

$$\mathrm{GTFP}_{it} = \alpha_i + \beta_1 * hc_{it} + \beta_2 * gf_{it} + \beta_3 * \mathrm{inno}_{it} + \beta_4 * st_{it}, \tag{4}$$

where i and t represent province i and time t, respectively, and α_i denotes the province effect, which captures the unobserved heterogeneity across provinces. GTFP is green total factor productivity, hc denotes human capital, gf stands for gross fixed capital formation, inno represents innovation capacity, and st refers to structural transformation.

2.3. Time-Varying Coefficient Panel Data Model.

The impact of economic growth driving forces on economic growth is time-varying. Economic growth driving forces conversion is a dynamic and continuous process, and the transformation process takes some time. Since the reform and opening up, China mainly relied on land, capital, and low-cost labour to promote economic development. Labour and capital have an important impact on economic growth. However, in recent years, China's economic development reached a certain extent and elements-driven economic growth gradually weakened, which cannot keep continuous high growth. At the same time, land tension, resource shortages, ecological environment deterioration, and other negative problems gradually appear [20]. In order to maintain sustained economic growth, it is necessary to improve the GTFP by improving the innovation capacity and structural transformation. That is to say, innovation capacity and structural

transformation exert a more crucial role in economic growth at present. In a word, various stages of economic development have different economic growth driving forces. Therefore, it is crucial for China to understand the time-varying characteristic of the conventional and new driving forces' contribution to economic growth.

Varying coefficient models were developed by Hastie and Tibshirani [30] and are an extension of classical linear regression models in the sense that the regression coefficients are replaced by functions in certain variables (often time t). Varying coefficient regression models are regarded as very useful tools for analysing the relationship between a response and a group of covariates [31]. The varying coefficient model allows coefficients to depend on some informative variables and model flexibility. Therefore, varying coefficient models have been popular in longitudinal data and in panel data studies and have been applied in fields such as finance and economics. In addition, because they take on spatiotemporal characteristics, panel data can well describe and depict the systematic and dynamic characteristics of the decision objects.

A varying coefficient panel data model includes the time- and individual-varying coefficient panel data models. The time-varying coefficient panel data model allows us to detect the time-varying impacts of the conventional and new driving forces on the GTFP. Since various responses to GTFP may be expected at different times, it is more suitable for us to employ the time-varying coefficient panel data model to detect the time-varying characteristics of the conventional and new driving forces' contributions to the GTFP. The estimation methods of the time-varying coefficient panel data model include the intragroup dispersion method and the least squares dummy variable method. In this paper, we employ the least squares dummy variable method to estimate the coefficients. Therefore, we construct a time-varying coefficient panel data model for province i at time t as follows:

$$\mathrm{GTFP}_{it} = D\alpha_i + \beta_{1t} * gf_{it} + \beta_{2t} * \mathrm{inno}_{it} + \beta_{3t} * st_{it} + \mu_{it}, \tag{5}$$

where D is the dummy variable, i and t represent province i and time t, respectively, GTFP denotes green total factor productivity, gf stands for gross fixed capital formation, $inno$ represents innovation capacity, and st refers to structural transformation; β denotes a vector of estimated parameters in the equation; and μ_{it} is the error term.

3. Variables Selection and Data Source

In this section, we introduce the variables selection and data source. In Section 3.1, we introduce the samples. In Section 3.2, we briefly describe the variables employed to measure GTFP. In Section 3.3, we select the driving forces.

3.1. Samples.

This paper is conducted in China and focuses on the economic growth driving force conversion. Considering the integrity and availability of data, the sample excludes Tibet, Hong Kong, Macao, and Taiwan. We divide

the sample into three subsamples: east, middle, and west areas. The east area (sample 1) includes Beijing, Shanghai, Tianjin, Jiangsu, Guangdong, Liaoning, Hainan, Hebei, Fujian, Shandong, and Zhejiang. The middle area (sample 2) includes Anhui, Hubei, Hunan, Henan, Jiangxi, Heilongjiang, Jilin, and Shanxi. The west area (sample 3) includes Gansu, Guangxi, Sichuan, Inner Mongolia, Ningxia, Qinghai, Shaanxi, Xinjiang, Yunnan, Chongqing, and Guizhou. Annual data are used for 30 Chinese provinces during the period from 1997 to 2015. The data are obtained from the Easy Professional Superior (EPS) macro database and the National Bureau of Statistics.

3.2. Variables to Measure GTFP.

This study measures the GTFP of 30 Chinese provinces. The measurement of GTFP integrated three dimensions: the economy, resources, and environment. According to the existing literature, we know that labour, capital, and energy are the most frequently used resource input indicators [10], while real gross domestic product (GDP) is the most frequently used economy indicator and desirable output [32], and waste water, waste gas, and solid waste are the most frequently used environmental indicators and undesirable outputs in measuring GTFP [33]. In this paper, because of the availability of data, we employ three resource indicators (labour input, capital input, and energy input) as input variables, one economy indicator (real GDP) as desirable output, and two environmental indicators (waste water and waste gas) as undesirable outputs to measure the GTFP. These inputs and outputs indicators are specified as follows.

3.2.1. Labour Input.

Labour input refers to the amount of labour. Due to labour mobility, the amount of labour input is different at different times in one year, so the number of year-end employed persons is taken as the indicator [20], where a unit is ten thousand persons. The data of the labour of China's 30 provinces from 1997 to 2015 were directly obtained from the EPS macro database.

3.2.2. Capital Input.

Capital input is measured by capital stock. We employ the perpetual inventory method to estimate capital stock, for which the basic equation is $K_{it} = I_{it} + (1 - \delta)K_{i(t-1)}$, in which K_{it} is the actual capital stock in province i in period t, and I_{it} is the gross fixed capital formation in area i in period t. Compared with the fixed assets investment of the whole society, the gross fixed capital formation is slightly better than the former when measuring capital stock [10]. Therefore, we use the gross fixed capital formation to calculate capital stock. Additionally, δ represents the depreciation rate. Different depreciation rates are adopted to calculate capital stock among different provinces. Because of the partial lack of data from the Chinese provinces, we follow the study of Zhang [34] to estimate the Chinese provinces' capital stock and adopt 9.6% as the depreciation rate. The base period is 1997. Capital stock is calculated in 100 million at the constant price in 1997. The corresponding data for 1997–2015 come from the National Bureau of Statistics.

3.2.3. Energy Input.

The total energy consumption, GDP divided by GDP per unit of energy consumption, is employed to measure energy input. Total energy consumption refers to the total consumption of various kinds of energy by the country's production sectors in a given period of time and shows the scale, composition, and pace of the increase in energy input [4, 35]. The unit of energy input is 10000 tce, and the data come from the EPS macro database.

3.2.4. Desirable Output.

Similar to much of the existing literature, we use the real GDP to represent the desirable output. To ensure the comparability of the data, it is converted based on the year 1997. The data originate from the National Bureau of Statistics.

3.2.5. Undesirable Output.

The existing literature concerning undesirable outputs mainly includes three forms: waste water, waste gas, and solid waste. However, due to the unavailability of data regarding provincial solid waste, it was excluded from this study. Therefore, we employ the total industrial waste water discharge and the total industrial waste gas discharge as undesirable outputs to calculate GTFP. A unit of waste water is 10,000 tons, and a unit of waste gas is 100 million cube meters. The data for the period from 1997 to 2015 were obtained from the EPS macro database.

3.3. Driving Forces.

In this paper, we select GTFP as the explained variable. Human capital and gross fixed capital formation are regarded as the conventional driving forces, while innovation capacity and structural transformation are regarded as the new driving forces.

First, human capital, which is calculated by the number of year-end employed people multiplied by their average years of schooling, is considered as an important driving force of economic growth [36]. Human capital is usually the sum of the knowledge, technical skills, capabilities, and qualities that can create economic and social value. The average years of schooling have been commonly used as the specification of the quantity of human capital stock empirically [37]. The proportion of employees at the primary, junior, middle, and higher education levels to the total employees calculates the average years of education of the labours in a certain area. The effect of human capital on GTFP involves multiple channels. On one hand, an increase in human capital directly affects GTFP by enhancing labour productivity in production. On the other hand, human capital is an important input into R&D and therefore indirectly increases labour productivity by accelerating technological change [38]. Human capital is conducive to producing and adopting new technologies, improving resource utilization efficiency and GTFP, thus promoting economic growth. The human capital data can be obtained from the EPS macro database.

Second, gross fixed capital formation is regarded as a crucial driving force of GTFP. Economic growth can be promoted by gross fixed capital formation through creating

massive benefits, increasing investments by creating enlarged markets and economies of scale, and transferring information, technology, and knowledge spillovers. To ensure the comparability of the data, it is converted based on the year 1997. The data originate directly from the National Bureau of Statistics.

Third, innovation capacity is an important driving force for productivity growth [1]. Technological innovation plays an important role in optimizing energy structure and in promoting economic growth. As pointed out by Pan et al. [39], the information on patent grants is covered in patent applications. Simultaneously, pendency from patent applications to grant is necessary, as patent grants cannot truly reflect the current level of regional innovation capacity. Therefore, in this paper, the natural logarithm of patent applications is used to calculate innovation capacity, and the original data was obtained from the EPS macro database.

Fourth, structural transformation is measured by the coefficient of industrial structure, which can reflect the industrial distribution. According to the principle of industrial evolution, we adopt the sum of the proportion of the added value of the three industries accounting for the added value of the primary industry to calculate the coefficient of industrial structure. Structural transformation results in changes in energy consumption and pollution emission, consequently, having great effects on GTFP. It is asserted that differences in industry structure may inevitably affect GTFP. The structural transformation data was collected from the EPS macro database.

The measurements and sources of the conventional and new driving forces are shown in Table 1.

4. Case Study

In this section, we mainly perform a case study. Firstly, we present the GTFP characteristics of 30 Chinese provinces in Section 4.1. In Section 4.2, we analyse the results of economic growth driving forces conversion. In Section 4.3, we further analyse the results of the time-varying characteristics of the conventional and new driving forces' contributions to the GTFP.

4.1. Stylized Facts of GTFP.
Based on the above data, the MaxDEA software is used to measure the GTFP of 30 Chinese provinces from 1998 to 2015, and the average GTFP of China and its east, middle, and west areas is presented in Figure 1.

There is rising and spatial heterogeneity concerning China's GTFP. On one hand, from a national perspective, China's GTFP increased from 0.983 to 1.159 during the period from 1998 to 2015, showing an overall upward trend. This means that during the sample period, China's GTFP development rose continuously. This is mainly due to the increasing awareness of the damage of environmental pollution and the importance of sustainable economic development in China. Consequently, China tried to accelerate the transformation of the mode of economic growth by enacting a series of policies on energy conservation and emission reduction, which can promote economic growth, reduce environmental pollution, and improve the GTFP.

On the other hand, from the perspective of regional distribution, there is a spatial heterogeneity of GTFP. First, in the east regions, GTFP shows a rising trend. This is in accordance with the fact that the provinces in the east area are almost developed provinces, such as Beijing, Shanghai, Tianjin, Jiangsu, and Guangdong, which possess strong economic power, rich educational resources, convenient transportation, and advanced environmental technology. As a result, their GTFP is maintained at a high level, a win-win situation for sustainable economic growth and environmental protection. However, the GTFP in the middle and west areas implies a trend of falling and then rising. The GTFP is lower than 1 during the period from 1998 to 2015, which is at a low level. This is mainly owing to the relatively weak economic foundation of the middle and west provinces and inadequacy in the introduction of technical talents, as well as shortages of technological innovation capabilities and pollution control equipment, which has led to environmental technology lagging behind the eastern coastal areas. At the same time, the central and western provinces also need to undertake the transfer of heavy pollution, high investment, and labour-intensive industries from the eastern provinces, thereby further aggravating environmental pollution and causing the low GTFP.

Therefore, the GTFP among the 30 Chinese provinces has spatial heterogeneity. This is in line with the current imbalance in China's regional economic development. High-tech industries are developed in developed areas, while high-pollution and high-energy industries are transferred to the less developed regions, resulting in more environmental pollution and decreases in the GTFP of these less developed provinces.

4.2. Conversion of Economic Growth Driving Forces.
Before investigating the economic growth driving force conversion, we conduct a panel unit root test. For the preliminary investigation, we perform a panel unit root test by applying the Levin–Lin–Chu (LLC) test proposed by Levin et al. [40] and the Fisher-augmented Dickey–Fuller (ADF) test proposed by Maddala and Wu [41]. The results of the panel unit root test are exhibited in Table 2. The decision criterion is that the variable is stationary if the unit root tests confirm nonrejection of the null hypothesis at a 10% level of significance. The results of the LLC and Fisher-ADF tests indicate that all variables reject the null hypothesis (nonstationary) at a 10% significance level. That is to say, all variables used in this study are stationary.

Subsequently, we use Hausman tests to choose the specific panel regression model. The results of the Hausman tests reject the null hypothesis in the full sample and sample 1, while they did not reject the null hypothesis in samples 2 and 3. In other words, we should employ a fixed effects (FE) model for the full sample and sample 1 and a random effects (RE) model for samples 2 and 3 to implement the regression analysis. The standardized coefficients of the panel regression model are documented in Table 3, where columns 2, 3, 4, and 5 report the impact of the driving forces on the GTFP in the full sample and in samples 1, 2, and 3, respectively.

TABLE 1: Conventional and new driving forces.

Driving force	Variable	Measurement	Source
Conventional	Human capital	Based on the years of education of the labours and measured by millions of persons	EPS macro database
	Gross fixed capital formation	Gross fixed capital formation	National Bureau of Statistics
New	Innovation capacity	Log (patent applications)	EPS macro database
	Structural transformation	Coefficient of industrial structure	EPS macro database

Notes: this table exhibits the measurement and source of conventional and new driving forces of economic growth.
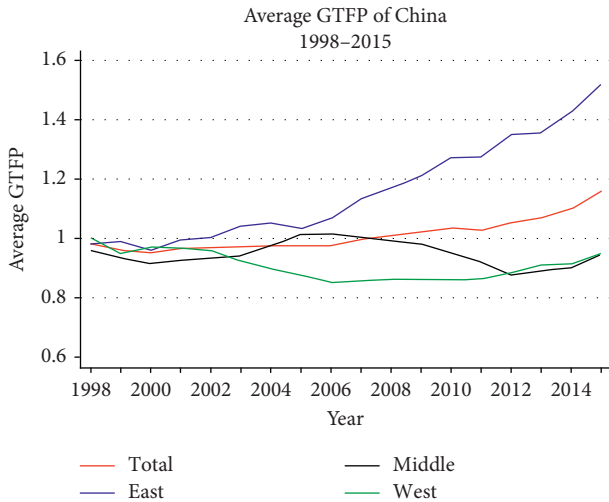
FIGURE 1: China's average GTFP. The red line represents China's average GTFP, while the blue, black, and green lines indicate the mean average GTFP of China's east, middle, and west areas, respectively. The sample period is from 1998 to 2015.

TABLE 2: Results of the panel unit root test.

Variables	LLC	Fisher-ADF test
GTFP	−15.615 (0.000)	160.161 (0.000)
hc	−2.409 (0.008)	145.170 (0.000)
gf	−3.398 (0.000)	167.332 (0.000)
inno	−5.810 (0.000)	161.382 (0.000)
st	−4.769 (0.000)	175.940 (0.000)

Notes: P values are shown in parentheses. The sample period is from 1998 to 2015.

TABLE 3: Results of the panel regression model.

	Full Sample	Sample 1	Sample 2	Sample 3
hc	0.110 (0.0599)	0.028 (0.109)	−0.110 (0.118)	0.696*** (0.0363)
gf	−0.719*** (0.0299)	−0.485*** (0.0608)	0.051 (0.0576)	−1.344*** (0.0196)
inno	0.283** (0.0290)	−0.316* (0.0634)	0.062 (0.0502)	0.758*** (0.0160)
st	1.293*** (0.0674)	2.825*** (0.122)	−0.127 (0.127)	0.211** (0.0493)
N	540	198	144	198
R^2	0.258	0.564		

Notes: this table presents the standardized coefficients of the panel regression model. Symbols ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively. Standard errors are shown in parentheses.

China's economy has undergone driving force conversion. It is apparent from the standardized results given in Table 3 that, whether in the full sample or in subsamples 1 and 2, new driving forces (innovation capacity and structural transformation) promote economic growth more than conventional driving forces (human capital and gross fixed capital formation), which means that China's economic growth driving forces have been transformed from traditional ones to new ones. This is in accord with the current economic growth driving force conversion in China's economic development.

The impacts of different driving forces on the GTFP are spatially heterogeneous. In the full sample, the conventional driving force affecting the GTFP is gross fixed capital formation, while the new driving forces are innovation capacity and structural transformation. The results imply that human capital has an insignificant impact on the GTFP, whereas gross fixed capital formation has a significant negative impact on the GTFP. In other words, in the full sample, the conventional driving force affecting the GTFP is only gross fixed capital formation. This negative impact may be closely related to the following phenomenon: with the development of the economy, the investment efficiency in China is decreasing. Furthermore, in the full sample, the new driving forces, innovation capacity and structural transformation, have significant and positive effects on the GTFP. The reasons may be that improving innovation capacity and transforming industrial structure can improve resource allocation efficiency and GTFP.

In the east area, gross fixed capital formation and innovation capacity have a significant negative impact on the GTFP, while structural transformation exerts a positive and significant effect on the GTFP. These results may be because the east provinces have always had abundant capital, advanced environmental technology, and strong innovation capacity. Consequently, the transformation of industrial structure will boost productivity. However, the improved productivity may not be environment friendly, which will cause more environmental pollution, thus reducing the GTFP. The transformation of industrial structure is beneficial to improving resource allocation efficiency and GTFP.

In the middle area, neither the conventional nor the new driving forces exert a significant impact on the GTFP, while both the conventional and the new driving forces exert a significant impact on the GTFP in the west area. The middle areas develop their economies through introducing capital and technology while simultaneously introducing a large amount of environmental pollution. Therefore, neither the conventional

TABLE 4: Results of the varying coefficient panel data model.

Year	gf	inno	is
1998	0.181	−0.292**	−0.382
1999	0.130	−0.265	−0.293***
2000	0.045**	−0.199*	−0.231***
2001	−0.031***	−0.130**	−0.224***
2002	−0.049**	−0.121**	−0.177***
2003	−0.096***	−0.084**	−0.143***
2004	−0.135***	−0.050**	−0.125***
2005	−0.141***	−0.044**	−0.108***
2006	−0.188***	0.002***	−0.107***
2007	−0.205***	0.007***	−0.045***
2008	−0.227***	0.014***	0.019***
2009	−0.232***	0.010**	0.076***
2010	−0.238***	0.006**	0.131***
2011	−0.249***	0.015**	0.144***
2012	−0.251***	−0.003**	0.244***
2013	−0.239***	0**	0.203***
2014	−0.258***	0.005**	0.268***
2015	−0.273***	0.020**	0.285***

Notes: symbols ***, **, and * indicate significance at 1%, 5%, and 10% levels, respectively.

FIGURE 2: Results of the varying coefficient panel data model. Notes: the sample period is from 1998 to 2015. The red dash line denotes the impact of the conventional driving force (gross fixed capital formation) on the GTFP. The solid blue and black line represents the impact of innovation capacity and structural transformation on the GTFP, respectively.

nor the new driving forces exert a significant impact on the GTFP. The west areas, which remain at a low level, boost their economies and GTFP by improving the human capital level and innovation capacity, introducing advanced technology, and accelerating industrial structure transformation. Overall, there are spatial heterogeneous impacts among the different areas. These heterogeneous impacts may be closely connected with the following phenomenon: different provinces have achieved different economic development levels and various economic development modes.

4.3. Time-Varying Impacts of the Conventional and New Driving Forces on GTFP. Table 4 presents the results of the time-varying coefficient panel data model for our 30 Chinese provinces in the period from 1998 to 2015. Columns 2, 3, and 4 in Table 4 report the time-varying impacts of gross fixed capital formation, innovation capacity, and structural transformation on the GTFP, respectively. Figure 2 intuitively presents the time-varying characteristics of the conventional and new driving forces' contributions to the GTFP.

The time-varying characteristics of the conventional and new driving forces that contribute to the GTFP are heterogeneous. As seen in Table 4 and Figure 2, the contributions to the GTFP among the different driving forces are inconclusive, which indicates that the time-varying impacts of the conventional and new driving forces that contribute to the GTFP are various. On one hand, the contribution of the conventional driving forces (gross fixed capital formation) is not promising. This result is consistent with the study of Li and Lin [5], which examined the impacts of an investment-driven economic
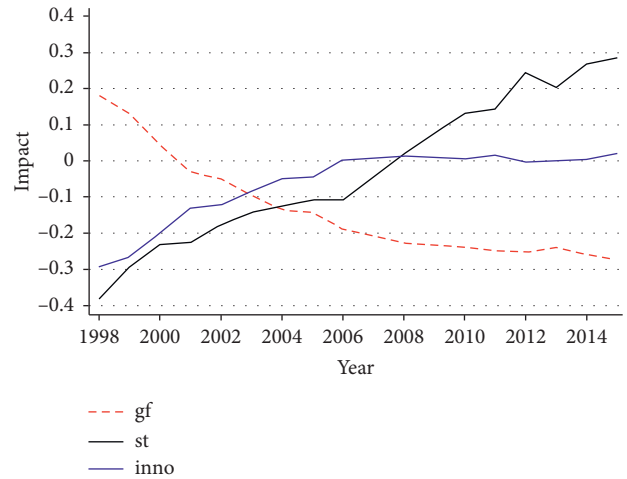
growth model on green productivity in 30 Chinese provinces and found that the investment-driven growth model, which usually leads to the overexploitation of resources and increases environmental pollution, produces a negative effect on green productivity. This is mainly because China's investment efficiency has declined since the twenty-first century.

On the other hand, the contributions of the new driving forces on the GTFP perform well. The contributions of innovation capacity and structural transformation are increasing, perhaps because they could optimize and upgrade industrial and energy structure and improve resource allocation efficiency, thus decreasing environmental pollution, improving the GTFP, and promoting economic growth.

5. Conclusions

The main objective of this study is to investigate the economic growth driving force conversion in China and the time-varying characteristics of driving forces. To achieve these objectives, we attempt to study economic growth from the perspective of GTFP, in which the previous research about economic growth mainly focuses on per capita income growth or national income [42, 43], using a GML index based on the SBM-DDF to measure the GTFP of 30 Chinese provinces during the period 1998–2015. Then, we construct a panel regression model to discover the economic growth driving force conversion, in which existing studies only explored the influencing factors of economic growth [36, 44]. Finally, we adopt a time-varying coefficient panel data model to further investigate the time-varying

characteristics of the conventional and new driving forces that contribute to the GTFP.

Based on the empirical results, several important conclusions are drawn as follows. First, China's GTFP is rising and spatially heterogeneous. On one hand, from a national perspective, the GTFP of China shows an overall upward trend. On the other hand, from the perspective of regional distribution, the GTFP in the east regions show a rising trend. However, the GTFP in the middle and west areas imply a trend of falling and then rising. These heterogeneous impacts are in line with the current imbalance in China's regional economic development.

Additionally, China's economy has undergone driving force conversion. Specifically, China's economic growth driving forces have been transformed from traditional ones (human capital and gross fixed capital formation) to new ones (innovation capacity and structural transformation). This is in accordance with the current economic growth driving force conversion in China's economic growth. Besides that, the impacts of the different driving forces on the GTFP are spatially heterogeneous. First, in the full sample, the conventional driving force affecting the GTFP is gross fixed capital formation, while the new driving forces are innovation capacity and structural transformation. Second, in the east area, gross fixed capital formation and innovation capacity have a significant negative impact on the GTFP, while industrial transformation exerts a positive and significant effect on the GTFP. Third, in the middle area, neither the conventional nor the new driving forces exert a significant impact on the GTFP, while in the west area, both the conventional and the new driving forces exert a significant impact on the GTFP. These heterogeneous impacts may be closely connected with the following phenomenon: different provinces have achieved different economic development levels and various economic development modes.

Furthermore, the time-varying characteristics of conventional and new driving forces imply that China is in a period of transformation from traditional driving forces to new driving forces. The contributions to the GTFP among the different driving forces are inconclusive, which indicates that the time-varying impacts of the conventional and new driving forces that contribute to the GTFP are various. On one hand, the contribution of the conventional driving forces is not promising. This is mainly because China's investment efficiency has declined since the twenty-first century. On the other hand, the impacts of the new driving forces on the GTFP perform well. The contributions of innovation capacity and structural transformation are increasing, perhaps because they could optimize and upgrade industrial and energy structure and improve resource allocation efficiency, thus decreasing environmental pollution, improving the GTFP, and promoting economic growth.

Accordingly, the following policy implications can be pursued to improve the GTFP of these 30 Chinese provinces. First, to foster economic growth, the Chinese government ought to enact various policies conducive to enhancing innovation capacity and accelerating structural transformation. Second, these results show us that the GTFP of the less developed provinces in China is not optimistic, and more action should be taken in practice by both industry and government to handle the undesirable outputs of industrial enterprise, thus improving economic growth and reducing the environmental pollution.

This paper has several limitations and can be expanded by further research in the following regards. First, although the GTFP can be measured by the GML index based on the SBM-DDF, the inputs and outputs cannot be completely represented, owing to the lack of data. Better data availability permitting, future work could improve upon the input and output indicators presented here and expand the coverage to other, equally important, innovation input, the use of natural resources such as land, sea, forests, or mineral resources, and environmental pollution, such as the emission of carbon dioxide, sulfide, and solid waste. Another limitation is that the empirical research in this paper is based on the analysis of macro data to analyse the driving forces of economic growth. Further research could expand on the micro perspectives.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

References

[1] L. Kogan, D. Papanikolaou, A. Seru, and N. Stoffman, "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, vol. 132, no. 2, pp. 665–712, 2017.

[2] J. Zeira, "Innovations, patent races and endogenous growth," *Journal of Economic Growth*, vol. 16, no. 2, pp. 135–156, 2011.

[3] O. Galor and D. N. Weil, "Population, technology, and growth: from malthusian stagnation to the demographic transition and beyond," *American Economic Review*, vol. 90, no. 4, pp. 806–828, 2000.

[4] A. Emrouznejad and G.-L. Yang, "A framework for measuring global malmquist-luenberger productivity index with CO_2 emissions on Chinese manufacturing industries," *Energy*, vol. 115, pp. 840–856, 2016.

[5] K. Li and B. Lin, "Economic growth model, structural transformation, and green productivity in China," *Applied Energy*, vol. 187, pp. 489–500, 2017.

[6] M. J. Herrerias, R. Joyeux, and E. Girardin, "Short-and long-run causality between energy consumption and economic growth: evidence across regions in China," *Applied Energy*, vol. 112, pp. 1483–1492, 2013.

[7] Y. J. Feng, S. Y. Zhong, Q. Y. Li, X. M. Zhao, and X. Dong, "Ecological well-being performance growth in China (1994–2014): from perspectives of industrial structure green

adjustment and green total factor productivity," *Journal of Cleaner Production*, vol. 236, 2019.

[8] S. Chen and J. Golley, ""Green" productivity growth in China's industrial economy," *Energy Economics*, vol. 44, pp. 89–98, 2014.

[9] B. Lin and Z. Chen, "Does factor market distortion inhibit the green total factor productivity in China?" *Journal of Cleaner Production*, vol. 197, pp. 25–33, 2018.

[10] Z. Liu and L. Xin, "Has China's belt and road initiative promoted its green total factor productivity?—evidence from primary provinces along the route," *Energy Policy*, vol. 129, pp. 360–369, 2019.

[11] R. M. Solow, "Technical change and the aggregate production function," *The Review of Economics and Statistics*, vol. 39, no. 3, pp. 312–320, 1957.

[12] B.-H. Gong and R. C. Sickles, "Finite sample evidence on the performance of stochastic frontiers and data envelement analysis using panel data," *Journal of Econometrics*, vol. 51, no. 1-2, pp. 259–284, 1992.

[13] D. W. Caves, L. R. Christensen, and W. E. Diewert, "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, vol. 50, no. 6, 1982.

[14] S. Managi and P. R. Jena, "Environmental productivity and kuznets curve in India," *Ecological Economics*, vol. 65, no. 2, pp. 432–440, 2008.

[15] Y. H. Chung, R. Färe, and S. Grosskopf, "Productivity and undesirable outputs: a directional distance function approach," *Journal of Environmental Management*, vol. 51, no. 3, pp. 229–240, 1997.

[16] J. T. Pastor and C. A. K. Lovell, "A global malmquist productivity index," *Economics Letters*, vol. 88, no. 2, pp. 266–271, 2005.

[17] D.-H. Oh, "A global malmquist-luenberger productivity index," *Journal of Productivity Analysis*, vol. 34, no. 3, pp. 183–197, 2010.

[18] H. Fukuyama and W. L. Weber, "A directional slacks-based measure of technical inefficiency," *Socio-Economic Planning Sciences*, vol. 43, no. 4, pp. 274–287, 2009.

[19] X. C. Xu, X. Q. Huang, J. Huang, X. Gao, and L. H. Chen, "Spatial-temporal characteristics of agriculture green total factor productivity in China, 1998–2016: based on more sophisticated calculations of carbon emissions," *International Journal of Environmental Research and Public Health*, vol. 16, no. 20, 2019.

[20] M. Song, J. Du, and K. H. Tan, "Impact of fiscal decentralization on green total factor productivity," *International Journal of Production Economics*, vol. 205, pp. 359–367, 2018.

[21] C. F. Chen, Q. X. Lan, M. Gao, and Y. W. Sun, "Green total factor productivity growth and its determinants in China's industrial economy," *Sustainability*, vol. 10, no. 4, 2018.

[22] S. Shao, R. Luan, Z. Yang, and C. Li, "Does directed technological change get greener: empirical evidence from Shanghai's industrial green development transformation," *Ecological Indicators*, vol. 69, pp. 758–770, 2016.

[23] Z. Huang, G. Liao, and Z. Li, "Loaning scale and government subsidy for promoting green innovation," *Technological Forecasting and Social Change*, vol. 144, pp. 148–156, 2019.

[24] N. Tripathy, "Does measure of financial development matter for economic growth in India?" *Quantitative Finance and Economics*, vol. 3, no. 3, pp. 508–525, 2019.

[25] P. Adjei Kwakwa, H. Alhassan, H. Alhassan, and S. Aboagye, "Environmental kuznets curve hypothesis in a financial development and natural resource extraction context: evidence

[26] B. Kahouli, "The short and long run causality relationship among economic growth, energy consumption and financial development: evidence from south mediterranean countries (SMCs)," *Energy Economics*, vol. 68, pp. 19–30, 2017.

[27] F. Wen, F. Min, Y. J. Zhang, and C. Yang, "Crude oil price shocks, monetary policy, and China's economy," *International Journal of Finance & Economics*, vol. 24, no. 2, pp. 812–827, 2019.

[28] M. Qamruzzaman and J. Wei, "Do financial inclusion, stock market development attract foreign capital flows in developing economy: a panel data investigation," *Quantitative Finance and Economics*, vol. 3, no. 1, pp. 88–108, 2019.

[29] Z. Li, Z. Huang, and H. Dong, "The influential factors on outward foreign direct investment: evidence from the "the belt and road"," *Emerging Markets Finance and Trade*, vol. 55, no. 14, pp. 3211–3226, 2019.

[30] T. Hastie and R. Tibshirani, "Varying-coefficient models," *Journal of the Royal Statistical Society: Series B*, vol. 55, no. 4, pp. 757–779, 1993.

[31] B. U. Park, E. Mammen, Y. K. Lee, and E. R. Lee, "Varying coefficient regression models: a review and new developments," *International Statistical Review*, vol. 83, no. 1, pp. 36–64, 2015.

[32] S. Wu, B. Li, Q. Nie, and C. Chen, "Government expenditure, corruption and total factor productivity," *Journal of Cleaner Production*, vol. 168, pp. 279–289, 2017.

[33] J. F. Hu, Z. Wang, Y. H. Lian, and Q. H. Huang, "Environmental regulation, foreign direct investment and green technological progress-evidence from Chinese manufacturing industries," *International Journal of Environmental Research and Public Health*, vol. 15, no. 2, 2018.

[34] J. Zhang, "Estimation of China's provincial capital stock (1952–2004) with applications," *Journal of Chinese Economic and Business Studies*, vol. 6, no. 2, pp. 177–196, 2008.

[35] Z. Li, H. Dong, Z. Huang, and P. Failler, "Impact of foreign direct investment on environmental performance," *Sustainability*, vol. 11, no. 13, 2019.

[36] G. Jones and W. J. Schneider, "Intelligence, human capital, and economic growth: a bayesian averaging of classical estimates (bace) approach," *Journal of Economic Growth*, vol. 11, no. 1, pp. 71–93, 2006.

[37] D. Cohen and M. Soto, "Growth and human capital: good data, good results," *Journal of Economic Growth*, vol. 12, no. 1, pp. 51–76, 2007.

[38] F. Cinnirella and J. Streb, "The role of human capital and innovation in economic development: evidence from post-malthusian prussia," *Journal of Economic Growth*, vol. 22, no. 2, pp. 193–227, 2017.

[39] X. Pan, B. Ai, C. Li, X. Pan, and Y. Yan, "Dynamic relationship among environmental regulation, technological innovation and energy efficiency based on large scale provincial panel data in China," *Technological Forecasting and Social Change*, vol. 144, pp. 428–435, 2019.

[40] A. Levin, C.-F. Lin, and C.-S. James Chu, "Unit root tests in panel data: asymptotic and finite-sample properties," *Journal of Econometrics*, vol. 108, no. 1, pp. 1–24, 2002.

[41] G. S. Maddala and S. Wu, "A comparative study of unit root tests with panel data and a new simple test," *Oxford Bulletin of Economics and Statistics*, vol. 61, no. 1, pp. 631–652, 1999.

[42] P. Kalaitzidakis, T. P. Mamuneas, A. Savvides, and T. Stengos, "Measures of human capital and nonlinearities in economic

growth," *Journal of Economic Growth*, vol. 6, no. 3, pp. 229–254, 2001.

[43] G. C. Chow, "Capital formation and economic growth in China," *The Quarterly Journal of Economics*, vol. 108, no. 3, pp. 809–842, 1993.

[44] A. Sorensen, "R&D, learning, and phases of economic growth," *Journal of Economic Growth*, vol. 4, no. 4, pp. 429–445, 1999.

Research Article

Large-Scale and Refined Green Space Identification-Based Sustainable Urban Renewal Mode Assessment

Rong Guo [ID],[1] **Xiaoya Song** [ID],[1] **Peiran Li,**[2] **Guangming Wu,**[2] **and Zhiling Guo**[2]

[1]*Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology,*
  Ministry of Industry and Information Technology, School of Architecture, Harbin Institute of Technology, Harbin 150006, China
[2]*Center for Spatial Information Science, The University of Tokyo, Kashiwa 277-8568, Japan*

Correspondence should be addressed to Rong Guo; hitarchit.gr@gmail.com and Xiaoya Song; xy.song8989@gmail.com

Urban sustainable renewal has received extensive attention in a wide range of fields, including urban planning, urban management, energy management, and transportation. Given that environmental resource conservation is critical to urban sustainability renewal, this study highlighted the imbalance among green space, urban development, and transportation accessibility. Here, a novel node-place-green model is presented to measure sustainable urban development; meanwhile, deep learning is utilized to identify and extract the green space to measure the environmental index. Based on the generated node, place, and green value, urban developing status could be classified into nine modes for further analysis of transportation, urban function, and ecological construction. The experimental results of Harbin reveal the feasibility of the proposed method in providing specific guidelines for urban planning and policies on sustainable development.

1. Introduction

A multitude of efforts towards low-carbon urban development foster concentration strategies around improving built environment, which is closely related to vehicular emissions [1]. Recently, a multitude of efforts towards sustainable performance assessment are focusing on strategies around improving urban renewal, which is closely related to transportation emissions [2, 3]. For instance, a study demonstrated that the sustainability research board reported that the conservation of resources and the environment gains the highest weight of 29% in the sustainable urban renewal model [4]. Also, a more recent study showed that improving neighborhood design leads to a 75% reduction in energy use and GHG emissions [5]. Besides, some extensive literatures explored the strategies of reducing GHG emissions by improving the green space during urban renewal [6]. Hence, combining urban renewal with urban green space is critical to sustainable development [7]. During urban renewal, the imbalance in urban resource allocation is inevitable and is mainly reflected when developing transportation infrastructure and the corresponding urban functional facilities [8].

The cities in the developing countries renovate things that have long been supported by land finance and marked by space excessive consumption; the following consequential environmental and social issues as follows are in need of being urgently addressed: low concentration degree of functions [9], lack of vitality [10], and the public space being unsuitable for the demand of development [11, 12]. Additionally, an array of functional contradictions like the fading nature ecological system are primarily attributed to the urban huge inner stock space as well as pretty low efficiency of land use [13]. At present, mainly featuring the partial sections used for the planned regions of renovation projects, the ways through which these old cities renovate and implement schemes remain destitute of an all-round research on the urban systematical level [14]. During the renovation, typically, developers only pass the benchmark of various technique standards and compress the public service facilities and opened spaces [15, 16]; they also raise the volume of development at their will and maximize their profits [17, 18]. As a result, this leads to an array of the Gordian knots, including excessive intensity of old city development,

destitution of public service facilities, uncoordinated environmental landscape, and unbalanced urban functions [19]. Old city renovation, namely, a crucial method to revitalize the urban stock land, holds the key to raising land intensity degree and boosting the metabolism of old urban functions [20]. Therefore, it is impressive to study and explore the problem of how to strike a balance between transportation and land use in the stock development to revitalize the city and coordinate with the urban sustainable redevelopment goals [21].

Regarding urban sustainable redevelopment with balance, the node-place model has always been one of the hottest topics. The node-place model establishes a methodological framework, the maximum value for studying the coordinated development between land use and transportation in the adjacent regions [22]. This model makes a hypothesis that there exists a sustainable and positive feedback mechanism between the "node value" relying on the transportation services and the "place value" relying on the land use. The balance between the "node value" and the "place value" of a certain area can be considered as an important reference for evaluating whether this area possesses the potentials of urban construction. Five development states are distinguished by the node-place model (Figure 1). Each situation reflects its development stage and state in terms of the balance between urban function and traffic. The "balanced" state means the coordinated development of transportation and urban functions. The "dependent" state at the bottom of the curve shows that both the transportation construction and urban functions are relatively weak. Therefore, external factors need to intervene to maintain and promote the development of this area. As opposed to the "dependent" state, both transportation and urban functions have been developed to the utmost intensity on the "stress" state, which means further redevelopment may cause larger conflicts. The remaining two are the "unbalanced" state. On this state, the development only focuses on one aspect (transportation or urban function). A large number of articles attempted to utilize the node-place model from the perspective of station classification [23], transportation network assessment [24], sustainable redevelopment [25], accessibility analysis [26], and planning policy [27]. The node-place model is used to determine the form and extent of redevelopment of the Tokyo station regions [28]. Vale compared related literatures and combined land use, transportation, and walkability conditions to evaluate and classify station regions of Lisbon [29]. Considering the complexity of urban problems, the node-place model has gradually begun to be extended to evaluate complex situations [30]. The design index was included in the evaluation system and applied to location-related city regulations and policies [31]. In this paper, the refined evaluation indexes of urban green space are incorporated into the node-place models to widen the decision-making dimension of sustainable urban renewal mode. Combination of transportation, land use, and ecological construction could identify the development stage of sustainable urban renewal more scientifically.

As for the measurement of urban green space, instead of the time-consuming survey and fieldwork, with the rapid development of remote sensing [32] as well as computer vision techniques [33], the automatic
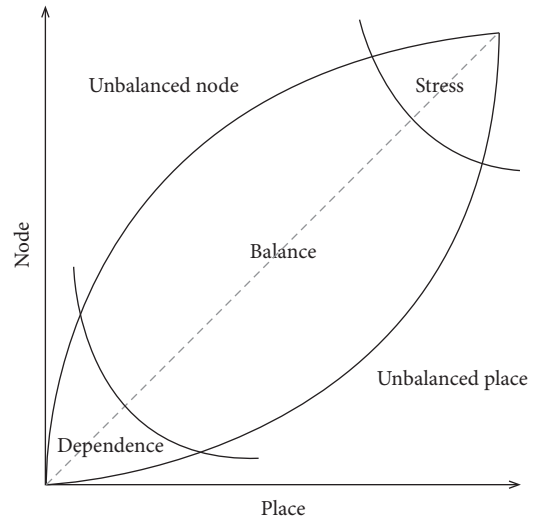
Figure 1: The node-place model by Bertolini [22].

extraction of the green space from remote sensing imagery based on machine learning methods [34] has emerged as a promising alternative strategy. In terms of remote sensing imagery for vegetation, although the normalized difference vegetation index (NDVI) [35] has been widely utilized in several empirical studies [36–38], the limitation of resolution and availability inevitably hinders its deployment in urban regions. Thanks to the dramatically increasing availability of new large-scale remote sensing data sources [39, 40], the ever-expanding high-resolution dataset with ground truth can be utilized for training the urban vegetation extraction model. On the other hand, the deep learning methods, here referred to as deep convolution neural networks [41], have shown the tremendous capability in pattern recognition tasks. Recent deep learning based studies for land feature extraction including vegetation segmentation [42, 43], which outperformed conventional machine learning based methods such as support vector machine (SVM) [44] and neural network (NN) [32], reveal its potential in different urban vegetation extraction tasks.

In this research, we chose the main urban districts of Harbin as our study area. First, we use U-Net [45] to identify and extract the green space, mainly including low vegetation and trees in Harbin; the ratio of vegetation to the region reveals and measures the environmental index of the corresponding region. Second, we expanded the node-place model, using those green indicators as the third dimension to jointly measure the state of urban construction. Finally, according to the evaluation results of the node-place-green model, we classified the sustainable renewal models in Harbin and proposed a targeted renewal strategy for each type of characteristic.

The main contributions of this study are fourfold:

We expanded the node-place model's dimensions combining transportation, urban function, and ecological construction to analyze the sustainable urban development mode

We utilized the U-Net model to identify urban vegetation and trees on a large scale in satellite images, which ensures the refinement of urban feature recognition

We analyzed the sustainable urban renewal model using the identified green space, which dynamically monitors the sustainable renewal status

We identified the state of the urban construction in Harbin and proposed a future sustainable development model for transportation, urban function, and ecological construction

The remainder of this paper is organized as follows. In Section 2, the workflow of the proposed method is illustrated, where the details of metrics as well as the algorithms are also presented. Then, the study area and data source are introduced. After that, experimental results and discussion are demonstrated in Section 3. Finally, the conclusions are drawn in Section 4.

2. Materials and Methods

2.1. Study Area and Data.
In this research, we choose the main urban districts of Harbin as our study area, endeavouring to identify which regions have the potential for further sustainable renewal. As the northernmost subprovincial city in China, Harbin is located in the Northeast Plain. In Harbin, the urban construction land is $458 \, \mathrm{km}^2$, including the old town of $298 \, \mathrm{km}^2$. With the development of the city, various contradictions in the old town have become more prominent. The high urban density, traffic congestion, environmental pollution, and insufficient green space have led to a decline in the overall functions of main urban districts and increased difficulty in urban renewal. Harbin is also a typical winter city with a long and cold winter. Therefore, there is urgency to improve the green space and enhance the quality of the outdoor environment. In order to analyze the place value in each region, in this research, we collect land use status drawn by GIS in basic regions at the town-and-street level in the overall land use plan of the main urban area of Harbin and PoI data with 27 features. As for urban green classification, the ISPRS benchmark [39] and Google Earth (GE) RGB imagery are selected to perform the training and testing data, respectively. Considering the resolution and color space difference between training and testing datasets, the preprocessing, including downscaling and color normalization, is applied.

2.2. U-Net for Urban Vegetation Area Extraction.
As one of the state-of-the-art models for deep learning based image semantic segmentation, U-Net has been successfully applied to perform different tasks with high accuracy and efficiency. Considering the characteristics of U-Net, which could perform pixel-to-pixel and end-to-end mapping from input to output while enabling precise localization for the vegetation segmentation results, in this study, we propose to adopt U-Net as a representative segmentation model for urban vegetation area extraction.

Concretely, the convolution operation performs element-wise multiplication via kernels. In U-Net, it can be formulated as follows:

$$z = \sum_{i=1}^{h_f} \sum_{j=1}^{w_f} \sum_{d=1}^{c_l} \Theta_{i,j,s,d'} \times x_{i,j,d} + b_{d'}, \tag{1}$$

where h_f and w_f represent the height and width of the kernel Θ, c_l is the number of channels for input x in layer l, and b in shape $1 \times 1 \times 1 \times d'$ denotes the bias.

Subsequently, Leaky ReLU [46] and batch-normalization [47] are used to generate the hypothesis from z and reduce internal covariate shift, respectively. The final classification of either vegetation or nonvegetation is predicted by using softmax function [48].

2.3. Node-Place-Green Model.
The node-place model is applied to balance the transportations and urban functions of the railway network and the railway station area [22]. Considering the similarities between the transportation system and urban construction, the node-place model is gradually being applied to the evaluation of large-scale urban development. The subway and bus system in Harbin exhibit the same transportation functions as the railway network and are densely distributed in urban regions. Therefore, the establishment principles and theories of node-location model are also applicable to the research objects of this article.

Urban green space refers to the vegetation system that exists in the urban area in natural or artificial form, including trees and low shrubs. Urban green space is prone to regulate the urban ecosystem, protect the environment, and provide residents with a good place to relax. Therefore, the advantages of the green space system in the construction of sustainable urban development are constantly highlighted. This paper broadens the dimensions of the node-place model and incorporates the evaluation indicators of the green space system into the evaluation system (Table 1).

The place value, node value, and green value, are respectively, Z-transformed, and the processed result (x_m', y_m', and g_m') represents the place value, node value, and green value of each basic region. Then, the three are used as three-dimensional coordinates to make a scatter plot, y, which can be denoted by equation (2). In the end, according to the division of balance, imbalance, dependence, and pressure, the location projected by the basic unit is judged as nine sustainable urban renewal modes. Each of these nine modes can be represented in a cube, as illustrated in Figure 2:

$$A_m = \Phi\left(x_m', \, y_m', \, g_m'\right), \tag{2}$$

where A_m denotes the coordinated development status of transportation, urban functions, and green space of region m, and x_m', y_m', and g_m', respectively, denote the place value, node value, and green value of region m.

TABLE 1: Indicators in node-place-green model.

Description	Calculation of the indices
Node index	
Number of directions served by subway	y_1 = number of subway services offered at station
Daily frequency of services by subway	y_2 = number of subways departing from station on working day
Number of directions served by other public transport	y_3 = number of public trfansport services offered at station
Daily frequency of services by other public transport	y_4 = number of buses, trams, and underground trains departing from station on working day
Proximity to CBD	y_5 = travel time to the CBD
Place index	
Restaurant, entertainment, and hotel within 15 minutes' walking distance	x_1 = number of POIs within 15 minutes' walking distance in restaurant, entertainment, and hotel
Industry and distribution within 15 minutes' walking distance	x_2 = number of POIs within 15 minutes' walking distance in transport distribution centers and warehouses, private warehouses, gas stations, factories, and crafts
Commerce, finance, and real estate within 15 minutes' walking distance	x_3 = number of POIs within 15 minutes' walking distance in commercial corporations, banks, real estate agency, communication, newspaper, and broadcasting
Education, health, and culture within 15 minutes' walking distance	x_4 = number of POIs within 15 minutes' walking distance in hospitals, social insurance, social welfare, all kinds of schools, research institute, skill classroom, and government office
Administration and services within 15 minutes' walking distance	x_5 = number of POIs within 15 minutes' walking distance in travel agencies, amusement parks, gardens, public bathhouses, and temples
Land use mixture	$x_6 = 1 - ((a-b)/d) - (a-c)/d)/2$ $a = \max(x1, x2, x3, x4, x5)$ $b = \min(x1, x2, x3, x4, x5)$ $c = \text{mean}(x1, x2, x3, x4, x5)$ $d = \sum x1 + x2 + x3 + x4 + x5$
Green index	
Low vegetation cover	g_1 = the ratio of low vegetation area to building area
Tree cover	g_2 = the ratio of tree area to building area

3. Results and Discussion

3.1. Place Value. First, the node-place-green model is utilized to calculate the place value of Harbin. The place value reflects the present status of public service facilities in urban functions. The larger the place value, the higher the completeness of urban functions in the region, which means living more conveniently. Deducing from the box plots (Figure 3), it can be stated that the place value of Harbin is mainly distributed between 0.05 and 0.38, and the median and mean value are lower than the node value and green value. In addition, it is demonstrated that there is a large gap in the place value in various regions of Harbin, and the overall development in urban function lags behind that of transportation and green space. The spatial distribution of Harbin place value is illustrated in Figure 4. On the whole, the Harbin place value presents a single-center structure. The place value gradually decreases with the old city as the center. Moreover, although individual residential regions in the outer urban regions have higher place value, the outer urban regions are generally lower than the old urban regions. In terms of central regions, regions with high place value are mainly concentrated in Nangang District and Daoli District, and they are distributed in multiple centers. The multicenter structure is consistent with the distribution of Harbin's main business districts. According to the distribution of the "three old" regions in Harbin, it can be stated that the urban functions in old southern Harbin are still relatively weak.

From the perspective of the spatial distribution of the place value, the urban renewal in the north, south, and west of the main urban area has potential.

3.2. Node Value. Second, the node value of Harbin is calculated. Node value reflects public transportation accessibility. The larger the node value, the more developed the public transportation, that is, the higher the regional accessibility. As shown in Figure 3, the Harbin node value is mainly distributed between 0.35 and 0.45, and the median and mean value are the highest of the three. More specifically, the development of public transportation in various regions of Harbin is relatively balanced, and the overall development level is higher than that of urban functions and green space. The spatial distribution of node value in Harbin is demonstrated in Figure 5. The node value is generally more scattered than the place value, showing one linear and multicenter structure. The node values in the southern part of the Songhua River are all higher than those in the northern part. In addition, the high value is more concentratedly distributed at the junction of the north and south urban regions. Deducing from the result, it can be stated that the urban transportation infrastructure construction focuses on the initial stage of strengthening the coordinated development of the northern and southern cities. In terms of the southern urban area, vertical transportation development is relatively mature and has excellent accessibility.
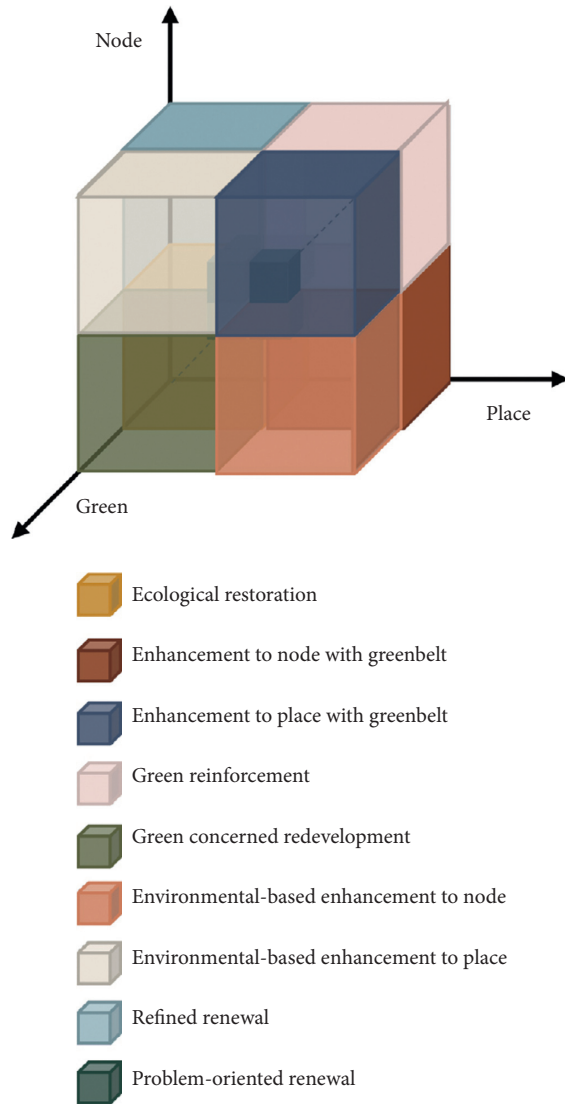
FIGURE 2: The node-place-green model and the nine sustainable urban renewal modes.
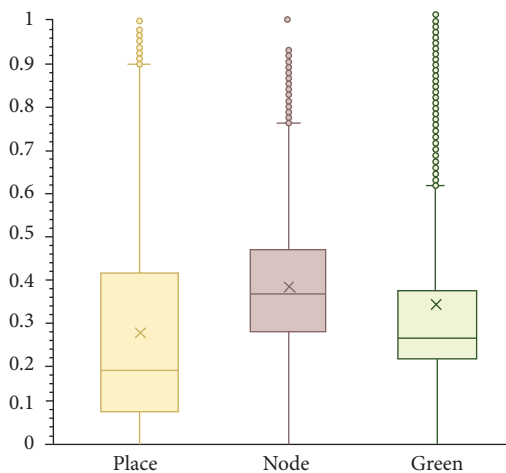
FIGURE 3: Box plots of the normalization of place, node, and green in Harbin.

From the perspective of node value, the northern parts of Harbin's main urban area have more significant potential for renewal.

3.3. Green Value. We use deep learning to identify the urban green space (low vegetation and trees) in Harbin. The ratio of vegetation to the region reveals and measures the environmental index. The larger the green value, the higher the green vegetation coverage in the region. As illustrated in Figure 3, most of the green value is distributed between 0.18 and 0.35, and the distribution is relatively concentrated. It shows that the gap in various regions of Harbin in green value is small, but the overall level is low. Harbin's green value is shown schematically in Figure 6. Opposed to the place value and node value, the green value distribution has no visible structural characteristics. Some regions on the edge of the city have high green value, and most of these regions are undeveloped farmland. The overall green value of the urban area is low, and there is no large-scale concentration. The green value in the old urban area with high place value and traffic value is generally low, mostly below 0.2, and this area has a high demand for greening.

3.4. Determination of Sustainable Urban Renewal Mode. As analyzed above, place value, node value, and green value, respectively, reveal the current status of urban construction in terms of urban function, transportation, and environment. These three comprehensively reflect the current situation of the city. For a comprehensive discussion on strategies for sustainable urban renewal, we performed a cluster analysis to identify the development status of the regions in Harbin. Based on the transformed place value, node value, and green value, nine unique clusters are interpreted in Table 2, and the corresponding structure is illustrated in Figure 7. The detailed analysis is as follows.

3.4.1. Mode 1: Ecological Restoration. Mode 1 comprises low place value, node value, and green value. That is, the region is relatively weak in transportation, urban functions, and the environment. 58% of regions in Harbin were identified in such a group.

This mode is named "ecological restoration" because these regions have a poor environment under low-intensity urban development. Except for the central area, such regions are widely distributed in Harbin's main urban area, which reveals the low overall development intensity and the lack of green space system. Due to geographical barriers, in the process of urban renewal, the development of the northern Songhua River has been lagging behind that of the southern part in the upgrading of transportation infrastructure or the improvement of urban public service facilities. Moreover, in the southern Songhua River, the development intensity of Qunli New District and residential quarters' density has gradually increased in recent years. Nevertheless, the result reveals the same problems exposed in the above two types of development states. Under the direction of sustainable renewal, regardless of the previous development status, the

0.00–0.20
0.20–0.40
0.40–0.60
0.60–0.80
0.80–1.00

Figure 4: Place value of the regions in Harbin.

region in this mode should focus on improving the green space system.

3.4.2. Mode 2: Enhancement to Node with Greenbelt.

9.1% of Harbin regions are identified with high place value and a low node value and green value. More specifically, the regions are relatively well developed in urban functions and weak in transportation and the environment.

Mode 2 is named "enhancement to node with greenbelt" because these regions have high-intensity urban functions but lack traffic accessibility and green space. These regions are concentrated in the old city of central Harbin. There are large-scale and high-density old buildings in these regions, with excellent urban historical and cultural characteristics. However, in the process of urban renewal, due to the obsolete road system and high building density, the development is still concentrated on the commercial development of existing buildings. This result reveals that the current development intensity of the area in terms of urban functions is tending towards perfection.

Consequently, it should focus on the redevelopment of public transportation and the recombination of green space and urban public space.

3.4.3. Mode 3: Enhancement to Place with Greenbelt.

This mode comprises high node value and low place value and green value. Besides, the region is relatively well developed in urban transportation and weak in function and environment. The regions under this mode are identified to account for 9.2% of Harbin.

This mode is called "enhancement to place with greenbelt" because these regions have developed public transportation systems, but urban public service facilities and green space systems still need to be improved. Such regions are distributed in the central area and the south of Harbin. With the construction of the subway connecting the north and south of Harbin, the renewal of the area around the subway is gradually accelerating. As illustrated in Figure 7, with the development of two subway lines in the
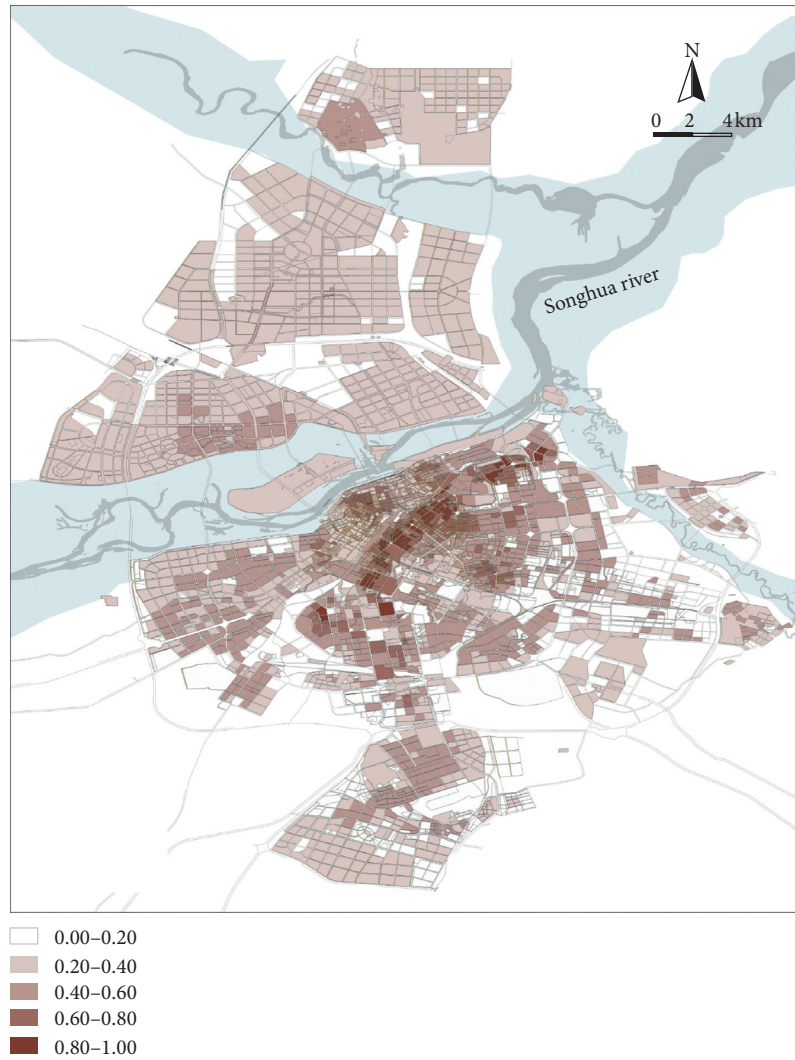
Figure 5: Node value of the regions in Harbin.

central area, the accessibility of these regions has been well improved. Additionally, urban functions cannot meet the current traffic conditions. Moreover, in the construction of transportation infrastructure, it is rarely combined with green space. It reveals that, in the future sustainable urban renewal, the redevelopment of urban commercial and public service facilities should be appropriately increased, and urban traffic nodes should be combined with relevant green space for systematic construction.

3.4.4. Mode 4: Green Reinforcement.

6.5% of regions in Harbin were identified with high node value and place value and low green value. The regions under mode 4 are relatively well developed in urban functions and transportation and are relatively weak in the environment.

This mode is named "green reinforcement" because these regions have complete urban public service facilities and a developed public transportation system with fewer low vegetation and trees and weak green space systems. Such regions are distributed in parts of the old city and

commercial districts in the city center. The development of transportation and urban functions is relatively balanced, and the intensity is relatively high. However, the area and amount of green space are small. Except for parks, these regions have a small green area and low vegetation and tree coverage. This result reveals that the future renewal of the area should slow down the development of traffic and urban functions, focusing on the maintenance of existing green space and the expansion of public space.

3.4.5. Mode 5: Green Concerned Redevelopment.

This mode incorporates low node value and space value and high green value. These regions are still in the infancy in urban functions and transportation, but the vegetation coverage rate is high. The regions under this mode are identified to account for 15.6% of Harbin.

This mode is named "green concerned redevelopment" because, compared with the development of transportation and urban functions, the vegetation coverage in this mode is higher, and the green environment is well. Such regions are

FIGURE 6: Green value of the regions in Harbin.

TABLE 2: Explanation for the sustainable urban renewal modes.

| Sustainable urban renewal mode | | Evaluation features | | |
		Place value	Node value	Green value
1	Ecological restoration	Low	Low	Low
2	Enhancement to node with greenbelt	High	Low	Low
3	Enhancement to place with greenbelt	Low	High	Low
4	Green reinforcement	High	High	Low
5	Green concerned redevelopment	Low	Low	High
6	Environmentally based enhancement to node	High	Low	High
7	Environmentally based enhancement to place	Low	High	High
8	Refined renewal	High	High	High
9	Problem-oriented renewal	Medium	Medium	Medium

scattered in various parts of Harbin, the outer area is large and concentrated, and the central area is small and scattered. Most of these regions distributed in the periphery of the urban area are undeveloped farmland. Additionally, most of these regions are urban green space. This reveals that the future renewal should focus on small-intensity redevelopment based on protecting the existing green environment.

3.4.6. Mode 6: Environmental-Based Enhancement to Node. 0.9% of the regions in Harbin are identified with high place value and green value and low node value. This indicates that the regions are well constructed in urban functions and green space, but the traffic accessibility is weak. This mode is named "environmentally based enhancement to node" because the urban functions and greening environment

Ecological restoration
Enhancement to node with greenbelt
Enhancement to place with greenbelt
Green reinforcement
Green concerned redevelopment
Environmental-based enhancement to node
Environmental-based enhancement to place
Refined renewal
Problem-oriented renewal
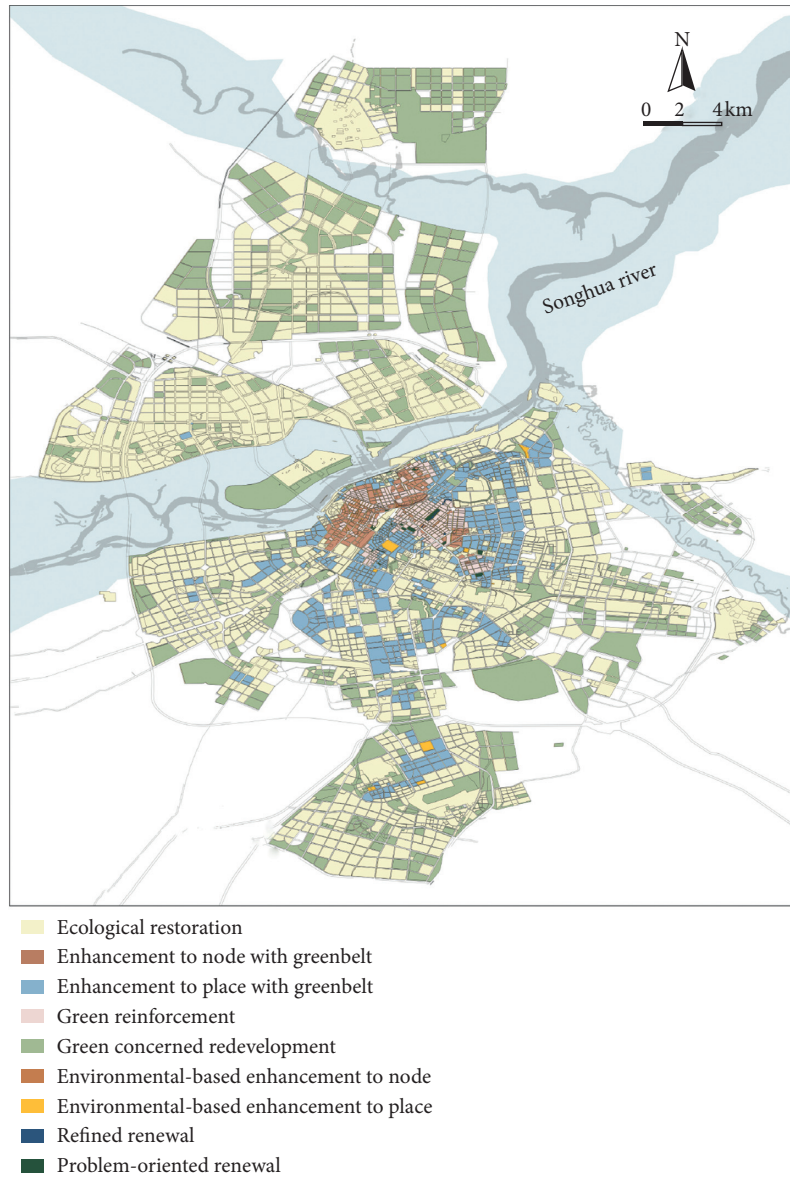
FIGURE 7: Sustainable urban renewal modes of Harbin.

construction are coordinated and perfect. Nevertheless, transportation is relatively inconvenient. Such regions are scattered in the center of Harbin, corresponding to urban green spaces with little traffic. Moreover, they are adjacent to Harbin's commercial district and have a high potential for redevelopment. Therefore, future renewal should focus on increasing public transport connectivity while maintaining the existing environment.

3.4.7. Mode 7: Environmental-Based Enhancement to Place. Mode 7 comprises high node value and green value and low place value. These regions are identified to account for 0.2% of Harbin and are well constructed in urban traffic and greening with weak urban function. This mode is named "environmentally based enhancement to place" because the urban public transportation in these regions is highly

accessible and the green environment is good, but the urban functions are insufficient. Such regions are scattered in the city center, corresponding to several urban parks and parks with single functions. These regions are generally adjacent to "enhancement to place with greenbelt"; so future updates should be coordinated with the redevelopment of "enhancement to place with greenbelt" and endeavor to protect the existing environment.

3.4.8. Mode 8: Refined Renewal. Mode 8 contains high node value, place value, and green value. More specifically, the development of these areas has higher intensity and better perfection in terms of urban functions, transportation, and green space. These regions are identified to account for 0.17% of Harbin. This mode is called "refined renewal" because, in Harbin, these regions have concentrated and

complete development in the above three aspects. Such regions are scattered in central urban regions, and the renewal should focus on refinement and quality.

3.4.9. Mode 9: Problem-Oriented Renewal. 0.69% of regions in Harbin are identified with medium node value, place value, and green value. Additionally, the development of these regions is coordinated in terms of urban functions, transportation, and greening. Also, there is still a certain amount of development space. Given the relative balance in the above three aspects, the renewal of this type of the regions should focus on the current demands within the region, that is, to adjust the focus of development based on the problem to maintain the current equilibrium state.

In brief, based on the place value, node value, and green value, nine sustainable urban renewal modes are identified. These nine renewal modes are feasible to guide the redevelopment of Harbin in the future to a considerable extent. Moreover, the regions in modes 1–4 need to focus on the lack of green space. Modes 1–8 need to focus on the balance of urban functions and transportation in future construction.

4. Conclusions

The imbalance in urban resource allocation is inevitable and is mainly reflected when developing transportation infrastructure and the corresponding urban functional facilities during urban renewal. The refined evaluation indexes of urban green space incorporated into the node-place models are prone to widen the decision-making dimension of sustainable urban renewal mode. Moreover, combination of transportation, land use, and ecological construction could identify the development stage of sustainable urban renewal more scientifically. In this study, we use U-Net to identify the urban green space (low vegetation and trees) in Harbin. The ratio of vegetation to the region reveals and measures the environmental index of the corresponding region, which ensures the refinement of urban feature recognition. Secondly, we expanded the node-place model, using those green indicators as the third dimension to jointly measure the state of urban construction, which dynamically monitors the sustainable renewal status. Since then, we separately analyzed the development status of urban functions, transportation, and green space in Harbin. Finally, we identified the state of the urban construction in Harbin and proposed a future sustainable development model for transportation, urban function, and ecological construction. We find that there is a large gap in the place value in various regions of Harbin, and the overall development in urban function lags behind that of transportation and green space. Furthermore, regions in northern and southwest Harbin have potential for renewal. Based on the place value, node value, and green value, nine sustainable urban renewal modes are identified. Although three dimensions of urban development imbalance are introduced in this work, the proposed method still has limitations. This study does not incorporate population movement data as well as the scenario simulation. In future research, we plan to explore a more comprehensive spatiotemporal analysis on the urban development imbalance. Such a study would provide more specific guidelines for sustainable urban renewal.

Data Availability

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

Conflicts of Interest

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

Authors' Contributions

Rong Guo and Xiaoya Song contributed equally to this work.

Acknowledgments

References

[1] H. Zhang, X. Song, Y. Long et al., "Mobile phone GPS data in urban bicycle-sharing: layout optimization and emissions reduction analysis," *Applied Energy*, vol. 242, pp. 138–147, 2019.

[2] Y. Long, Y. Yoshida, K. Fang, H. Zhang, and M. Dhondt, "City-level household carbon footprint from purchaser point of view by a modified input-output model," *Applied Energy*, vol. 236, pp. 379–387, 2019.

[3] E. Zarghami and D. Fatourehchi, "Comparative analysis of rating systems in developing and developed countries: a systematic review and a future agenda towards a region-based sustainability assessment," *Journal of Cleaner Production*, vol. 254, Article ID 120024, 2020.

[4] S. Yildiz, S. Kivrak, and G. Arslan, "Contribution of built environment design elements to the sustainability of urban renewal projects: model proposal," *Journal of Urban Planning and Development*, vol. 145, no. 1, Article ID 04018045, 2019.

[5] C. Hachem, "Impact of neighborhood design on energy performance and GHG emissions," *Applied Energy*, vol. 177, pp. 422–434, 2016.

[6] Y. Long, Y. Yoshida, Q. Liu, H. Zhang, S. Wang, and K. Fang, "Comparison of city-level carbon footprint evaluation by applying single- and multi-regional input-output tables," *Journal of Environmental Management*, vol. 260, Article ID 110108, 2020.

[7] V. Bansal, S. Bansal, and J. Sen, "Assessing the role of urban design in a rapidly urbanizing historical city and its contribution in restoring its urban ecology: the case of Varanasi, India," *Urban Ecology*, pp. 413–434, 2020.

[8] X. Song, R. Guo, T. Xia et al., "Mining urban sustainable performance: millions of GPS data reveal high-emission travel attraction in Tokyo," *Journal of Cleaner Production*, vol. 242, Article ID 118396, 2020.

[9] Y. Wang, J. Li, G. Zhang, Y. Li, and M. H. Asare, "Fuzzy evaluation of comprehensive benefit in urban renewal based on the perspective of core stakeholders," *Habitat International*, vol. 66, pp. 163–170, 2017.

[10] X. Li, E. C. M. Hui, T. Chen, W. Lang, and Y. Guo, "From Habitat III to the new urbanization agenda in China: seeing through the practices of the "three old renewals" in Guangzhou," *Land Use Policy*, vol. 81, pp. 513–522, 2019.

[11] Y. Long, Y. Yoshida, J. Meng, D. Guan, L. Yao, and H. Zhang, "Unequal age-based household emission and its monthly variation embodied in energy consumption—a cases study of Tokyo, Japan," *Applied Energy*, vol. 247, pp. 350–362, 2019.

[12] G. Liu, L. Wei, J. Gu, T. Zhou, and Y. Liu, "Benefit distribution in urban renewal from the perspectives of efficiency and fairness: a game theoretical model and the government's role in China," *Cities*, vol. 96, Article ID 102422, 2020.

[13] R. C. Estoque and M. Ooba, X. T. Seposo, Heat health risk assessment in Philippine cities using remotely sensed data and social-ecological indicators," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.

[14] L. W. C. Lai and F. T. Lorne, "Sustainable urban renewal and built heritage conservation in a global real estate revolution," *Sustainability*, vol. 11, no. 3, p. 850, 2019.

[15] Z. Qiao, L. Liu, Y. Qin, X. Xu, B. Wang, and Z. Liu, "The impact of urban renewal on land surface temperature changes: a case study in the main city of Guangzhou, China," *Remote Sensing*, vol. 12, no. 5, p. 794, 2020.

[16] Q. Yu, H. Zhang, W. Li et al., "Mobile phone data in urban bicycle-sharing: market-oriented sub-area division and spatial analysis on emission reduction potentials," *Journal of Cleaner Production*, vol. 254, Article ID 119974, 2020.

[17] Y. Peng, J. Qian, F. Ren, W. Zhang, and Q. Du, "Sustainability of land use promoted by construction-to-ecological land conversion: a case study of Shenzhen city, China," *Sustainability*, vol. 8, no. 7, p. 671, 2016.

[18] H. Leng, X. Chen, Y. Ma, N. H. Wong, and T. Ming, "Urban morphology and building heating energy consumption: evidence from Harbin, a severe cold region city," *Energy and Buildings*, vol. 224, Article ID 110143, 2020.

[19] H. Leng, S. Li, S. Yan, and X. An, "Exploring the relationship between green space in a neighbourhood and cardiovascular health in the winter city of China: a study using a health survey for harbin," *International Journal of Environmental Research and Public Health*, vol. 17, no. 2, p. 513, 2020.

[20] T. Zhou, Y. Zhou, and G. Liu, "Key variables for decision-making on urban renewal in China: a case study of Chongqing," *Sustainability*, vol. 9, no. 3, p. 370, 2017.

[21] J. A. Adedeji and O. Arayela, "Urban renewal strategies and economic growth in Ondo State, Nigeria: a case study," *Journal of Contemporary Urban Affairs (JCUA)*, vol. 2, no. 1, pp. 76–83, 2017.

[22] L. Bertolini, "Spatial development patterns and public transport: the application of an analytical model in The Netherlands," *Planning Practice and Research*, vol. 14, no. 2, pp. 199–210, 1999.

[23] Y. Zhang, S. Marshall, and E. Manley, "Network criticality and the node-place-design model: classifying metro station areas in Greater London," *Journal of Transport Geography*, vol. 79, p. 102485, 2019.

[24] R. Falconer, C. Babb, and D. Olaru, "Cities as systems: node and place conflict across a rail transit network,""Cities as systems: node and place conflict across a rail transit network," in *Planning Boomtown and Beyond*, S. Biermann, D. Olaru, and V. Paül, Eds., pp. 446–475, UWA Publishing, Perth, WA, USA, 2016.

[25] P. Pezeshknejad, S. Monajem, and H. Mozafari, "Evaluating sustainability and land use integration of BRT stations via extended node place model, an application on BRT stations of

Tehran," *Journal of Transport Geography*, vol. 82, Article ID 102626, 2020.

[26] F. Caset, D. S. Vale, and C. M. Viana, "Measuring the accessibility of railway stations in the Brussels Regional Express Network: a node-place modeling approach," *Networks and Spatial Economics*, vol. 18, no. 3, pp. 495–530, 2018.

[27] D. Olaru, S. Moncrieff, G. McCarney et al., "Place vs. node transit: planning policies revisited," *Sustainability*, vol. 11, no. 2, p. 477, 2019.

[28] P. Chorus and L. Bertolini, "An application of the node place model to explore the spatial development dynamics of station areas in Tokyo," *Journal of Transport and Land Use*, vol. 4, no. 1, pp. 45–58, 2011.

[29] D. S. Vale, "Transit-oriented development, integration of land use and transport, and pedestrian accessibility: combining node-place model with pedestrian shed ratio to evaluate and classify station areas in Lisbon," *Journal of Transport Geography*, vol. 45, pp. 70–80, 2015.

[30] Z. Li, Z. Han, J. Xin, X. Luo, S. Su, and M. Weng, "Transit oriented development among metro station areas in Shanghai, China: variations, typology, optimization and implications for land use planning," *Land Use Policy*, vol. 82, pp. 269–282, 2019.

[31] D. S. Vale, C. M. Viana, and M. Pereira, "The extended node-place model at the local scale: evaluating the integration of land use and transport for Lisbon's subway network," *Journal of Transport Geography*, vol. 69, pp. 282–293, 2018.

[32] B. Höfle, M. Hollaus, and J. Hagenauer, "Urban vegetation detection using radiometrically calibrated small-footprint full-waveform airborne LiDAR data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 134–147, 2012.

[33] M. Nixon and A. S. Aguado, *Feature Extraction and Image Processing for Computer Vision*, Academic Press, Cambridge, MA, USA, 2019.

[34] E. Alpaydin, *Introduction to Machine Learning*, MIT press, Cambridge, MA, USA, 2020.

[35] C. J. Tucker, J. E. Pinzon, M. E. Brown et al., "An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data," *International Journal of Remote Sensing*, vol. 26, no. 20, pp. 4485–4498, 2005.

[36] G. M. Gandhi, S. Parthiban, N. Thummalu, and A. Christy, "Ndvi: vegetation change detection using remote sensing and Gis—a case study of Vellore district," *Procedia Computer Science*, vol. 57, pp. 1199–1210, 2015.

[37] Q. Wang and J. D. Tenhunen, "Vegetation mapping with multitemporal NDVI in North Eastern China transect (NECT)," *International Journal of Applied Earth Observation and Geoinformation*, vol. 6, no. 1, pp. 17–31, 2004.

[38] R. Geerken, B. Zaitchik, and J. P. Evans, "Classifying rangeland vegetation type and coverage from NDVI time series using Fourier Filtered Cycle Similarity," *International Journal of Remote Sensing*, vol. 26, no. 24, pp. 5535–5554, 2005.

[39] F. Rottensteiner, G. Sohn, J. Jung et al., "The ISPRS benchmark on urban object classification and 3D building reconstruction," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, no. 1, pp. 293–298, 2012.

[40] Q. Chen, L. Wang, Y. Wu, G. Wu, Z. Guo, and S. L. Waslander, "Aerial imagery for roof segmentation: a large-scale dataset towards automatic mapping of buildings," 2018, https://arxiv.org/abs/1807.09532.

[41] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT press, Cambridge, MA, USA, 2016.

[42] Z. Guo, G. Wu, X. Song et al., "Super-resolution integrated building semantic segmentation for multi-source remote sensing imagery," *IEEE Access*, vol. 7, pp. 99381–99397, 2019.

[43] G. Wu, Y. Guo, X. Song et al., "A stacked fully convolutional networks with feature alignment framework for multi-label land-cover segmentation," *Remote Sensing*, vol. 11, no. 9, p. 1051, 2019.

[44] C. Iovan, D. Boldo, M. Cord, and M. Erikson, "Automatic extraction and classification of vegetation areas from high resolution images in urban areas," in *Proceedings of the Scandinavian Conference on Image Analysis*, pp. 858–867, Springer, Aalborg, Denmark, June 2007.

[45] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical image computing and computer-assisted*, pp. 234–241, Munich, Germany, October 2015.

[46] X. Zhang, Y. Zou, and W. Shi, "Dilated convolution neural network with LeakyReLU for environmental sound classification," in *Proceedings of the 2017 22nd International Conference on Digital Signal Processing (DSP)*, IEEE, London, UK, pp. 1–5, August 2017.

[47] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, https://arxiv.org/abs/1502.03167.

[48] H. Peng, J. Li, Y. Song, and Y. Liu, "Incrementally learning the hierarchical softmax function for neural language models," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CF, USA, February 2017.

Research Article

Competitiveness Evaluation of Chinese Dairy Industry Based on Accelerated Genetic Algorithm Projection Pursuit Model

Chen Hongli [ID],[1,2,3] Wu Yanyan,[2,3] Liu Xiuli [ID],[1,4] and Zhang Wenju [ID][2]

[1]*University of Chinese Academy of Sciences, Beijing 100049, China*
[2]*College of Animal Science and Technology, Shihezi University, Shihezi 832000, China*
[3]*China Xinjiang Tianshan Junken Animal Husbandry Co., Ltd., Shihezi 832000, China*
[4]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China*

Correspondence should be addressed to Liu Xiuli; xiuli.liu@amss.ac.cn and Zhang Wenju; zhangwj1022@sina.com

Scientific evaluation of the competitiveness of China's dairy industry is of great significance to the development of regional dairy industry. In view of the fact that the existing evaluation model cannot solve the dynamic nonlinear optimization problem, this paper established a projection pursuit model based on accelerating genetic algorithm and carried on the comprehensive evaluation and dynamic analysis on the competitiveness of dairy industry in China's 31 provinces and cities (not including Hong Kong, Macao, and Taiwan) from five dimensions, namely, basic factors of production, the industry demand conditions, the enterprise competitiveness and growth force, related and supporting industries, and the natural environmental elements. It was found that firstly, from the perspective of time, the development trend of China's dairy industry basically shows a positive trend. With the improvement of national living standards, the competitiveness of the dairy industry is also constantly improved. Secondly, from the perspective of space, regional differences in the competitiveness of the dairy industry are significant. Inner Mongolia, Heilongjiang, Shandong, and Hebei are at the forefront of the development of the industry, while the development of the dairy industry in the rest of the region has a big gap with those areas. From the perspective of regional layout, the advantageous region of China's dairy industry has been preliminarily formed, and the competitive advantage is mainly concentrated in the northern region. Accordingly, the countermeasures and suggestions to enhance the competitiveness of China's dairy industry were put forward.

1. Introduction

Milk and dairy products are very important parts of healthy diet in the world, which are closely related to people's life and health. In recent years, the global milk production and consumption have shown a continuous growth trend. In 2018, the global milk production and consumption were 506 million tons and 177 million tons, respectively, 1.47% and 0.25% higher than those in 2017. Although the supply and demand of milk has slowed down statistically, milk is still the most important source for the global milk production. With the improvement of people's living standard, the supply and demand of Chinese dairy products is increasing rapidly. Over the past 20 years, China's dairy industry has developed rapidly. The number of dairy cows in stock has increased from 4.887 million yuan RMB in year 2000 to 7.2 million yuan in 2018, and the total production of dairy products has increased from 829,000 tons in 2000 to 26.871 million tons in 2018. The total industrial output value of China's dairy industry also keeps increasing. In 2018, the total industrial output value of China's dairy industry reached 393.7 billion yuan RMB, about 20 times of the 19.5 billion yuan RMB in 2000 (as is shown in Figure 1).

China's dairy industry gradually presents a certain scale effect and has formed well-known brands such as Mengniu, Yili, and Guangming. However, the development of dairy industry in different regions of China is not balanced. From the distribution of milk sources, as shown in Figure 2, the top
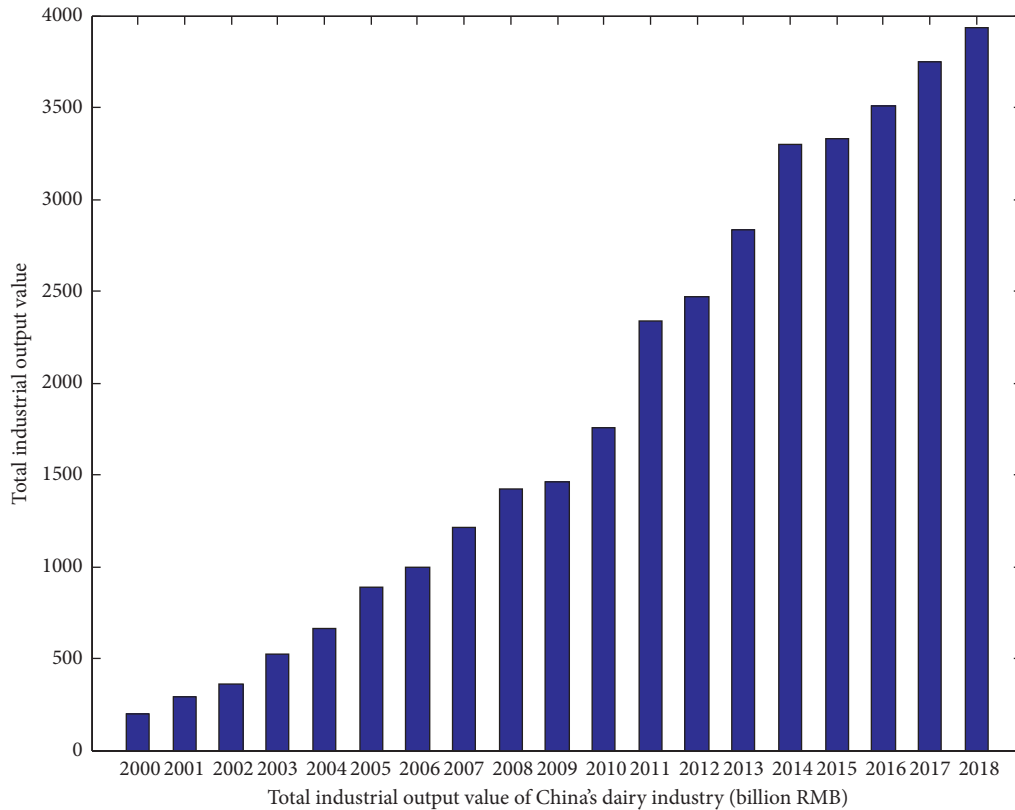
FIGURE 1: Total industrial output value of China's dairy industry from 2000 to 2018. Data source: China dairy yearbook 2001–2018.
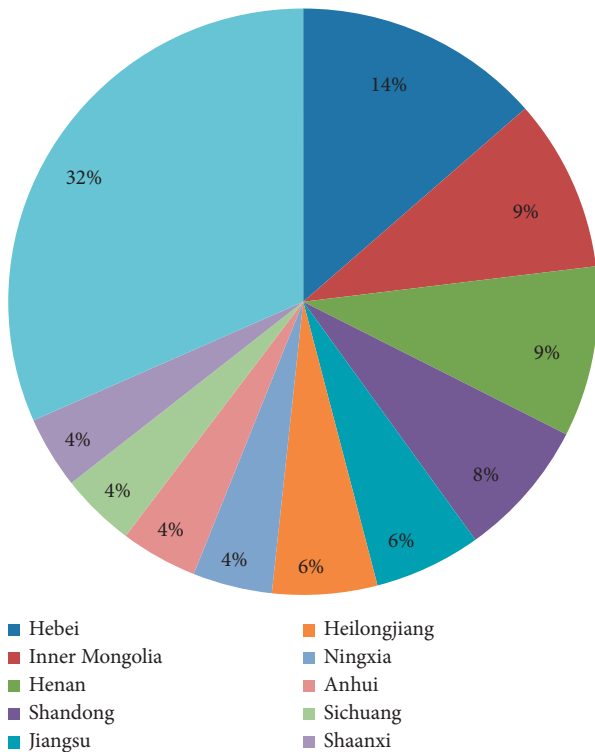
FIGURE 2: Distribution of milk sources in China in year 2018. Data source: China dairy yearbook 2018.

10 regions in terms of dairy production in year 2018 are Hebei, Inner Mongolia, Henan, Shandong, Jiangsu, Heilongjiang, Ningxia, Anhui, Sichuan, and Shaanxi. Among them, Hebei province ranked first in the output of dairy products in year 2018, accounting for 13.59 percent of the country's total output. Hebei produced 3.6529 million tons of dairy products in year 2018. In year 2018, Inner Mongolia ranked second with a production of 2.54882 million tons of dairy products. In year 2018, Henan ranked third in dairy production with 2.516,200 tons. The root cause of unbalanced development of dairy industry in different regions is that, in addition to regional support for the industry, it is also closely related to natural environmental factors and basic industrial conditions. In terms of sales revenue, Mengniu, Yili, and Guangming were 68.977 billion RMB, 78.976 billion RMB, and 20.986 billion RMB in year 2018, respectively, accounting for 59.24% of the national total output. The industry concentration degree is increasingly high (as is shown in Figure 2).

At present, the dairy industry has become an important industry for China to promote regional industrial restructuring and increase farmers' income. Even in some regions (such as Xinjiang and Inner Mongolia), the dairy industry has increasingly become the pillar of local industrial development. Relevant ministries and commissions of the state have also paid close attention to the development of China's dairy industry, providing a favorable environment for the development of China's dairy industry. In year 2018, the

General Office of the State Council issued Opinions on Promoting the Revitalization of the Dairy Industry to ensure the quality and safety of dairy products and comprehensively accelerate the revitalization of the dairy industry to ensure the quality and safety of dairy products. Opinions on Further Promoting the Revitalization of the Dairy Industry jointly issued by nine ministries and commissions put forward that the national dairy output should reach 45 million tons by 2025 and the quality, efficiency, and competitiveness of China's dairy industry should be effectively improved [1].

Despite the rapid development of China's dairy industry, the corresponding demand for dairy products in China is also very large, so the total amount of dairy imports in China continues to increase. According to the statistics of the General Administration of Customs [2], in 2019, China's dairy imports totaled 2.973 million tons, an increase of 12.8% year-on-year, and the import value was 11.13 billion US dollars, an increase of 10.6% year-on-year. With the influence of novel coronavirus pneumonia, China is now consuming 34.3 kilograms of dairy products per capita, and the consumption of dairy products per capita is only 1/3 of the world average, 1/2 of developing countries, and 1/7 of developed countries. The total import volume of dairy products may increase significantly. However, China's dairy trade deficit continues to expand, seriously affecting the rapid development of China's dairy enterprises.

In a word, China's dairy industry has made some achievements, but the international competitiveness of China's dairy industry is far behind that of other countries in the world and does not have comparative advantages. Although there is fierce competition between different countries, the competitiveness of dairy industry among various regions in China is also very worthy of attention. Professor Porter's research on the practice of various countries shows that the fierce domestic market competition plays an important role in the cultivation, formation, and maintenance of the international competitive advantage of the industry. Therefore, it is of great significance for the rapid promotion and sustainable development of the competitiveness of China's dairy industry to build a scientific evaluation index system and to analyze and tap into the overall level and regional differences of the competitiveness of China's dairy industry.

2. Literature Review of Regional Industrial Competitiveness

The studies of regional dairy industry competitiveness are mainly based on the industrial competitiveness theory. At present, domestic and foreign researches mainly focus on the following aspects:

(1) The proposal of industrial competitiveness and its formation mechanism: The study of industrial competitiveness can be traced back to the theory of comparative advantage in the period of classical economics and the theory of factor endowments developed by Heckscher-Ohlin and Vernon's dynamic comparative advantage theory. Based on the theory of comparative advantage, the authors in [3] proposed the "diamond model," which has become a relatively perfect analytical framework for the theory of industrial competitiveness. Many experts revised and supplemented the theory of industrial competitiveness based on factor endowment theory, dynamic comparative advantage theory, or "diamond model" [4–10].

(2) Key factors affecting industrial competitiveness: The authors of [4] believed that the industrial competitive advantage mainly depends on the production factor structure of a country and the production factor input structure of the product. The authors of [5] believed that, in addition to national production factor structure, product life cycle was also an important factor affecting industrial competitive advantage. According to [3], there are five factors that affect the industrial competitive situation, which are the threat of new entrants, the bargaining power of buyers (customers), the threat of substitutes (or services), the bargaining power of suppliers, and the competitive situation of existing competitors. The authors in [6] believed that economies of scale, product differentiation, and global competitive strategy are important factors affecting the international competitive advantage of industries. At the same time, domestic scholars mainly study the key factors of its competitiveness from the industrial level. The authors in [11] believed that the advantages of factor endowment, division of labor innovation, and transaction efficiency can improve the comprehensive competitive advantages of China's high-tech manufacturing industry. The authors in [12] proposed that intellectual capital will exert profound and extensive influence on industrial competitiveness. The authors in [13] believed that independent innovation has a great impact on the international competitiveness of China's manufacturing industry. The authors in [14] believed that knowledge production, knowledge capitalization, and manufacturing basis can enhance the industrial competitiveness of a country.

(3) The evaluation methods to study the industrial competitiveness of a specific country or region: For national or regional industrial competitiveness evaluation, experts and scholars at home and abroad built industry competitiveness evaluation index system and used quantitative evaluation model to analyze the industrial competitiveness evaluation. The main evaluation models include analytic hierarchy process (AHP) [15–19], principal component analysis (PCA) [20–22], fuzzy comprehensive evaluation [23–25], grey correlation method [26–29], SEM method [30–32], and location entropy model [33–35]. The competitive power evaluation method evaluates the industrial or regional competitive power from different angles and various fields and finally puts forward targeted countermeasures and suggestions.

(4) Relevant research on dairy industry competitiveness: In terms of competitiveness of dairy industry, relevant literatures, based on the theory of comparative advantage, compared the competitive advantages of regional dairy industry and its influencing factors, evaluated the sustainable development level of regional dairy industry by constructing relevant indicators, and conducted comprehensive evaluation and dynamic analysis on the development of competitiveness of dairy industry [36–42].

According to the above literature review, domestic and foreign experts and scholars have studied the industrial competitiveness from various perspectives and fields, but there are few studies on the competitiveness of regional dairy industry. Existing literature on the competitiveness of regional dairy products is based on the static analysis of data in a certain region or at a certain time. The relevant evaluation methods are relatively subjective and do not pay attention to the changes in the competitiveness of China's dairy industry in the development process. Aiming at the above problems, this paper adopted projection pursuit model (PP for short) to comprehensively evaluate and dynamically analyze the competitiveness of dairy industry of 31 provinces and regions in China from 2007 to 2016 and studied the panel data of 31 provinces in China for 10 years, which is comprehensive and can reflect the development and change process of regional dairy competitiveness. Projection pursuit model can determine the weighing values of assessment index system by means of dimension reduction method, which outperforms AHP and DEA methods. However, the projection pursuit model has the problems of nonlinear solution problem. Therefore, this paper introduced the accelerating genetic algorithm for real number coding global solution, which can guarantee the stability of the model results.

The basic idea of projection pursuit model is to project the high-dimensional data into the low-dimensional subspace through some combination, find the projection that can reflect the high-dimensional data structure or characteristics by maximizing or minimizing a projection index, and analyze the data structure in the low-dimensional space to achieve the purpose of analyzing and studying the high-dimensional data [43, 44]. However, for high-dimensional nonlinear problems, the accuracy model of PP is not high, the optimal projection direction is unstable, the algorithm parameters are uncertain, it is easy to fall into local optimum, and there are premature convergence and other additional problems [45]. Therefore, the accelerated genetic algorithm can simply deal with the global optimization problem in PP model and obtain the best projection with fast convergence speed, high accuracy, and good robustness. As China's dairy industry plays an important role in the national agricultural development strategy, the research in this paper is based on 31 provinces and cities in China, which can quantify the characteristics of each region and guide the development of China's dairy industry. The main objectives of this paper are as follows: (1) to build a scientific and reasonable evaluation index system of China's dairy industry competitiveness and to improve the nonlinear and discrete defects of the PP model by accelerating genetic algorithms; (2) to construct the projection pursuit model with relatively simple algorithm architecture and parameters, which runs an efficient solution algorithm; (3) to achieve comprehensive evaluation and dynamic analysis of the competitiveness of China's dairy industry from both horizontal and vertical aspects and to find out the characteristics of regional spatial distribution and the main impact factor.

3. Model Construction

3.1. Construction of Evaluation Index System. There are many factors involved in the competitiveness of dairy industry. Google Academy, China HowNet, and related websites are searched with keywords as "dairy products" and "competitiveness" and some competitiveness evaluation indexes are collected. In combination with the views of experts and scholars such as Porter [45] and Cho and Chang Moon [46] the relevant summary of the competitiveness evaluation index system of dairy industry in this paper is constructed, as shown in Table 1.

According to the research results in recent years and combined with the statistical data characteristics of China's dairy products, five factors are used to evaluate the competitiveness of the dairy industry, namely, production factors, industry demand conditions, enterprise competitiveness and growth force, related and supporting industries, and natural environment factors. This paper, based on the related literature and analysis, selected 20 individual indicators and constructed an evaluation system reflecting the competitiveness of the Chinese dairy industry (as is shown in Table 2).

3.2. Evaluation Model Construction. In view of the fact that AHP, fuzzy comprehensive evaluation, and other models that are greatly influenced by subjective factors are adopted in most relevant evaluation methods, this paper constructed a projection pursuit model (PP), in order to evaluate the competitiveness of dairy industry in 31 provinces and cities in China in a more objective scientific and reasonable way. The values of multidimensional evaluation index were projected into one-dimensional projection data. Real coded accelerating genetic algorithm (RAGA) was introduced into the PP model for the calculation of the optimal projection direction. The impacts on the dairy industry competitiveness were analyzed according to the size of projection value of each index, and the competitiveness level of regional dairy industry was determined according to the projection values [52]. The construction process of PP model is as follows:

Step 1. Nondimensional index data: due to different evaluation index data, each evaluating index is nondimensional in order to reduce the influence of raw data for the final evaluation results. It is assumed that the sample set at the kth year is $\{x_k(i, j)$, $i = 1, 2, \ldots, n; j = 1, 2, \ldots, m\}$, where n represents the number of evaluated areas, m is the number of evaluation indexes, and k is the year. In this paper, the

TABLE 1: Summary of competitiveness evaluation indexes of dairy industry.

Scholar (year)	Evaluation index system
Wijnands et al. (2010) [47]	Growth in export share on the world market, annual growth in labor productivity of the daily industry, and annual growth in real value added of the food industry
Nica (2017) [48]	Net turnover, average number of employees, work productivity, gross profit, and profit rate
Ramphul (2019) [49]	Cost of milk production, farm size, infrastructure for milk collection, milk processing capacity, and quality of dairy products
Seyiti and Deng (2013) [42]	Raw milk share, capital share, location quotient, market share, total labor productivity, asset profit rate, per capita GDP index, and technical market turnover index
Ma and Yang (2017) [38]	The number of cows on hand, the output of raw milk, the area of pasture land, the average number of all employees, the annual per capita consumption of dairy products, the total profit, the annual disposable income, sales revenue, the average price of fresh milk, the total assets, the asset liability ratio, the number of dairy enterprises, the number of breeding bull stations selected for premium seed subsidy, the number of feed processing enterprises, the number of technicians, and the credit environmental science
Han and Cheng (2015) [41]	Share of raw milk, share of capital, proportion of regional output value to total output value, relative specialization coefficient, market share, product sales rate, cost profit rate, asset profit rate, sales profit rate, total labor productivity, average growth rate of total assets, and increase of sales revenue rate
Xu (2019) [50]	Milk production, per capita share of rattan, total assets of dairy industry, profit, market share of dairy enterprises, value-added tax payable by dairy enterprises, and asset liability ratio of dairy industry

TABLE 2: Competitiveness evaluation index system of Chinese dairy industry.

System layer	Objective layer	Index layer	Direction
Competitiveness evaluation index system of Chinese dairy industry	Production factors	Cow inventory (ten thousands) X_1	Positive
		Raw milk yield (ten thousand tons) X_2	Positive
		Total area of grassland (ten thousands mu) X_3	Positive
		Number of dairy enterprises X_4	Positive
	Industry demand conditions	Household consumption of dairy products per capita (kg/person) X_5	Positive
		Milk per capita (kg/person) X_6	Positive
		Annual per capita disposable income (RMB) X_7	Positive
	Enterprise competitiveness and growth force	Total profit (100 million RMB) X_8	Positive
		Industrial sales output value (100 million RMB) X_9	Positive
		Market share (%) X_{10}	Positive
		Number of loss-making enterprises X_{11}	Negative
		Number of bulls (head) X_{12}	Positive
	Related and supporting industries	Industrial sales output value of feed enterprises (100 million RMB) X_{13}	Positive
		Mixed feed output (ten thousand tons) X_{14}	Positive
		Feed enterprise profit (100 million RMB) X_{15}	Positive
	Natural environment factors	Annual mean relative humidity (%) X_{16}	Fit index (FI)
		Annual average temperature (°C) X_{17}	Fit index (FI)
		Annual extreme maximum temperature (°C) X_{18}	Negative
		Annual extreme minimum temperature (°C) X_{19}	Positive
		Annual sunshine hours (hours) X_{20}	Positive

number of loss-making enterprises and annual extreme maximum temperatures is negative index; the other indexes are positive indicators. Therefore, positive index, negative index, and fit index are nondimensional.

Nondimensional positive indicators of evaluation index data in the kth year: there is

$$x_k'(i, j) = \frac{x_k(i, j) - x_{k\min}(j)}{x_{k\max}(j) - x_{k\min}(j)}. \tag{1}$$

Nondimensional negative indicators of evaluation index data in the kth year: there is

$$x_k'(i, j) = \frac{x_{k\max}(j) - x_k(i, j)}{x_{k\max}(j) - x_{k\min}(j)}. \tag{2}$$

Nondimensional fit indicators of evaluation index data in the kth year: there is

$$x_k'(i, j) = \frac{1}{\left| x_k(i, j) - \text{fitindx} \right|}, \tag{3}$$

where $x_k'(i, j)$ are the nondimensional results of the evaluation index data in the k the year in (1)–(3) and $x_{k\max}(j)$ and $x_{k\min}(j)$ represent the maximum and minimum values of the evaluation index j.

Step 2. Construction of projection index function: the projection pursuit model forms the optimal projection direction $a = \{a(1), a(2), \ldots, a(m)\}$ by projecting the mth dimensional data $\{x_k(i, j), j = 1, 2, \ldots, m\}$ into the low-dimensional subspace. Then, the projection value of the evaluation index sample i on the optimal projection direction is as follows:

$$z(i) = \sum_{j=1}^{m} a_k(j) x_k(i, j), \quad i = 1, 2, \ldots, n. \tag{4}$$

According to this, one-dimensional scatter diagram of projected values is drawn, which requires the projected values to disperse as a whole and gather locally. Therefore, the projection index function related to standard deviation and in-class coefficients is constructed as follows:

$$Q(a) = S_z D_z, \tag{5}$$

wherein S_z is the standard deviation of the projected value of the evaluation index $z(i)$ and D_z is the in-class coefficient of the projected value of the evaluation index $z(i)$. There are

$$S_z = \sqrt{\frac{\sum_{i=1}^{n}(z(i) - E(z))^2}{n - 1}},$$

$$D_z = \sum_{i=1}^{n} \sum_{j=1}^{m} (R - r(i, j) \times u(R - r(i, j))), \tag{6}$$

where $E(z)$ is the mean value of the sequence of projected values on the optimal projection direction of the evaluation index, $r(i, j) = |z(i) - z(j)|$ is the difference of the projected values on the optimal projection direction, R is the window radius of local density (usually determined according to the experiment), and $u(R - r(i, j))$ is the unit step function. When $R \gg r(i, j)$, there is $u(R - r(i, j)) = 1$; otherwise it is 0.

Step 3. Optimized indicator functions: the optimal projection direction is the vector corresponding to the maximum projection index function. For the determined index set $a = \{a(1), a(2), \ldots, a(m)\}$, the change of the optimal projection direction of the evaluation index will make the projection index function $Q(a)$ change. Therefore, the estimation of the optimal projection direction is actually to solve the problem of maximizing the projection index function. The target function is built as follows:

$$\begin{cases} \max & Q(a) = S_Z D_z \\ \text{s.t.} & \sum_{j=1}^{m} a^2(j) = 1, \quad a(j) \in [0, 1]. \end{cases} \tag{7}$$

Step 4. The optimal projection direction of the evaluation index $a^* = \{a^*(1), a^*(2), \ldots, a^*(m)\}$ can be obtained by solution. The optimal projection direction can be understood as the weight. First, it needs to be

normalized, and then the normalized value of the optimal projection direction after normalization is substituted into (7) to obtain the optimal projection value of each index $z^*(i)$, as follows:

$$z^*(i) = \frac{a^*(j) x(i, j)}{\sum_{j=1}^{m} a^*(j)}, \quad i = 1, 2, \ldots, n. \tag{8}$$

Similarly, the optimal projection value of the spatial angle can be obtained according to the above calculation.

4. Results and Discussion

4.1. Vertical Evaluation of Competitiveness of Chinese Dairy Enterprises. In terms of time dimension, this paper selected data from 2007 to 2016 to vertically evaluate the competitiveness of China's dairy industry through the above index system. The data were obtained from China Dairy Yearbook and China Statistical Yearbook from 2008 to 2018. Based on time dimension of 10 years and 20 index numbers, the optimal projection direction is gotten from accelerating genetic algorithm. Set the selected parent generation of initial population size $s = 400$, crossover probability $P_c = 0.8$, mutation probability $P = 0.2$, the outstanding individual number is 20, and the number of acceleration times is 7, the maximum number of iterations is 200, and the optimal projection directions are gotten based on the time dimension of each evaluation index, as is shown in Table 3.

In terms of the vertical evaluation of the competitiveness of Chinese dairy enterprises, according to the characteristics of the accelerating genetic algorithm proposed in this paper, the accelerating genetic algorithm (AGA) and the standard genetic algorithm (GA) are solved and simulated on the MATLAB platform, and then the two algorithms are solved iteratively, with the negative value of the objective function as the fitness function and the fitness obtained by projecting the pursuit curve, as shown in Figure 3. It shows that, with the increase of the number of iterations, the convergence speed of AGA is significantly faster than that of GA and the AGA algorithm close to the horizontal coordinate does not change with the increase of the number of iterations, reflecting its convergence stability and convergence reliability. In a word, the performance of AGA is obviously better than that of GA.

The optimal projection direction of each evaluated index in the time dimension is shown in Table 3: the optimal projection direction of eight evaluation indexes is above 0.25, namely, the industrial sales value, number of loss-making businesses, feed corporate profits, the sunshine hours throughout the year, mixed feed production, households per capita consumption of dairy products, raw milk production, and the annual per capita disposable income, while there are certain differences but it can be concluded that the factors of production factors, the industry demand conditions, the enterprise competitiveness and growth force, related and supporting industries, and the natural environmental elements have relatively consistent importance in the Chinese dairy industry competitiveness evaluation. Among them, enterprise competitiveness and growth force

TABLE 3: Optimal projection direction of evaluation index based on time dimension.

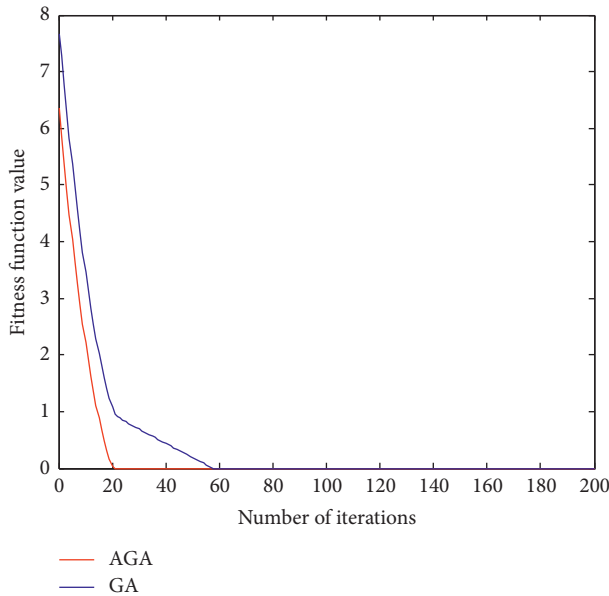| Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Index	X_9	X_{11}	X_{15}	X_{20}	X_{14}	X_5	X_2	X_7	X_{12}	X_{10}	X_{13}	X_{18}	X_6	X_8	X_1	X_{16}	X_3	X_4	X_{19}	X_{17}
Optimal projection direction	0.3337	0.3143	0.3089	0.2837	0.2751	0.2738	0.2629	0.25	0.2416	0.2247	0.2115	0.2089	0.1808	0.1786	0.17	0.1363	0.1195	0.0598	0.0561	0.0379

FIGURE 3: Comparison of the fitness curve of projection vector.

are the main factors affecting the competitiveness of China's dairy industry. According to formula (8), the projected value of Chinese dairy industry index based on time dimension is obtained, as shown in Table 4.

According to Table 4, the time dimension weighted value of competitiveness of Chinese dairy industry can be calculated, as shown in Table 5.

According to the results in Table 5, the comprehensive evaluation results of competitiveness of Chinese dairy industry based on time dimension are ranked as follows: 2016 > 2015 > 2014 > 2013 > 2012 > 2011 > 2010 > 2009 > 2008 > 2007. It can be seen that the development trend of China's dairy industry basically shows a positive trend. With the improvement of the national living standard, the competitiveness of the dairy industry is also constantly improved, from 0.1216 in 2007 to 0.8267 in 2016.

As for the factors of production factor base dimension, it can be seen that although the number of the Chinese dairy industries decreases in recent years, the prairie area does not have too much change, and the cow restock remains the same which reached its peak in 2010 and 2011. With China's dairy industry promoting technological progress and its competitiveness, the final output of raw milk increased from 576,600 tons in 2007 to 965,500 tons in 2016. Therefore, from 2007 to 2016, the production factor base of China's dairy industry was continuously strengthened, and its competitiveness evaluation value increased from 0.0190 to 0.1040.

In regard to the dimension of industry demand conditions, with the improvement of China's national living standard, residents' annual per capita disposable income rose gradually, more and more consumer dairy products were considered as a daily life necessity, and household per capita consumption of dairy products increased gradually, and, at the same time as the production of dairy products increased year by year, dairy per capita is also increasing. Therefore, from 2007 to 2016, the demand conditions level of China's dairy industry rose steadily, and its competitiveness evaluation value increased from 0.0129 to 0.1731.

From the perspective of enterprise competitiveness and growth force dimension, due to the national rectification of dairy enterprises, some unqualified enterprises gradually withdrew from the market, and the total profit and industrial sales value of enterprises increased significantly. The number of bulls as the evaluation index of competitiveness of basic enterprises also showed a trend of increasing year by year with the development of technology. From 2007 to 2016, the overall competitiveness and growth power of enterprises have been rising, and the evaluation value of their competitiveness has increased from 0.0381 to 0.3152.

From the dimension of related and supporting industries, feed industries, as related and supporting industries of China's dairy industries, have seen a growth at the aspects of mixed feed production and corporate profits expected in year 2015 when the industrial sales value of feed industries suffered a brief decline, which laid a solid industrial foundation for increasing the competitiveness of China's dairy industry. From year 2007 to year 2016, the foundation of relevant and supporting industries was gradually consolidated, and the evaluation value of its competitiveness increased from 0 to 0.1427.

From the perspective of natural environmental elements dimensions, dairy production, to a great extent, is affected by the natural environmental elements, and regional differences in natural environment factor endowments directly affect the production of dairy products enterprise decision-making and dairy production quality. From the point of historical data, the annual average relative humidity, annual average temperature, annual extreme minimum temperature and extreme maximum temperature, and annual sunshine hours throughout the year can present different degree of volatility, but these factors do not affect the competitiveness of China's dairy industry from overall increase year by year. From 2007 to 2016, the evaluation value of the competitiveness of natural environmental factors increased from 0.0516 to 0.0916.

4.2. Spatial Competitiveness Evaluation of China's Dairy Enterprises. From the spatial dimension, the data from 31 provinces and cities in China (excluding Hong Kong, Macao, and Taiwan regions) in 2007 to 2016 are regarded as

TABLE 4: Projected value of Chinese dairy industry index based on time dimension.

Index layer	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Cow inventory (ten thousands) X_1	0.0094	0.0138	0.0191	0.0267	0.0412	0.0072	0.0000	0.0079	0.0090	0.0090
Raw milk yield (ten thousand tons) X_2	0.0000	0.0012	0.0078	0.0197	0.0317	0.0400	0.0481	0.0457	0.0526	0.0637
Total area of grassland (ten thousands mu) X_3	0.0000	0.0272	0.0272	0.0272	0.0272	0.0272	0.0272	0.0272	0.0272	0.0289
Number of dairy enterprises X_4	0.0097	0.0145	0.0145	0.0121	0.0024	0.0024	0.0024	0.0000	0.0024	0.0024
Household consumption of dairy products per capita (kg/person) X_5	0.0000	0.0125	0.0159	0.0186	0.0365	0.0469	0.0516	0.0570	0.0614	0.0663
Milk per capita (kg/person) X_6	0.0033	0.0000	0.0034	0.0074	0.0157	0.0270	0.0327	0.0333	0.0371	0.0438
Annual per capita disposable income (RMB) X_7	0.0000	0.0104	0.0179	0.0281	0.0425	0.0573	0.0293	0.0396	0.0499	0.0606
Total profit (100 million RMB) X_8	0.0000	0.0059	0.0179	0.0309	0.0258	0.0277	0.0315	0.0398	0.0433	0.0433
Industrial sales output value (100 million RMB) X_9	0.0000	0.0024	0.0111	0.0211	0.0396	0.0472	0.0604	0.0792	0.0808	0.0808
Market share (%) X_{10}	0.0000	0.0544	0.0544	0.0544	0.0544	0.0544	0.0544	0.0544	0.0544	0.0544
Number of loss-making enterprises X_{11}	0.0381	0.0000	0.0381	0.0381	0.0761	0.0571	0.0761	0.0761	0.0761	0.0761
Number of bulls (head) X_{12}	0.0000	0.0000	0.0128	0.0476	0.0494	0.0567	0.0567	0.0530	0.0585	0.0585
Industrial sales output value of feed enterprises (100 million RMB) X_{13}	0.0000	0.0070	0.0115	0.0186	0.0294	0.0372	0.0446	0.0512	0.0011	0.0012
Mixed feed output (ten thousand tons) X_{14}	0.0000	0.0062	0.0059	0.0126	0.0202	0.0292	0.0354	0.0421	0.0432	0.0666
Feed enterprise profit (100 million RMB yuan) X_{15}	0.0000	0.0000	0.0100	0.0392	0.0402	0.0533	0.0631	0.0688	0.0748	0.0748
Annual mean relative humidity (%) X_{16}	0.0236	0.0307	0.0330	0.0165	0.0330	0.0141	0.0189	0.0118	0.0024	0.0000
Annual average temperature (°C) X_{17}	0.0092	0.0046	0.0046	0.0023	0.0000	0.0023	0.0046	0.0046	0.0046	0.0069
Annual extreme maximum temperature (°C) X_{18}	0.0000	0.0506	0.0506	0.0506	0.0506	0.0506	0.0506	0.0506	0.0506	0.0506
Annual extreme minimum temperature (°C) X_{19}	0.0136	0.0000	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012	0.0136	0.0012
Annual sunshine hours (hours) X_{20}	0.0052	0.0171	0.0300	0.0138	0.0215	0.0000	0.0687	0.0141	0.0146	0.0329

TABLE 5: Evaluation value of competitiveness of Chinese dairy industry based on time dimension.

Index layer	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Factor base	0.0190	0.0567	0.0685	0.0856	0.1025	0.0768	0.0777	0.0807	0.0912	0.1040
Industrial demand condition	0.0129	0.0374	0.0517	0.0662	0.0972	0.1337	0.1160	0.1299	0.1508	0.1731
Enterprise competitiveness and growth power	0.0381	0.0731	0.1395	0.1726	0.2385	0.2437	0.2518	0.2891	0.3046	0.3152
Related and supporting industries	0.0000	0.0132	0.0274	0.0704	0.0898	0.1196	0.1431	0.1621	0.1191	0.1427
Natural environment factor	0.0516	0.1030	0.11a94	0.0845	0.1063	0.0683	0.1440	0.0823	0.0858	0.0916
Comprehensive level	0.1216	0.2833	0.4066	0.4794	0.6343	0.6422	0.7326	0.7442	0.7514	0.8267

the foundation; a comparative study was conducted to find out the competitiveness level difference between the dairy industries in different regions. The accelerating genetic algorithm was solved with the 31 sample numbers and 20 indexes. Set the selected parent generation of initial population size $s = 400$, crossover probability $P_c = 0.8$, mutation probability $P = 0.2$, the outstanding individual number is 20, the maximum number of iterations is 200, the times of acceleration are 7, and the optimal projection direction of each evaluation index on the spatial dimension is gotten as shown in Table 6.

For the horizontal evaluation of competitiveness of Chinese dairy enterprises, according to the characteristics of the accelerated genetic algorithm proposed in this paper, the accelerated genetic algorithm (AGA) and the standard genetic algorithm (GA) are solved and simulated on the MATLAB platform, and then the two algorithms are solved iteratively, with the negative value of the objective function as the fitness function, and the fitness obtained by projecting the pursuit curve is shown in Figure 4. It shows that, with the increase of the number of iterations, the convergence speed of AGA is significantly faster than that of GA, and the AGA algorithm close to the horizontal coordinate does not change with the increase of the number of iterations, reflecting its convergence stability and convergence reliability. In a word, the performance of AGA is obviously better than that of GA.

It can be seen from Table 6 that the number of loss-making businesses, raw milk production, market share, cow inventory, and sunshine hours throughout the year rank top 5, which means that production factors base, enterprise competitiveness and growth, and natural environmental elements are the main factors influencing China's dairy industry competitiveness in 31 provinces. According to the optimal projection direction, the projected value of dairy competitiveness of each region was obtained, as shown in Table 7.

In order to clearly analyze the spatial differences in the competitiveness of the dairy industry and explore the status and level of the regional dairy industry in different dimensions, this paper classified and sorted out the regions according to the evaluation index value of the spatial dimension of the competitiveness of the dairy industry. The specific division method is as follows: firstly, the evaluation index values of the spatial dimension of dairy industry competitiveness are arranged in ascending order. The maximum value $z(i)_{max} = 0.681796$ is the optimal competitiveness, and the minimum value $z(i)_{min} = 0.044230$ is the weakest competitiveness; secondly, according to the regional characteristics of China's administrative regions, it

is divided into four categories, the evaluation index values are correspondingly converted into four grey categories, and then the comprehensive decision-making measures are calculated by calculating the clustering coefficient of each region; finally, the categories of all regions are determined according to the comprehensive decision-making measures, and the corresponding calculation steps are referred to as [51, 52]. The result is shown in Figure 5.

As can be seen from Figure 5, there are significant differences in the competitiveness of China's dairy industry in 31 provinces and cities, and there is no overall regional correlation. The most competitive provinces and cities in China's dairy industry include Inner Mongolia, Heilongjiang, Shandong, and Hebei, among which Inner Mongolia dairy industry competition is the leading one. The best projection value of its comprehensive competitiveness is 0.6818, which is quite obvious compared with the second place 0.5803 in Shandong province. Particularly with the improvement of the concentration degree of Inner Mongolia dairy industry in recent years, its production factor base, industrial demand conditions, enterprise competitiveness and growth, related and supporting industries, and natural environment factors are in a dominant position.

According to the evaluation index value of spatial dimension of competitiveness of dairy industry in Table 7, 31 regions are divided into 4 grades, as is seen in Tables 8–10. As shown in Table 8, the areas ranking high in overall competitiveness have advantages (ranking top 4) in the aspects of production factors base and enterprise competitiveness and growth performance. However, related and supporting industries and the natural environmental elements in these areas have relatively obvious gap, especially the natural environmental elements, only Inner Mongolia ranking above the national average level. Famous dairy enterprises such as Mengniu group, Yili group, Feihe dairy, and Sanyuan dairy are mainly located in such provinces as Inner Mongolia, Heilongjiang, and Hebei. Strong enterprise strength and sufficient milk sources and pastures make these regions highly competitive. Therefore, the provinces with higher overall ranking need to maintain factors of production base and strong force of enterprise competitiveness and growth. Heilongjiang and Hebei at the same time also need to continue exploring industry demand conditions, Inner Mongolia and Heilongjiang should pay more attention to the related and supporting industries, and Shandong, Heilongjiang, and Hebei need to improve the dairy industry needs, natural environmental elements, and sustainable regional dairy competitiveness level.

TABLE 6: Optimal projection direction of evaluation index based on spatial dimension.

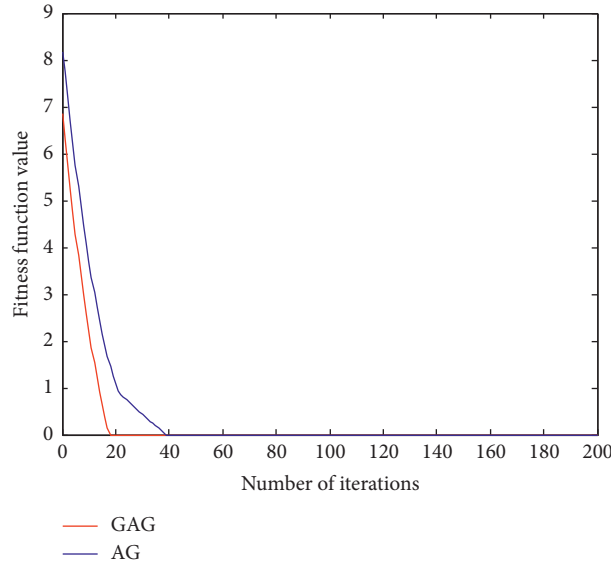| Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Index	X_{11}	X_2	X_{10}	X_1	X_{20}	X_9	X_{12}	X_4	X_8	X_{15}	X_{17}	X_7	X_5	X_{14}	X_6	X_3	X_{19}	X_{13}	X_{18}	X_{16}
Optimal projection direction	0.3317	0.3279	0.3061	0.2870	0.2812	0.2660	0.2595	0.2568	0.2361	0.2333	0.1927	0.1865	0.1853	0.1742	0.1621	0.1497	0.1027	0.0842	0.0774	0.0644

FIGURE 4: Comparison of the fitness curve of projection vector.

According to Table 9, it can be seen that the corporate competitiveness and growth power in all the regions ranked at the second level, but other factors influencing indicators of competitiveness levels show the instability characteristics, especially the natural environmental elements. Except that Xinjiang and Liaoning are still in the second level, the rest are in a low level. Therefore, in the process of late industry development, the natural environment should be focused to make up for the shortage. Of course, Shanghai, Jiangsu, and Beijing are the provinces with higher level of industrial demand. Due to the relatively developed economy in these areas, people's living standards are relatively high, and the per capita occupancy and demand for dairy products are relatively large. The per capita consumption of dairy products in these three regions accounts for 15.62% of the national average, and the per capita disposable income accounts for 16.4% of the national average. The relatively large scale of demand and relatively high per capita disposable income can promote the rapid development of the dairy industry. Therefore, regions at the second level should continue to maintain their competitive advantages of enterprise competitiveness and growth power, and, at the same time, they need to learn from the regions at the first level and strive to get closer to the first level by improving the industry's own ability.

As shown in Table 10, only Chongqing, Hainan, and Guizhou are at the fourth level with evaluation index value of 0–0.1, so the third level and the fourth level are analyzed together. For the areas whose dairy industry competitiveness ranked relatively lower, their influence factor index competitiveness is also relatively backward. Of course, there are also some areas that have certain advantages, such as Ningxia and Fujian, which have certain industrial competitive advantage on the natural environment factors and industrial demand. That is to say, if the two regions continue to improve the factors of production factor base and enterprise competitiveness and growth, the whole competitiveness level will be updated. The improvement of the competitiveness of dairy industry in other regions requires long-term accumulation of production factors, mining of industrial demand, expansion of enterprise competitiveness and growth power, development of relevant and supportive industries, and optimization of natural environment factors to stimulate the improvement of regional dairy industry competitiveness.

4.3. Comparative Analysis of Evaluation Results. The projection values of time and space dimensions of competitiveness of China's dairy industry obtained from GAG-PP are compared with the calculation results of AHP (subjective evaluation) and entropy weight evaluation (objective evaluation), as is shown in Tables 11 and 12. This paper uses Yaahp software to determine the weight in AHP by [42] and expert consultation. Entropy weight evaluation is used to judge the dispersion degree of an index by entropy value. The smaller the information entropy value is, the greater the dispersion degree of the index is, and the greater the influence (weight) of the index on the comprehensive evaluation is [54].

Table 7: Projected values of dairy industry indicators of 31 provinces and cities in China based on spatial dimensions.

Index layer	Beijing	Tianjin	Hebei	Shanxi	Inner Mongolia	Liaoning	Jilin	Heilongjiang	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong	Henan	Hubei	Hunan	Guangdong	Guangxi	Hainan	Chongqing	Sichuan	Guizhou	Yunnan	Xizang	Shanxi	Gansu	Qinghai	Ningxia	Xinjiang
Cow inventory (ten thousands) X_1	0.0038	0.0039	0.0524	0.0096	0.0689	0.0094	0.0084	0.0576	0.0016	0.0053	0.0015	0.0026	0.0012	0.0014	0.0338	0.0239	0.0017	0.0024	0.0014	0.0009	0.0000	0.0005	0.0049	0.0009	0.0046	0.0070	0.0145	0.0065	0.0064	0.0092	0.0604
Raw milk yield (ten thousand tons) X_2	0.0132	0.0112	0.0662	0.0122	0.0787	0.0228	0.0027	0.0434	0.0108	0.0269	0.0097	0.0176	0.0042	0.0059	0.0556	0.0393	0.0161	0.0064	0.0133	0.0059	0.0000	0.0032	0.0174	0.0013	0.0097	0.0001	0.0342	0.0049	0.0030	0.0105	0.0082
Total area of grassland (ten thousands mu) X_3	0.0001	0.0000	0.0020	0.0020	0.0345	0.0015	0.0025	0.0033	0.0000	0.0001	0.0014	0.0007	0.0009	0.0019	0.0007	0.0019	0.0028	0.0028	0.0014	0.0038	0.0004	0.0009	0.0089	0.0018	0.0067	0.0359	0.0023	0.0078	0.0159	0.0013	0.0251
Number of dairy enterprises X_4	0.0085	0.0116	0.0331	0.0146	0.0516	0.0170	0.0085	0.0509	0.0069	0.0239	0.0162	0.0100	0.0069	0.0062	0.0617	0.0331	0.0108	0.0116	0.0208	0.0108	0.0015	0.0031	0.0046	0.0023	0.0108	0.0000	0.0355	0.0146	0.0046	0.0139	0.0231
Household consumption of dairy products per capita (kg/person) X_5	0.0363	0.0200	0.0067	0.0093	0.0116	0.0123	0.0009	0.0028	0.0445	0.0090	0.0167	0.0216	0.0043	0.0088	0.0156	0.0075	0.0076	0.0002	0.0129	0.0045	0.0000	0.0196	0.0046	0.0023	0.0029	0.0330	0.0155	0.0079	0.0071	0.0114	0.0090
Milk per capita (kg/person) X_6	0.0081	0.0097	0.0111	0.0041	0.0389	0.0063	0.0011	0.0138	0.0056	0.0041	0.0021	0.0034	0.0013	0.0015	0.0069	0.0050	0.0033	0.0011	0.0014	0.0014	0.0000	0.0012	0.0025	0.0004	0.0024	0.0005	0.0111	0.0022	0.0065	0.0197	0.0044
Annual per capita disposable income (RMB) X_7	0.0406	0.0233	0.0064	0.0059	0.0108	0.0122	0.0061	0.0040	0.0448	0.0215	0.0306	0.0063	0.0076	0.0057	0.0137	0.0055	0.0073	0.0077	0.0217	0.0065	0.0063	0.0093	0.0050	0.0015	0.0046	0.0009	0.0054	0.0000	0.0021	0.0050	0.0025
Total profit (100 million RMB) X_8	0.0010	0.0022	0.0197	0.0029	0.0567	0.0097	0.0011	0.0407	0.0107	0.0053	0.0030	0.0047	0.0011	0.0032	0.0298	0.0118	0.0034	0.0041	0.0208	0.0035	0.0000	0.0008	0.0036	0.0005	0.0025	0.0006	0.0091	0.0014	0.0007	0.0035	0.0024
Industrial sales output value (100 million RMB) X_9	0.0117	0.0081	0.0305	0.0053	0.0639	0.0142	0.0032	0.0472	0.0174	0.0110	0.0062	0.0080	0.0019	0.0039	0.0353	0.0138	0.0065	0.0066	0.0196	0.0037	0.0000	0.0032	0.0064	0.0010	0.0038	0.0003	0.0173	0.0025	0.0010	0.0056	0.0043
Market share (%) X_{10}	0.0122	0.0082	0.0367	0.0082	0.0735	0.0163	0.0041	0.0531	0.0204	0.0122	0.0082	0.0082	0.0041	0.0041	0.0408	0.0163	0.0082	0.0082	0.0204	0.0041	0.0000	0.0041	0.0082	0.0000	0.0041	0.0000	0.0204	0.0041	0.0000	0.0041	0.0041
Number of loss-making enterprises X_{11}	0.0065	0.0109	0.0338	0.0185	0.0578	0.0196	0.0109	0.0589	0.0076	0.0284	0.0185	0.0131	0.0087	0.0076	0.0796	0.0447	0.0120	0.0142	0.0240	0.0131	0.0011	0.0022	0.0175	0.0022	0.0098	0.0000	0.0415	0.0153	0.0065	0.0131	0.0218
Number of bulls (head) X_{12}	0.0376	0.0174	0.0485	0.0109	0.0496	0.0112	0.0268	0.0368	0.0371	0.0090	0.0000	0.0097	0.0000	0.0057	0.0377	0.0623	0.0092	0.0040	0.0017	0.0216	0.0000	0.0001	0.0091	0.0018	0.0231	0.0000	0.0066	0.0015	0.0111	0.0126	0.0505
Industrial sales output value of feed enterprises (100 million RMB) X_{13}	0.0015	0.0013	0.0061	0.0013	0.0026	0.0158	0.0045	0.0007	0.0011	0.0071	0.0036	0.0042	0.0055	0.0077	0.0202	0.0105	0.0067	0.0102	0.0165	0.0073	0.0009	0.0020	0.0094	0.0006	0.0015	0.0020	0.0006	0.0000	0.0002	0.0010	0.0012
Mixed feed output (ten thousand tons) X_{14}	0.0031	0.0028	0.0175	0.0035	0.0086	0.0247	0.0110	0.0038	0.0017	0.0102	0.0069	0.0085	0.0127	0.0186	0.0418	0.0249	0.0153	0.0217	0.0331	0.0209	0.0029	0.0041	0.0188	0.0016	0.0041	0.0000	0.0043	0.0019	0.0001	0.0005	0.0036
Feed enterprise profit (100 million RMB yuan) X_{15}	0.0071	0.0038	0.0189	0.0009	0.0122	0.0452	0.0100	0.0046	0.0051	0.0234	0.0074	0.0100	0.0026	0.0230	0.0560	0.0509	0.0176	0.0234	0.0422	0.0172	0.0016	0.0054	0.0300	0.0011	0.0034	0.0000	0.0083	0.0012	0.0001	0.0006	0.0043
Annual mean relative humidity (%) X_{16}	0.0013	0.0066	0.0037	0.0023	0.0005	0.0019	0.0125	0.0023	0.0007	0.0007	0.0007	0.0004	0.0006	0.0006	0.0033	0.0066	0.0003	0.0005	0.0003	0.0002	0.0001	0.0003	0.0003	0.0002	0.0010	0.0000	0.0155	0.0023	0.0039	0.0008	0.0029
Annual average temperature (°C) X_{17}	0.0023	0.0022	0.0031	0.0012	0.0004	0.0005	0.0002	0.0000	0.0086	0.0060	0.0100	0.0065	0.0463	0.0194	0.0034	0.0048	0.0079	0.0137	0.0110	0.0151	0.0046	0.0156	0.0074	0.0033	0.0052	0.0008	0.0037	0.0006	0.0001	0.0010	0.0004
Annual extreme maximum temperature (°C) X_{18}	0.0012	0.0016	0.0002	0.0031	0.0050	0.0056	0.0063	0.0041	0.0011	0.0017	0.0005	0.0016	0.0008	0.0015	0.0011	0.0000	0.0017	0.0008	0.0020	0.0021	0.0023	0.0004	0.0029	0.0082	0.0116	0.0186	0.0004	0.0038	0.0088	0.0028	0.0020
Annual extreme minimum temperature (°C) X_{19}	0.0038	0.0036	0.0047	0.0026	0.0012	0.0005	0.0004	0.0000	0.0074	0.0062	0.0078	0.0061	0.0126	0.0089	0.0041	0.0057	0.0069	0.0087	0.0141	0.0121	0.0246	0.0121	0.0090	0.0077	0.0094	0.0044	0.0056	0.0019	0.0015	0.0021	0.0010
Annual sunshine hours (hours) X_{20}	0.0484	0.0429	0.0377	0.0511	0.0548	0.0508	0.0517	0.0396	0.0227	0.0315	0.0202	0.0275	0.0078	0.0310	0.0389	0.0289	0.0261	0.0242	0.0215	0.0201	0.0288	0.0000	0.0019	0.0056	0.0481	0.0675	0.0312	0.0503	0.0537	0.0585	0.0631
Comprehensive level	0.2484	0.1915	0.4392	0.1694	0.6818	0.2975	0.1731	0.4683	0.2561	0.2536	0.1712	0.1708	0.1710	0.1667	0.5803	0.3975	0.1712	0.1723	0.3003	0.1748	0.0752	0.0880	0.1925	0.0442	0.1694	0.1715	0.2828	0.1308	0.1333	0.1771	0.2945

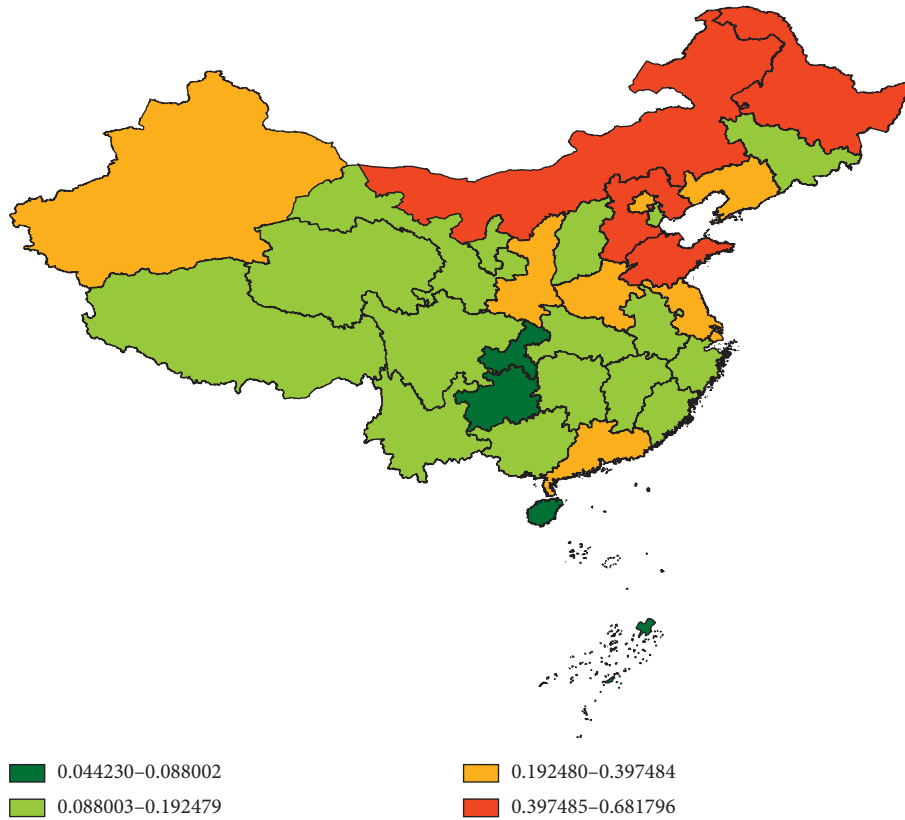| 0.044230–0.088002 | 0.192480–0.397484 |
| 0.088003–0.192479 | 0.397485–0.681796 |

FIGURE 5: Spatial differentiation of competitiveness of Chinese dairy industry.

TABLE 8: Regions with first-level competitiveness.

Areas	Evaluation index value ranking [>0.4]	Factor base ranking	Industrial demand condition ranking	Enterprise competitiveness and growth power ranking	Relevant and supporting industries	Natural environment factor ranking
Inner Mongolia	1	1	3	1	14	8
Shandong	2	4	7	3	1	17
Heilongjiang	3	2	18	2	20	24
Hebei	4	3	16	4	9	19

TABLE 9: Regions with second-level competitiveness.

Areas	Evaluation index value ranking [>0.4]	Factor base ranking	Industrial demand condition ranking	Enterprise competitiveness and growth power ranking	Relevant and supporting industries	Natural environment factor ranking
Henan	5	6	21	5	3	23
Guangdong	6	13	9	8	2	20
Liaoning	7	10	14	10	4	12
Xinjiang	8	5	23	9	21	5
Shaanxi	9	7	12	6	17	16
Shanghai	10	26	1	7	24	27
Jiangsu	11	8	6	12	11	22
Beijing	12	22	2	11	18	14

TABLE 10: Regions with second-level and fourth-level competitiveness.

Area	Evaluation index value ranking [0–0.3]	Factor base ranking	Industrial demand condition ranking	Enterprise competitiveness and growth power ranking	Relevant and supporting industries	Natural environment factor ranking
Sichuan	13	10	17	17	5	31
Tianjin	14	21	4	13	24	15
Ningxia	15	14	8	21	29	7
Guangxi	16	25	25	16	8	18
Jilin	17	24	29	15	13	5
Hunan	18	23	28	23	6	21
Xizang	19	11	10	31	30	1
Hubei	20	17	20	20	11	25
Zhejiang	21	20	5	23	16	28
Fujian	22	28	11	27	12	2
Anhui	23	19	13	18	15	26
Shaanxi	24	12	19	17	25	11
Yunnan	25	16	27	20	22	3
Jiangxi	26	27	22	25	8	9
Qinghai	27	19	24	27	31	6
Gansu	28	15	26	25	28	14
Chongqing	29	29	15	28	19	29
Hainan	30	31	30	30	26	10
Guizhou	31	30	31	29	27	30

TABLE 11: Comparison of competitiveness evaluation results of China's dairy industry based on time dimension.

Comprehensive level	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
GAG-PP	0.1216	0.2833	0.4066	0.4794	0.6343	0.6422	0.7326	0.7442	0.7514	0.8267
AHP	0.1725	0.3954	0.4870	0.5759	0.7066	0.5001	0.5053	0.5635	0.5846	0.6297
Entropy weight	0.2741	0.2484	0.3521	0.4171	0.5001	0.4980	0.6054	0.5951	0.6701	0.6218

As shown in Figures 6 and 7, the results of AHP and entropy weight of time dimension and space dimension data of China's dairy industry competitiveness are quite different. The main reason is that AHP is based on the main experience of experts, while entropy weight is only based on the weight of data. Once the data distribution is relatively uniform, entropy weight loses its significance.

According to the comparison of the evaluation results of the competitiveness of China's dairy industry based on the time dimension in Table 11, it can be seen that there are certain differences in the calculation results of the three, but the overall competitiveness development trend from 2007 to 2016 is basically the same, showing a great increase in competitiveness. Combined with the relevant research in this paper and the actual situation of China's dairy industry competitiveness, GAG-PP shows that the competitiveness of China's dairy industry in the time dimension presents an upward trend year by year. The AHP shows that the competitiveness of China's dairy industry presents an inverted "U" shape in the time dimension, while the entropy weight shows that the competitiveness of China's dairy industry fluctuated greatly from 2013 to 2016. Combined with the research results of [38, 41, 42] and consulting relevant experts, it is considered that only GAG-PP results can more accurately reflect the actual evaluation of dairy industry competitiveness, which shows that GAG-PP can improve the accuracy and scientificity of evaluation.

The comparison of the calculated time dimension competitiveness evaluation value competitiveness evaluation value results is shown in Table 12. Although the final classification results of the three models are relatively close, there are also small fluctuations. For example, in the regions in the first level of dairy industry competitiveness, the ranking results of GAG-PP and entropy weight are Inner Mongolia > Shandong > Heilongjiang > Hebei, the ranking results of AHP are Inner Mongolia > Heilongjiang > Shandong > Hebei, from the actual situation, and the ranking results of AHP are not very consistent with the actual situation, which shows that the results of AHP are subjective. The ranking results of GAG-PP are Henan > Guangdong > Liaoning > Xinjiang > Shaanxi > Shanghai > Jiangsu > Beijing, the ranking results of AHP are Xinjiang > Henan > Tibet > Liaoning > Shaanxi > Guangdong > Jiangsu > Shanghai, and the ranking results of entropy weight are HeNan > Guangdong > Xinjiang > Shaanxi > Liaoning > Shanghai > Jiangsu > Tibet. Although, in the second level of regional ranking, both entropy weight and AHP have Tibet region, from the actual situation of the competitiveness of Tibetan dairy products, it is not particularly prominent, but the results caused by the model are not very consistent with the reality, so the ranking results of GAG-PP can more truly reflect the actual situation of the competitiveness of China's dairy industry.

Table 12: Comparison of evaluation results of competitiveness of China's dairy industry based on spatial dimension.

Region	Beijing	Tianjin	Hebei	Shanxi	Inner Mongolia	Liaoning	Jilin	Heilongjiang	Shanghai	Jiangsu	Zhejiang	Anhui	Fujian	Jiangxi	Shandong	Henan	Hubei	Hunan	Guangdong	Guangxi	Hainan	Chongqing	Sichuan	Guizhou	Yunnan	Xizang	Shaanxi	Gansu	Qinghai	Ningxia	Xinjiang
GAG-PP	0.2484	0.1915	0.4392	0.1694	0.6818	0.2975	0.1731	0.4683	0.2561	0.2536	0.1712	0.1708	0.1710	0.1667	0.5803	0.3975	0.1712	0.1723	0.3003	0.1748	0.0752	0.0880	0.1925	0.0442	0.1694	0.1715	0.2828	0.1308	0.1333	0.1771	0.2945
AHP	0.1784	0.1422	0.4493	0.1378	0.7564	0.2359	0.1334	0.4859	0.1826	0.1838	0.1293	0.1240	0.1283	0.1208	0.4781	0.3124	0.1223	0.1277	0.2180	0.1192	0.0444	0.0763	0.1737	0.0368	0.1247	0.2409	0.2354	0.1201	0.1437	0.1529	0.4492
Entropy weight	0.1808	0.1599	0.3632	0.1202	0.6096	0.2553	0.1803	0.3794	0.2055	0.1945	0.1367	0.1325	0.1677	0.1510	0.4768	0.3432	0.1450	0.1583	0.2758	0.1621	0.0624	0.0880	0.1913	0.0557	0.1486	0.1938	0.2621	0.0978	0.1346	0.1247	0.2649
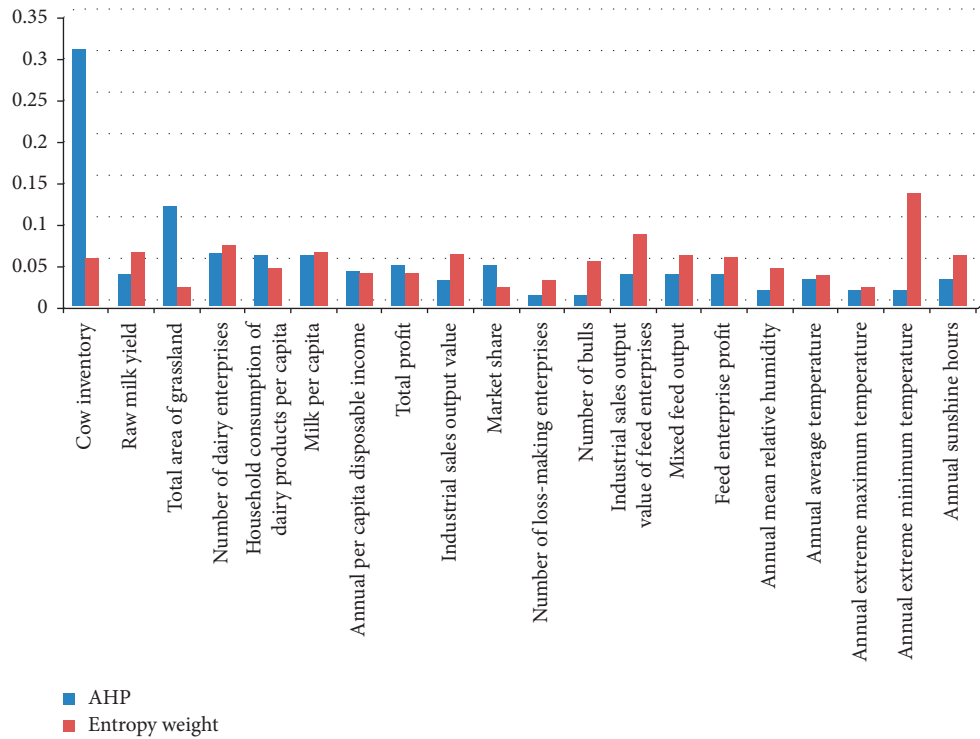
Figure 6: AHP and entropy weight of time dimension data of China's dairy industry competitiveness.
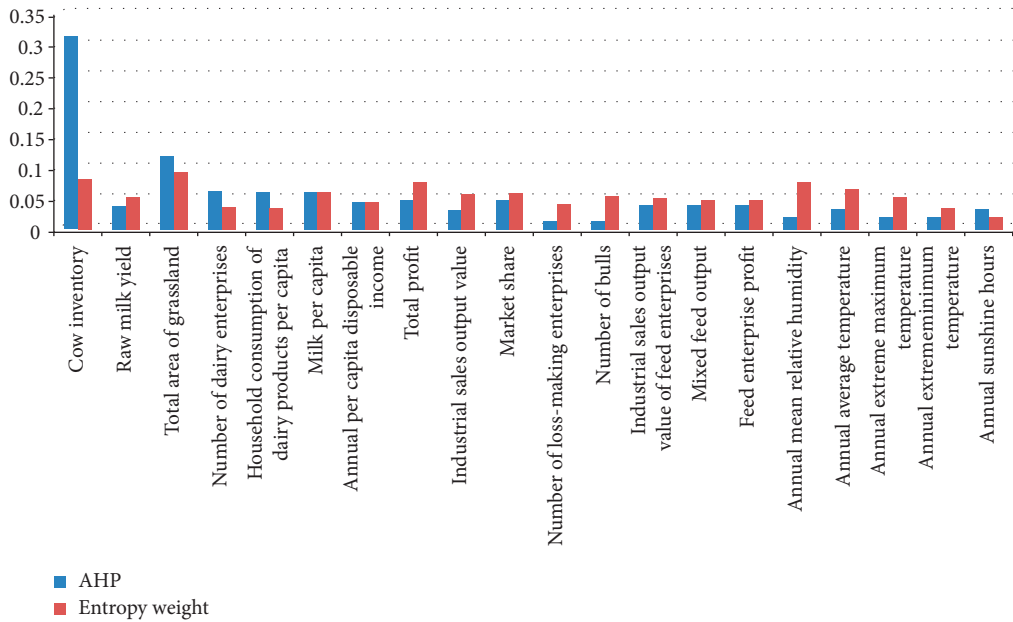
Figure 7: AHP and entropy weight of spatial dimension data of China's dairy industry competitiveness.

5. Conclusions and Recommendations

This paper selected the factors of production factor base, industry demand conditions, enterprise competitiveness and growth force, related and supporting industries, and natural environmental elements as five dimensions and evaluated the competitiveness of China's dairy industry by means of accelerating genetic projection pursuit model from two perspectives of time and space. The results show that competitiveness of China's dairy industry in time presents the overall good development momentum, but from the point of view on the space it shows a larger difference. At the same time, the optimal projection direction of each evaluation index based on time and space dimensions indicates that the number of loss-making enterprises, the number of hours of sunshine in a year, and raw milk output are

relatively consistent in their importance to the competitiveness of China's dairy industry, among which the most important is to increase raw milk output. Therefore, the dairy breeding technology innovation can be enhanced to improve the dairy breed structure by improving dairy breed. Meanwhile, forage and feed varieties can be optimized and standardized to improve the yield of milk. In addition, learning the advanced technology and advanced management experience from other countries can promote the overall strength of the dairy industry.

Data Availability

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

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] Minstry of Agriculture and Rural Affairs of the People's Republic of China, *Nine Ministries and Commissions Including the Ministry of Agriculture and Rural Areas Jointly Issued Several Opinions on Further Promoting the Revitalization of Dairy Industry*, Minstry of Agriculture and Rural Affairs of the People's Republic of China, Beijing, China, 2018, http://www.moa.gov.cn/xw/zwdt/201812/t20181226_6165640.htm. 2020.6.5.

[2] Prospective Industry Research Institute, *China Dairy Industry Market Demand Forecast and Investment Strategic Planning Analysis Report*, Prospective Industry Research Institute, Beijing, China, 2019, https://www.qianzhan.com/analyst/detail/220/200514-be029970.html. 2020.6.5.

[3] M. Porter, *National Competitive Advantage*, Huaxia Press, Beijing, China, 2002.

[4] M. Peter, "Ricardian–Heckscher–Ohlin comparative advantage: theory and evidence," *Journal of International Economics*, vol. 82, no. 2, pp. 137–151, 2010.

[5] B. R. Copeland and A. Kotwal, "Product quality and the theory of comparative advantage," *European Economic Review*, vol. 40, no. 9, pp. 1745–1760, 1996.

[6] K. Paul, *Pop Internationalism*, MIT Press, Boston, MA, USA, 1996.

[7] X. Luo, "Coastal tourism commodity industry cluster based on diamond model and ecological Niche," *Journal of Coastal Research*, vol. 94, no. 1, p. 828, 2019.

[8] K. Fang, Y. Zhou, S. Wang, R. Ye, and S. Guo, "Assessing national renewable energy competitiveness of the G20: a

[9] revised Porter's diamond model," *Renewable and Sustainable Energy Reviews*, vol. 93, pp. 719–731, 2018.

[9] T. W. Chung, "A study on logistics cluster competitiveness among Asia main countries using the Porter's diamond model," *The Asian Journal of Shipping and Logistics*, vol. 32, no. 4, pp. 257–264, 2016.

[10] S. Fainshmidt, A. Smith, and W. Q. Judge, "National competitiveness and Porter's diamond model: the role of MNE penetration and governance quality," *Global Strategy Journal*, vol. 6, no. 2, pp. 81–104, 2016.

[11] G. Li and S. Lu, "Transformation advantages of China's high-tech manufacturing industry—empirical analysis based on comprehensive advantage strategy theory," *Journal of Shanxi University of Finance and Economics*, vol. 41, no. 11, pp. 55–68, 2019.

[12] C. Tan and L. Liu, "The differential effect of knowledge capital on the intensive and extensive complexity of national advantageous industrial portfolio," *International Economics and Trade Research*, vol. 35, no. 7, pp. 40–53, 2019.

[13] Y. He, Q. Han, and Y. Zeng, "A research on the impact of indigenous innovation on international competitiveness of China's manufacturing industry," *Science Research Management*, vol. 40, no. 7, pp. 33–46, 2019.

[14] T. Zhang, "Sources of national competitive advantages—knowledge production, knowledge capitalization and manufacturing base," *Exploration and Free Views*, no. 7, pp. 136–146, 2019.

[15] F. Zhou, "Study on evaluation index system of regional cross-border electronic commerce industry," *Journal of Chongqing University of Technology (Natural Science)*, vol. 33, no. 1, pp. 201–209, 2019.

[16] M. K. Singh, H. Kumar, M. P. Gupta, and J. Madaan, "Analyzing the determinants affecting the industrial competitiveness of electronics manufacturing in India by using TISM and AHP," *Global Journal of Flexible Systems Management*, vol. 19, no. 3, pp. 191–207, 2018.

[17] J. Xu and Y. Liu, "Evaluation of industrial competitiveness based on BMK and AHP analysis," *Soft Science*, vol. 25, no. 3, pp. 70–73, 2011.

[18] B. Yan, R. Ren, and M. Wang, "AHP and fuzzy evaluation of competitiveness of China's auto industrial clusters," *Journal of Southwest Jiaotong University*, vol. 8, no. 4, pp. 12–18, 2007.

[19] S. Liu, X. Ye, and W. Chen, "Analysis of competitiveness of characteristic industrial cluster based on AHP—a case study of ceramic industrial cluster in Jiangxi, Guangdong and Fujian," *China Soft Science*, no. 8, pp. 141–146, 2005.

[20] H. Li, S. Wang, and Q. Xie, "Research on oil industrial competitiveness in Shandong based on principal component analysis," *Modern Management*, vol. 1, pp. 125–130, 2011.

[21] L. Cui and L. Zhang, "the evaluation of the competitiveness of high-tech industry in coastal provinces in China," *Emergence and Transfer of Wealth*, vol. 4, no. 3, pp. 31–40, 2014.

[22] D. Chen, D. Yang, H. Li, and L. Chen, "Dynamic competitiveness of segmentation of construction industry based on principal component analysis," *Journal of Engineering Management*, vol. 31, no. 2, pp. 38–42, 2017.

[23] S. Yan, "Assessment of Beijing's new energy industry based on AHP-FCE comprehensive evaluation," *Science and Technology Management Research*, vol. 37, no. 7, pp. 93–97, 2017.

[24] W. Yu, "Research on the comprehensive evaluation model of regional sports industry competitiveness based on PCA and NNS," *Journal of Sichuan Normal University*, vol. 38, no. 4, pp. 624–628, 2015.

[25] C. Yang, X. Li, and S. Zhang, "Comparative research on competitiveness of Suzhou service outsourcing industry based

on multi-step fuzzy comprehensive evaluation," *Science and Technology Management Research*, vol. 33, no. 17, pp. 65–71, 2013.

[26] J. Du and Y. Liu, "Regional dierences of development level of creative cultural industry in Jiangsu province based on grey relational clustering model," *Generalized Virtual Economy*, vol. 10, no. 2, pp. 67–74, 2019.

[27] G. Wang, W. Chen, and Q. Cao, "An empirical study on influencing factors and impact mechanism of forestry industry competitiveness based on the structural equation model," *Journal of Ocean University of China*, no. 2, pp. 72–78, 2019.

[28] Y. Hu and T. Wu, "Research on the evaluation of distribution industry competitiveness for region based on SEM," *Journal of UESTC (Social Sciences Edition)*, pp. 1–7, 2019.

[29] Q. Guo and T. Liu, "Synergistic effect of technological innovation, industrial structure and financial development," *Journal of Shandong University of Science and Technology*, vol. 19, no. 5, pp. 61–67, 2017.

[30] M. Zu and W. Zhou, "Grey correlation analysis between the technological innovation ability and automobile industry competitiveness," *Journal of Anhui University of Technology*, vol. 33, no. 6, pp. 12–14, 2016.

[31] X. Ruan and R. Shi, "Study on the evaluation of competitiveness of new energy automobile industry based on grey correlation model," *Mathematics in Practice and Theory*, vol. 46, no. 21, pp. 72–79, 2016.

[32] H. Zhang and Y. Zhang, "Evaluation of tourism industry competitiveness in western Hunan based on structural equation model," *Journal of Jishou University*, vol. 35, no. S2, pp. 32–35, 2014.

[33] H. Zhou, "An empirical analysis of the influence of industrial agglomeration on the competitiveness of Zhejiang manufacturing industry," *Productivity Study*, no. 9, pp. 132–134, 2013.

[34] L. Chen and Y. Liang, "Research on the relevance between the degree of mineral industrial agglomeration and industrial competitiveness in China's western region," *China Population, Resources and Environment*, vol. 21, no. 5, pp. 31–37, 2011.

[35] T. Hui and X. Yang, "Research on the level of specialization competitiveness of cultural industry based on location entropy measurement," *Gansu Finance*, no. 8, pp. 63–67, 2019.

[36] C. Liu, L. Han, and Y. Zhang, "International comparison and development suggestions of China's dairy industry competitiveness," *Chinese Rural Economy*, no. 7, pp. 130–144, 2018.

[37] X. Cheng, "Study on the influence of Chinese dairy industry agglomeration on industrial competitiveness," *Economic Forum*, no. 4, pp. 95–98, 2018.

[38] Z. Ma and Y. Yang, "Competitiveness evaluation and strategy selection of dairy industry in China," *Reformation and Strategy*, vol. 33, no. 8, pp. 158–162, 2017.

[39] Y. Bian, "Analysis on international competitiveness of Chinese dairy industry," *Journal of Hebei University of Economics and Business (Comprehensive Edition)*, vol. 16, no. 4, pp. 91–95, 2016.

[40] Y. Zhang, A. Ren, and J. Zhao, "Analysis on milk powder international competitiveness based on diamond model," *Chinese Journal of Agricultural Resources and Regional Planning*, vol. 37, no. 2, pp. 84–90, 2016.

[41] Y. Han and X. Cheng, "An empirical study on regional competitiveness of China's dairy industry-a dynamic analysis based on multidimensional panel data," *Forward Position*, no. 8, pp. 63–66, 2015.

[42] S. Seyiti and F. Deng, "Evaluation of regional dairy industry competitiveness based on AHP," *Statistics and Decision*, no. 14, pp. 67–69, 2013.

[43] J. H. Friedman and W. Stuetzle, "Projection pursuit regression," *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 817–823, 1981.

[44] Y. Gong, Y. Zhang, F. Ding, J. Hao, H. Wang, and D. Zhang, "Projection pursuit model for assessment of groundwater quality based on firefly algorithm," *Journal of China University of Mining & Technology*, vol. 44, no. 3, pp. 566–572, 2015.

[45] M. Porter, "Competitive Strategy Techniques for Analyzing Industries and Competitors," The Free Press, New York, NY, USA, 1980.

[46] D.-S. Cho and H. Chang Moon, "A nation's international competitiveness in different stages of economic development," *Advances in Competitiveness Research*, vol. 6, pp. 5–19, 1998.

[47] J. H. M. Wijnands, H. J. Bremmers, B. M. J. Meulen, and K. J. van der Poppe, *Competitiveness and Innovativeness of the EU Dairy Industry*, Wageningen Academic Publishers, Wageningen, Netherlands, 2010.

[48] M. Nica, "Analysis of competitiveness on the market of milk and dairy products in Romania," in *Proceedings of the Agrarian Economy and Rural Development—Realities and Perspectives for Romania*, Bucharest, Romania, November 2017.

[49] O. Ramphul, "Global competitiveness in dairy sector," *Social Science Electronic Publishing*, vol. 59, no. 5, pp. 257–264, 2016.

[50] Q. Xu, *Analysis on the Regional Competitiveness of China's Dairy Industry*, Southwestern University of Finance and Economics, Chengdu,China, 2014.

[51] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. C-23, no. 9, pp. 881–890, 1974.

[52] Q. Cheng, "Structure entropy weight method to confirm the weight of evaluating index," *System Engineering Theory and Practice*, vol. 30, no. 7, pp. 1225–1228, 2010.

Research Article

A Seismic Hazard Prediction System for Urban Buildings Based on Time-History Analysis

Yongmei Zhai and Shenglong Chen [ORCID]

Shanghai Institute of Disaster Prevention and Relief, Tongji University, Shanghai 200092, China

Correspondence should be addressed to Shenglong Chen; zensenlon@yeah.net

Seismic hazard prediction is essential for earthquake preparation in urban areas. Prediction methods based on refined the finite element method (FEM) models and time-history analysis require higher calculation, which is hard to apply universally. In this paper, by applying two simplified models and time-history analysis, the authors were able to develop a refined seismic hazard prediction system for urban buildings using MATLAB and OpenSees. The system is modular in design with a unified data interface to transfer data, where only the macro parameters of buildings are required to complete the computation work. Taking the campus of Tongji University as a test area, the system was applied to predict seismic hazard and compare it with the results of the vulnerability method by applying the seismic damage index to verify its reliability. The results show that the system can meet the needs of accuracy, efficiency, and visualization of seismic hazard prediction in urban areas.

1. Introduction

With the rapid development of the economy and expansion of urbanization in China, along with increased population density and urban building, a sudden unforeseen earthquake would undoubtedly magnify the resulting hazard [1]. In order to improve overall seismic capacity and minimize seismic hazards in urban areas, it is of considerable significance to analyze and study every weakness in earthquake resistance structures and formulate corresponding disaster prevention plans.

Seismic hazard prediction for buildings is an essential work for urban earthquake mitigation [2]. It refers to the quantity distribution of different damage states in a specific area (sample space) under the given seismic intensity; this distribution, called the seismic hazard matrix, is obtained through the calculation and statistics of typical samples [3]. The purpose is to examine the global seismic hazard level of area and the difference of seismic performance among various buildings. Due to the destructive nature of an earthquake, it is impossible to conduct physical experiments during the event; thus, numerical

simulations have become an essential means of seismic hazard prediction. The most widely used methods are based on a damage vulnerability matrix [4] and the capacity spectrum [5, 6].

The basic principle of the damage vulnerability matrix is to select various factors that affect the seismic performance of buildings based on historical seismic data and then obtain the damage vulnerability index of the target area according to the actual situation of each factor [7]. This method has been widely applied due to its simplicity and usability [8–10]. However, the method ignores the duration and the spectral characteristics of ground motions and cannot reflect the damage state of specific buildings [11]. The method based on the capacity spectrum, also known as pushover analysis, uses the earthquake response spectrum as an input to investigate the vulnerability of specific buildings [12]. The excess probability of different damage states is calculated by the intersection of the capacity curve and the demand spectrum (performance point), and then the seismic vulnerability curve expressed by the spectral displacement is constructed to determine the seismic vulnerability. Based on the mature theory and detailed

structural parameters, many countries and regions, including America, Canada, China, Israel, and others, have developed seismic damage prediction systems based on the capacity spectrum [13–17]. However, this method also has limitations. It is challenging to consider the influence of high-order mode shape on the seismic response and the time-domain characteristics of ground motions, such as the velocity pulse.

In order to avoid these disadvantages, with the in-depth study of performance-based seismic theory [18], the following two improvements have emerged: (1) using a refined model (such as the finite element model, FEM, and discrete element model, DEM) to simulate the structure and (2) applying time-history analysis. Compared with the previous methods, the property of materials and the characteristics of ground motion can be considered more reasonable. Due to the higher accuracy in the theory, there was an extensive application on seismic hazard prediction for a single building. Castaldo et al. [19] performed incremental dynamic analysis (IDA) to evaluate the seismic fragility of a nonlinear hardening and softening structure equipped with friction pendulum isolators. Hajipour et al. [20] constructed FEM models for seven steel buildings with RC shear wall in OpenSees considering the nonlinear geometric effects of materials and employed time-history analysis to attain the seismic performance. Recently, with the rapid improvement of the computational ability of computers, scholars have begun to conduct time-history analysis with refined models in regional seismic hazard prediction gradually. Mahin et al. [21] applied the computational modeling and simulation center (Simcenter) to simulate the entire seismic process and evaluate the seismic loss for over 1.8 million buildings in the San Francisco Bay Area. The Institute of Earthquake Engineering (ERI) [22] developed an integrated earthquake simulation (IES) system and predicted the seismic hazard of Tokyo based on time-history analysis with the fiber beam models and DEM models. Abdurrahman et al. [23] developed a new version in MATLAB, including site response analysis and structural analysis, and applied the system in Istanbul. On the other hand, due to a large number of buildings in urban areas, such systems require extensive computation, although with the emergence of high-performance computing methods such as parallel computation [24], to meet the timeliness of prediction, it needs to be conducted by supercomputers. Moreover, it requires detailed information of buildings in component level to establish the refined models, which takes lots of manpower and time, and it is hard to collect all the data needed for thousands of buildings in the urban area.

Therefore, based on the above background, the authors of this study developed a refined seismic prediction system suitable for urban buildings. The system uses the Geographic Information System (GIS) to acquire and manage urban building information and visualize prediction results. For different buildings, appropriate simplified models are applied for time-history analysis, which dramatically improves computational efficiency. Based on

OpenSees and MATLAB, the automation of hazard prediction and data processing is realized, which reduces manual intervention. Finally, in order to verify the reliability of the system, the campus of Tongji University, Shanghai, was selected as the test area for this study's new seismic prediction system. It was due to the university's urban setting and the authors' affiliation with the Shanghai Institute of Disaster Prevention and Relief located on campus. A comparison between prediction results and results of the vulnerability method showing vulnerability indicates that the system can realize rapid, refined, and visualized seismic hazard prediction in urban areas and can support relevant departments and organizations in disaster prevention planning and emergency rescue decision-making.

Specifically, in Section 2, the frame of the proposed methodology is presented, and the simplified models are introduced briefly. In Section 3, the technical framework and data flow of each module in the prediction system are introduced. Finally, a regional seismic hazard prediction was performed for the test area by applying the proposed system as a case study, and the validation of the system is discussed in Section 4.

2. Methodology

2.1. Methodology Framework. There are three main steps for the seismic hazard prediction based on time-history analysis, as illustrated in Figure 1.

Step 1. Data collection.

It is the essential work of seismic hazard prediction. The most effective way to obtain and manage the building data is to apply the Geographic Information System (GIS). Numerous cities have already implemented their GIS databases, where the macro building parameters (such as height, area, and construction time) could be directly acquired. Otherwise, data could be supplemented by field investigations and CAD drawings.

Step 2. Seismic response analysis.

After obtaining the macro building parameters, the calculation model of buildings in the target area can be established. Two simplified models are utilized in the paper in accordance with building characteristics, which will be introduced in Section 2.2. Subsequently, a nonlinear time-history analysis is conducted to obtain the seismic response of each building under earthquake scenarios in different return periods.

Step 3. Seismic hazard assessment.

Following the seismic response analysis, the limit of each damage state is firstly calculated. Based on the damage limit, the probability of different damage states of each building is determined according to the seismic response of each story. Finally, the prediction results of the target area are input into the GIS database and visualized to illustrate the seismic hazard intuitively for those who lack relevant knowledge.
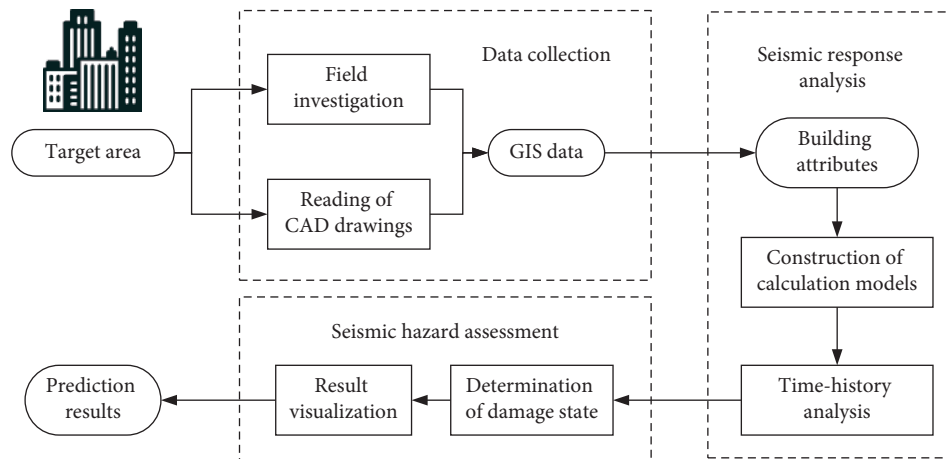
Figure 1: Method frame.

2.2. Calculation Model.

2.2. Calculation Model. As mentioned in the Introduction, in order to balance calculation accuracy and efficiency, two multiple degrees of freedom (MDOF) models are applied in the system, the parameters of which can be determined by macro building attributes.

For a large number of regular multistory buildings in the urban area, mainly including the RC frame and masonry structure, the deformation characteristic under the lateral load mainly consists of shear deformation. Therefore, a building can be simplified into a nonlinear MDOF shear (NMS) model [25], as illustrated in Figure 2(a). Compared with the single-degree-of-freedom model, the influence of the high-order mode shape on multistory buildings can be considered with the NMS model. The models assume a rigid story and that the mass of each story is concentrated so that each story can be simplified into a mass point, which are connected by a shear spring between them. Notably, the skeleton line of the spring adopts the trilinear skeleton line recommended in the HAZUS report [26], as illustrated in Figure 2(c). Previous studies [27, 28] have demonstrated that the shear spring of the NMS model could reflect the elastoplastic characteristics of the structure, which can simulate the seismic performance of multistory buildings.

For high-rise buildings, where the effects of bending deformation are especially critical, a nonlinear MDOF flexural-shear coupling (NMFS) model was chosen [29], as illustrated in Figure 2(b). The NMFS model discretizes each story into a nonlinear bending spring and a shear spring. The bending spring is utilized to simulate the bending deformation of the shear wall, while the shear spring is used to simulate the shear deformation of the frame. Two springs are connected with a rigid link to ensure horizontal displacement coordination. Similar to the NMS model, the trilinear skeleton line is selected as the skeleton line of springs. In order to verify the accuracy of the simplified models, an NMFS model, an NMS model, and a refined FEM model of a 13-story reinforced concrete frame-shear structure were established and time-history analysis was performed. As shown in Figure 2(d), the

result of the NMFS model is in good agreement with that of the FEM model, but the result of the NMS model is far from the above two. It is because the NMS model cannot consider bending deformation. Therefore, the NMFS model is suitable for the seismic hazard prediction of high-rise buildings.

The particular calibration methods of the two models can be found in [25, 29].

3. Function Realization

Taking OpenSees 2.5 and MATLAB 2017a as the primary development platforms, the system was divided into three modules: (1) data preprocessing module, (2) structural computing module, and (3) postprocessing module, as shown in Figure 3. The three modules are independent of each other and use a unified data interface to transfer data through files. Adopting such an overall framework has the following advantages: (1) different modules are not subject to the limits of other modules, which facilitates function expansion and upgrade; (2) a unified data interface between different modules ensures consistency of data transfer between modules and facilitates data viewing and processing; and (3) using files for data transfer reduces system memory consumption and improves computational efficiency.

3.1. Data Preprocessing Module. The function of the data preprocessing module is to obtain building attributes from GIS data and calculate parameters of finite element models. The module is mainly composed of three submodules: GIS data-calling, model parameter generation, and data output, as shown in Figure 4.

The GIS data-calling submodule reads the building attributes from GIS data and stores them in local variables, which are called up by MATLAB later. An .xlsx file is used as the exported data format. On the one hand, the .xlsx format is highly compatible and easy to read and has less information loss during the conversion. On the other hand, some fields are often not needed for data preprocessing in GIS
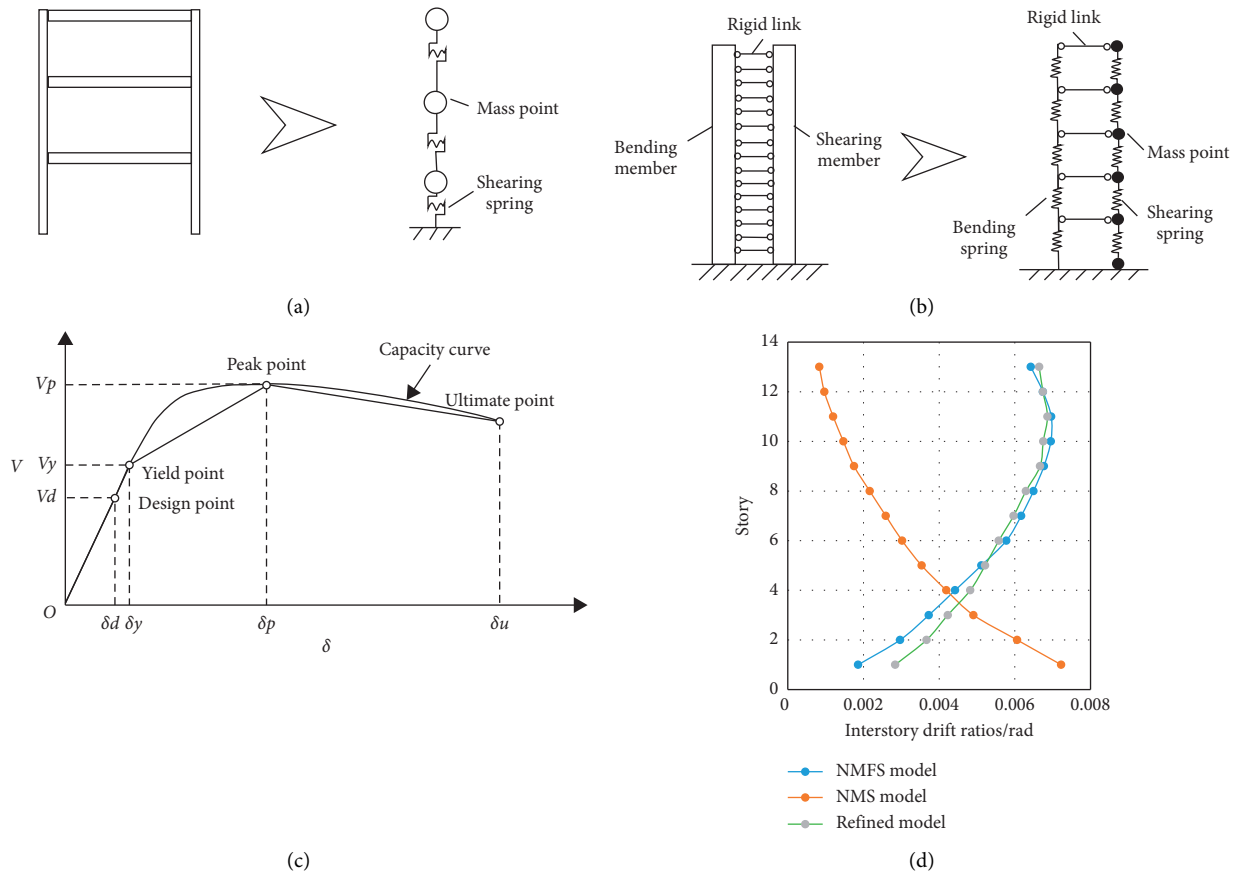
(a)

(b)

(c)

(d)

Figure 2: Illustration and validation of simplified models. (a) Nonlinear MDOF shear (NMS) model. (b) Nonlinear MDOF flexural-shear coupling (NMFS) model. (c) Trilinear skeleton line. (d) Envelope curve of interstory drift ratios of three models.
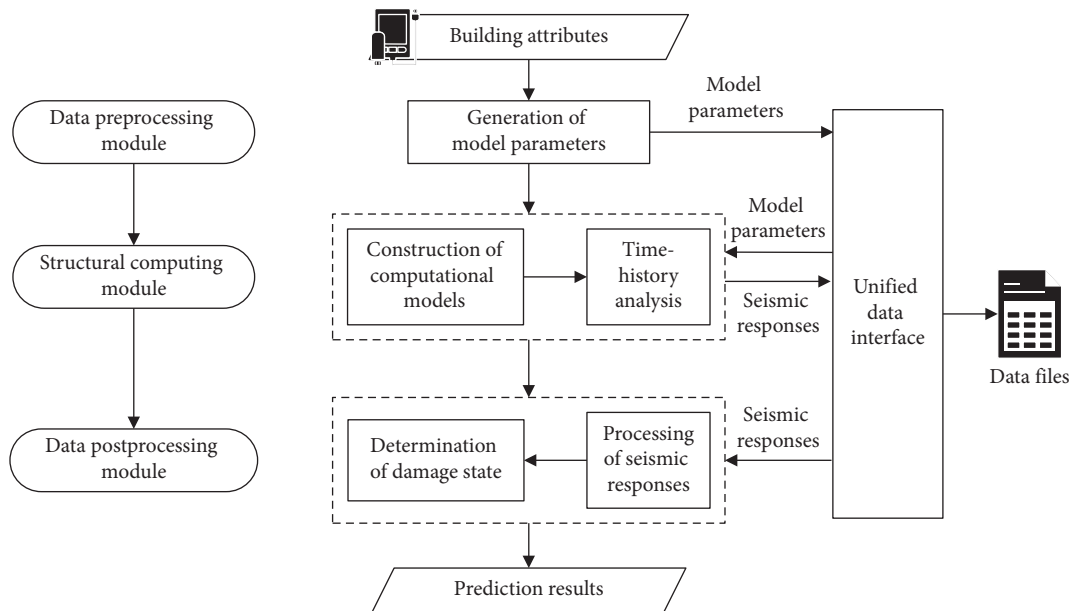

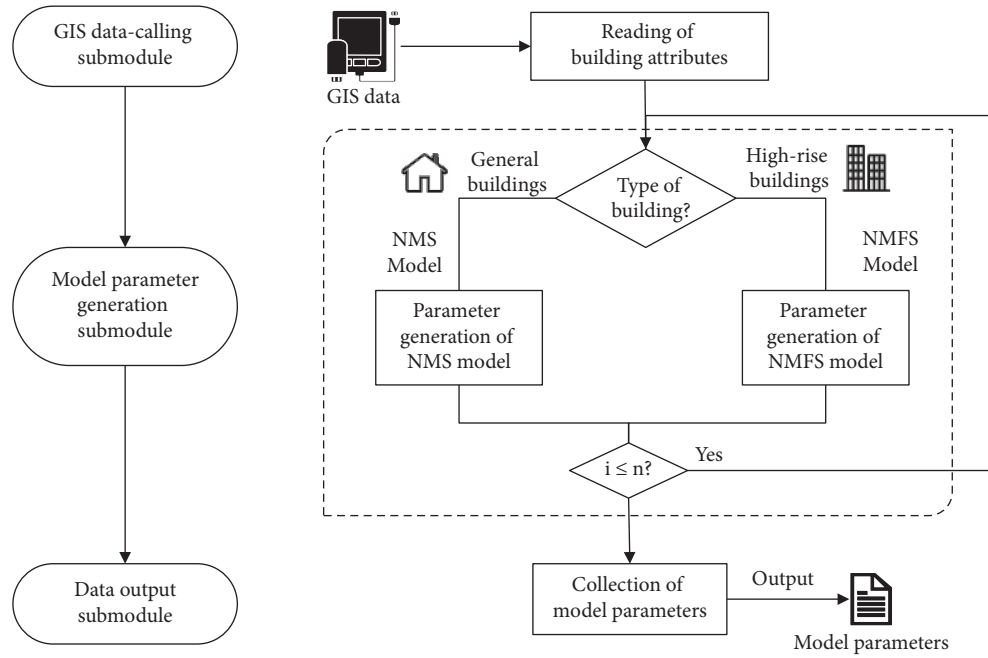Figure 3: Overall framework and data flow of the system.

FIGURE 4: Framework of data preprocessing module.

attribute tables. The use of Excel also facilitates the modification and management of building attributes. Selected .xlsx files were checked and modified. They generally contain fields such as building code, construction time, and structure type.

The primary function of the model parameter generation submodule is to calculate the resilience parameters required to calibrate the MDOF model. It is composed of several different functions, where the NMS and NMFS functions are used to calibrate the NMS and NMFS models. The NMS function first calculates the shear stiffness of the model and then chooses concrete or masonry function to calculate the resilience parameter according to the structure type. The mainframe of the concrete/masonry function is a loop command that determines the parameters of each story. In the loop body, the base-shear function is first called to calculate the design bearing capacity through the base-shear method. Then, the yield/peak/ultimate function is called to calculate the yield/peak/ultimate capacity and displacement of the skeleton line. Finally, the degradation function is called to determine the hysteresis parameter by the lookup table method [30]. The particular calibration methods of NMS model can be found in [25]. The data flow of the NMS function is shown in Figure 5.

The framework of the NMFS function is more complicated. The elastic parameters of the model should be determined first, which is done by the elastic function. The shear stiffness GA and bending stiffness EI can be calculated with formulas (1) and (2). Since these are multivariate nonlinear equations, the Newton iteration method can be used to find their approximate solutions:

$$\frac{T_i}{T_1} = \frac{\gamma_1}{\gamma_i} \sqrt{\frac{\gamma_1^2 + \alpha_0^2}{\gamma_i^2 + \alpha_0^2}}, \tag{1}$$

$$2 + \left[2 + \frac{\alpha_0^4}{\gamma_i^2 \left(\gamma_i^2 + \alpha_0^2\right)}\right] \cos\left(\gamma_i\right) \cosh\left(\sqrt{\gamma_i^2 + \alpha_0^2}\right)$$
$$+ \frac{\alpha_0^2}{\gamma_i \sqrt{\gamma_i^2 + \alpha_0^2}} \sin\left(\gamma_i\right) \sinh\left(\sqrt{\gamma_i^2 + \alpha_0^2}\right) = 0. \tag{2}$$

where T_i is the i order period of the structure and γ_i is the eigenvalue coefficient related to the i order of vibration, which is the function of bending-shear stiffness ratio α_0. The definition of α_0 is as follows, where H is the story height of the model:

$$\alpha_0 = H\left(\frac{GA}{EI}\right)^{(1/2)}. \tag{3}$$

It should be noted that the mode decomposition method is applied to calculate the yield parameters of models. Therefore, first, the "! OpenSees.exe modal-analysis.tcl" command in the NMFS function is run to call OpenSees to perform modal analysis. After obtaining the model shape and period of models, the design bearing capacity can be calculated through the mode-superposition response spectrum method. Finally, the parameters of the skeleton line and degradation are determined. The particular calibration methods of NMFS model can be found in [29]. The data flow of the NMFS function is shown in Figure 6.

The data output submodule summarizes the calculated resilience parameter matrix, generates separate submatrices for each building, and saves them in files for the structural calculation module to read.
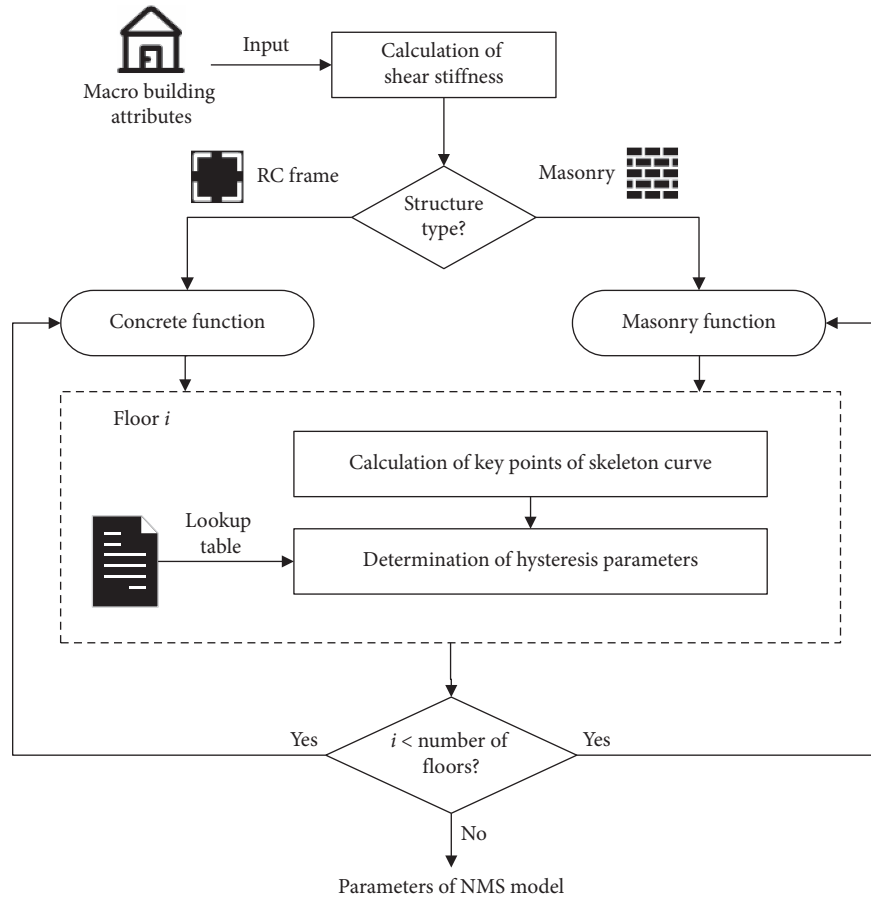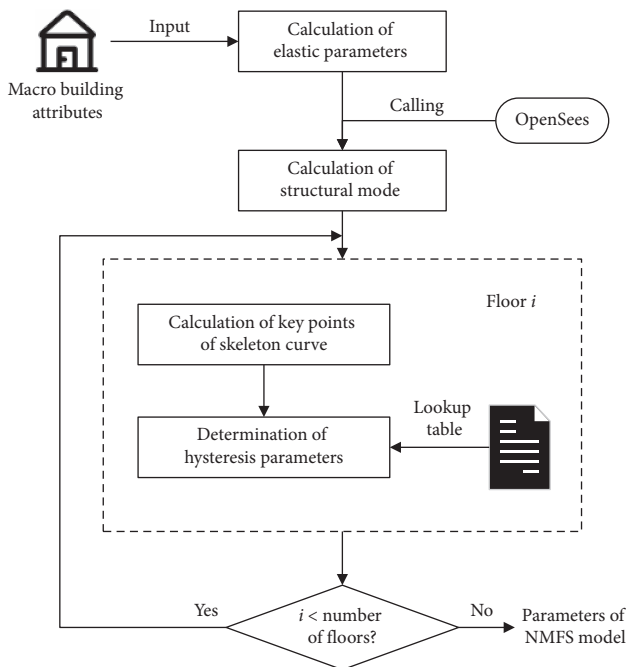
Figure 5: Data flow of NMS function.

Figure 6: Data flow of the NMFS model.

3.2. Structural Calculation Module. The structural calculation module based on OpenSees uses nonlinear time-history analysis to calculate the seismic response of the NMS or NMFS model under different seismic intensities. The module is the core calculation and analysis part of the system, including the following four submodules:

(1) The information reading submodule reads the model parameters of each building and ground motion data generated by the data preprocessing module through files according to the number of buildings. After that, parameters are stored in the matrices and then assigned to global variables by the "lassign" command.

(2) The modeling submodule retrieves model parameters from memory; generates the nodes, elements, and boundary conditions of the calculation model according to the modeling rules; and defines the mass, stiffness matrix, and material recovery force characteristics (skeleton line and hysteresis model) of each node. Since the calculation model assumes that the stiffness and mass of each story are evenly distributed along with the height, determining the loop body controls each parameter in the modeling submodule.

(3) The analysis submodule defines the static analysis conditions of the self-heavy load and the time-history analysis conditions of different ground motion periods and determines the selection of the degree of freedom in the calculation of the constraint function, iterative algorithm, control algorithm, and system equation. The dynamic analysis uses the Newmark implicit integration method. Each seismic wave is analyzed in 5000 steps, and each step is 0.01 s.

(4) The recording submodule records the seismic response of structural models, such as node displacement, inter-story displacement angle, and mode shapes of each order, and saves it as a file through the unified data interface.

For the calculation of a single model, the above four modules are connected in a series. When the calculation of a single model is completed, the entire memory storing the model variables and temporary data is released, and the information reading submodule is returned to the next model until all calculations are completed. The framework of the structural calculation module is shown in Figure 7.

In the modeling submodule, both of the simplified models apply the uniaxial hysteretic material to simulate the inter-story resilience characteristics of the structure and the two-node link element to connect the concentrated mass nodes of each story. Uniaxial hysteretic materials are often used to construct constitutive models of uniaxial trilinear hysteresis materials. The inter-story force-displacement relationships and hysteresis parameters of the hysteretic material are defined in Figure 8 [31]. The two-node link element is a widely used joint element, which can directly use the uniaxial stress-strain material relationship model as the force-displacement relationship model of the structure [32]. However, this element has many features that are difficult to grasp. Two points require special attention: (1) for a non-zero length element, the local x-axis points are from node i to node j, and (2) the local degrees of freedom 2 and 3 (shear and flexural) of the two-node link elements are coupled, and the shear distance controls the stiffness matrix. The shear distance is defined as the ratio of the distance between the shear center point and the node i to the unit length. In the default state, the shear distance is defined as the middle of the element, as shown in Figure 9. For the meanings of the variables in Figures 8 and 9, refer to [31, 32].

3.3. Data Postprocessing Module.

The primary function of the data postprocessing module is to process the enormous seismic response data generated by the structural calculation module and determine the seismic damage state of the structure according to the calculated damage limit. The module consists of three submodules: response processing, damage limit calculation, and seismic damage result. The framework of the module is shown in Figure 10.

The response processing submodule first reads the seismic response of each building and then calculates the response envelope of each story under all seismic conditions. The response parameter required for damage discrimination is divided into two types: (1) interstory displacement angle $\Delta\delta_i$ and (2) interstory slope $\Delta\theta_i$. The recorder command in OpenSees can directly record the interstory displacement angle, and the interstory slope needs to be calculated from the displacement parameters of each story, as shown in formulas (4) and (5):

$$\theta_i = \frac{u_i}{h_i}, \tag{4}$$

$$\Delta\theta_i = \theta_i - \theta_{i-1}, \tag{5}$$

where u_i is the displacement, h_i is the height, and θ_i is the slope of story i (when $i = 1$, $\theta_{i-1} = 0$).

The damage limit calculation submodule needs to call different functions according to structural types and calculation models when determining the damage limit of each story. Among them, the NMS RC and NMS M functions are used to calculate the damage limit of the RC frame structure and multistory masonry structure in the NMS model, respectively. The NMFS RC function is used to calculate the damage limit of the RC frame shear/cylinder structure in the NMFS model. Two main methods are used to determine the damage limit through force-based and deformation-based failure criteria [33]. One is to directly read the prewritten file and look up the table according to structure types to obtain the limit of displacement. The second is to call the key points of the building capacity curve calculated by the data preprocessing module from memory. The data flow of the damage limit calculation submodule is shown in Figure 11.

The seismic damage submodule determines the damage state of each story according to the response envelope and damage limit and calculates the regional seismic hazard. This study applies the seismic hazard matrix proposed by Yin to characterize the seismic performance level of an area [34]. The seismic damage matrix describes the probability of a certain level of damage occurring from an earthquake, which can be expressed as

$$P[D_j] = \sum P[D_j \mid I]P[I], \tag{6}$$

where $P[D_j \mid I]$ indicates the probability of level j damage when an earthquake of intensity i is experienced by a building in a region and $P[I]$ indicates the probability of an earthquake of intensity i occurring over a while.

4. Case Study

4.1. Overview of the Test Area.

To verify the reliability of the system, the authors of this study selected the campus of Tongji University, Shanghai, which has a complete computer-aided design (CAD) planning map and GIS data for the sections of the test area, to conduct seismic hazard prediction. Missing or insufficient data of building materials were obtained through field surveys and by
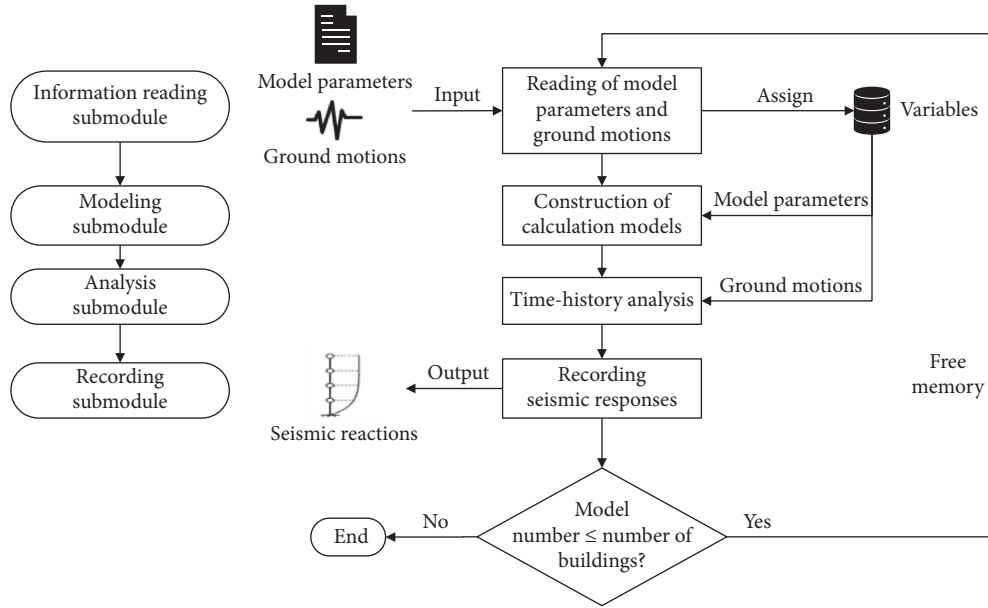
FIGURE 7: Framework of structure calculation module.
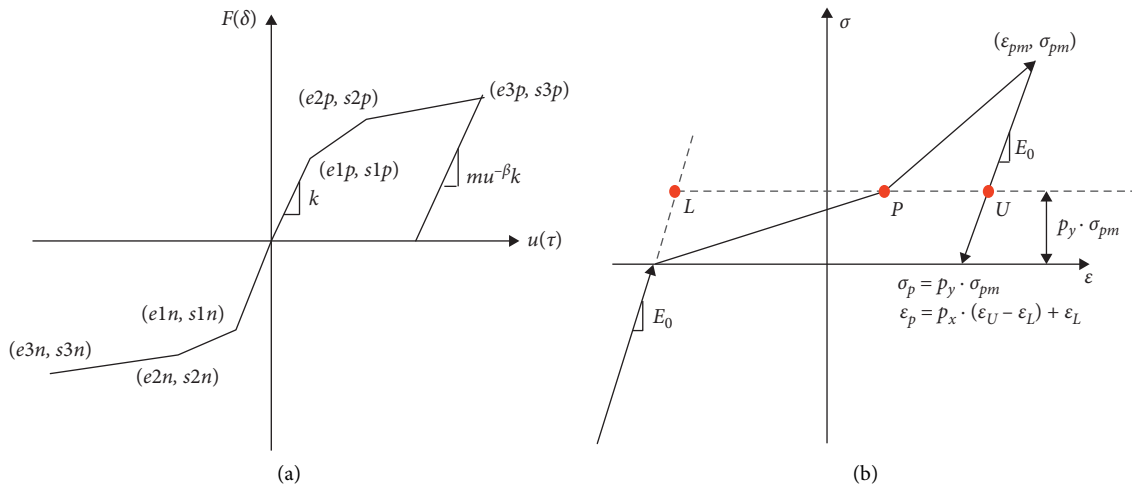

(a)

(b)

FIGURE 8: Illustration of hysteresis parameters of hysteretic material.

reviewing building design drawings in the university archives. The collected building data were stored and managed using ArcGIS 10.4. Through GIS data and field surveys, information on a total of 421 buildings was obtained. There is a variety of building types in the test area, most of which are typical multistory masonry houses in urban areas, RC frame office buildings, and RC frame shear/cylinder high-rise buildings, as shown in Figure 12.

The test area has experienced several developments and transformations throughout its history. The construction times of different buildings are diverse, and the distribution of structure types varies from generation to generation. The system developed in this study considers the influence of when the construction took place. The time of construction

determines the design specifications adopted by the structure, which affects the value of the bearing capacity and the damage limit of the model. The distribution of construction time for different structural types in the test area is shown in Figure 13. This figure shows many of RC structures built since the 1980s, while the proportion of masonry structures gradually decreased.

4.2. Selection of Ground Motion.
Collecting ground motion data is a prerequisite for regional seismic hazard prediction. Due to the lack of detailed geological data on the test area, the ground motion parameter plot of Shanghai was used as the basis of ground motion data, as shown in Figure 14 [35]. The response spectrum applied for the test
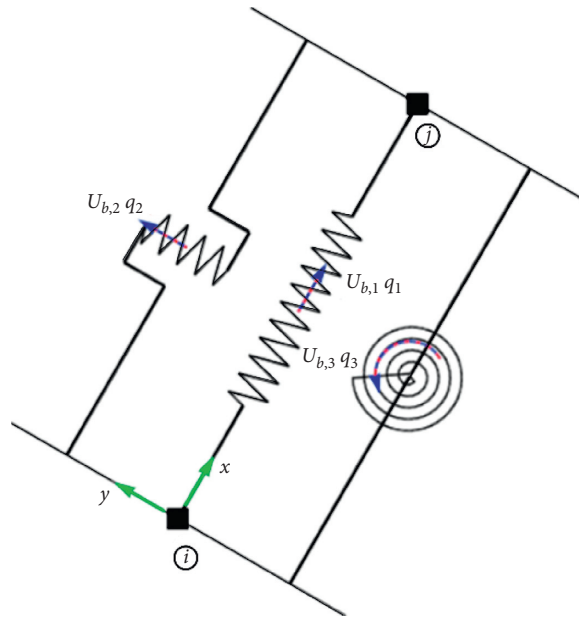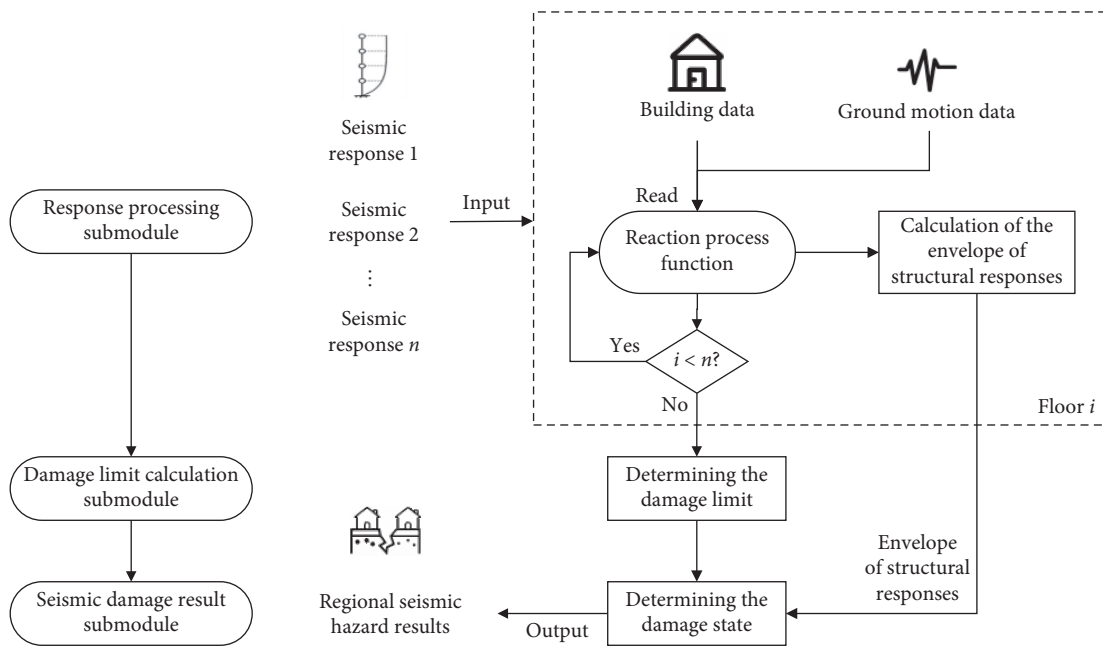
Figure 9: Illustration of the two-node link element.

Figure 10: Framework of data postprocessing module.oor.

area is shown in Figure 15. For the meanings of variables, refer to [36]. According to the contour map of the surface peak acceleration provided by the plot and the regulations for seismic design of Shanghai buildings (J10284-2013) [36], the highest surface peak acceleration in the test area with a 50-year probability of surpassing 63%, 10%, and 3% was scaled to peak ground-level intensities of 35 gal, 100 gal, and 200 gal; the site feature cycle T_g takes 0.9 s, 0.9 s, and 1.1 s; and the maximum earthquake

influence coefficient $\alpha_{\max}$ takes values of 0.09, 0.23, and 0.45.

The predominant period of the input ground motion should align as closely as possible with the characteristic period of the test area, and the response spectrum should also align well with the design response spectrum within a specified period. According to the design response spectrum and considering the site conditions in Shanghai, 28 ground motions (14 for frequency and moderate earthquakes and 14
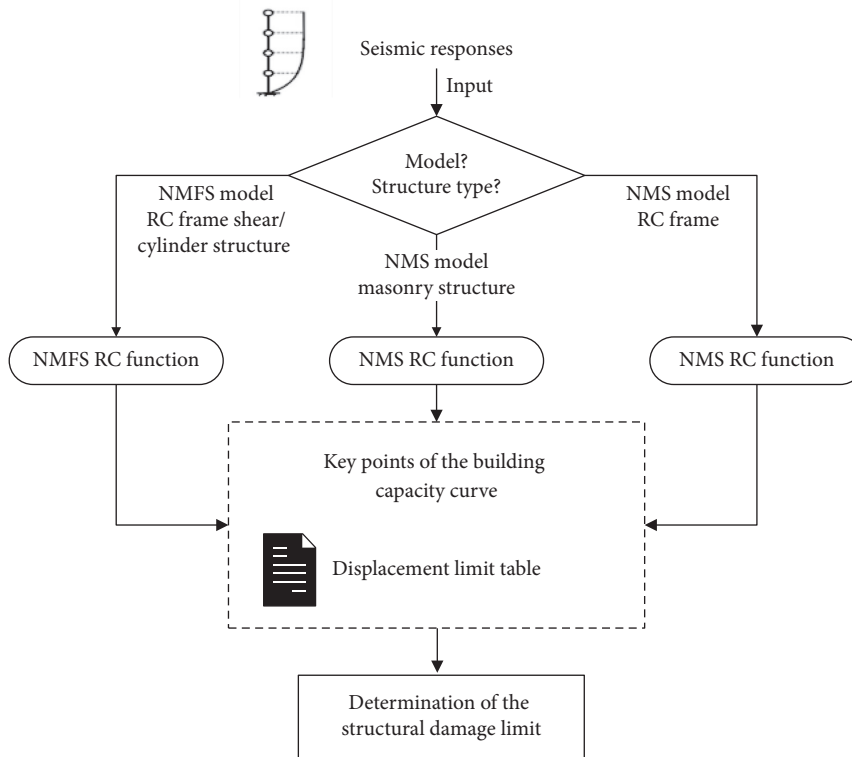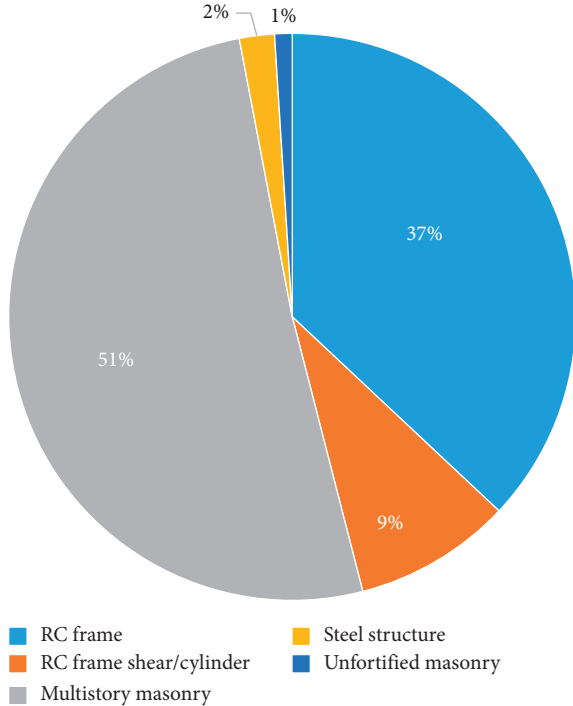
FIGURE 11: Data flow of damage limit submodule.

FIGURE 12: Distribution of building structure types in the test area. RC, reinforced concrete.

FIGURE 13: Construction time distribution of different structural types in the test area.

for rare earthquakes) were selected from the strong earthquake database of the Pacific Earthquake Engineering Research Center (PEER) [37] and the KiK-net Digital Strong
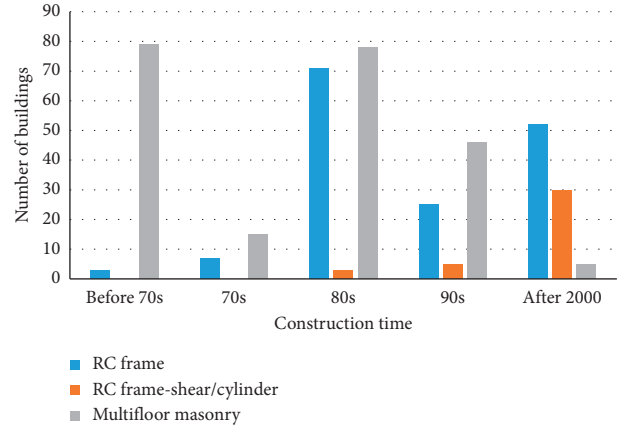
Earthquake Recording System of the National Research Institute for Earth Science and Disaster Resilience (NIED) [38]. A comparison of the average response spectrum for the 28 ground motions and the recorded response spectra for this project is shown in Figure 16. The average response spectrum of the selected ground motions is in good agreement within the period 0 to 6 s. Since the natural period of most buildings in the test area was within 6 s, the 28 ground motions were selected to conduct the elastoplastic time-history analysis to obtain the seismic response and damage state of each building.
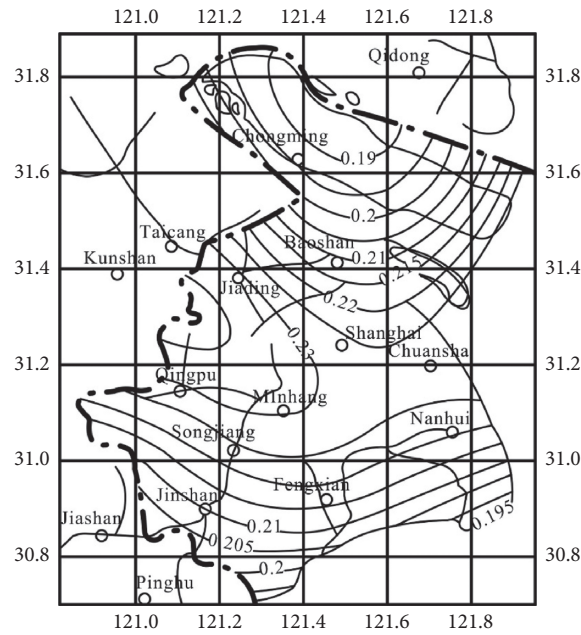
FIGURE 14: Ground motion parameter plot of Shanghai.
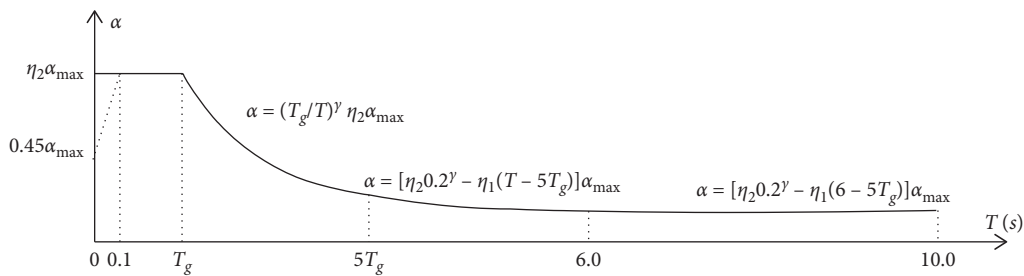

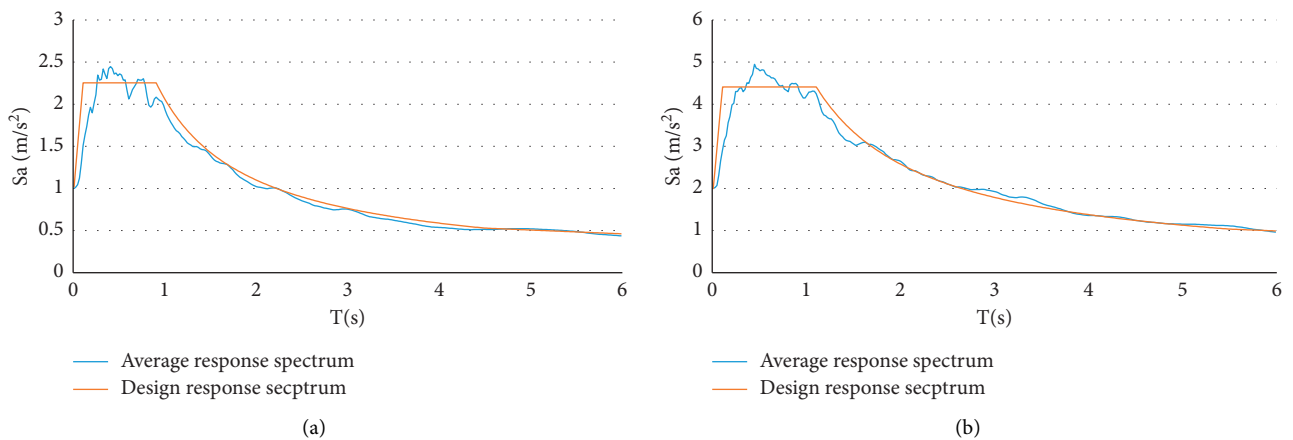FIGURE 15: Response spectrum of the test area.

FIGURE 16: Comparison between average response and design response spectrum. (a) Moderate earthquake. (b) Rare earthquakes.

4.3. Prediction Results.

By using the system developed in this study, 421 buildings in the test area were investigated based on seismic hazard prediction under frequent earthquakes (intensity VI 0.05 g), moderate earthquakes (intensity VII 0.1 g), and rare earthquakes (intensity VIII, 0.2 g). A total of 17,682 calculations with a laptop (i7 7700 HQ, 32 g memory)
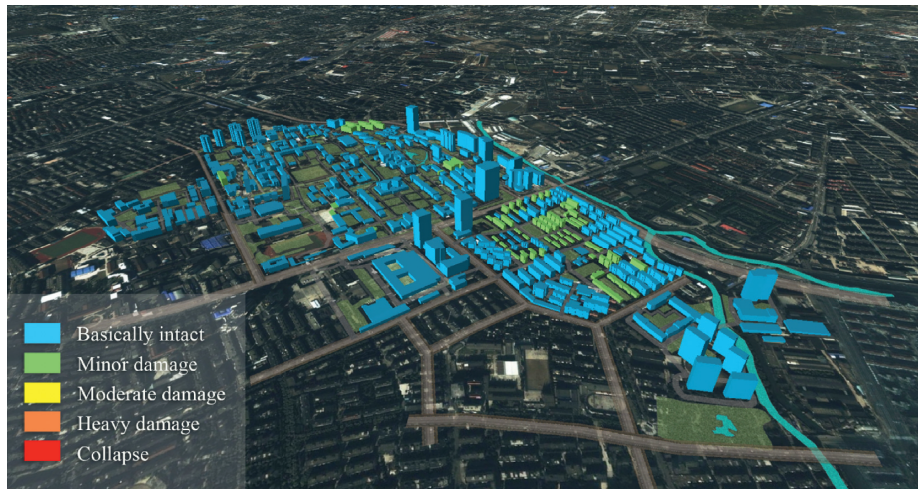
(a)
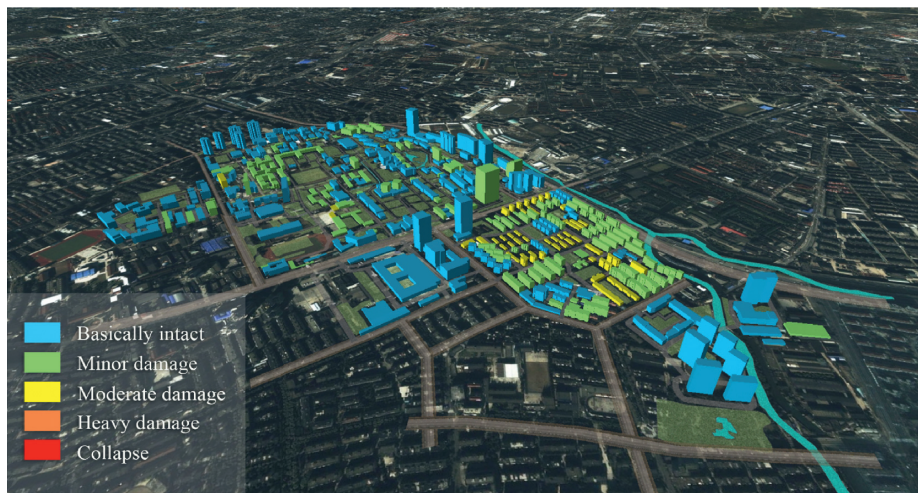
(b)

(c)

FIGURE 17: Seismic hazard prediction results of the test area under different earthquake levels. (a) Frequent earthquake. (b) Moderate earthquake. (c) Rare earthquake.
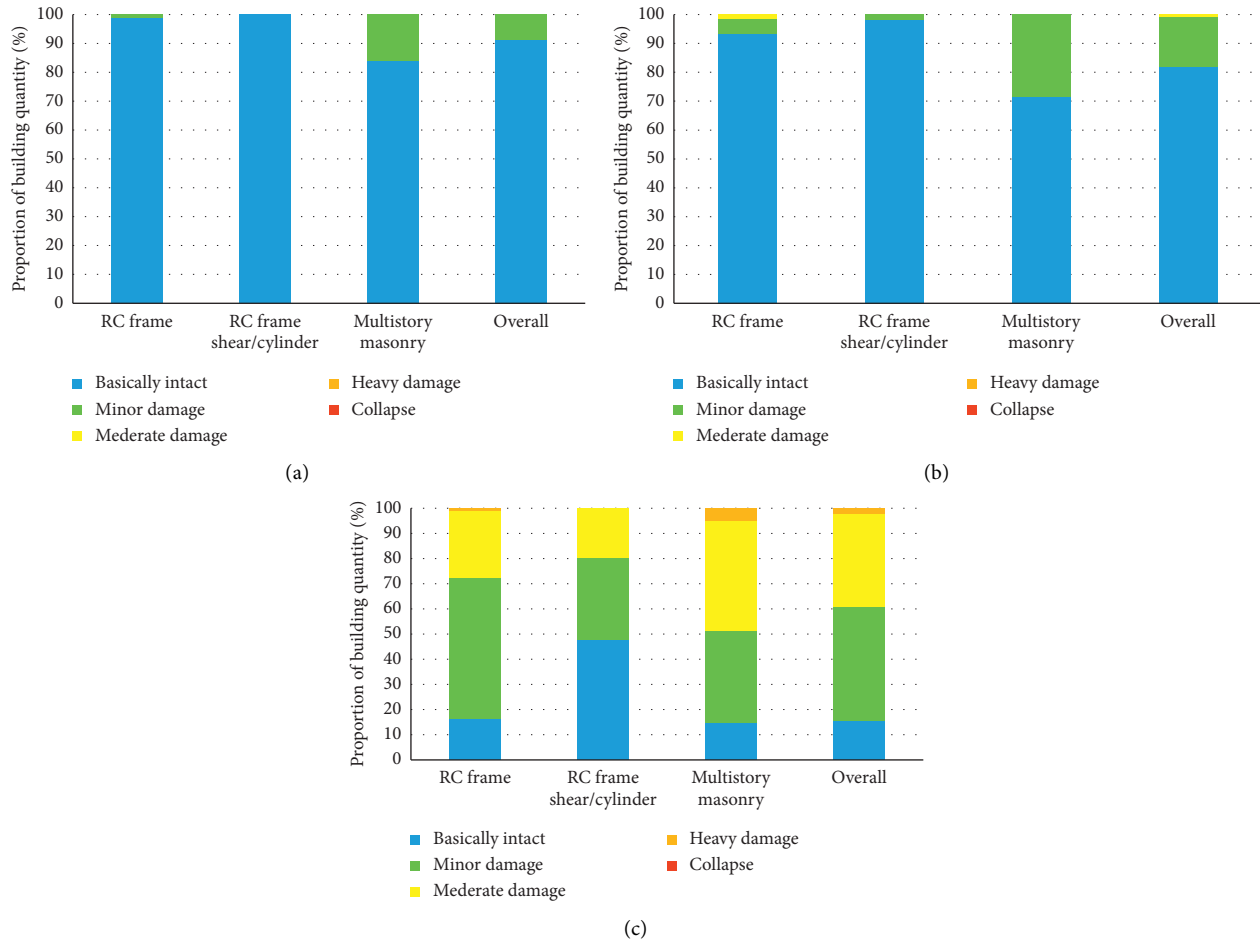
(a)

(b)

(c)

Figure 18: Damage state of different structural types under different earthquake levels. (a) Frequent earthquake. (b) Moderate earthquake. (c) Rare earthquake.

took about 58 minutes. The seismic hazard prediction results of the test area under different earthquake levels were visualized on the ArcScene module in ArcGIS 10.4, as shown in Figure 17.

The damage states of different structural types are shown in Figure 18. From the prediction results, the seismic capacity of various structural types in the test area was determined to be uneven due to construction time, design level, and structural problems. The seismic capacity of the RC frame and RC frame-shear/cylinder structure with a seismic design is the best. Most buildings were intact under moderate earthquake prediction evaluation and moderately damaged under rare earthquake prediction evaluation. Because the construction time of multistory masonry buildings represents the oldest buildings and their seismic capacity is slightly worse than that of RC structures, about half of the masonry buildings showed moderate damage under rare earthquakes. However, in general, the damage assessment of buildings in the test area met the target levels of the three-level seismic design.

4.4. Discussion. In order to verify the reliability of the prediction results in this system, the results were compared with results based on the vulnerability matrix [7]. Comparison data were selected from the disaster prevention planning report on Weifang Street in the Pudong New Area of Shanghai. Ouyang et al. [39] conducted a sample survey of the target area and applied a semi-experiential and semi-theoretical method to obtain the seismic damage matrices of different structural types in Weifang Street under different construction years. Due to the sample of buildings, this section only compares the results of a reinforced concrete frame and multistory masonry structures, as shown in Figures 19 and 20. In figures, THA means the results based on time-history analysis, and VM means the results based on a vulnerability matrix.

To facilitate the comparative analysis, the seismic damage index was used to indicate the seismic performance of buildings of different ages assessed under the vulnerability matrix used in [39] and the time-history analysis method used in this paper. The seismic damage index is a
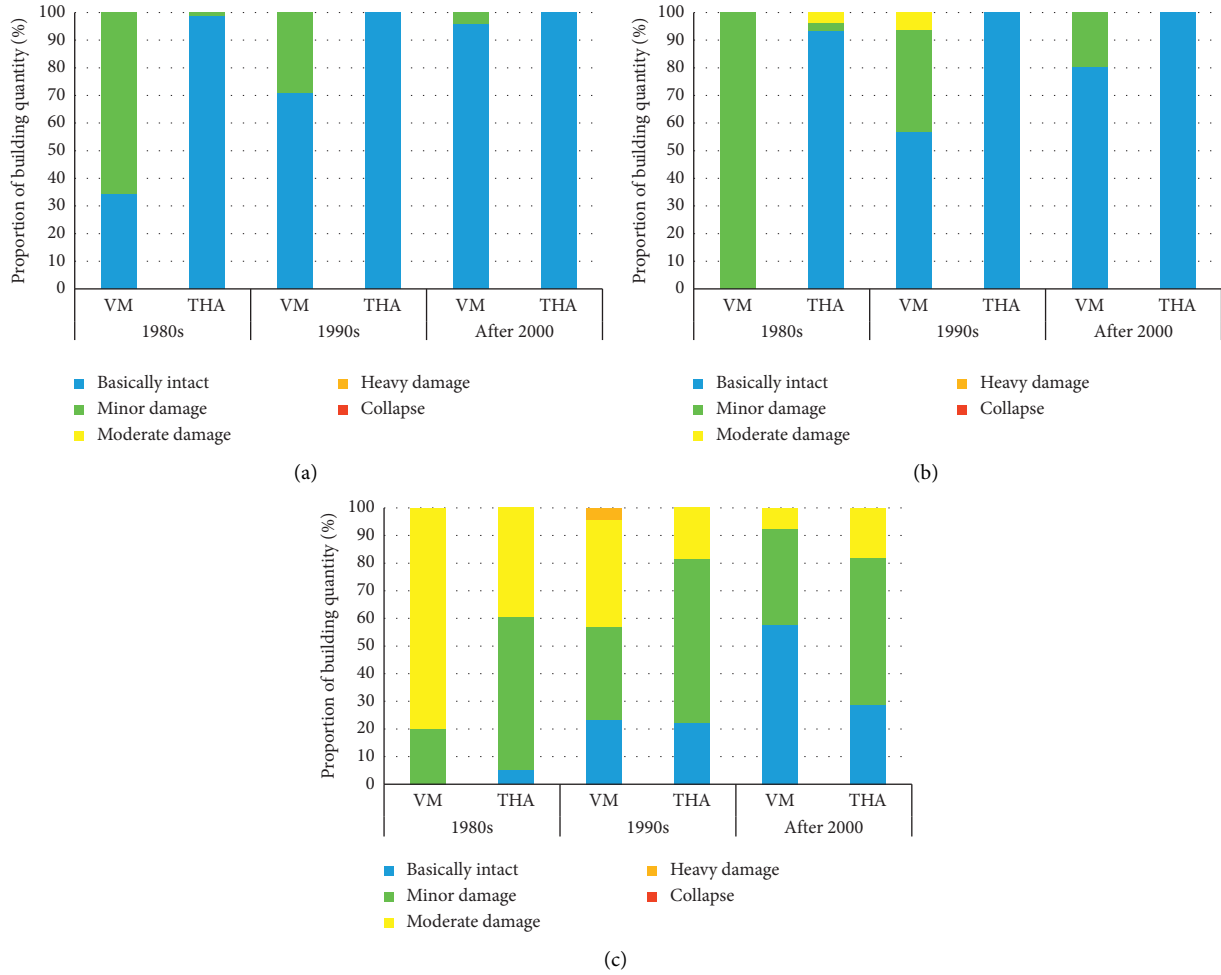
(a)

(b)

(c)

FIGURE 19: Comparison results of RC frame structures under different earthquake levels. (a) Frequent earthquake. (b) Moderate earthquake. (c) Rare earthquake. THA, time-history analysis; VM, vulnerability matrix.

dimensionless index ranging from 0 to 1. The smaller the index value, the lighter the damage [40]. The average seismic damage index of an area can be calculated as

$$D_z = \sum \left(D_j \times P_j \right), \tag{7}$$

where D_z indicates the regional average seismic damage index under certain intensity, D_j indicates the seismic damage index corresponding to level j damage under certain seismic intensity, and P_j indicates the probability of level j damage level under certain seismic intensity. The division of damage states and corresponding seismic damage indices are shown in Table 1.

A comparison of seismic damage indices of the vulnerability matrix and time-history analysis is shown in Figure 21. The seismic damage index of buildings built in the same era increases with increased seismic intensity; the older the construction time, the lower the earthquake damage index under the same seismic intensity. Due to differences in building properties in the two areas (Weifang

Street in the Pudong New Area and the Tongji campus), a certain degree of dispersion was found in the results, but the trend of the seismic damage index proved consistent; moreover, in newer buildings, a smaller difference between the two methods verified the accuracy of the results to some extent.

Notably, seismic hazard prediction based on a vulnerability matrix developed in [39] used seismic intensity as an evaluation index, which is too rough. The system developed in this study is based on a nonlinear time-history analysis, which can fully consider the time-domain and frequency-domain characteristics of ground motion. Besides, although construction age is taken into account to improve the accuracy in the vulnerability-matrix-based approach, the results are based on statistical methods that cannot reflect the damage state of a specific building. Therefore, there is a tendency to overestimate the damage of older buildings, especially under high seismic intensity. The system applies multiple degrees of freedom (MDOF), finite element models
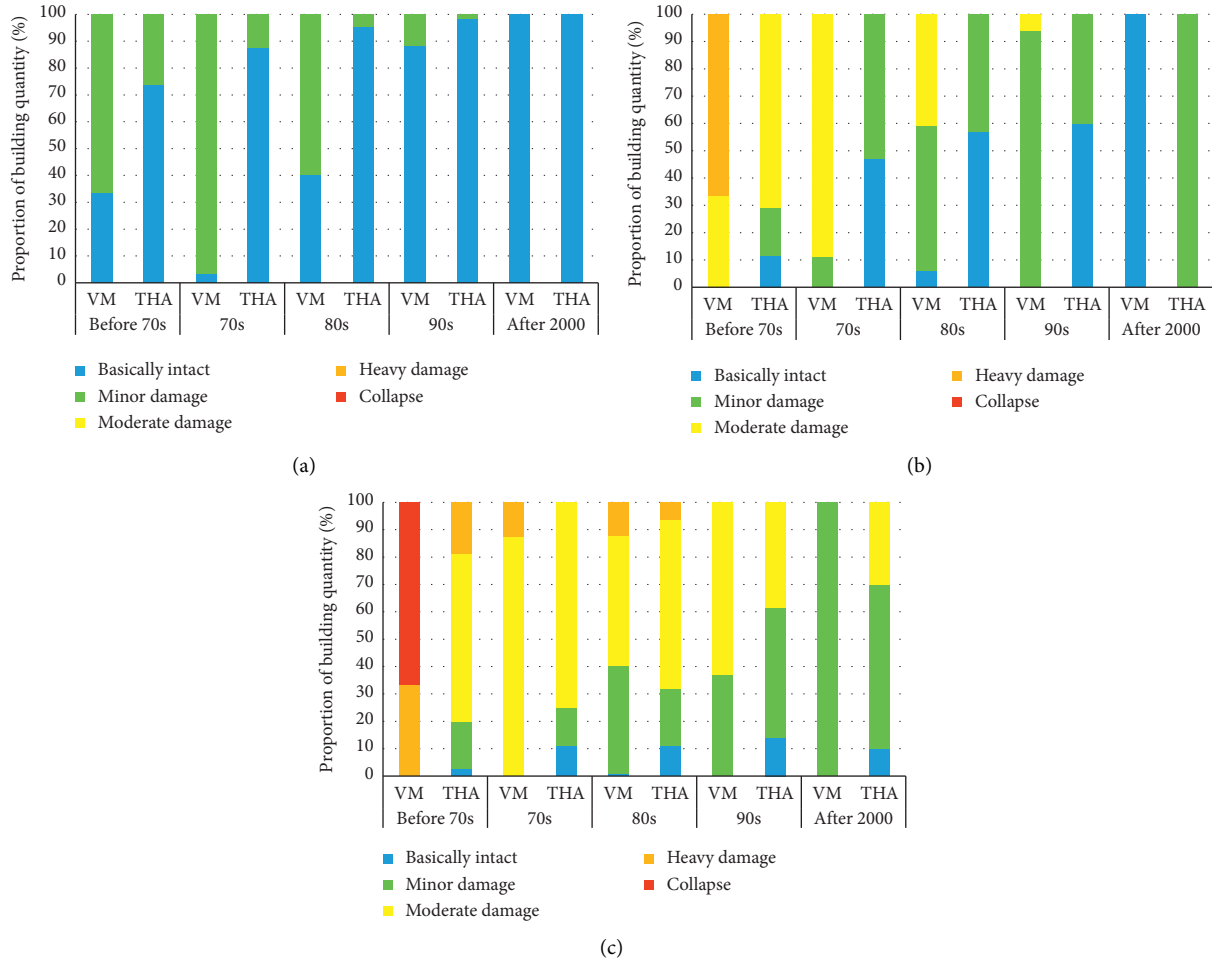
(a)

(b)

(c)

Figure 20: Comparison results of multistory reinforced masonry under different earthquake levels. (a) Frequent earthquake. (b) Moderate earthquake. (c) Rare earthquake.

Table 1: Division of damage states and corresponding seismic damage indices.

Damage state	Range of seismic damage index	Midvalue
Intact	$D \leq 0.1$	0
Minor damage	$0.1 < D \leq 0.3$	0.2
Moderate damage	$0.3 < D \leq 0.55$	0.4
Heavy damage	$0.55 < D \leq 0.85$	0.7
Collapse	$D > 0.85$	1

to simulate buildings, which can obtain the damage of each story and predict seismic hazard more accurately. Moreover, the seismic hazard results are more refined, which can provide a data foundation for future economic loss analysis and seismic resilience evaluation. In summary, the method used in this paper is better able to take into account the building specificities of the target area than the vulnerability matrix method. For predicting seismic damage in areas where there is a lack of historical seismic data, the proposed method is preferred.
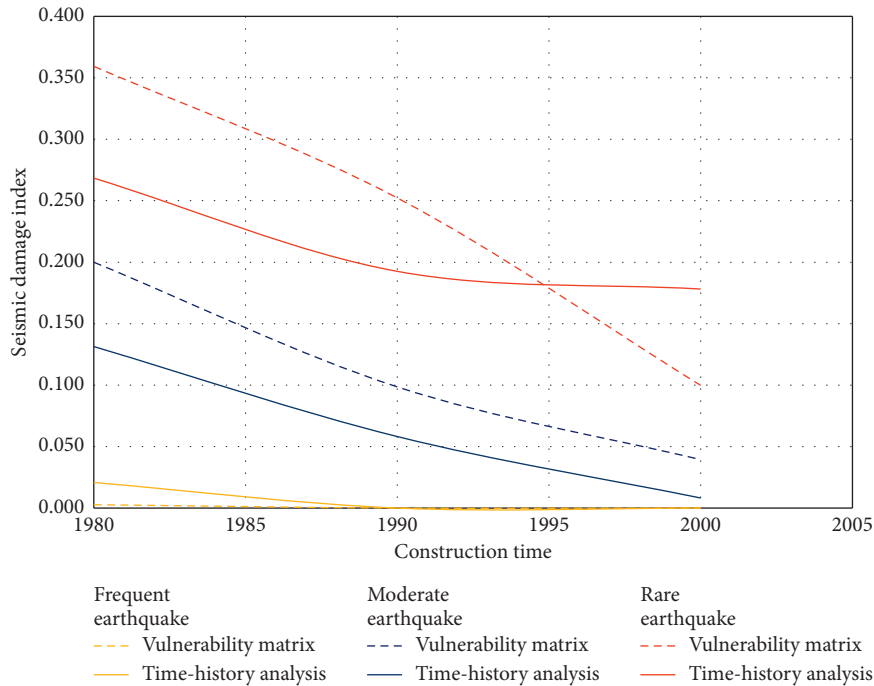
(a)
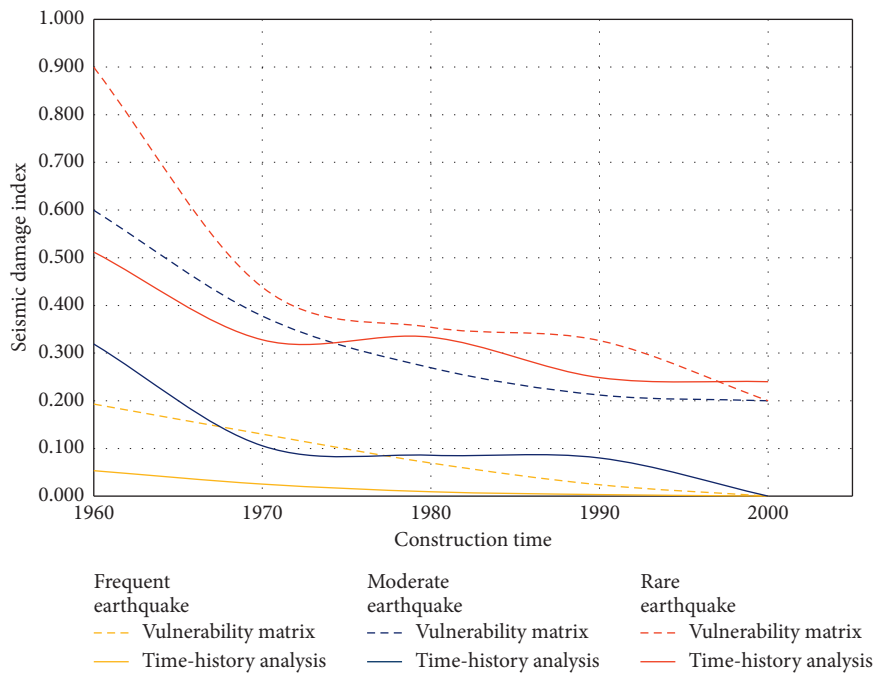
(b)

FIGURE 21: Comparison of seismic damage index between vulnerability matrix and time-history analysis. (a) RC structure. (b) Multistory masonry structure.

5. Conclusions

In this paper, two simplified models using a time-history analysis method were applied. Based on MATLAB and OpenSees, a refined seismic hazard prediction system for urban buildings was developed. The campus of Tongji University, Shanghai, was chosen as the test area to demonstrate the application effects of the system. Conclusions concerning the results of this study are as follows:

(1) The system adopts a unified data interface, which can be seamlessly connected between the modules based on MATLAB and OpenSees and realizes the automatic calculation of seismic hazard prediction.

(2) The system applies different simplified models for different structures. Computation work can be done only by the macro parameters of buildings in GIS data. While this ensures calculation accuracy, calculation efficiency is improved, and the manual operation and hardware requirements of the prediction are reduced.

(3) The prediction results of this time-history analysis system were compared with the results of the vulnerability method, and the trend of seismic hazard predictions was consistent in the results of both, which verified the reliability of the system. At the same time, the system can obtain the specific seismic response and damage state of each story of a building, which achieves the refined prediction necessary to assess the results of seismic hazards accurately.

With the continuous expansion of the city scale and higher requirements for disaster risk response, there is a broad application prospect and research demand for seismic hazard prediction in urban areas based on the time-history analysis. The following issues should be studied in greater depth in the future:

(1) The calculation accuracy of the method applied in the paper is theoretically higher than the traditional methods, but the applicability of the method and the rationality of the parameters are to be tested. Therefore, it is necessary to compare the prediction results with the actual seismic damage in the future.

(2) Although the simplified models used in this paper can greatly reduce the computation time required for modeling and time-history analysis, it is still necessary to adopt methods such as parallel computing and machine learning to reduce the computation time to meet the timeliness of seismic hazard prediction for urban-scale in the future.

(3) Due to the lack of time and detailed building data, the vulnerability matrix of the test area was not obtained, so the prediction results were compared to the results of another area. In the future, more study cases in the same or similar areas are needed to verify the system's reliability further.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] J. Jiang and Z. Xu, *Disaster Prevention and Mitigation Engineering*, Mechanical Industry Press, Beijing, China, 2005.

[2] S. Hu, *The Research on Earthquake Damage Prediction of Building Stock and the Discussion of Direct Economic Loss of Infrastructure*, China Earthquake Administration, Beijing, China, 2007.

[3] N. Lantada, J. Irizarry, A. H. Barbat et al., "Seismic hazard and risk scenarios for Barcelona, Spain, using the risk-UE vulnerability index method," *Bulletin of Earthquake Engineering*, vol. 8, pp. 201–229, 2010.

[4] C. Rojahn, R. L. Sharpe, R. E. Scholl, A. S. Kiremidjian, R. V. Nutt, and R. R. Wilson, "Earthquake damage and loss evaluation for California," *Earthquake Spectra*, vol. 2, pp. 767–782, 1986.

[5] FEMA, *Earthquake Loss Estimation Methodology-HAZUS97*, Federal Emergency Management Agency——National Institute of Building Sciences, Washington, D.C., USA, 1997.

[6] FEMA, *Earthquake Loss Estimation Methodology-HAZUS99. Technical Manual*, Federal Emergency Management Agency——National Institute of Building Sciences, Washington,D.C., USA, 1999.

[7] G. M. Calvi, R. Pinho, G. Magenes, J. J. Bommer, L. F. Restrepo-Vélez, and H. Crowley, "Development of seismic vulnerability assessment methodologies over the past 30 years," *ISET Journal of Earthquake Technology*, vol. 43, pp. 75–104, 2006.

[8] A. Eleftheriadou and A. Karabinis, "Development of damage probability matrices based on Greek earthquake damage data," *Earthquake Engineering and Engineering Vibration*, vol. 10, pp. 129–141, 2011.

[9] S. Hu, B. Sun, and D. Wang, "A method for earthquake damage prediction of building group based on building vulnerability classification," *Earthquake Engineering and Engineering Vibration*, vol. 30, pp. 96–101, 2010.

[10] A. K. Eleftheriadou and A. I. Karabinis, "Evaluation of damage probability matrices from observational seismic damage data," *International Journal of Earthquakes and Structures*, vol. 4, 2013.

[11] Q. Chen, X. Zeng, C. Xiong, Z. Xu, and X. Lu, "Comparison of seismic damage simulation methods for different regional buildings," *Engineering Mechanics*, vol. 34, pp. 105–110, 2017.

[12] H. Sucuoğlu and M. S. Günay, "Generalized force vectors for multi-mode pushover analysis," *Earthquake Engineering & Structural Dynamics*, vol. 40, pp. 55–74, 2011.

[13] C. A. Kircher, R. V. Whitman, and W. T. Holmes, "HAZUS earthquake loss estimation methods," *Natural Hazards Review*, vol. 7, no. 2, pp. 45–59, 2006.

[14] J. W. F. Remo and N. Pinter, "Hazus-MH earthquake modeling in the central USA," *Natural Hazards*, vol. 63, no. 2, pp. 1055–1081, 2012.

[15] S. K. Ploeger, G. M. Atkinson, and C. Samson, "Applying the HAZUS-MH software tool to assess seismic risk in downtown Ottawa, Canada," *Natural Hazards*, vol. 53, no. 1, pp. 1–20, 2010.

[16] T. Levi, B. Tavron, O. Katz et al., *Earthquake Loss Estimation in Israel Using the New HAZUS-MH Software: Preliminary Implementation*, Ministry of National Infrastructures Geological Survey of Israel, Jerusalem, Israel, 2010.

[17] H. Chen, B. Sun, X. Chen, and Y. Zhong, "Research on China earthquake loss estimation system based on cloud computing," *Dizhen Gongcheng Yu Gongcheng Zhendong (Earthquake*

Engineering and Engineering Vibration), vol. 33, pp. 198–203, 2013.

[18] M. C. Schmidtlein, J. M. Shafer, M. Berry, and S. L. Cutter, "Modeled earthquake losses and social vulnerability in Charleston, South Carolina," *Applied Geography*, vol. 31, pp. 269–281, 2011.

[19] P. Castaldo, B. Palazzo, G. Alfano, and M. F. Palumbo, "Seismic reliability based ductility demand for hardening and softening structures isolated by friction pendulum bearings," *Structural Control and Health Monitoring*, vol. 25, p. e2256, 2018.

[20] M. R. Mohammadizadeh, "Seismic vulnerability assessment of simple steel buildings with RC shear wall," *Modares Civil Engineering Journal*, vol. 18, pp. 71–83, 2018.

[21] S. Mahin, G. Deierlein, S. Govindjee, and L. Lowes, "SimCenter computational framework for simulating and mitigating the effects of earthquakes," in *Proceedings of 11th U.S. National Conference on Earthquake Engineering*, Los Angeles, CA, USA, June 2018.

[22] H. Muneo and I. Tsuyoshi, "Current state of integrated earthquake simulation for earthquake hazard and disaster," *Journal of Seismology*, vol. 12, pp. 307–321, 2008.

[23] A. Sahin, R. Sisman, A. Askan, and M. Hori, "Development of integrated earthquake simulation system for Istanbul," *Earth, Planets and Space*, vol. 68, p. 115, 2016.

[24] M. L. L. Wijerathne, M. Hori, T. Kabeyazawa, and T. Ichimura, "Strengthening of parallel computation performance of integrated earthquake simulation," *Journal of Computing in Civil Engineering*, vol. 27, pp. 570–573, 2013.

[25] C. Xiong, X. Lu, X. Lin, Z. Xu, and L. Ye, "Parameter determination and damage assessment for THA-based regional seismic damage prediction of multi-story buildings," *Journal of Earthquake Engineering*, vol. 21, no. 3, pp. 461–485, 2017.

[26] X. Zhen, X. Lu, G. Hong, H. Bo, and A. Ren, "Seismic damage simulation in urban areas based on a high-fidelity structural model and a physics engine," *Natural Hazards*, vol. 71, pp. 1679–1693, 2014.

[27] A. S. Veletsos and W. P. Vann, "Response of ground-excited elastoplastic systems," *Journal of the Structural Division*, vol. 97, pp. 1257–1281, 1971.

[28] O. Diaz, E. Mendoza, and L. Esteva, "Seismic ductility demands predicted by alternate models of building frames," *Earthquake Spectra*, vol. 10, no. 3, pp. 465–487, 1994.

[29] X. Chen, X. Lu, G. Hong, and Z. Xu, "A nonlinear computational model for regional seismic simulation of tall buildings," *Bulletin of Earthquake Engineering*, vol. 14, pp. 1047–1069, 2016.

[30] J. S. Steelman and J. F. Hajjar, "Influence of inelastic seismic response modeling on regional loss estimation," *Engineering Structures*, vol. 31, no. 12, pp. 2976–2987, 2009.

[31] M. Scott, F. Filippou, and S. Mazzoni, "Opensees wiki: hysteretic material. availabe online," 2016, http://opensees.berkeley.edu/wiki/index.php/Hysteretic_Material.

[32] A. Schellenberg, "Two node link element," 2019, http://opensees.berkeley.edu/wiki/index.php/Two_Node_Link_Element.

[33] C. Xiong, Z. Xu, X. Zeng, X. Lu, and L. Ye, "Damage prediction method suitable for regional seismic damage simulation of concrete highrise building structures," *Journal of Natural Disasters*, vol. 25, pp. 69–78, 2016.

[34] Y. Zhiqian, L. Shuzhen, Z. Zhi, and Y. Shuwen, "A method for prediction and classification of earthquake disasters," *Earthquake Research In China*, vol. 7, pp. 1–19, 1991.

[35] S. S. Bureau and T. University, *Shanghai Ground Motion Parameter Zoning*, Seismological Press, Beijing, China, 2004.

[36] S. Urban and R. Committee, *Shanghai Building Seismic Design Code*, Shanghai Municipal Commission of Housing and Urban-Rural Development, Shanghai, China, 2013.

[37] PEER, "NGA strong-motion databases," 2019, https://ngawest2.berkeley.edu/.

[38] NIED, "Strong-motion seismograph networks (K-NET, KiK-net)," 2019, http://www.kyoshin.bosai.go.jp/.

[39] Y. Zhai, S. Chen, and Q. OuYang, "GIS-based seismic hazard prediction system for urban earthquake disaster prevention planning," *Sustainability*, vol. 11, p. 2620, 2019.

[40] J. Hu, *Earthquake Engineering*, Seismological Press, Beijing, China, 2006.