

Impact of Extreme Weather Events and Climate Change on Roadway Transportation Systems

Lead Guest Editor: Yajie Zou

Guest Editors: Lingtao Wu and Hailong Liu





Impact of Extreme Weather Events and Climate Change on Roadway Transportation Systems

Journal of Advanced Transportation

**Impact of Extreme Weather Events
and Climate Change on Roadway
Transportation Systems**

Lead Guest Editor: Yajie Zou





Guest Editors: Lingtao Wu and Hailong Liu



Copyright © 2021 Hindawi Limited. All rights reserved.
























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

Associate Editors

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

Academic Editors

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


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



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

Contents

Commuting in the Storm: Adaptation of Transit Riders and Measures for Transit Operator—A Case in Shanghai

Kun Bo , Jing Teng , Cen Zhang , and Dahai Han 

Research Article (15 pages), Article ID 6622461, Volume 2021 (2021)

Resilience-Based Optimization of Postdisaster Restoration Strategy for Road Networks

Xinhua Mao , Jibiao Zhou , Changwei Yuan, and Dan Liu






Research Article (15 pages), Article ID 8871876, Volume 2021 (2021)

Developing Roadway Safety Models for Winter Weather Conditions Using a Feature Selection Algorithm

Bryce Hallmark  and Jing Dong 


Research Article (13 pages), Article ID 8824943, Volume 2020 (2020)

Origin-Destination-Based Travel Time Reliability under Different Rainfall Intensities: An Investigation Using Open-Source Data

Qi Zhang , Hong Chen , Hongchao Liu , Wei Li , and Yibin Zhang 


Research Article (12 pages), Article ID 8816020, Volume 2020 (2020)

Predicting Wet-Road Crashes Using the Finite-Mixture Zero-Truncated Negative Binomial Model

Ying Chen  and Zhongxiang Huang



Research Article (9 pages), Article ID 8828939, Volume 2020 (2020)

The Effect of Regret-Based Risky Route Choice on the Traffic Equilibrium for Emergency Evacuation

Ze Wang , Haiqiang Yang, and Linglin Ni

Research Article (8 pages), Article ID 8858302, Volume 2020 (2020)

Exploring the Application of the Linear Poisson Autoregressive Model for Analyzing the Dynamic Impact of Traffic Laws on Fatal Traffic Accident Frequency

Yue Zhang, Yajie Zou , Lingtao Wu, Jinjun Tang , and Malik Muneeb Abid

Research Article (9 pages), Article ID 8854068, Volume 2020 (2020)

Research Article

Commuting in the Storm: Adaptation of Transit Riders and Measures for Transit Operator—A Case in Shanghai

Kun Bo ^{1,2,3} Jing Teng ^{1,2,4} Cen Zhang ⁵ and Dahai Han ⁶

¹College of Transportation Engineering, Tongji University, Shanghai 201804, China

²Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China

³College of Higher Technical, Shanghai Maritime University, Shanghai 201306, China

⁴Shanghai Key Laboratory of Rail Infrastructure Durability and System Safety, Shanghai 201804, China

⁵Department of Urban Management, Kyoto University, Kyoto 615-8540, Japan

⁶Department of Civil and Environmental Engineering, University of Wisconsin-Milwaukee, PO Box 784, Milwaukee, WI 53211, USA

Correspondence should be addressed to Jing Teng; tengjing@tongji.edu.cn and Cen Zhang; zhang.cen@trans.kuciv.kyoto-u.ac.jp

Received 7 October 2020; Revised 7 February 2021; Accepted 11 February 2021; Published 24 March 2021

Academic Editor: Lingtao Wu

Copyright © 2021 Kun Bo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Adverse weather has been recognized as an important factor to affect travelers' activity plans in departure time, transport mode, route taken, or cancellation. In the storm, road waterlogging degrades the capacity of road networks and the service quality of transit systems, which further may affect the supply and demand for transit. Based on a typical case in Shanghai that commuters have no easy access to metro service, this paper aims to explore how transit passengers adapt to different situations in the storm and what emergency plans should be taken accordingly. Derived from the revealed preference (RP) and stated preference (SP) survey results for experienced transit commuters, a nested logit (NL) model was developed to describe the travel behavior of transit commuters. Six alternatives, Direct Bus, Bus + Bus, Metro, Bus + Metro, Taxi, and Cancel Trip, and three storm scenarios were set for transit commuters in this case. Estimated parameters indicate that, in storm weather, crowdedness is less considered by transit commuters, and transfer times, walking time, in-vehicle time, and waiting time have negative effects on the selection of the corresponding alternative, whereas the impact of taxi fare is positive since the higher fare is usually accompanied by worse weather and traffic condition. Sensitivity analysis shows that walking time to metro station, in-vehicle time, and waiting time at a bus stop are the most critical factors leading to transit ridership reduction in the storm. According to this, three possible plans for the transit operator, shuttle bus to the metro station (P1), information announcement (P2), and route adjustment (P3), are simulated and compared. We recommend adopting P2, P2 + P3, and P1 + P2 + P3 in turn with the increase of road waterlogging. These findings have important practical implications for developing transit emergency management plans and serve as references for the transit agencies and operators.

1. Introduction

Weather is recognized as a critical factor in transportation system operation. Inclement weather can significantly decrease traffic speed and road capacity, cause traffic congestion or disruption, and affect traffic safety. Public transport, a vital component of urban transport systems serving commuting trips, may also be largely affected by the inclement weather. It has been proved that inclement weather can degrade the bus service, affect passengers' travel choices, and reduce transit ridership [1]. As reported in the

Intergovernmental Panel on Climate Change (IPCC) [2], a significant increase in extreme weather and climate events has been observed since the 1950s on a global scale. Facing increasing days with inclement weather, transit operators and government should provide powerful emergency management plans to guarantee commuting travel.

It is of great importance to understand the impact of transport system performance in adverse weather and how travelers adapt to it. There exist some studies investigating the influence of weather on the travel behavior of passengers, including travel demand, modal shift, departure time, and

route/destination choices [3–6]. The most common method of examining inclement weather's impact on travel has involved collecting and analyzing traffic data and ridership data in conjunction with weather variables, such as temperature, rainfall, and wind speed. Recently, more temporally and spatially disaggregated analyses are proposed to increase the explanatory power of results [7–9]. Besides, some studies pay attention to identify the impact of more issues in adverse weather, such as setting of infrastructure, traveler segments, and trip purpose [10–13]. The results of studies show the heterogeneous impacts of adverse weather on travel behavior due to discrepancies in the overall settings of studies.

According to our understanding, the influence of adverse weather on travel behavior can be decomposed into two parts: the negative impacts of extreme weather on transportation services and the corresponding adjustment of travel behavior by travelers. In essence, ridership is directly affected by the performance of the travel services (e.g., accessibility, reliability, convenience, comfort, and travel time) rather than weather. Several studies have investigated the influence of weather on the performance of transport systems, such as road capacity, vehicle moving speed, bus dwell time, and transit service reliability [14, 15]. From our point of view, in addition to the settings of infrastructure and meteorological elements, the management level is also a decisive factor in the impact of adverse weather on transport services. We believe that good measures can reduce the negative impact and improve performance effectively. For transit operators, to know what measures should be taken, it is necessary to understand the influence of the transit service performance on travel choices under various situations. However, the limited research on adaptation of travelers to the change of transit service quality in adverse weather may be not supportive to conduct effective emergency measures.

To fill this gap, this paper aims to explore the adaptation of transit commuters to different situations and provide feasible measures for the transit operators accordingly. Firstly, the revealed preference (RP) and stated preference (SP) surveys were conducted, and then the relationship between travel mode choice and transit service performance in the storm was established. Derived from the statistic and sensitivity analysis, three possible plans, shuttle bus to the metro station, information announcement, and routine adjustment, are simulated and compared. These results could have implications for developing bus emergency management plans and serve as references for the public transit agency.

The remainder of this paper is organized as follows. Section 2 reviews related studies and provides the literature background for the present study. Section 3 introduces the data and the study area. Analytical approaches are also elaborated on in this section. Section 4 presents and discusses the analysis results and research findings. Section 5 proposes and then simulates some emergency plans for the transit operator. Section 6 summarizes the research findings and points out limitations and future work.

2. Literature Review

The impacts of weather on transportation have received substantial research attention. Studies can be identified as two major parts: impacts on physical and operational aspects and impacts on travel demand and traveler behavior. In this section, we shall review the relevant studies, outline the major findings, and identify research gaps.

2.1. Impacts on Physical and Operational Aspects.

Weather has a direct impact on the performance of transportation infrastructure [6]. Heavy rainfall, storms, snow, and fog all can result in deterioration of traffic conditions, like low visibility, wet roads, and waterlogging. Regarding different travel modes, previous studies have shown a negative impact on travel speeds [3, 16–19]. Some studies indicate that wet roads can reduce the average vehicle speed by 6–7% in urban areas and even 8–12% when it rains [20]. And the reduction of travel speed can even lead to a 20% decrease in road capacity in the urban network [21], resulting in traffic congestion or disruptions [1], the unreliability of the transportation system [22], and more traffic accidents [23]. For transit services, the impact of weather is more complex than private and nonmotorized transport. In one aspect, it has negative impacts on bus operation in terms of service frequency, headway regularity, and travel time variability [1]; in the other aspect, it increases the difficulty and time on the way to/from transit stations for passengers [7].

2.2. Impact on Travel Demand and Traveler Behavior.

Aside from its impact on the physical and operational aspects of transport systems, adverse weather may affect travelers' decisions. A lot of studies highlighted travel behavior changes due to weather conditions, like model shift [4, 6], changes in destination and route [3], and put-off or cancel trips [24]. Transit riders are often directly subject to adverse weather while waiting or walking to/from the station, and they are indirectly affected by the deterioration of in-vehicle transit services [7]. Both direct and indirect effects influence riders' behavior as well as transit ridership.

In most previous studies, adverse weather, such as rain, has been found to exert a negative influence on transit ridership [1, 25]. Nevertheless, some empirical studies report different results on the rain-transit ridership relationship, which shows that it is positively associated with public transit share [17, 26]. Further studies indicate that transit ridership has a slight decline during light rainfall, then increases with rain becomes heavier, and eventually drops sharply in heavy rain [27]. The heterogeneous impacts show the complexity of the relationship between adverse weather and transit ridership. Therefore, some studies pay attention to identify specific impacts of adverse weather with more detailed issues. For example, senior passengers [28], lower-income travelers [29], and occasional users [10] are more sensitive to adverse weather. Additionally, adverse weather has more impact on recreational travel than commuter travel [12], bus trips than

metro trips [13], weekend trips than weekday trips [7], and long commuting trips in winter than that in summer [11]. Besides the adverse weather's impact on transit, ridership also varies on the type, location, facilities, and routes of stations [8] and time of day [9].

In short, most of these studies examined inclement weather's impact on ridership in conjunction with amount variables, weather elements (e.g., temperature, rainfall, and wind speed), time (day of week, time of day, and seasonality), facility (e.g., weather protection of palm, accessibility of station, and connecting bus routes), and individual attributions (e.g., age, gender, income, experience, and travel purpose). Less attention was paid to explore how transit passengers adapt to transit service in storm according to service performance, such as travel time variability, service frequency, reliability, accessibility, and comfort. We believe that it helps improve transit management for infrastructure as well as transit service adjustment for transit operator and agency.

3. Data and Analytical Methods

In this section, we will introduce a case study, including data collection and analytical methods, to describe the dynamics of travel choice behavior in adverse weather.

3.1. Data Collection. As shown in Figure 1, the data adopted in this study are originated from surveys in a largely residential area with over 15,000 residents near the middle ring roads in Shanghai, China. This residential area covers nearly 10 bus stops, but no metro station. A passenger from the center of this area needs to walk over 1 km to the nearest metro station. There exist large commuting demands between the residential area and the central business district nearly 6 km away. However, road waterlogging and disruption occur recurrently in storm weather within this area, causing terrible bus service.

It is a representative case to reflect the dilemmas of transit commuters without convenient metro service. They have to make a compromise between time, comfort, and economic loss since there is no convenient and reliable transit service in adverse weather. Based on this case, later we will have a detailed analysis of travel behavior in different conditions and propose some possible measures accordingly.

The questionnaire, designed for RP and SP survey, was conducted in the weekday morning peak hours in September 2016, targeting passengers leaving their homes to take public transit. The RP survey collected sociodemographic information, for example, gender, age, occupation, and monthly income, as well as travel-related information, for example, travel purpose, travel mode, travel time and transfer time, and commuting experience in the storm. The SP survey asked respondents to make travel choices in hypothetical scenarios.

Three hypothetical road waterlogging scenarios (S1–S3) are assumed to have different impacts on road conditions and transit operation in storm weather (see Figure 2):

S1: in the slight storm (beyond 20 mm rainfall), the small-scale ponding occurs, causing walking time increase, vehicle speed decline, and slight bus delay

S2: in the medium storm (around 50 mm rainfall), several parts of roads are waterlogging, leading to inconvenient walking, traffic congestion, and serious bus delays

S3: in the heavy storm (over 100 mm rainfall), roads within some regions are waterlogging seriously, resulting in walking difficulties, partial road closure, and bus detours

Through on-the-spot investigation in a storm, we found that in this case there are six reasonable travel alternatives for transit commuters, which specifies five possible combinations of travel modes and another option to cancel the trip. The performance of services under six travel alternatives is settled according to the investigation data. The details are shown as follows (see Figure 1 and Table 1):

A1: taking a bus with no transfer (Direct Bus Service, DB)

A2: riding a bus and then transferring to another bus in a different routine (Bus + Bus, B + B)

A3: walking to a metro station and then taking a metro (Metro, M)

A4: riding a bus and then transferring to a metro (Bus + Metro, B + M)

A5: taking a taxi (Taxi, Tx)

A6: cancelling the trip (Cancel Trip, CT)

Based on these, we design the SP survey, which contained twelve different conditions derived from three hypothetical road waterlogging scenarios and four crowded scenarios. For each case, respondents were offered the above six travel alternatives and their performances (see Table 1). It is assumed that the respondents have enough knowledge from experience and can make choices accordingly.

In our designed scenarios, Direct Bus, the most convenient way in normal weather, was largely affected by the storm in walking time, waiting time, and in-vehicle travel time. Metro operation is less affected by storm, whereas the need for long time walking to the metro station may reduce their willingness. Taking a bus and then transferring to metro or bus, there could be alternatives to keep short walking time and relatively low delay. Traveling by taxi can avoid long walking time, large travel delay, and possible crowdedness in transit service effectively, and yet the fare of taxi, increasing with road waterlogging due to congestion and detouring, is much higher than that of all alternatives in public transportation in all scenarios.

To ensure that the samples were representative, some trap questions and logical judgments were used in preliminary selection. 185 questionnaires that have enough knowledge of commuting by transit in storm weather are selected, and 162 among them are transit commuters choosing bus or metro as their major travel mode. Totally,



FIGURE 1: Shanghai, the study context.



FIGURE 2: Hypothetical road waterlogging scenarios in the storm. (a) Scenario 1, (b) Scenario 2, and (c) Scenario 3.

1944 valid travel choices gathered from these 162 experienced transit commuters in the SP survey are used to calibrate the parameters of the model in this study. Table 2 presents the descriptive statistics of these participants. Some key features of the participants in this study are as follows:

- Most of the respondents are working-age adults in the normal sense, ranging from 16 to 60 years old. Since the government is trying to create incentives for people to work longer, the elderly also make up a large proportion of commuters.
- The monthly income of more than 80% of respondents ranges from 2000 yuan (about 300 US dollars) to 10,000 yuan (about 1500 US dollars). The proportion of high-income people is just over 10%,

which is only half of 20.6% in the Shanghai Statistical Yearbook [30].

- More than 90% of respondents who have transit commuting experience in a storm choose bus or metro as their major travel mode.
- About half of respondents' travels in the RP survey, which was carried in the weekday morning peak, are commuting trips.
- The expected travel time of near 80% of the respondents is over 60 minutes.
- Above 40% of the respondents need transfers.

In sum, the data revealed that the majority of the respondents are transit passengers with long commuting time,

TABLE 1: Performance of services in scenarios.

Performance of services	Scenarios	Alternatives				
		A1	A2	A3	A4	A5
Walking time (minute)	S0	9	5	15	6	0
	S1	10	5	20	7	0
	S2	12	6	25	8	0
	S3	15	8	30	10	0
Waiting time (minute)	S0	15	12	3	15	5
	S1	20	15	3	15	10
	S2	30	25	5	20	20
	S3	45	40	9	30	30
In-vehicle travel time (minute)	S0	25	25	15	25	15
	S1	30	30	15	30	20
	S2	40	40	15	35	25
	S3	60	60	20	45	30
Overall travel time (minute)	S0	49	42	33	46	20
	S1	60	50	38	52	30
	S2	82	71	45	63	45
	S3	120	108	59	85	60
Fare (yuan)	S0					35
	S1					40
	S2	2	3	4	5	60
	S3					100
Transfer	S0-S3	0	1	0	1	0

Note: S0 is the normal weather; S1-S3 are three hypothetical scenarios of the storm.

TABLE 2: Survey profile.

Respondent characteristic	Variables	Percentage	Trip characteristic	Variables	Percentage
Gender	Male	41.98	Travel purpose	Commute	49.38
	Female	58.02		Business	6.17
Age	16-	0.62		Leisure	11.73
	16-24	14.20		Hospital	7.41
	25-34	33.95		Visiting friend	10.49
	35-44	16.05		Others	14.81
	45-60	11.73	Bus	51.85	
	60+	23.46	Metro	41.36	
Monthly income (yuan)	2000-	9.26	Travel mode	Taxi	1.85
	2000-5000	38.89		Bike	1.85
	5000-10,000	41.36	Walk	3.09	
	10000+	10.49	0-30	9.88	
Occupation	Government official	1.23	Travel time (minutes)	30-60	12.35
	Company staff	37.65		60-90	20.37
	Student	4.94		90+	57.41
	Freelancer	8.64	Transfer	One or more transfer	43.21
Others	47.53	None transfer		56.79	

medium income, and working age, which make sure that they are our targeted group in this study.

3.2. Analytical Methods. Generally, researchers adopt the discrete choice model (DCM) [31-35] or structural equation model (SEM) [17] to estimate the travel mode choice using the RP and SP survey data. The combination of RP and SP data covers both the existing absolute attribute levels and a much wider range of attributes; thus, it is beneficial to build a more robust model. In this study, the RP data can help us select our target respondents and decision variables. And the SP data is used to estimate the critical factors of that in the

storm weather with road waterlogging. Multinomial logit (MNL) model has been proved to be suitable for modeling discrete choice outcomes under mixed traffic conditions [31]. One inherent assumption of the MNL model is the independence of irrelevant alternatives (IIA), which means that the alternatives are uncorrelated. Considering the similarity of A1~A4, it may lead to the fact that IIA assumption cannot hold; in this case, the nested logit (NL) models were adopted to link the probabilities of choice for commuting travelers to explanatory variables.

In this study, we aim to explore the relationship between the performance of alternatives and the choice probability of transit commuters in different storm scenarios. It is different

from most previous studies, linking weather variables, such as temperature, rainfall, and wind speed with travel choices. To achieve it, three main types of effects are considered in our model. The first one is the direct impact of weather on the accessibility of stations, mainly affected by walking time to/from stations or between transfer stations. The second one is the indirect impact of weather, mainly affected by the degradation of service performance, such as variations of waiting time, in-vehicle travel time, and crowdedness. The third one is the travel cost; travel fare varies with different transport modes and route choices. Besides, individual attributes, for example, age, gender, income, and travel purpose, may differ in sensitivity to the three effects. Travel purpose is set as commuting in this study.

Overall, the details of the decision variables are as follows:

- (a) Crowdedness (CD) refers to the crowding level in terms of transit passenger's feeling (4 levels: 0, 0.3, 0.6, and 0.9)
- (b) Walking time (WT) refers to the time spent by the transit passenger to walk from the departure point to the transit station
- (c) Transfer times (TT) refers to the times of a passenger moving from one vehicle to another, which is fixed in one alternative
- (d) Waiting time (OWT) refers to the time spent for the arrival of the vehicle
- (e) In-vehicle travel time (IT) refers to the time from boarding to alighting
- (f) Travel fare (TF) refers to the fare paid for alternatives
- (g) Monthly income (MI) refers to the scaled parameter which reflects the ability to pay (4 levels, 0, 1, 2, and 3)
- (h) Gender (GE) refers to the gender of the commuter (0, male; 1, female)
- (i) Age (AG) refers to the age group the commuter belongs to (6 levels, 0, 1, 2, 3, 4, and 5)

Among these variables in various scenarios, MI is fixed for a specific commuter, and TT is fixed for a specific alternative. Travel time, including WT, OWT, and IT, varies with scenarios and alternatives. Fare in A1 to A4 (public transportation alternatives) is fixed, while fare in A5 (taxi alternative) varies according to scenarios. CD is an uncertain variable related to the supply and demand of transit service; for taxi mode, it usually can be treated as 0. A1 to A4 belong to public transportation nest (nest 1) which have low fare cost, A5 is taxi nest (nest 2) with high fare cost, and A6 which stops traveling is in cancel nest (nest 3). Suppose that the choice set includes j alternatives ($j = 1, 2, \dots, 6$) belonging to nest i ($i = 1, 2, 3$) and the utility that individual n ($n = 1, 2, \dots, N$) gains from alternative j is formulated as

$$U_{j,n}^k = \alpha_j + \beta_j X_n^k + \varepsilon_{jn}, \quad (1)$$

where $U_{j,n}^k$ is the utility of individual n for choosing alternative j in scenario k , X_n^k is the vector of observable attributes of individual n in scenario k , β_j is the coefficient associated with X_n^k for alternative j , α_j is the intercept of utility function of alternative j , and ε_{jn} is the random error term.

The probability formulation of the NL model (a two-level NL model) can be expressed as follows:

$$\begin{aligned} P_{j,n}^k &= P_{i,n}^k \cdot P_{j|i,n}^k, \\ P_{j|i,n}^k &= \frac{\exp(U_{j,n}^k/\lambda_i)}{\sum_{j' \in S_i} \exp(U_{j',n}^k/\lambda_i)}, \\ V_{i,n}^k &= \ln \left(\sum_{j' \in S_i} \exp(U_{j',n}^k/\lambda_i) \right), \\ P_{i,n}^k &= \frac{\exp(\lambda_i V_{i,n}^k)}{\sum_{i'} \exp(\lambda_{i'} V_{i',n}^k)}, \end{aligned} \quad (2)$$

where $P_{j,n}^k$ is the probability that individual n chooses alternative j in a scenario k , $P_{i,n}^k$ is the probability that individual n chooses nest i in a scenario k , $P_{j|i,n}^k$ is the probability that individual n chooses alternative j if nest i is chosen, S_i is the set of all alternatives included in nest i in a scenario k , $V_{i,n}^k$ is the log sum variable of nest j in scenario k , and λ_i is the dissimilarity parameter for nest i .

4. Results Analysis

In this section, to better understand the behavior changes of transit commuters in adverse weather, we further analyze the statistical results, estimated parameters, and sensitivity based on the choice model and collected data in Section 3.

4.1. Statistical Results. According to Table 3, there are three nests in alternatives: Public transportation, Taxi, and Cancel. From the SP survey in designed scenarios, one may expect that the proportion of public transportation mode decreases sharply from S1 to S3 while choosing probabilities of taxi and cancel increase significantly. It is obvious that storm weather has a huge effect on the travel choice of commuters, and overall, with walking time, waiting time, and in-vehicle time increasing, the willingness to choose public transportation will diminish. Specifically, the probability of DB and B+B modes in S1–S3 has fallen sharply, while increasing from S1 to S2 in M mode and S2 to S3 in B+M mode occurred. It suggests that metro service is less affected and considered as a good substitution for bus service in the storm, which is consistent with previous studies [36]. Even if taxi fare exceeds 20 times of bus fare in S3, the number of people who choose to commute by public transport was almost the same as people choosing taxi, which shows that the travel fare is not decisive for transit commuters in the storm. Meanwhile, over one-third of commuters decide to

TABLE 3: Selection of transit commuters in the storm.

Scenarios	Alternatives					Cancel CT (%)
	Public transportation				Taxi	
	DB (%)	B + B (%)	M (%)	B + M (%)	Tx (%)	
S1	22.53	31.02	19.29	9.26	11.42	6.48
		82.10				
S2	14.66	16.20	23.92	6.17	22.38	16.67
		60.96				
S3	1.23	4.94	14.97	9.88	30.86	38.12
		31.02				

cancel trips and be absent from work which may have a great impact on travel demand and bring huge losses to society.

4.2. Estimated Parameters. The nested logit model was used to test the effects of walking time, waiting time, in-vehicle travel time, crowdedness, and income on commute choices of transit passengers in the storm. Walking time (WT), waiting time (OWT), in-vehicle travel time (IT), fare (TF), and income (MI) were kept in the final model. Though slight impacts of crowdedness (CD), gender (GE), and age (AG) existed in tests, they are not significant at a 95% confidence level. Adding these variables will lead to a rise of *R*-square but the decline of AIC, BIC, and adjusted *R*-square. Further, in addition, some interaction items, such as MI * WT, MI * OWT, MI * IT, and MI * TF, were also tested, but none was significant even at 90% confidence level. We emphasized that the value of time and cost are treated as the same for a certain individual; the coefficients of the time and cost variables were set as generic in our model. The weights and significances of decision variables and performance of the final model are shown in Table 4. Next, we would like to have a detailed analysis of that.

4.2.1. Intercept. Initially, we set different intercepts for all alternatives (the cancel nest as a fixed item is 0). However, we found that alternatives in public transportation nest (A1–A4) have similar estimated values of the intercept. According to our tests, a universal intercept for alternatives in public transportation nest, which is adopted in the final model, can improve the AIC and BIC but will not significantly reduce the *R*-square. A generic intercept for public transportation nest can improve the AIC and BIC but not decrease the *R*-square obviously. It shows that transit commuters have no obvious preference difference in alternatives in public transportations nest.

4.2.2. Transfer. The coefficient of transfer for public transportation nest is -1.88 ; that means the transfer largely reduces the choice probability for alternatives. In taxi and cancel nests, no transfer behavior is considered in the travel process. We suggest that usually multitransfer travel plans are not attractive for commuters, and even in an emergency, transit agency should avoid providing passengers plans more than one transfer.

4.2.3. Income. The coefficients of income for A1, A3, and A4 are negative, whereas for A2, A5, and A6 are positive. It is reasonable that the high-income group has a preference for A5 and A6. If one chooses taxi (A5), he/she needs to pay much more on travel fare, and if one chooses to cancel trip (A6), he/she needs to bear economic loss due to the absence from work. For A2 in our case, the passenger has the shortest walking distance. Therefore, commuters with higher income may be more concerned with the performance of walking time. Overall, it seems that high-income individuals may care about comfort issues more than economic issues and travel time.

4.2.4. Waiting Time and In-Vehicle Time. The coefficients of waiting time and in-vehicle time are both negative, which shows that longer travel time will decrease the choice probability for alternatives. From estimated parameters, in this case, it seems that in-vehicle time has a greater influence than waiting time, which may be a little different from normal weather. It can be explained that, as in storm weather, wetted travelers' crowd in vehicles may feel more uncomfortable than individuals waiting at the station.

4.2.5. Walking Time. Compared with a coefficient of waiting time (-0.059) and in-vehicle time (-0.042), the coefficient value of *walking time* (-0.174) is 3 to 4 times larger, which shows the huge influence of walking time on travel choice in storm. It seems that walking difficulty due to the rain and waterlogging significantly decreases the travel willingness. When walking time exceeds personal tolerance, passengers have to change travel options. It is the decisive factor of commuting in the storm weather.

4.2.6. Fare. The coefficient of fare is positive; it seems to be not reasonable since usually higher cost means less attractiveness to traveler. However, our designed scenario is highly based on real condition; transit fare is constant in different scenarios, while taxi fare is based on travel time and distance which is highly related to storm and road condition. When the traffic condition and weather get worse, the taxi costs more, and to reduce the impact of the storm, taxi, in turn, has a higher attraction. For taxi mode, traffic gets worse, and fare goes higher. When the negative impacts exceed the tolerance, commuter becomes more eager to guarantee commuting even at a higher cost.

4.2.7. Crowdedness. Unexpectedly, there is no obvious effect of crowdedness on all passengers. One possible explanation is that transit commuters in a metropolis, like Shanghai, are used to the crowdedness environment in bus or metro in daily commuting. Compared with normal days, the obvious changes, such as heavy rain, road waterlogging, bus delay, longer walking, and waiting time, are more likely to be concerned in the storm. Therefore, crowdedness is not the key factor affecting commuting in storm weather.

TABLE 4: Estimation results.

Explanatory variables	Alternatives					
	Direct Bus	Public transportation nest Bus + Bus	Metro	Bus + Metro	Taxi nest Taxi	Cancel nest Cancel Trip
Intercept			6.48*		1.31*	0 (fixed)
Transfer			-1.88*		—	—
Dissimilarity of nest			0.787*		1 (fixed)	1 (fixed)
Income	-0.477*	1.06*	-0.501*	-0.256	0.306*	0.58*
In-vehicle travel time			-0.0587*			—
Walking time			-0.174*			—
Waiting time			-0.0418**			—
Fare			0.028*			—
Sample size	1944					
Number of parameters	14					
Final log-likelihood	-3134					
Likelihood ratio	698					
R-square	0.201					
Adjusted R-square	0.196					
AIC	6296					
BIC	6374					

**Significant at 0.01 level; *significant at 0.05 level.

4.3. Sensitivity Analysis

4.3.1. Definition of Sensitivity. In the NL model, the estimated coefficients, which are the odds ratio of the specific travel mode to the reference level, cannot reflect the overall impact of a particular variable directly since it also depends on the magnitudes of all other variables. Therefore, a “strict impact” for a given variable cannot be determined due to the diversity of combinations with other variables. The objective is to anticipate the influence of value changes of variables on the choice of certain individual and subsequently on the share of alternatives. In this case study, sensitivity is defined based on the elasticity with an infinitesimal change, which is called point elasticity. Since the variables of operation performance are continuous, we assume that the relative change of one variable is the same for every individual in the population and the disaggregate direct point elasticity of the model with respect to the variable $x_{m,k}^n$ is defined as

$$\begin{aligned}
 \Delta p_{jmk}^n &= \frac{\partial p_{jmk}^n}{\partial x_{m,k}^n} \cdot x_{m,k}^n, \\
 e_{mjk}^n &= \frac{\Delta p_{jmk}^n}{p_{jmk}^n}, \\
 p_{jmk}^n &= f_j(x_n^k, \hat{\gamma}), \\
 E_{mjk} &= \frac{\sum_{n \in A} \Delta p_{jmk}^n}{\sum_{n \in A} p_{jmk}^n} = \frac{\sum_{n \in A} e_{mjk}^n \cdot p_{jmk}^n}{\sum_{n \in A} p_{jmk}^n} \\
 &= \frac{\sum_{n \in A} (\partial f_j(x_n^k, \hat{\gamma}) / \partial x_{m,n}^k) \cdot x_{m,n}^k}{\sum_{n \in A} f_j(x_n^k, \hat{\gamma})},
 \end{aligned} \tag{3}$$

where E_{mjk} is the aggregate point elasticity of the model of travel mode j in choice set A. $p_{i,n}$ is the estimated probability of individual n choosing travel mode j with variables X_n^k in a scenario k . A is the set containing all relevant samples. $f_j(\cdot)$ is the probability mass function for evaluating $p_{i,n}$, which is obtained from the estimated NL model. $\hat{\gamma}$ is the estimated coefficients from the NL.

4.3.2. Sensitivity Analysis of Single Variable. Table 5 reports the sensitivity for each variable according to the estimated NL model. The numbers in the tables present the percentage change in the probability of an alternative with respect to the changes in one variable in a certain situation. As shown in Table 5, red values indicate an increase in the probability, whereas blue values indicate the opposite. Five variables, for example, transfer times, walking time, waiting time, in-vehicle time, and fare, indicate the performance of transport operation. In storm weather with road waterlogging, the quantitative values of indicators can directly reflect the bus service quality at that time and indirectly reflect the severity level of impact by weather. The sensitivity analysis results were interpreted from the mode share shift responding to single variable change.

4.3.3. Walking Time. In our case, we assume that passengers can be taken to their destination by taxi without walking. Therefore, WT5 is set as 0, and sensitivity is also estimated as 0 here. For WT1 to WT4, with walking time of certain alternative getting longer, the probabilities of this alternative are expected to decrease while the probabilities of other alternatives all increase. Specifically, in all scenarios, WT3 has the largest impacts (0.7 to 0.9) on A5 and A6, which prove the key influence of walking time to metro station.

TABLE 5: Sensitivity for each variable.

Variables	Scenario	A1	A2	A3	A4	A5	A6	Variables	Scenario	A1	A2	A3	A4	A5	A6
WT1	1	-1.626	0.520	0.584	0.575	0.417	0.410	IT1	1	-1.642	0.525	0.590	0.580	0.421	0.414
	2	-2.319	0.297	0.332	0.327	0.222	0.219		2	-2.927	0.375	0.420	0.413	0.281	0.276
	3	-3.164	0.135	0.149	0.147	0.077	0.075		3	-4.260	0.182	0.201	0.198	0.103	0.102
WT2	1	0.311	-0.734	0.310	0.319	0.255	0.261	IT2	1	0.627	-1.481	0.625	0.643	0.515	0.528
	2	0.301	-1.183	0.300	0.309	0.229	0.235		2	0.579	-2.279	0.578	0.595	0.442	0.453
	3	0.133	-1.604	0.133	0.137	0.076	0.078		3	0.336	-4.050	0.335	0.346	0.192	0.196
WT3	1	1.018	0.904	-3.398	1.002	0.727	0.714	IT3	1	0.257	0.228	-0.858	0.253	0.183	0.180
	2	1.450	1.293	-4.070	1.428	0.969	0.952		2	0.293	0.261	-0.822	0.288	0.196	0.192
	3	1.515	1.368	-5.110	1.495	0.779	0.765		3	0.340	0.307	-1.147	0.335	0.175	0.172
WT4	1	0.121	0.113	0.121	-1.426	0.088	0.087	IT4	1	0.175	0.162	0.175	-2.057	0.127	0.126
	2	0.225	0.210	0.225	-1.544	0.153	0.152		2	0.331	0.309	0.332	-2.274	0.226	0.223
	3	0.238	0.225	0.239	-1.972	0.124	0.123		3	0.361	0.341	0.361	-2.987	0.188	0.186
WT5	1	0.000	0.000	0.000	0.000	0.000	0.000	IT5	1	0.145	0.148	0.145	0.145	-1.028	0.147
	2	0.000	0.000	0.000	0.000	0.000	0.000		2	0.304	0.310	0.304	0.305	-1.161	0.308
	3	0.000	0.000	0.000	0.000	0.000	0.000		3	0.728	0.732	0.728	0.729	-1.619	0.730
OWT1	1	-0.780	0.249	0.280	0.276	0.200	0.197	TF1	1	0.052	-0.017	-0.019	-0.018	-0.013	-0.013
	2	-1.390	0.178	0.199	0.196	0.133	0.131		2	0.062	-0.008	-0.009	-0.009	-0.006	-0.006
	3	-2.277	0.097	0.107	0.106	0.055	0.054		3	0.068	-0.003	-0.003	-0.003	-0.002	-0.002
OWT2	1	0.223	-0.528	0.223	0.229	0.183	0.188	TF2	1	-0.030	0.071	-0.030	-0.031	-0.025	-0.025
	2	0.258	-1.015	0.257	0.265	0.197	0.202		2	-0.021	0.082	-0.021	-0.021	-0.016	-0.016
	3	0.160	-1.924	0.159	0.164	0.091	0.093		3	-0.008	0.097	-0.008	-0.008	-0.005	-0.005
OWT3	1	0.037	0.033	-0.122	0.036	0.026	0.026	TF3	1	-0.033	-0.029	0.109	-0.032	-0.023	-0.023
	2	0.070	0.062	-0.195	0.069	0.046	0.046		2	-0.037	-0.033	0.105	-0.037	-0.025	-0.024
	3	0.109	0.098	-0.368	0.108	0.056	0.055		3	-0.032	-0.029	0.109	-0.032	-0.017	-0.016
OWT4	1	0.062	0.058	0.062	-0.733	0.045	0.045	TF4	1	-0.014	-0.013	-0.014	0.164	-0.010	-0.010
	2	0.135	0.126	0.135	-0.926	0.092	0.091		2	-0.023	-0.021	-0.023	0.155	-0.015	-0.015
	3	0.172	0.162	0.172	-1.419	0.090	0.088		3	-0.019	-0.018	-0.019	0.158	-0.010	-0.010
OWT5	1	0.052	0.053	0.052	0.052	-0.366	0.052	TF5	1	-0.138	-0.141	-0.138	-0.139	0.982	-0.140
	2	0.173	0.177	0.173	0.174	-0.662	0.176		2	-0.291	-0.296	-0.290	-0.291	1.109	-0.294
	3	0.389	0.391	0.389	0.390	-0.865	0.390		3	-0.696	-0.699	-0.696	-0.696	1.546	-0.697

Besides, in S1, when WT1 increased by 1%, the selection probability of A5 and A6 increased by 0.417% and 0.410%, respectively. It also shows that, in slight storm, walking time to bus stop cannot be ignored. Overall, to reduce ridership decline of public transportation, WT3 in all scenarios and WT1 in S1 are of significant importance.

4.3.4. Waiting Time. Similar to walking time, the probability of one alternative declines as its waiting time becomes longer. However, compared with walking time, the impact would be much less. Specifically, for a 1% increase in waiting time (OWT), the probabilities of A5 and A6 are expected to increase by 0.197% and 0.39% at most, while for 1% increase in walking time (WT), the increase can reach 0.97% and 0.95%, respectively. OWT1 in S1, OWT2 in S1 and S2, and OWT5 in S3 which obviously affect the Cancel Trip should be of concern.

4.3.5. In-Vehicle Time. Similar to walking time and waiting time, the rise of in-vehicle time decreases the choice probability of the respective alternative and increases that of CT. From the results in Table 5, the choice probability of CT is sensitive to changes of IT1 and IT2 in S1, IT2 and IT5 in S2, and IT5 in S3. These results showed that transit commuters can benefit most from reducing the in-vehicle time of A1 and A2 before the weather becomes terrific.

4.3.6. Fare. Since fare of A1~A4 in our designed scenarios is constant, the sensitivity for them does not make any sense. The fare of taxi which is related to travel distance and in-vehicle time road reflect the severity of road waterlogging, leading detouring, and congestion. It can be treated as a variable combination of environment and cost. The impacts of fare are opposite to other variables. Specifically, a 1% increase in taxi fares will increase the selection probability of taxi by 1.5%, while other alternatives will decrease by nearly 0.7% in S3.

Overall, when we compared the absolute direct elastic which measures the impact of a change of an attribute of alternative i on the choice probability of the same alternative, for OWT and IT, there is the same order from high to low: A1, A2, A4, A5, and A3, while for WT, the order becomes A3, A1, A4, A2, and A5. For bus services A1 and A2, the performance of the waiting time and in-vehicle time is needed to be improved. For metro service A3, the key point changes to the walking time.

Since one of our goals is to reduce the choice of Cancel Trip in storm weather, a comparison of Absolute Cross Elasticities to Cancel Trip (A6) is made. We found that the key factors (value > 0.15) vary with scenarios. In S1, the main factors with the order from high to low are WT3, WT1, IT5, IT2, IT1, OWT5, OWT2, and OWT1, while in S2, it becomes WT3, IT2, IT5, IT1, IT4, WT2, WT1, and OWT2, and in S3, it changes to WT3, IT5, BF5, OWT5, IT2, IT4, and IT3. It shows that WT3 is the most critical issue in all scenarios. Besides, WT, IT, and OWT impacts are under balance in S1, but when the condition gets worse, influence of IT of bus

may take the major part, and later when it becomes a terrible storm, attributes of taxi IT5, BF5, and OWT5 have a great influence on Cancel Trip.

In summary, increase of attributes of one alternative, such as walking time, waiting time, and in-vehicle time, has negative impacts on probability choice but positive impacts on other alternatives, while fare is the opposite. Walking time is the key factor for all alternatives. In-vehicle time is similar to waiting time to some extent, with a larger value of sensitivity. When the weather is not so bad, reducing time in all aspects can take effects. When it gets worse, reducing in-vehicle time may still work. If the weather becomes terrible, metro and taxi rather than bus are the preferred choices for most commuters. Reducing the walking time to metro or guaranteeing high-quality taxi service may be feasible.

5. Simulation

In this section, based on information of the case and data analysis, we propose possible plans for transit operators accordingly and simulate the ridership share dynamics in different scenarios. According to the simulation results, we recommend suitable measures in this case for transit operators in different conditions.

5.1. Simulation Method. Once the choice model has been estimated, we hope to use it to simulate the response of transit commuters to emergency measures and evaluate the performance of schemes in promoting transit ridership. The method of simulation is as follows.

Consider a choice model p_{jk}^n providing the probability that individual n chooses alternative j within the choice set C_n in Scenario k , given the explanatory variables X_n^k . To calculate the ridership share in the population of size N , a sample of N_s individuals is drawn. As it is rarely possible to draw from the population with equal sampling probability, it is assumed that stratified sampling has been used and that each individual in the sample is associated with a weight w_n correcting for sampling biases. To achieve it, w_n can be presented by a ratio of proportion in sample and target population for personal attributes of individual n . In this case, we use monthly income distribution in the population derived from official demographic information in Shanghai Statistical Yearbook to correct bias. If $MI_n = c$ for individual n , w_n can be estimated as

$$w_n = \frac{\rho'_c}{\rho_c} \quad (4)$$

$$MI_n = c,$$

where ρ'_c and ρ_c are proportion target population and in sample, respectively. In this case, ρ'_c came from the official demographic information in Shanghai Statistical Yearbook 2016.

The weights are normalized such that

$$N_s = \sum_{n=1}^{N_s} W_n. \quad (5)$$

The normalized weight of individual n , W_n , is formulated as

$$W_n = \frac{w_n N_s}{\left[\sum_{n'=1}^{N_s} w_{n'} \right]}. \quad (6)$$

An estimator of the ridership share of alternative j in the population R_j^k is

$$R_j^k = \sum_{n=1}^{N_s} W_n P_n^k(j|X_n^k). \quad (7)$$

5.2. Simulation of Possible Plans. According to the above analysis, it was found that increase in walking time to metro station, waiting time at bus stop, and in-vehicle time are the main obstacles for transit commuters in storm weather. To cope with the impacts of the storm on transit service, three feasible emergencies are proposed:

P1: shuttle bus, connecting bus stations and metro station near the residential area

P2: information announcement, such as real-time arrival and transfer information for bus service

P3: route adjustment, slightly adjusting the bus routine to avoid serious section with waterlogging and congestion

Parameters after taking emergence plan are different from original plan. Therefore, reasonable setting of all alternatives is shown in Table 6 and explained as follows:

P1. Due to shuttle bus between bus stops and metro station, there exist new choice A7 for commuters which is taking shuttle bus and then transferring by metro, noted as S+M. Compared to A3, passenger who chooses A7 must transfer one time which increases IT and OWT to gain shorter WT. Compared to the existing transfer plan A4, since shuttle only services for connecting nearby bus stops and metro station and can avoid road waterlogging, choosing A7 can have shorter OWT and IT and similar WT for commuters. In emergence plan P1, passengers have alternatives A1 to A7.

P2. If real-time information for bus service is available for passengers, according to previous studies, the expected waiting time can decrease up to 30% [37]. Therefore, OWT in A1, A2, and A4 is settled as 70% of values after the information announcement is taken.

P3. Temporarily change the route of bus lines to avoid road sections with serious road waterlogging section, but lengthen bus line slightly. In this case, according to investigation and data analysis, nearly 50% of delay occurs in road waterlogging sections; we may expect that adjust bus line can reduce 50% delay. Therefore, IT of A1 and A2 will have a significant decline in P3.

Besides, we combined the changes due to different plans together to form four combined plans: P1 + P2, P1 + P3, P2 + P3, and P1 + P2 + P3. Therefore, in the next, 7 plans are simulated and compared in each scenario.

5.3. Simulation Results Analysis. The detailed simulation results of ridership share of alternatives with adoption of plans are shown in Table 7. The aggregation of passengers who choose the main mode as bus (DB, B + B), metro (M, B + M, S + M), and public transportation, is listed as Bus, Metro, and PT, respectively. Value of Bus/Metro reflects the ridership ratio of Bus to Metro in different situations. Further, Table 8 exhibits the ridership share changes to normal weather and the benefits of strategies with a combination of plans. Red and green indicate positive values and negative values, respectively.

Generally, emergency plans will increase the ridership of public transportation and decrease that of taxi and cancel. In single plan strategy (1-P), P1 provides S + M mode which decreases walking time compared to M and waiting time compared to B + M, while P2 aims to decrease bus waiting time and P3 tries to reduce in-vehicle time. Obviously, P1 may increase the attraction of Metro, and P2 and P3 can improve bus ridership, resulting in low Bus/Metro in P1. Besides, we note that, with the deterioration of weather, more benefits can be gained for one certain plan, but the effect varies on plans. Considering a single plan, P1, P2, and P3 perform best in S3, S1, and S2, respectively. Analysis indicates that P2 has a general effect on all scenarios, P1 and P3 are more suitable in a worse situation in the storm.

In 2-Plan strategy, P2 + P3 which is concentrated in bus service performs better than the balanced solutions P1 + P2 and P1 + P3 in all scenarios. We must emphasize that P2 + P3 can greatly increase bus ridership in storm weather, but the cost of reducing metro passengers needs to be carefully considered in the final decision. When it comes to multiplan strategy, as one may expect, the more the plans adopt, the better the performance is. 3-P strategy adopting all three plans is more effective than 2-P strategy, which is also better than 1-P strategy. However, with the adoption of more plans, the margin benefit that brought by one more plan declines. Taking S3 as an example, 1-P strategy can bring 11.42% ridership share rise compared to no plan taking; the margin benefit of 1-P is 11.42%. If one more plan is adopted, in 2-P strategy the rise can reach 19.44% which is much higher than 1-P, but the margin benefit of 2-P decreases to 8.02%. In 3-P, the value further declines to 4.79%. Therefore, considering the margin benefit of emergency plans, we recommend 1-P strategy P2 or original plan P0 in scenario S1, 2-P strategy P2 + P3 in scenario S2, and 3P strategy P1 + P2 + P3 in scenario S3.

Overall, all emergency plans are effective in increasing public transportation ridership and decreasing cancel choice probability. In scenario S1, the impact of the light storm on public transportation is not serious. Therefore, taking no measure or just publishing bus arrival information can be acceptable in this condition. When weather gets worse, P2 + P3 can guarantee bus service and maintain bus

TABLE 6: Parameters of emergence plans.

Performance of services	Scenarios	Alternatives												
		A1	A1-P2	A1-P3	A1-P23	A2	A2-P2	A2-P3	A2-P23	A3	A4	A4-P2	A7-P1	A5
Walking time (minute)	S0	9	9	9	9	5	5	5	5	15	6	6	6	0
	S1	10	10	10	10	5	5	5	5	20	7	7	7	0
	S2	12	12	12	12	6	6	6	6	25	8	8	8	0
	S3	15	15	15	15	8	8	8	8	30	10	10	10	0
Waiting time (minute)	S0	15	10	15	10	12	8	12	8	3	15	11	8	5
	S1	20	14	20	14	15	11	15	11	3	15	11	8	10
	S2	30	21	30	21	25	18	25	18	5	20	16	10	20
	S3	45	32	45	32	40	28	40	28	9	30	23	14	30
In-vehicle travel time (minute)	S0	25	25	30	30	25	25	30	30	15	25	25	25	15
	S1	30	30	32	32	30	30	32	32	15	30	30	25	20
	S2	40	40	35	35	40	40	35	35	15	35	35	27	25
	S3	60	60	42	42	60	60	42	42	20	45	45	35	30
Overall travel time (minute)	S0	49	44	54	49	42	38	47	43	33	46	42	39	20
	S1	60	54	62	56	50	46	52	48	38	52	48	40	30
	S2	82	73	77	68	71	64	66	59	45	63	59	45	45
	S3	120	107	102	89	108	96	90	78	59	85	78	59	60
Transfer	S0-S3	0	0	0	0	1	1	1	1	0	1	1	1	0

TABLE 7: Simulation results in different scenarios.

Scenario	Plan	Public transportation								Taxi	Cancel	
		DB (%)	B + B (%)	Bus (%)	M (%)	B + M (%)	S + M (%)	Metro (%)	Bus/ Metro	PT (%)	TX (%)	CT (%)
S1	P0	24.12	28.82	52.94	21.02	7.28	—	28.29	1.871	81.24	12.42	6.34
	P1	21.82	26.26	48.08	19.01	6.59	8.88	34.48	1.394	82.56	11.55	5.90
	P2	27.74	30.55	58.30	17.59	7.73	—	25.32	2.303	83.61	10.85	5.54
	P3	26.15	31.09	57.24	18.92	6.54	—	25.46	2.248	82.70	11.45	5.84
	P1 + P2	25.73	28.57	54.30	16.31	7.17	6.69	30.18	1.799	84.48	8.95	4.57
	P1 + P3	26.15	29.01	55.16	16.58	5.74	6.81	29.13	1.894	84.29	9.11	4.66
	P2 + P3	29.78	32.66	62.44	15.68	6.88	—	22.57	2.767	85.01	8.96	4.57
	P1 + P2 + P3	27.84	30.77	58.60	14.66	6.44	6.02	27.11	2.161	85.72	8.23	4.20
S2	P0	10.70	18.58	29.28	22.39	11.03	—	33.43	0.876	62.70	20.89	16.41
	P1	8.64	15.20	23.84	18.08	8.92	16.97	43.97	0.542	67.81	18.02	14.18
	P2	14.17	22.59	36.75	18.41	12.45	—	30.86	1.191	67.61	18.14	14.24
	P3	17.59	25.10	42.69	17.51	8.61	—	26.12	1.634	68.81	17.47	13.71
	P1 + P2	12.08	19.59	31.67	15.70	10.64	12.94	39.28	0.806	70.95	16.26	12.79
	P1 + P3	12.48	20.20	32.69	16.22	8.00	13.37	37.59	0.870	70.27	16.64	13.09
	P2 + P3	22.11	29.01	51.12	13.66	9.23	—	22.89	2.233	74.01	14.56	11.42
	P1 + P2 + P3	19.63	26.11	45.74	12.13	8.21	10.00	30.33	1.508	76.07	13.40	10.53
S3	P0	2.94	5.42	8.36	14.91	7.14	—	22.05	0.379	30.42	31.07	38.51
	P1	2.14	4.00	6.14	10.82	5.19	19.69	35.70	0.172	41.84	25.94	32.22
	P2	5.13	8.69	13.82	12.72	9.81	—	22.53	0.614	36.35	28.42	35.23
	P3	8.32	15.03	23.35	11.49	5.49	—	16.98	1.376	40.33	26.66	33.01
	P1 + P2	4.02	6.98	11.01	9.97	7.71	15.95	33.64	0.327	44.64	24.69	30.67
	P1 + P3	4.18	7.24	11.42	10.36	4.97	16.58	31.91	0.358	43.34	25.27	31.39
	P2 + P3	12.94	21.46	34.40	8.74	6.72	—	15.46	2.226	49.86	22.41	27.73
	P1 + P2 + P3	11.00	18.62	29.62	7.43	5.72	11.88	25.03	1.183	54.65	20.25	25.11

P0 is the plan without taking any emergency plans.

ridership share in commuting in S2. If things get worse, in S3, one way is trying our best to guarantee both metro and bus service, adopting all three plans at huge cost and difficult to carry out in practice. Besides, since the metro service is less affected than bus service in the storm, giving up the main

bus service and providing shuttle buses to connect residential areas and metro stations to guarantee the accessibility of metro could be another possible choice in S3. However, it has some limitations, only suitable for areas with highly developed metro networks.

TABLE 8: Performances of plans in ridership compared with original plan.

Scenario	Plan	Bus (%)	Metro (%)	Public transportation (%)	Taxi (%)	Cancel (%)	Strategy	Transit share increase (%)	Margin benefit (%)
S1	P1	-4.86	6.18	1.32	-0.88	-0.44	1-P	2.37	2.37
	P2	5.35	-2.98	2.37	-1.57	-0.80			
	P3	4.30	-2.83	1.46	-0.97	-0.50			
	P1 + P2	1.35	1.88	3.24	-3.47	-1.77	2-P	3.77	1.40
	P1 + P3	2.22	0.83	3.05	-3.31	-1.68			
	P2 + P3	9.50	-5.73	3.77	-3.46	-1.77	3-P	4.48	0.71
P1 + P2 + P3	5.66	-1.18	4.48	-4.19	-2.14				
S2	P1	-5.44	10.54	5.10	-2.87	-2.23	1-P	6.11	6.11
	P2	7.48	-2.56	4.91	-2.75	-2.16			
	P3	13.41	-7.30	6.11	-3.42	-2.70			
	P1 + P2	2.39	5.85	8.25	-4.63	-3.62	2-P	11.31	5.20
	P1 + P3	3.41	4.16	7.57	-4.25	-3.32			
	P2 + P3	21.84	-10.53	11.31	-6.33	-4.99	3-P	13.37	2.06
P1 + P2 + P3	16.46	-3.09	13.37	-7.49	-5.88				
S3	P1	-2.23	13.65	11.42	-5.13	-6.29	1-P	11.42	11.42
	P2	5.46	0.47	5.93	-2.65	-3.28			
	P3	14.99	-5.08	9.91	-4.41	-5.50			
	P1 + P2	2.64	11.58	14.22	-6.39	-7.84	2-P	19.44	8.02
	P1 + P3	3.06	9.86	12.92	-5.80	-7.12			
	P2 + P3	26.04	-6.60	19.44	-8.66	-10.78	3-P	24.23	4.79
P1 + P2 + P3	21.25	2.98	24.23	-10.82	-13.40				

The values in the tables are presented by the ratio to transit ridership under normal weather.

6. Discussion and Conclusions

Our paper aims to describe travel behavior dynamics of transit commuters in the storm and provide possible emergency plans for transit operator/agency to guarantee transit trips of commuters. First of all, this study conducted RP and SP surveys on the impact of storms on travel behavior, focusing on the relationship between travel choice changes of transit commuters and transit service performance. To achieve it, we established a choice model by considering gender, age, income, walking time, waiting time, in-vehicle time, crowdedness, transfer, and fare. From results' analysis, it shows that, in storm weather, walking time, waiting time, and in-vehicle time have obvious negative impacts on the choice probability of alternatives; high-income commuters prefer Bus + Bus, Taxi, and Cancel Trip; age, gender, and crowdedness have limited impacts on storm weather. Through sensitivity analysis, we further found that, in a light storm, reducing travel time, including walking time, waiting time, in-vehicle time, can have effects. When it gets worse, the decline of in-vehicle time may be more sensitive, and therefore Metro becomes the most popular choice. When the weather is terrible, walking becomes more difficult, and thus most commuters give up Metro and choose Taxi or Cancel Trip. Accordingly, three possible emergency plans, information announcement, route adjustment, and shuttle bus to metro station, are simulated in different scenarios. The simulation results can provide references for public transit agencies by suggesting important implications for future public transport development. A strategic proposal of transit agency is to work out detailed emergency plans in conjunction with meteorological departments, road management bureaus, telecommunications companies, and transit

operators, including information release, operation adjustment, and traffic management. Specifically, real-time information should be provided, including weather condition, the emergency state and duration, temporary route plan and timetable, and estimated delay or arrival time at stops, to guide passengers to adjust their travel plans under all scenarios in the storm. Further, according to weather data from the meteorological department and risk analysis of road congestion by road management department, bus operators can adjust bus routes in time to avoid high-risk sections. When weather getting worse and maintenance of normal bus operation becomes difficult, metro system which is less affected by heavy storms can be a reliable substitute for bus service. Thus, it is necessary to put forward a temporary feeder bus scheme, including temporary route, bus stops, and schedule, to bridge the residential areas and metro stations.

Despite these promising implications, there are still some limitations that need to be addressed. Firstly, due to lack of enough data to assess time reliability of alternatives in storm, this study does not directly consider the reliability which may play important roles in making travel decisions. Secondly, to narrow the scope of the study, recreational travel which is more likely to be affected than commuter travel is out of consideration in this study. Thirdly, to reduce the complexity of methodology and focus on transit service, private car, which has some differences from the alternative Taxi, was not listed as an alternative for transit commuters. Thus, more issues should be taken into account to provide a refined profile of how adverse weather affects travel choice. Finally, this study exhibits the adaptation of transit commuter to the storm and provides possible countermeasures for transit agency by the case study in a specific small zone. A general application would require further verification in more areas. Further research should be undertaken to

investigate the spatiotemporal heterogeneity of the influence by adverse weather, simulate transit ridership dynamics in different areas over time, and evaluate the performance of feasible emergency plans.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Teng, Bo, and Zhang contributed to study conception and design. Teng and Bo contributed to data collection. Zhang and Bo contributed to methodology, analysis and interpretation of results, and draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

This work was supported by the Ministry of Transport of the People's Republic of China Science and Technology Research Foundation for Transportation (no. 2015318221020).

References

- [1] M. Hofmann and M. O'Mahony, "The impact of adverse Weather Conditions on Urban Bus Performance measures," in *Proceedings of the 2005 IEEE Intelligent Transportation Systems*, pp. 84–89, Vienna, Austria, 2005.
- [2] C. B. Field, V. R. Barros, and D. J. Dokken, *IPCC, 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Publication Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 2014.
- [3] M. Cools, E. Moons, L. Creemers, and G. Wets, "Changes in travel behavior in response to weather conditions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2157, no. 1, pp. 22–28, 2010.
- [4] E. Wets, B. van Wee, and K. Maat, "Commuting by bicycle: an overview of the literature," *Transport Reviews*, vol. 30, no. 1, pp. 59–96, 2010.
- [5] A. J. Khattak and A. De Palma, "The impact of adverse weather conditions on the propensity to change travel decisions: a survey of Brussels commuters," *Transportation Research Part A: Policy and Practice*, vol. 31, no. 3, pp. 181–203, 1997.
- [6] M. J. Koetse and P. Rietveld, "The impact of climate change and weather on transport: an overview of empirical findings," *Transportation Research Part D: Transport and Environment*, vol. 14, no. 3, pp. 205–221, 2009.
- [7] Z. Guo, N. H. M. Wilson, and A. Rahbee, "Impact of weather on transit ridership in Chicago, Illinois," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2034, no. 1, pp. 3–10, 2007.
- [8] A. Singhal, C. Kamga, and A. Yazici, "Impact of weather on urban transit ridership," *Transportation Research Part A: Policy and Practice*, vol. 69, pp. 379–391, 2014.
- [9] S. Tao, J. Corcoran, F. Rowe, and M. Hickman, "To travel or not to travel: 'Weather' is the question. Modelling the effect of local weather conditions on bus ridership," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 147–167, 2018.
- [10] L. Böcker, M. Dijst, and J. Faber, "Weather, transport mode choices and emotional travel experiences," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 360–373, 2016.
- [11] M. Hyland, C. Frei, A. Frei, and H. S. Mahmassani, "Riders on the storm: exploring weather and seasonality effects on commute mode choice in Chicago," *Travel Behaviour and Society*, vol. 13, pp. 44–60, 2018.
- [12] S. Datla and S. Sharma, "Impact of cold and snow on temporal and spatial variations of highway traffic volumes," *J. Journal of Transport Geography*, vol. 16, no. 5, pp. 358–372, 2008.
- [13] J. Anta, J. B. Pérez-López, A. Martínez-Pardo, A. Novales, and M. Orro, "Influence of the weather on mode choice in corridors with time-varying congestion: a mixed data study," *Transportation*, vol. 43, no. 2, pp. 337–355, 2016.
- [14] M. Kyte, Z. Khatib, P. Shannon, and F. Kitchener, "Effect of weather on free-flow speed," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1776, no. 1, pp. 60–68, 2001.
- [15] B. L. Kitchener, K. G. Byrne, R. B. Copperman, S. M. Hennessy, and N. J. Goodall, "An investigation into the impact of rainfall on freeway traffic flow," in *Proceedings of the 83rd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2004.
- [16] R. B. Chen and H. S. Mahmassani, "Let it rain: weather effects on activity stress and scheduling behavior," *Travel Behaviour and Society*, vol. 2, no. 1, pp. 55–64, 2015.
- [17] C. Liu, Y. O. Susilo, and A. Karlström, "Investigating the impacts of weather variability on individual's daily activity-travel patterns: a comparison between commuters and non-commuters in Sweden," *Transportation Research Part A: Policy and Practice*, vol. 82, pp. 47–64, 2015.
- [18] Y. Motoaki and R. A. Daziano, "A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand," *Transportation Research Part A: Policy and Practice*, vol. 75, pp. 217–230, 2015.
- [19] S. Saneinejad, M. J. Roorda, and C. Kennedy, "Modelling the impact of weather conditions on active transportation travel behaviour," *Transportation Research Part D: Transport and Environment*, vol. 17, no. 2, pp. 129–137, 2012.
- [20] D. Akin, V. P. Sisiopiku, and A. Skabardonis, "Impacts of weather on traffic flow Characteristics of urban freeways in Istanbul," *Procedia-Social and Behavioral Sciences*, vol. 16, pp. 89–99, 2011.
- [21] H. Perrin, P. T. Martin, and B. G. Hansen, "Modifying signal timing during inclement weather," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1748, no. 1, pp. 66–71, 2001.
- [22] X. Zhang and M. Chen, "Quantifying the impact of weather events on travel time and reliability," *Journal of Advanced Transportation*, vol. 2019, 2019.
- [23] Y. Zou, Y. Zhang, and K. Cheng, "Exploring the impact of climate and extreme weather on fatal traffic accidents," *Sustainability*, vol. 13, no. 1, p. 390, 2021.
- [24] J.-L. Madre, K. W. Axhausen, and W. Brög, "Immobility in travel diary surveys," *Transportation*, vol. 34, no. 1, pp. 107–128, 2007.
- [25] P. Arana, S. Cabezudo, and M. Peñalba, "Influence of weather conditions on transit ridership: a statistical study using data

- from Smartcards,” *Transportation Research Part A: Policy and Practice*, vol. 59, pp. 1–12, 2014.
- [26] M. Sabir, J. Van Ommeren, M. Koetse, and P. Rietveld, “Adverse weather and commuting speed,” *Networks and Spatial Economics*, vol. 11, no. 4, pp. 701–712, 2011.
- [27] S. Rietveld, J. Corcoran, M. Hickman, and R. Stimson, “The influence of weather on local geographical patterns of bus usage,” *Journal of Transport Geography*, vol. 54, pp. 66–80, 2016.
- [28] M. Stimson, C. Morency, B. Agard, E. Descoimps, and J. S. Marcotte, “Using smart card data to assess the impacts of weather on public transport user behavior,” in *Proceedings of the Conference on Advanced Systems for Public Transit-CASPT12*, pp. 23–27, Santiago, Chile, July 2012.
- [29] N. S. Ngo, “Urban bus ridership, income, and extreme weather events,” *Transportation Research Part D: Transport and Environment*, vol. 77, pp. 464–475, 2019.
- [30] J. Wang and Z. Zhu, *Shanghai Statistical Yearbook*, China Statistics Press, Beijing, China, 2016.
- [31] R. Ashalatha, V. S. Manju, and A. B. Zacharia, “Mode choice behavior of commuters in Thiruvananthapuram city,” *Journal of Transportation Engineering*, vol. 139, no. 5, pp. 494–502, 2012.
- [32] P. C. Devarasetty, M. Burriss, and W. Douglass Shaw, “The value of travel time and reliability-evidence from a stated preference survey and actual usage,” *Transportation Research Part A: Policy and Practice*, vol. 46, no. 8, pp. 1227–1240, 2012.
- [33] C. Thrane, “Examining tourists’ long-distance transportation mode choices using a Multinomial Logit regression model,” *Tourism Management Perspectives*, vol. 15, pp. 115–121, 2015.
- [34] A. M. Zanni, M. Goulden, and T. Ryley, “Improving scenario methods in infrastructure planning: a case study of long distance travel and mobility in the UK under extreme weather uncertainty and a changing climate,” *Technological Forecasting and Social Change*, vol. 115, pp. 180–197, 2017.
- [35] A. M. Dingwall and T. J. Ryley, “The impact of extreme weather conditions on long distance travel behaviour,” *Transportation Research Part A: Policy and Practice*, vol. 77, pp. 305–319, 2015.
- [36] J. Wu and H. Liao, “Weather, travel mode choice, and impacts on subway ridership in Beijing,” *Transportation Research Part A: Policy and Practice*, vol. 135, pp. 264–279, 2020.
- [37] H. Lu, P. Burge, C. Heywood et al., “The impact of real-time information on passengers’ value of bus waiting time,” *Transportation Research Procedia*, vol. 31, pp. 18–34, 2018.

Research Article

Resilience-Based Optimization of Postdisaster Restoration Strategy for Road Networks

Xinhua Mao ¹, Jibiao Zhou ², Changwei Yuan,³ and Dan Liu¹

¹College of Transportation Engineering, Chang'an University, Xi'an 710064, China

²College of Transportation Engineering, Tongji University, Shanghai 201804, China

³Engineering Research Center of Highway Infrastructure Digitalization, Ministry of Education, Xi'an 710064, China

Correspondence should be addressed to Jibiao Zhou; zhoujibiao@tongji.edu.cn

Received 11 June 2020; Revised 5 September 2020; Accepted 2 February 2021; Published 12 February 2021

Academic Editor: Yuan Gao

Copyright © 2021 Xinhua Mao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work proposes a framework for the optimization of postdisaster road network restoration strategies from a perspective of resilience. The network performance is evaluated by the total system travel time (TSTT). After the implementation of a postdisaster restoration schedule, the network flows in a certain period of days are on a disequilibrium state; thus, a link-based day-to-day traffic assignment model is employed to compute TSTT and simulate the traffic evolution. Two indicators are developed to assess the road network resilience, i.e., the resilience of performance loss and the resilience of recovery rapidity. The former is calculated based on TSTT, and the latter is computed according to the restoration makespan. Then, we formulate the restoration optimization problem as a resilience-based bi-objective mixed integer programming model aiming to maximize the network resilience. Due to the NP-hardness of the model, a genetic algorithm is developed to solve the model. Finally, a case study is conducted to demonstrate the effectiveness of the proposed method. The effects of key parameters including the number of work crews, travelers' sensitivity to travel time, availability of budget, and decision makers' preference on the values of the two objectives are investigated as well.

1. Introduction

Road infrastructure forms the backbone of transport activities, which plays an important role in boosting economic development and increasing accessibility. Due to extreme weather, road networks are inclined to suffer from the disruptions caused by natural disasters such as floods, typhoons, and landslides. Hence, restoration activities are needed to recover the networks as soon as possible. However, the available budget in a short time after the disaster cannot afford to repair all the disrupted road segments. There is an increasing demand for making a cost-effective postdisaster road network restoration strategy (RNRS) [1, 2], which refers to determining the road segments to be repaired and the restoration time sequence.

In recent years, resilience has attracted growing attention in the road infrastructure management field [3–6]. How to keep the infrastructure networks at a high level of resilience

has become a challenge for transportation agencies. The concept of road network resilience is defined as the ability to absorb disruptive events and recover to normal operation within a reasonable period of time [7, 8]. Based on the definition, a variety of metrics have been introduced to assess the resilience of an infrastructure network. This assessment is indispensable for the optimization of a resilient infrastructure system. Despite the wide range of studies focusing on the resilience-based RNRS, there are still two gaps to be filled. Firstly, it is rare in the literature to investigate a joint optimization of budget allocation and restoration scheduling for an effective RNRS. Secondly, most of the previous studies apply user equilibrium (UE) models to evaluate the total system travel time (TSTT) during the restoration process. UE models assume that the network flow patterns achieve an equilibrium state overnight given new network conditions, which cannot reflect the day-to-day traffic dynamics.

In view of this, this study aims to investigate the optimal postdisaster restoration problem for road networks from the perspective of resilience considering the day-to-day network flow fluctuation. The main contribution of our work is to build a resilience-based bi-objective mixed integer programming model combined with a link-based day-to-day traffic assignment model to determine the optimal RNRS based on the tradeoff between the maximal resilience of performance loss and the maximal resilience of recovery rapidity. With our proposed method, decision makers would determine a set of prioritized road segments to be restored and the optimal time sequence of the restoration tasks.

The remainder of this paper is organized as follows. Section 2 introduces the literature review. Section 3 presents two resilience metrics, proposes a link-based day-to-day dynamics model, and develops a resilience-based optimization model for the postdisaster RNRS. Section 4 proposes a genetic algorithm to solve the optimization model. In Section 5, we employ a case study to validate our proposed method. Conclusions and future work are discussed in Section 6.

2. Literature Review

Due to the increasing natural hazards, recovering damaged road networks in a resilient manner has attracted growing attention in recent years. The existing studies in this realm can be classified into two categories, i.e., the budget allocation problem and the restoration scheduling problem. The budget allocation problem aims to determine the road segments to be repaired from a set of damaged road segments with budget limits. For instance, Liu et al. [9] developed a two-stage stochastic programming model to allocate limited retrofit budget over multiple road bridges to maximize the resilience and robustness of the entire road network. In order to mitigate the predisaster risk and improve network resilience, Zhang and Wang [10] proposed a resilience-based optimization model to identify the road network retrofit projects. As for the restoration scheduling problem, it focuses on identifying the time sequence of restoration activities. Bocchini and Frangopol [11] developed a multiobjective optimization model aiming at maximal resilience, minimal restoration time, and minimal restoration cost to formulate the restoration scheduling problem for road-bridge networks after an earthquake. Li et al. [12] established a resilience-based bilevel programming model to investigate the optimization of the road network recovery strategy under uncertainty aiming at the maximization of network resilience. Generally, traffic dynamics is an important issue that needs to be considered in the RNRS problem. Most of the previous studies employ the classic UE traffic assignment model to estimate TSTT assuming that the traffic flow patterns across the road network are always in an equilibrium state [13]. The only exception is Nogal et al. [14], where a new dynamic equilibrium-restricted assignment model is presented to simulate the postdisaster day-to-day flow evolution process. According to the work of De Palma and Rochat [15], travelers are highly sensitive in their route choice behaviors to the occurrence of an event in the

road network. Hence, travelers reselect their routes shortly after the restoration activity of any disrupted road segment is completed, which makes the network flows evolve constantly to reach a new equilibrium state within a period of time from the old equilibrium state. Simulation on a 3×3 grid network with 9 nodes, 12 links, and 6 routes by He et al. [16] indicated that all links have some flow fluctuations after a 50% capacity reduction and it takes about 20 days for the network flow patterns to achieve a new equilibrium state. This fluctuating traffic flow pattern is defined as partial UE (PUE) by Sumalee and Watling [17], namely, the network flows in a certain period of time are not on an equilibrium state. Therefore, it is problematic to use the UE model to evaluate TSTT during the network restoration process. The day-to-day model can better capture the PUE [18], which is appropriate for dealing with TSTT calculation in the road network restoration context.

Day-to-day traffic assignment models are capable of predicting day-to-day traffic fluctuations and the evolution process itself when the traffic network is perturbed by unexpected events, construction actions, and traffic controls. As noted by Watling and Hazelton [19], day-to-day traffic models have great flexibility, which accommodates a wide range of behavior rules, levels of aggregation, and traffic modes. The first effort can be attributed to Horowitz [20], who proposed a discrete time day-to-day dynamic traffic model for a two-link network from the perspective of system-optimal principle. Generally, two types of day-to-day models have been studied in the literature, i.e., continuous time models and discrete time models, and each type can be subcategorized into two groups, i.e., deterministic models and stochastic models [21, 22]. The continuous time models utilize differential equations to describe traffic evolution based on the assumption that travelers have a perfect perception of travel cost, which can capture the mathematical features in traffic transition [23–25]. The discrete time models assume that travelers repeat their route choice behavior each day following the traffic condition, which is more suited to the real world as mentioned by Watling and Hazelton [19]. Since the uncertainty associated with the random nature exists in the traffic evolution process, stochastic day-to-day traffic models get more attention from scholars than the deterministic day-to-day models. Most stochastic day-to-day traffic models follow Markov processes, which predict the traffic state by calculating transition matrices based on the previous traffic flow patterns [26–29]. Due to the advantages claimed above, day-to-day traffic models recently have been used in combination with other methods to solve some practical transportation problems. For example, Liu et al. [30] combined a path-based day-to-day traffic model with a robust optimization method to investigate distance-based congestion pricing problems. Faturechi and Miller-Hooks [18] developed a methodology framework composed of three-stage stochastic mathematical programming and a day-to-day traffic model to simulate postdisaster travel time resilience of roadway networks. However, most of the previous day-to-day traffic models are path-based and assume that travelers have infinite memories.

3. Formulation of the Problem

3.1. *Assumptions and Notations.* This study makes the following assumptions:

- (1) The travel demands between each origin-destination (OD) pair keep constant during the restoration process.
- (2) Each restoration activity is implemented by a single work crew. Once a restoration activity begins, the work crew has to finish the restoration activity prior to conducting the next restoration activity.
- (3) Each restoration activity only begins once during the entire restoration process, i.e., the restoration strategy is non-preemptive.
- (4) The capacity of each damaged road segment is reduced when the disaster occurs, and the capacity is not restored to the predisaster level until the damaged road segment is completely restored.
- (5) Due to sufficient preparations, the decision makers have a complete understanding of the damage state information for the road network; thus, the restoration time, the restoration cost, and other parameter values are known.

The notations used in this study are presented as follows.

3.1.1. Indices.

- (1) d : time period
- (2) i : road segment to be restored or restoration activity
- (3) r : work crew

3.1.2. Parameters.

- (1) I : total number of road segments to be restored or restoration activities
- (2) e_i : duration of restoration activity i , $i = 1, 2, \dots, I$
- (3) c_i : cost of restoration activity i , $i = 1, 2, \dots, I$
- (4) R : total number of work crews
- (5) B : availability of budget
- (6) M_{\max} : maximum acceptable makespan of the restoration

3.1.3. Variables.

- (1) C : total costs of the restoration
- (2) M : makespan of the restoration
- (3) $\pi(d)$: network performance on day d , $d = 1, 2, \dots, M_{\max}$
- (4) $T(d)$: total system travel time on day d , $d = 1, 2, \dots, M_{\max}$
- (5) R_p : resilience of performance loss
- (6) R_r : resilience of recovery rapidity

- (7) x_{id} : binary variable that represents $x_{id} = 1$ if road segment i is to be restored on day d ; otherwise, $x_{id} = 0$, $i = 1, 2, \dots, I$, $d = 1, 2, \dots, M_{\max}$
- (8) y_{ir} : binary variable that means $y_{ir} = 1$ if road segment i is to be restored by work crew r ; otherwise, $y_{ir} = 0$, $i = 1, 2, \dots, I$, $r = 1, 2, \dots, R$.

3.2. *Resilience Metrics.* Consider a road network, denoted as $G = (N, A)$, where N is the set of nodes and A is the set of road segments (or links). A disaster caused by extreme weather occurs on day $d = d_s$, which damages I road segments in the network. Hence, the network performance $\pi(d)$ drops to $\pi(d_s)$ from the predisaster network performance $\pi(d_0)$. Figure 1 plots the postdisaster network performance recovery process. Assume that the restoration is conducted immediately on day $d = d_s$, which aims to recover $\pi(d)$ to $\pi(d_0)$. Since the restoration budget cannot cover all the disrupted road segments, $\pi(d)$ is recovered to $\pi(d_s + M)$ when the budget is exhausted on day $d = d_s + M$. The restoration activities have to be completed with M_{\max} days. $\pi(d)$ will be further improved if more investments are available in the future.

Since there is a negative relationship between the road network performance and TSTT, we define $\pi(d)$ as the reciprocal of TSTT on day d divided by the reciprocal of the predisaster TSTT on day d_0 , as in equation (1). $\pi(d)$ is the ratio of $T(d_0)$ to $T(d)$, which reveals efficiencies in the use of a road network. The lower TSTT is, the higher $\pi(d)$ is.

$$\pi(d) = \frac{T(d)^{-1}}{T(d_0)^{-1}} = \frac{T(d_0)}{T(d)}. \quad (1)$$

It is clear that $\pi(d) \in (0, 1]$ and the predisaster network performance $\pi(d_0)$ equals to 1. When all the damaged road segments are restored, $\pi(d)$ will be recovered to 1.

Then, we develop two resilience metrics, i.e., the resilience of performance loss and resilience of recovery rapidity to evaluate the postdisaster network resilience [12, 31].

3.2.1. *Resilience of Performance Loss.* The shaded area in Figure 1 represents the performance loss. Figure 1 shows that the performance loss still exists after the restoration activities are completed on day $d = d_s + M$. For simplicity, we calculate the total performance loss TPL from day $d = d_s$ to day $d = d_s + M_{\max}$, which is formulated in the following equation:

$$\begin{aligned} \text{TPL} &= \int_{d_s}^{d_s + M_{\max}} [\pi(d_0) - \pi(x)] dx \approx M_{\max} \times \pi(d_0) \\ &\quad - \sum_{d=d_s}^{d_s + M_{\max}} \pi(d). \end{aligned} \quad (2)$$

Then, we formulate the resilience of performance loss as

$$R_p = 1 - \frac{M_{\max} \times \pi(d_0) - \sum_{d=d_s}^{d_s + M_{\max}} \pi(d)}{M_{\max} \times \pi(d_0)}, \quad (3)$$

where R_p represents the proportion of residual performance (i.e., overall performance minus lost performance) in overall

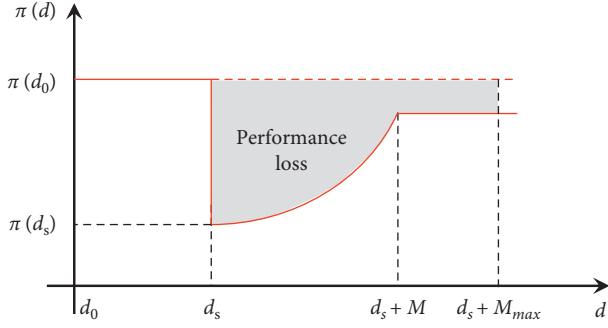


FIGURE 1: Postdisaster network performance recovery process.

performance within M_{\max} days and R_p reflects the function of the road network. The higher R_p is, the lower the total performance loss is and the less the traffic congestion is. $R_p \in [0, 1]$.

3.2.2. Resilience of Recovery Rapidity. Since recovery rapidity has a great priority during the restoration process, we employ the resilience of recovery rapidity as another resilience metric. We normalize and formulate the resilience of recovery rapidity as in the following equation:

$$R_r = \begin{cases} 1 - \frac{M}{M_{\max}}, & M \leq M_{\max}, \\ 0, & M > M_{\max}, \end{cases} \quad (4)$$

where R_r represents the speed of recovery. If $M > M_{\max}$, the recovery rapidity is too low, which is unacceptable. If $M \leq M_{\max}$, the higher R_r , the faster the road network can be repaired. $R_r \in [0, 1]$.

3.3. Link-Based Day-to-Day Dynamics Model. Before modeling the optimal RNRS problem, we first propose a link-based day-to-day dynamics model. For simplicity, we assume that restoration activities are conducted on day $d = d_s = 1$. The restoration activities will result in changes in the travelers' route choice behaviors, giving rise to day-to-day flow fluctuations. In order to estimate TSTT, a mechanism is needed to simulate the postdisaster network flow evolution trajectory [32]. Since the initial link flow pattern is easily observed by a well-designed field survey and each restoration activity is involved with a road segment, a link-based traffic model is more suitable to predict

the network flow evolution trajectory in the context of the road network restoration.

In a link-based day-to-day dynamics model, link flows on any day d , denoted by $\mathbf{q}(d) = (q_i, i \in N)^T$, tend to evolve towards the "target" link flows $\hat{\mathbf{q}}(d+1) = (\hat{q}_i, i \in N)^T$ on day $d+1$ at a rate of ν [16]. Hence, $\mathbf{q}(d+1)$ can be formulated as a weighted combination in the following equation:

$$\mathbf{q}(d+1) = \mathbf{q}(d) + \nu \cdot [\hat{\mathbf{q}}(d+1) - \mathbf{q}(d)]. \quad (5)$$

For a given $\mathbf{q}(d)$, $\hat{\mathbf{q}}(d+1)$ solves the following minimization problem:

$$\min_{\mathbf{Q} \in \Omega_d} \lambda \cdot \mathbf{h}(\mathbf{q}, d+1)^T \mathbf{Q} + (1-\lambda) \cdot D[\mathbf{q}(d), \mathbf{Q}], \quad (6)$$

$$\Omega_d = \{\mathbf{q}(d) | \mathbf{q}(d) = \Delta \mathbf{f}(d), \Lambda \mathbf{f}(d) = \Theta, \mathbf{f}(d) \geq 0\}, \quad (7)$$

where Ω_d is a feasibility vector set of link flows on day d . Δ is the link-path incidence matrix, $\Delta = (\Delta_{i,k,w}, i \in N, k \in K_w, w \in W)$. $\Delta_{i,k,w} = 1$ if link $i \in N$ lies on path $k \in K_w$, and $\Delta_{i,k,w} = 0$ otherwise. Λ is the OD pair-path incidence matrix, $\Lambda = (\Lambda_{k,w}, k \in K_w, w \in W)$. $\Lambda_{k,w} = 1$ if path $k \in K_w$ connects OD pair $w \in W$, and $\Lambda_{k,w} = 0$ otherwise. $\mathbf{f}(d)$ is the vector of path flows on day d . $\mathbf{f}(d) = (f_{k,w}(d), k \in K_w, w \in W)^T$. Θ is the vector of traffic demands between OD pairs. $\Theta = (\Theta_w, w \in W)^T$. λ ($0 \leq \lambda \leq 1$) is travelers' sensitivity to travel time. The larger the parameter value of λ , the more travelers will change their routes. $\mathbf{h}(\mathbf{q}, d+1)$ is the travel time perceived by the travelers on day $d+1$. Equation (6) is a weighted combination of $\min_{\mathbf{Q} \in \Omega_d} \mathbf{h}(\mathbf{q}, d+1)^T \mathbf{Q}$ and $\min_{\mathbf{Q} \in \Omega_d} D[\mathbf{q}(d), \mathbf{Q}]$. The former aims to minimize the total travel time of $\mathbf{q}(d+1)$, and the latter guarantees the minimal distance between $\mathbf{q}(d)$ and \mathbf{Q} .

$D[\mathbf{q}(d), \mathbf{Q}]$ can be calculated by

$$D[\mathbf{q}(d), \mathbf{Q}] = \sum_{i \in N} \int_{q_{i,d}}^{Q_{i,d+1}} [t_{i,d}(u) - t_{i,d}(q_{i,d})] du, \quad (8)$$

where $t_{i,d}(u)$ is the link travel time function.

As for $\mathbf{h}(\mathbf{q}, d+1)$ in a link-based day-to-day dynamics model, it can be formulated as a weighted average between their perceived travel time on day d and the experienced travel time on day d [33].

$$\mathbf{h}(\mathbf{q}, d+1) = \mu \cdot \mathbf{t}(\mathbf{q}, d) + (1-\mu) \cdot \mathbf{h}(\mathbf{q}, d). \quad (9)$$

Then, we expand equation (9) recursively as follows:

$$\begin{aligned} \mathbf{h}(\mathbf{q}, d+1) &= \mu \cdot \mathbf{t}(\mathbf{q}, d) + (1-\mu) \cdot \mathbf{h}(\mathbf{q}, d) \\ &= \mu \cdot \mathbf{t}(\mathbf{q}, d) + (1-\mu) [\mu \cdot \mathbf{t}(\mathbf{q}, d-1) + (1-\mu) \cdot \mathbf{h}(\mathbf{q}, d-1)] \\ &= \mu \cdot \mathbf{t}(\mathbf{q}, d) + \mu \cdot (1-\mu) \cdot \mathbf{t}(\mathbf{q}, d-1) + (1-\mu)^2 \cdot \mathbf{h}(\mathbf{q}, d-1) \\ &= \mu \cdot \mathbf{t}(\mathbf{q}, d) + \mu \cdot (1-\mu) \cdot \mathbf{t}(\mathbf{q}, d-1) + (1-\mu)^2 \cdot [\mu \cdot \mathbf{t}(\mathbf{q}, d-2) + (1-\mu) \cdot \mathbf{h}(\mathbf{q}, d-2)] \\ &\quad \dots \dots \dots \\ &= \mu \cdot \mathbf{t}(\mathbf{q}, d) + \mu \cdot \sum_{s=2}^{d-1} (1-\mu)^{s-1} \cdot \mathbf{t}(\mathbf{q}, d-s+1) + (1-\mu)^d \cdot \mathbf{h}(\mathbf{q}, 1). \end{aligned} \quad (10)$$

where $\mathbf{t}(\mathbf{q}, d)$ is the travelers' experienced travel time on day d . The initial perceived travel time $\mathbf{h}(\mathbf{q}, 1)$ can be estimated by reassigning the traffic flows following the network capacity reduction on day $d = 1$. $\mathbf{q}(1)$ is an optimal solution to the following minimization model:

$$\min_{\mathbf{q}(1) \in \Omega_1} \sum_{i \in N} \int_0^{q_i(1)} t_i[u, C_i(1)] du, \quad (11)$$

where $C_i(1)$ is the capacity of link i on day $d = 1$. Ω_1 can be obtained using equation (7).

Equation (10) reveals that travelers' perceived travel time is largely dependent on all of their previous experienced travel time. Equation (10) assumes that travelers have infinite memories, which is not realistic. As mentioned by Cascetta [27], travelers' perceived travel time is mainly affected by their finite memory, namely, travelers cannot remember all of their experiences in the past. Thus, we make the following assumption:

Assumption 1. Travelers' perceived travel time on day d is affected by their most recent m days' experienced travel time. m is defined as the travelers' memory length.

Based on Assumption 1, equation (10) can be rewritten as follows:

$$\begin{aligned} \mathbf{h}(\mathbf{q}, d+1) &= \mu \cdot \mathbf{t}(\mathbf{q}, d) + \mu \cdot \sum_{s=2}^m (1-\mu)^{s-1} \\ &\quad \cdot \mathbf{t}(\mathbf{q}, d-s+1) + (1-\mu)^d \cdot \mathbf{h}(\mathbf{q}, 1). \end{aligned} \quad (12)$$

It is easy to know that the sum of the two coefficients, i.e., μ and $\mu \cdot \sum_{s=2}^m (1-\mu)^{s-1}$ in equation (12) does not equal to 1. Hence, we employ a scaling factor to make the coefficients sum to 1. Equation (12) is further transformed to the following:

$$\begin{aligned} \mathbf{h}(\mathbf{q}, d+1) &= \frac{\mu}{1-(1-\mu)^{m+1}} \cdot \mathbf{t}(\mathbf{q}, d) + \frac{\mu}{1-(1-\mu)^{m+1}} \\ &\quad \cdot \sum_{s=2}^m (1-\mu)^{s-1} \cdot \mathbf{t}(\mathbf{q}, d-s+1) \\ &\quad + (1-\mu)^d \cdot \mathbf{h}(\mathbf{q}, 1), \end{aligned} \quad (13)$$

where $(\mu/1-(1-\mu)^{m+1}) + (\mu/1-(1-\mu)^{m+1}) \cdot \sum_{s=2}^m (1-\mu)^{s-1} = 1$.

Hence, TSTT on day d in equation (1) can be calculated as

$$T(d) = \mathbf{q}(d) \cdot \mathbf{t}(d). \quad (14)$$

It is easy to know that the postdisaster road network capacity will change when every single road segment is restored, thus giving rise to a new day-to-day flow evolution, namely, there are multiple "new evolutions" during the restoration process. Therefore, when a "new evolution" occurs, i.e., at the end of each restoration activity, the value of d in the link-based day-to-day dynamics model should be reset to 1.

3.4. Optimization Model. In this subsection, we formulate the postdisaster RNRS problem as a resilience-based bi-objective mixed integer programming model under resource constraints, which makes a tradeoff between the maximal resilience of performance loss and the maximal resilience of recovery rapidity. For simplicity, this model assumes $d_s = 1$.

$$\max R_p = 1 - \frac{M_{\max} \times \pi(d_0) - \sum_{d=d_s}^{d_s+M_{\max}} \pi(d)}{M_{\max} \times \pi(d_0)}, \quad (15)$$

$$\max R_r = \begin{cases} 1 - \frac{M}{M_{\max}}, & M \leq M_{\max}, \\ 0, & M > M_{\max}, \end{cases} \quad (16)$$

which subject to the following:

$$\max_i \left\{ \sum_{d=1}^{M_{\max}} d \cdot x_{i,d} + e_i \right\} = M \leq M_{\max}, \quad (17)$$

$$\sum_{d=1}^{M_{\max}} d \cdot x_{i,d} = 1, \quad i = 1, 2, \dots, I, \quad (18)$$

$$\sum_{r=1}^R y_{i,r} = 1, \quad i = 1, 2, \dots, I, \quad (19)$$

$$\begin{aligned} \sum_{\tau=\max\{1, d-e_i+1\}}^d \sum_{i=1}^I x_{i,\tau} \cdot y_{i,r} &\leq 1, \\ d &= 1, 2, \dots, M_{\max}, \quad r = 1, 2, \dots, R, \end{aligned} \quad (20)$$

$$\begin{aligned} \sum_{\tau=\max\{1, d-e_i+1\}}^d \sum_{i=1}^I x_{i,\tau} &\leq R, \\ d &= 1, 2, \dots, M_{\max}, \end{aligned} \quad (21)$$

$$\sum_{i=1}^I \sum_{d=1}^{M_{\max}} x_{i,d} \cdot c_i = C \leq B, \quad (22)$$

$$x_{i,d} = \{0, 1\}, \quad (23)$$

$$y_{i,r} = \{0, 1\}. \quad (24)$$

Equations (15) and (16) are two objectives, which maximize the resilience of performance loss and the resilience of recovery rapidity, respectively. Equation (17) defines the makespan of the restoration schedule, i.e., the finish time of the last restoration activity, which cannot exceed M_{\max} . Equation (18) ensures that each restoration activity is implemented only once during the restoration period, namely, the restoration strategy is non-preemptive. Equation (19) guarantees that each restoration activity is carried out by a single work crew. Equation (20) ensures that each work crew can only conduct one restoration activity at most every single day. Equation (21) makes sure that the number of ongoing restoration activities every single day cannot

exceed the number of work crews. Equation (22) is the budget constraint, i.e., the total restoration costs cannot exceed the budget. Equations (23) and (24) define the type of decision variables.

4. Model Solution

Bi-objective models are usually solved to obtain a set of Pareto optimal solutions [34]. Although all the obtained Pareto optimal solutions are feasible, the best solution cannot be determined because of the different scales and bounds of the two objectives. In order to solve the issue, we first transform the bi-objective model into a single objective model using the weighted combination method as follows [12]. A genetic algorithm (GA) is then adopted to deal with the single objective model.

$$F = \omega \cdot R_t + (1 - \omega) \cdot R_r, \quad (25)$$

where F is the weighted value and ω is the weighting factor, which means the decision makers' preference for the two resilience metrics.

The mixed integer programming model has been widely employed to solve network design and network restoration problems, which are known to be NP-hard. Due to the NP-hardness, finding the exact solution is an intractable issue even for small problems; thus, heuristic algorithms are more suitable than exact algorithms for solving these problems. GA has been well recognized as an effective tool to solve the optimization problem [35–37]. GA searches for the optimal solution by simulating the natural evolution process; specifically, it simulates the solution of problems as a process similar to the crossover and mutation of chromosome genes in biological evolution [38]. GA is also used in combination with other algorithms (e.g., simulated annealing algorithms [39, 40], swarm intelligence algorithms [41], neural network algorithms [42], and tabu search algorithm [43]) to solve optimization problems. Compared with the previous algorithms, GA in this study represents the chromosome with two line sections using the integer coding method, which can reduce the possibility of generating a large number of infeasible solutions and avoid unnecessary searches. The flow chart of GA is presented in Figure 2.

4.1. Chromosome Encoding and Decoding. The RNRS problem in this study consists of two subproblems, i.e., the selection of work crews for road segments to be restored and the time sequence of restoration activities. We apply the integer coding method to represent the chromosome with two line sections shown in Figure 3. Each line section has I genes. The gen value g_i^1 ($i = 1, 2, \dots, I$) of line section 1 represents the work crew that will restore the corresponding link. It is noting that the gen value 0 in line section 1 means that the corresponding link will not be restored. $g_i^1 = 0, 1, 2, \dots, R$. The gen value g_i^2 ($i = 1, 2, \dots, I$) of line section 2 indicates the precedence relationship among all restoration activities. $g_i^2 = 1, 2, \dots, I$. The smaller g_i^2 , the higher the priority of the restoration activity. All gen values in line section 2 are different.

Let MA_r ($r = 1, 2, \dots, R$) be the set of restoration activities assigned to work crew r , where restoration activities are ranked by their priorities from high to low, and the number of restoration activities in MA_r is Num_r . It is noting that these restoration activities with $g_i^1 = 0$ are not included in MA_r . We denote EA_r as the set of eligible restoration activities, namely, the unscheduled restoration activities that have all predecessor restoration activities scheduled, and we let ST_{rj} ($j = 1, 2, \dots, \text{Num}_r$) represent the start time of the j th restoration activity in MA_r . Then, the following decoding procedure can be used to transform the solution into a restoration schedule.

- (i) *Step 1.* Set $r = 1$, $MA_r = \{1\}$, and $ST_{r1} = 0$.
- (ii) *Step 2.* Update EA_r , i.e., remove the scheduled restoration activities from EA_r and put the new eligible restoration activities into EA_r .
- (iii) *Step 3.* Judge whether $EA_r = \emptyset$ or not. If the answer is true, output all ST_{rj} and go to step 5; otherwise, go to step 4.
- (iv) *Step 4.* Calculate ST_{rj} of all the eligible restoration activities, $ST_{rj} = ST_{r,j-1} + d_j$, and return to step 2.
- (v) *Step 5.* Judge whether $r = R$ or not. If the answer is false, let $r = r + 1$ and return to step 2; otherwise, terminate the procedure.

4.2. Population Initialization. The initial population consists of popSize feasible initial solutions. Each feasible initial solution can be generated using the following steps:

- (i) *Step 1.* Randomly generate a value for g_i^1 , $g_i^1 = 0, 1, 2, \dots, R$, $i = 1, 2, \dots, N$, namely, select a work crew for each restoration activity randomly, generating line section 1.
- (ii) *Step 2.* Randomly generate a value for g_i^2 , $g_i^2 = 1, 2, \dots, N$, $i = 1, 2, \dots, N$, namely, set a priority value for each restoration activity randomly, generating line section 2.
- (iii) *Step 3.* Transform the generated g_i^1 and g_i^2 into a restoration schedule using the decoding procedure presented in Section 4.1. Judge whether $\max_{i,r} \{ST_{ri} + d_i\} \leq M_{\max}$ or not. If the answer is true, a feasible solution has been generated; otherwise, return to step 1 and repeat the procedure until a feasible solution is generated.

4.3. Selection. We use equation (25) as the fitness function in this algorithm. The fitness value of the parent individual determines the probability that the child individual is selected. Specifically, the higher the fitness value of the parent individuals, the higher the probability that they will be selected to generate offspring individuals. The roulette wheel method is employed to select individuals. Denote F_k ($k = 1, 2, \dots, \text{popSize}$) as the fitness value of the k th individual. The total fitness value F_{total} of a population of size popSize can be calculated as

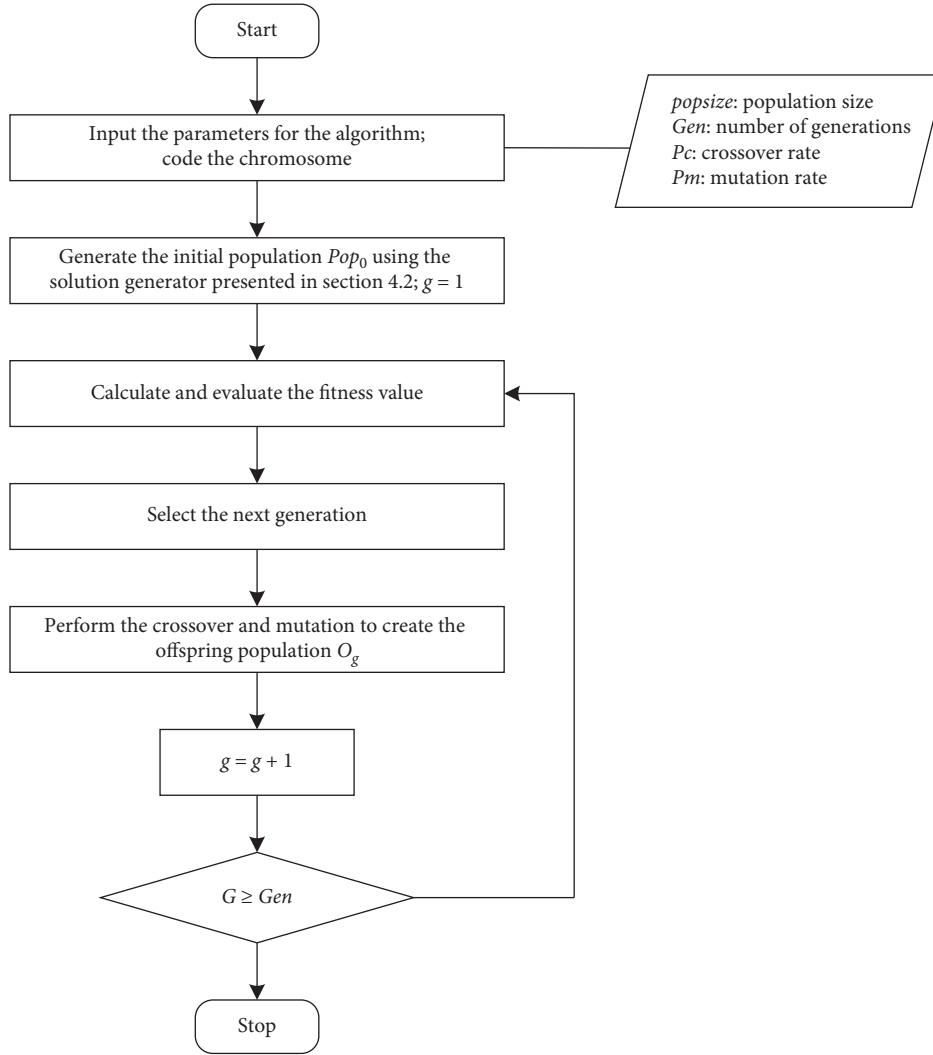


FIGURE 2: Flow chart of GA.

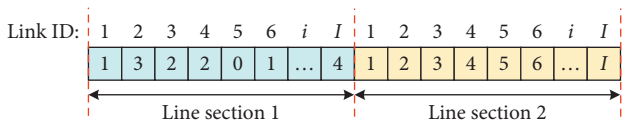


FIGURE 3: The chromosome representation.

$$F_{\text{total}} = \sum_{k=1}^{\text{popSize}} F_k. \quad (26)$$

The selection probability p_k for each individual k is formulated in the following equation [44]:

$$p_k = \frac{F_{\text{total}} - F_k}{F_{\text{total}} \cdot (\text{popSize} - 1)}. \quad (27)$$

Then, a random number $\varphi \in (0, 1]$ is generated. If $p_{k-1} < \varphi < p_k$, the k th individual is selected.

4.4. Crossover Operator. A crossover operator is used to generate two new offspring individuals from two parent

chromosomes by exchanging their genetic information [45], which can maintain population diversity. The crossover operator is performed based on a crossover rate that determines the probability that two parent individuals will be selected to exchange their genetic information. A single point crossover method is applied in this study. It is noting that only line section 1 is selected to crossover so as to change the job sequence of work crews. The crossover operation is performed by choosing a position i in line section 1 of the chromosome randomly and swapping all the gene values before that position. Thus, the first i gene values in line section 1 of an offspring chromosome are selected from one parent chromosome and the remaining gene values inherit the other parent chromosome. Figure 4 illustrates an example of a single point crossover operation.

4.5. Mutation Operator. The mutation operation is performed to a single chromosome, which can help the algorithm to avoid local optima. Based on a mutation rate, the mutation operation randomly changes one or more gene values of a chromosome. The widely used swap mutation

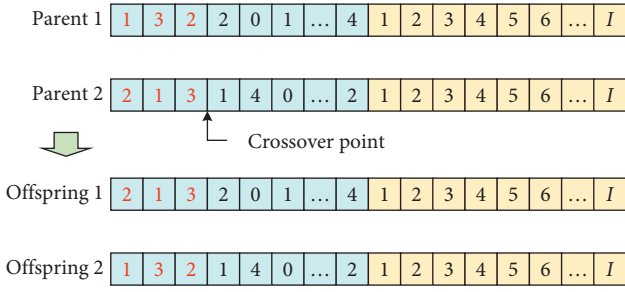


FIGURE 4: An example of a single point crossover operation.

strategy [46] that randomly exchanges two gene values of a chromosome is applied to the two line sections. Figure 5 shows an example of a swap mutation strategy.

5. Numerical Experiment

5.1. Road Network. A road network with 19 nodes and 36 links shown in Figure 6 is applied to validate the proposed method. The attributes of all links are presented in Table 1. There are two OD pairs, i.e., 1 → 17 and 8 → 19, the daily OD demands of which are 40,000 and 60,000, respectively. For simplicity, the predisaster link flows are obtained using a UE model.

The link travel time is estimated by the Bureau of Public Roads (BPR) function:

$$t_i(q_i) = t_i^0 \cdot \left[1 + \alpha \cdot \left(\frac{q_i}{C_i} \right)^\beta \right], \quad \forall i \in N, \quad (28)$$

where t_i^0 is the free flow travel time on link i , C_i is the capacity of link i , and α and β are two parameters, where $\alpha = 0.15$ and $\beta = 4$.

Disasters usually cause two types of damage to the road infrastructure: (1) complete damage, which means that the capacity of each damaged road segment drops to zero and the damaged road segments are out of service, and (2) partial damage, which indicates that the capacity of each damaged road segment is reduced to a certain extent and the damaged road segments have some partial passage. For simplicity, we assume that a disaster, which occurs on day $d = 1$, completely damages 21 road segments (in the dashed-line circle in Figure 6). It is worth noting that our proposed method can be easily extended to the cases with partially damaged road segments. The damaged road segments are marked by red arrows in the dashed-line circle in Figure 6. The restoration duration and cost of each disrupted link are listed in Table 2, where the unit of c_i is fund-unit. The total budget is 5,000 fund-unit. The other parameters are valued as in Table 3. The procedure is coded in MATLAB R2018b (version 11.4). All experiments are conducted on a Windows Server 2012 R2 server with an Intel Xeon E5-2640v4 CPU (2.4GHZ) and 64 GB DDR4 RAM.

5.2. Results. We define the solution with the highest fitness value after convergence is the optimal solution. Figure 7 illustrates the optimal RNRS generated by the proposed

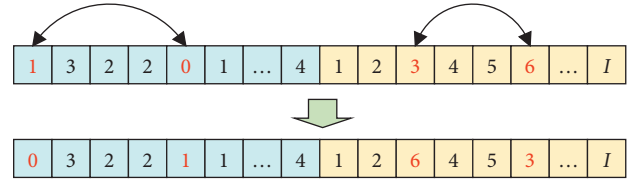


FIGURE 5: An example of a swap mutation strategy.

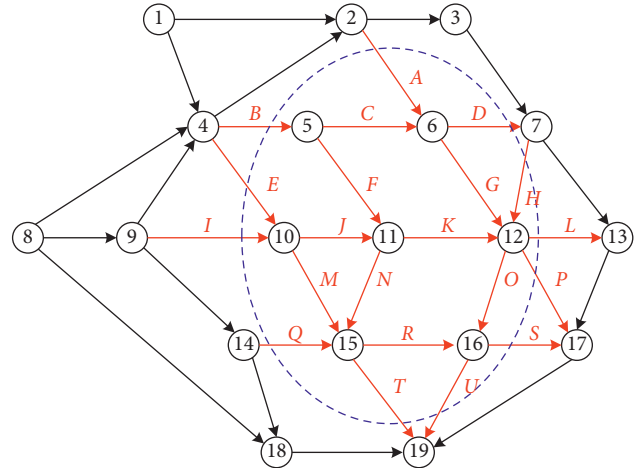


FIGURE 6: Layout of the testing road network.

TABLE 1: Link attributes of the road network.

Link	t_i^0 (h)	C_i (Veh/h)	Link	t_i^0 (h)	C_i (Veh/h)
1-2	3.6	1,600	9-10	2.2	2,000
1-4	1.5	1,700	9-14	2.8	1,600
2-3	2.4	1,700	10-11	1.8	2,200
2-6	2.1	2,200	10-15	2.5	1,700
3-7	1.8	1,900	11-12	1.5	2,400
4-2	3.4	1,600	11-15	1.9	2,000
4-5	2.3	2,000	12-13	1.7	2,200
4-10	2.4	1,500	12-16	1.8	2,200
5-6	1.8	2,200	12-17	2.3	2,400
5-11	2.1	1,600	13-17	1.9	1,600
6-7	2.0	2,400	14-15	1.7	1,700
6-12	2.0	1,400	14-18	1.7	2,200
7-12	2.4	2,200	15-16	1.7	2,000
7-13	1.9	1,600	15-19	2.0	1,600
8-4	3.3	2,000	16-17	1.3	2,000
8-9	1.4	2,100	16-19	1.8	1,700
8-18	4.4	1,700	17-19	2.9	2,000
9-4	2.1	2,200	18-19	2.6	2,200

method. Figure 7(a) shows the start time and completion time of the restoration activity for each link, where the number on each bar is the duration of each restoration activity. Figure 7(b) presents the restoration activities assigned to each work crew, where the letter and number on each bar represent the link ID and restoration duration, respectively. This optimal RNRS covers 12 out of 21 disrupted links to be restored by three work crews, given that the restoration budget is 5,000 fund-unit. The makespan is

TABLE 2: Restoration duration and cost of each disrupted link.

Link ID	e_i (day)	c_i	$q_{i,0}$ (Veh)	Link ID	e_i (day)	c_i	$q_{i,0}$ (Veh)
A	6	360	193	L	7	360	206
B	7	440	548	M	6	290	364
C	4	280	304	N	8	600	353
D	6	370	201	O	5	460	458
E	6	360	255	P	6	500	302
F	7	450	280	Q	8	510	268
G	8	640	260	R	8	480	630
H	5	170	175	S	5	450	423
I	9	590	752	T	7	420	515
J	5	240	615	U	5	350	832
K	6	340	478				

TABLE 3: Parameter values used in this study.

Parameters	Value
Maximum allowed makespan	$M_{\max} = 60$
Link flow evolution rate	$v = 1$
Travelers' sensitivity to travel time	$\lambda = 0.5$
Travelers' memory length	$m = 3$
Total number of work crews	$R = 3$
Population size	popSize = 50
Number of generations	Gen = 300
Crossover probability	$P_c = 0.9$
Mutation probability	$P_m = 0.6$
Decision maker's preference	$\omega = 0.5$

26 days. Figure 8 depicts the convergence process of the fitness values, which converges after 197 iterations. The best fitness value in the initial GA population is 0.231, and the fitness value reaches 0.612 after convergence.

In practice, the empirical restoration strategy, i.e., flow-first strategy (FFS), is commonly used. FFS determines the time sequence of restoration activities by link flows from highest to lowest and covers as many links as possible until the budget is exhausted. According to the link flows listed in Table 2, we generate a restoration scheme shown in Figure 9 following the FFS. In order to assess the efficiency of the optimal RNRS, we compare this optimal RNRS with the FFS. Table 4 lists the restoration results of the two strategies. Both strategies exhaust almost the same portion of the budget and cover 12 disrupted links. However, the optimal RNRS has a shorter makespan and produces much higher R_p and R_r . Obviously, the optimal RNRS outperforms the FFS, which indicates that the proposed method can recover the disrupted road network more quickly and generate fewer traffic delays.

Figure 10 presents the postdisaster TSTT evolution trajectories of the two restoration strategies. It is clear that TSTT has a significant fluctuation following the day-to-day traffic dynamics. Under the optimal RNRS, the postdisaster TSTT increases evidently and reaches the maximum equal to 3.12×10^6 on day $d = 2$. As more links are recovered, TSTT gradually decreases. It is noting that although all the restoration activities are completed on day $d = 26$, TSTT does not stop immediately and converges to a stationary (equilibrium) state on day $d = 31$. Under the FFS, it takes 7 days for the network flows to reach the equilibrium state after the

disaster and the equilibrium state continues for 14 days. This is because the FFS does not recover a new route till day $d = 20$; thus, the network capacity does not increase until day $d = 20$.

Since the road network flows cannot reach an equilibrium state overnight due to the network capacity variation, it is problematic to adopt UE models to simulate the traffic dynamics in this case. TSTT during the restoration makespan (i.e., 26 days) obtained by a day-to-day dynamics model and UE model for the two restoration strategies is compared in Figure 11. From the comparison of results, it is known that TSTT is less under a UE model than under a day-to-day dynamics model.

More damage scenarios are generated to verify our proposed method and analyze the two objectives. We consider 3 damage scenarios, where 18–20 links in the dashed-line circle in Figure 6 are randomly selected as the damaged road segments. For each scenario, there are C_{21}^i ($i = 18, 19, 20$) cases when i road segments are selected randomly to be damaged from the 21 road segments. The box plot for R_p and R_r of each scenario is presented in Figure 12. It can be found from Figure 12(a) that the maximum, minimum, median, upper quartile, and lower quartile of R_p all decrease as more road segments are damaged. As shown in Figure 12(b), the maximum and upper quartile of R_r decrease as more road segments are damaged. This phenomenon is in line with the fact that a network with more damaged road segments has lower resilience. It is also observed from Figure 12(b) that the lower quartile and the minimum of R_r keep constant in the three damage scenarios, which reveals that the three damage scenarios have the same worst-case optimal RNRS.

5.3. Sensitivity Analysis. In this subsection, we discuss the effects of key parameters including the number of work crews, travelers' sensitivity to travel time, availability of budget, and decision makers' preference on the restoration results. The sensitivity of every single parameter is analyzed by assuming other parameters are constant [47,48].

5.3.1. Number of Work Crews. Table 5 presents three optimal RNRSs for three different R . It is clear that the variation in R only affects the restoration time sequence but does not

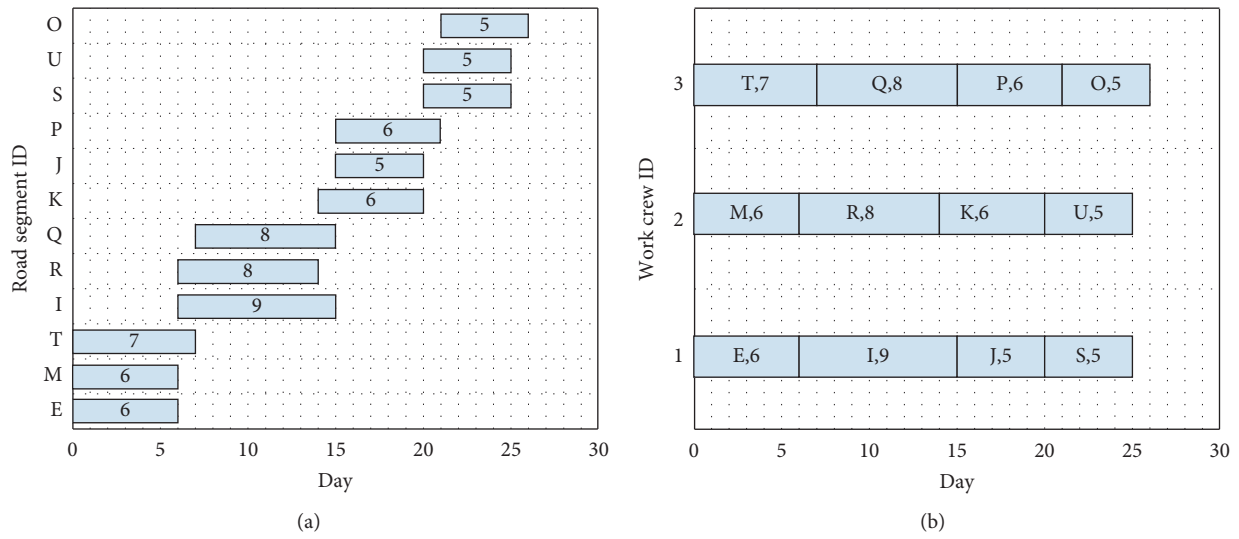


FIGURE 7: Optimal restoration strategy: (a) time sequence of maintenance; (b) job sequence of work crews.

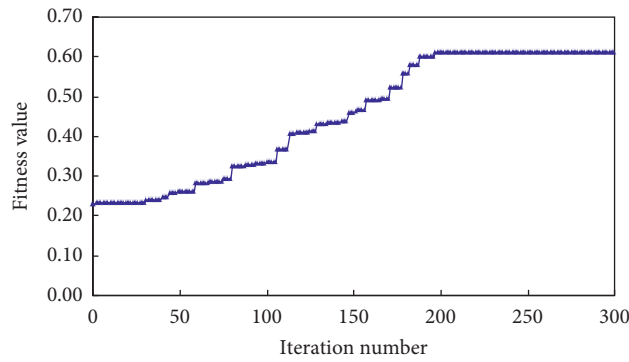


FIGURE 8: Convergence process of fitness values.

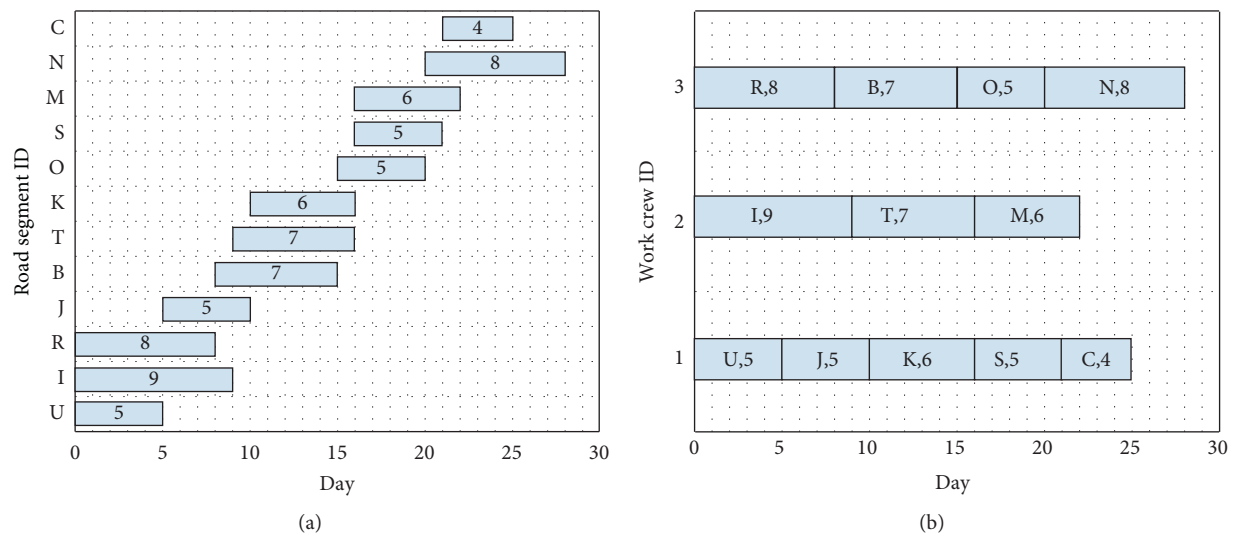


FIGURE 9: Flow-first restoration strategy: (a) time sequence of maintenance; (b) job sequence of work crews.

TABLE 4: Restoration results of the optimal RNRS and FFS.

Evaluation index	Optimal RNRS	FFS
R_p	0.472	0.404
R_r	0.567	0.533
M	26	28
C	4,990	4,940

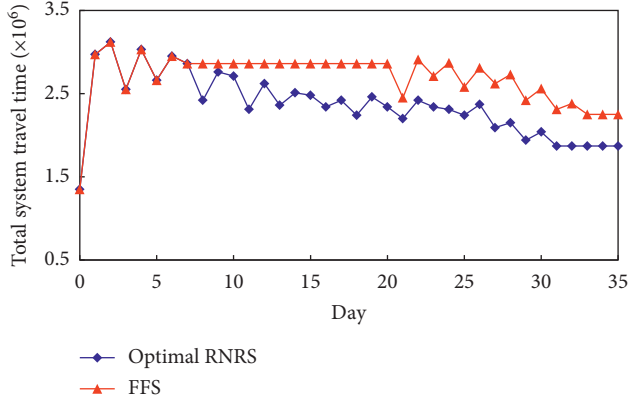


FIGURE 10: Total system travel time evolution trajectories of the optimal RNRS and FFS.

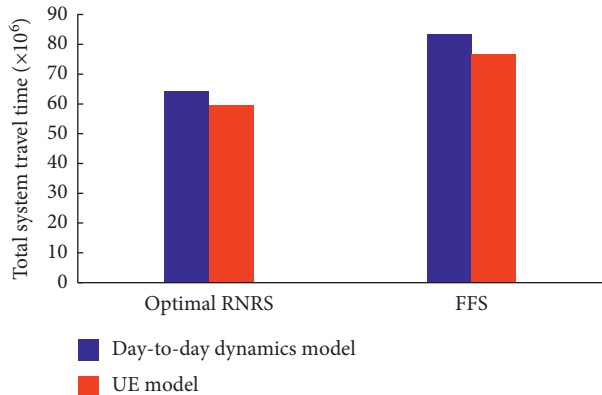


FIGURE 11: Total system travel time under a day-to-day dynamics model and UE model.

change the damaged links to be restored. The optimal RNRS of $R = 1$ has the minimum R_p and R_r , which indicates this restoration schedule is the worst. When $R = 1$, $M = 76$, which is more than 60; thus $R_r = 0$. Obviously, a higher R can reduce M , but the relationship between R and M is nonlinear. Generally, the restoration results are gradually improved with the increase in R . However, the growth rates of R_p and R_r decrease as R increases, i.e., the marginal benefits of manpower decrease. Additional experiments indicate that the marginal benefits of R_p and R_r equal to 0 when $R = 12$, which reveals that overmuch manpower cannot improve the restoration results but produces resource waste. Additionally, insufficient manpower will delay

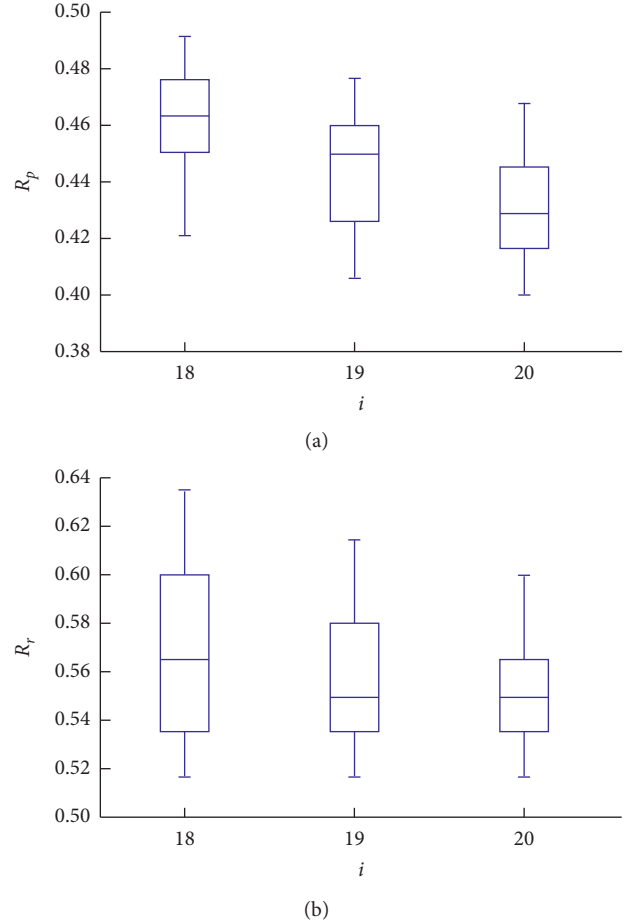


FIGURE 12: Box plot for (a) R_p and (b). R_r .

the restoration period. Thus, our proposed methods can be employed to evaluate manpower allocation plans in real-life road network recovery.

5.3.2. *Travelers' Sensitivity to Travel Time.* Table 6 shows three optimal RNRSs for three different λ . λ has a significant impact on the optimal RNRS including the restoration time sequence and the damaged links to be restored. As λ increases, the makespan is shortened, and both R_p and R_r increase. The larger the parameter value of λ , the more travelers will change their routes and the faster the link flows evolve to the new equilibrium state, thus causing less TSTT. Hence, in the actual RNRS, timely release of restoration information and effective traffic control strategies should be applied to help travelers to select the optimal route timely, which can improve the restoration schedule.

5.3.3. *Availability of Budget.* Table 7 lists the optimal RNRSs under three different scenarios of budget constraints. The optimal RNRSs of $B = 5,000$ and $B = 7,000$ differ greatly both in the restoration time sequence and the damaged links

TABLE 5: Effect of R on the optimal RNRS.

R	1	3	5
R_p	0.364	0.472	0.519
R_r	0.000	0.567	0.717
M	76	26	17
C	4,990	4,990	4,990
Optimal RNRS	Work crew 1: E-M-T-I-R-Q-K-J-P-S-U-O	Work crew 1: E-I-J-S	Work crew 1: E-Q
—	—	Work crew 2: M-R-K-U	Work crew 2: M-K-U
—	—	Work crew 3: T-Q-P-O	Work crew 3: T-J-O
—	—	—	Work crew 4: I-S
—	—	—	Work crew 5: R-P

TABLE 6: Effect of λ on the optimal RNRS.

λ	0.3	0.5	0.7
R_p	0.357	0.472	0.505
R_r	0.533	0.567	0.600
M	28	26	24
C	4,940	4,990	4,480
Optimal RNRS	Work crew 1: A-K-E-J-U	Work crew 1: E-I-J-S	Work crew 1: A-I-E
—	Work crew 2: D-B-T-L	Work crew 3: P-F-M-O	Work crew 2: M-R-K-U
—	Work crew 3: T-Q-P-O	Work crew 2: G-Q-O	Work crew 3: P-T-M-U

TABLE 7: Effect of B on the optimal RNRS.

B	5,000	7,000	8,000
R_p	0.472	0.517	0.517
R_r	0.567	0.383	0.383
M	26	37	37
C	4,990	6,890	6,890
Optimal RNRS	Work crew 1: E-I-J-S	Work crew 1: E-A-Q-J-H-D	Work crew 1: E-A-Q-J-H-D
—	Work crew 2: M-R-K-U	Work crew 2: M-G-S-I-L	Work crew 2: M-G-S-I-L
—	Work crew 3: T-Q-P-O	Work crew 3: T-P-R-K-O-U	Work crew 3: T-P-R-K-O-U

TABLE 8: Effect of ω on the optimal RNRS.

ω	0.3	0.5
R_p	0.413	0.472
R_r	0.617	0.567
M	23	26
C	4,390	4,990
Optimal RNRS	Work crew 1: E-A-B-C	Work crew 1: E-I-J-S
—	Work crew 2: M-G-I	Work crew 2: M-R-K-U
—	Work crew 3: T-P-Q	Work crew 3: T-Q-P-O

to be restored. Compared with $B = 5,000$, the optimal RNRS under $B = 7,000$ covers 5 more damaged links, which extends the makespan by 9 days and has higher R_p but lower R_r . It is noting that the restoration schedule does not change when B increases from 7,000 to 8,000. This is because that the manpower is insufficient. Thus, simply increasing the budget and keeping the workforce unchanged cannot improve the restoration schedule. Therefore, the monetary resources and the manpower should be matched.

5.3.4. Decision Maker's Preference. Table 8 indicates two optimal RNRSs for two different ω . Obviously, ω has a significant effect on the restoration schedule including the restoration time sequence and the damaged links to be restored. Compared with $\omega = 0.5$, the optimal RNRS with $\omega = 0.3$ is involved with 2 less damaged links, thus resulting in a lower R_p , M , and C but higher R_r , which represents more traffic delays and a shorter makespan. Hence, decision makers should make an optimal tradeoff between R_p and R_r .

6. Conclusions

This paper focuses on the resilience-based optimization of postdisaster road network restoration strategy. A TSTT-based function is used as the network performance indicator. Since the equilibrium-based methods cannot capture partial user equilibrium, a link-based day-to-day traffic model is employed to compute TSTT. We develop two resilience metrics to evaluate the effectiveness of the restoration strategy, i.e., the resilience of performance loss and the resilience of recovery rapidity. The former is calculated based on the network performance, and the latter is developed according to the restoration makespan. The restoration optimization problem is formulated as a resilience-based bi-objective mixed integer programming model, which aims to maximize the network resilience considering resource constraints. Then, a genetic algorithm is applied as the model solution.

The proposed method is validated through a case study. The results show that our method can provide an effective reference for transportation agencies to schedule post-disaster restoration activities. Compared with FFS, the optimal restoration strategy can reduce traffic congestion and shorten makespan. The comparison between the computational results using the day-to-day dynamics model and UE model explains why a link-based day-to-day dynamics model is preferred in this study. The sensitivity analyses of several key parameters reveal that the increase in the number of work crew or budget can improve the restoration schedule, but the marginal benefits of these two kinds of resources decrease. When either of the two resources exceeds a certain level, the restoration schedule will not be further improved. Travelers' sensitivity to travel time or decision makers' preference has a significant effect on the restoration schedule including the restoration time sequence and the damaged links to be restored. As travelers' sensitivity to travel time increases, the resilience of performance loss and the resilience of recovery rapidity will increase.

Future work should (1) consider some uncertainties, e.g., duration of restoration activities, traffic demands, and the restoration cost during the decision process; (2) develop a path-based day-to-day traffic model to simulate the PUE during the restoration process; (3) investigate the joint optimization of road network restoration and traffic control strategy; and (4) develop more intelligent algorithms and compare the efficiency of these algorithms to find a more suitable model solution.

Abbreviations

UE: User equilibrium
 TSTT: Total system travel time
 PUE: Partial user equilibrium
 OD: Origin-destination
 GA: Genetic algorithm
 BPR: Bureau of Public Roads
 FFS: Flow-first strategy.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the Humanities and Social Science Research Program of Ministry of Education in China (no. 18YJAZH120), National Natural Science Foundation of China (no. 52002282), Natural Science Foundation of Zhejiang Province (no. LQ19E080003), Fundamental Research Funds for the Central Universities (nos. 300102230624 and 300102238614), and Natural Science Research Program of Shaanxi Province (no. 2020JQ-360).

References

- [1] R. Fatouche and E. Miller-Hooks, "A mathematical framework for quantifying and optimizing protective actions for civil infrastructure systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 8, pp. 572–589, 2014.
- [2] S. Li and K. L. Teo, "Post-disaster multi-period road network repair: work scheduling and relief logistics optimization," *Annals of Operations Research*, vol. 283, no. 1-2, pp. 1345–1385, 2019.
- [3] A. Reggiani, "Network resilience for transport security: some methodological considerations," *Transport Policy*, vol. 28, pp. 63–68, 2013.
- [4] A. Kaviani, R. G. Thompson, and A. Rajabifard, "Improving regional road network resilience by optimised traffic guidance," *Transportmetrica A: Transport Science*, vol. 13, no. 9, pp. 794–828, 2017.
- [5] W. Zhang, N. Wang, and C. Nicholson, "Resilience-based post-disaster recovery strategies for road-bridge networks," *Structure and Infrastructure Engineering*, vol. 13, no. 11, pp. 1404–1413, 2017.
- [6] D. L. Alderson, G. G. Brown, W. M. Carlyle, and R. K. Wood, "Assessing and improving the operational resilience of a large highway infrastructure system to worst-case losses," *Transportation Science*, vol. 52, no. 4, pp. 1012–1034, 2018.
- [7] K. Barker, J. E. Ramirez-Marquez, and C. M. Rocco, "Resilience-based network component importance measures," *Reliability Engineering & System Safety*, vol. 117, pp. 89–97, 2013.
- [8] L.-G. Mattsson and E. Jenelius, "Vulnerability and resilience of transport systems - a discussion of recent research," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 16–34, 2015.
- [9] C. Liu, Y. Fan, and F. Ordóñez, "A two-stage stochastic programming model for transportation network protection," *Computers & Operations Research*, vol. 36, no. 5, pp. 1582–1590, 2009.
- [10] W. Zhang and N. Wang, "Resilience-based risk mitigation for road networks," *Structural Safety*, vol. 62, pp. 57–65, 2016.
- [11] P. Bocchini and D. M. Frangopol, "Restoration of bridge networks after an earthquake: multicriteria intervention optimization," *Earthquake Spectra*, vol. 28, no. 2, pp. 427–455, 2012.

- [12] Z. Li, C. Jin, P. Hu, and C. Wang, "Resilience-based transportation network recovery strategy during emergency recovery phase under uncertainty," *Reliability Engineering & System Safety*, vol. 188, pp. 503–514, 2019.
- [13] D. Rey, H. Bar-Gera, V. V. Dixit, and S. T. Waller, "A branch-and-price algorithm for the bilevel network maintenance scheduling problem," *Transportation Science*, vol. 53, no. 5, pp. 1455–1478, 2019.
- [14] M. Nogal, A. O'Connor, B. Caulfield, and B. Martinez-Pastor, "Resilience of traffic networks: from perturbation to recovery via a dynamic restricted equilibrium model," *Reliability Engineering & System Safety*, vol. 156, pp. 84–96, 2016.
- [15] A. De Palma and D. Rochat, "Understanding individual travel decisions: results from a commuters survey in Geneva," *Transportation*, vol. 26, no. 3, pp. 263–281, 1999.
- [16] X. He, X. Guo, and H. X. Liu, "A link-based day-to-day traffic assignment model," *Transportation Research Part B: Methodological*, vol. 44, no. 4, pp. 597–608, 2010.
- [17] A. Sumalee and D. P. Watling, "Partition-based algorithm for estimating transportation network reliability with dependent link failures," *Journal of Advanced Transportation*, vol. 42, no. 3, pp. 213–238, 2008.
- [18] R. Fatouche and E. Miller-Hooks, "Travel time resilience of roadway networks under disaster," *Transportation Research Part B: Methodological*, vol. 70, pp. 47–64, 2014.
- [19] D. Watling and M. L. Hazelton, "The dynamics and equilibria of day-to-day assignment models," *Networks and Spatial Economics*, vol. 3, no. 3, pp. 349–370, 2003.
- [20] J. L. Horowitz, "The stability of stochastic equilibrium in a two-link transportation network," *Transportation Research Part B: Methodological*, vol. 18, no. 1, pp. 13–28, 1984.
- [21] G. E. Cantarella and D. P. Watling, "A general stochastic process for day-to-day dynamic traffic assignment: formulation, asymptotic behaviour, and stability analysis," *Transportation Research Part B: Methodological*, vol. 92, pp. 3–21, 2016.
- [22] D. P. Watling and G. E. Cantarella, "Model representation & decision-making in an ever-changing world: the role of stochastic process models of transportation systems," *Networks and Spatial Economics*, vol. 15, no. 3, pp. 843–882, 2015.
- [23] T. L. Friesz, D. Bernstein, N. J. Mehta, R. L. Tobin, and S. Ganjalizadeh, "Day-to-day dynamic network disequilibria and idealized traveler information systems," *Operations Research*, vol. 42, no. 6, pp. 1120–1136, 1994.
- [24] M. J. Smith, "The stability of a dynamic model of traffic assignment—an application of a method of Lyapunov," *Transportation Science*, vol. 18, no. 3, pp. 245–252, 1984.
- [25] D. Zhang and A. Nagurney, "On the local and global stability of a travel route choice adjustment process," *Transportation Research Part B: Methodological*, vol. 30, no. 4, pp. 245–262, 1996.
- [26] G. E. Cantarella and E. Cascetta, "Dynamic processes and equilibrium in transportation networks: towards a unifying theory," *Transportation Science*, vol. 29, no. 4, pp. 305–329, 1995.
- [27] E. Cascetta, "A stochastic process approach to the analysis of temporal dynamics in transportation networks," *Transportation Research Part B: Methodological*, vol. 23, no. 1, pp. 1–17, 1989.
- [28] M. L. Hazelton and D. P. Watling, "Computation of equilibrium distributions of Markov traffic-assignment models," *Transportation Science*, vol. 38, no. 3, pp. 331–342, 2004.
- [29] T. Rambha and S. D. Boyles, "Dynamic pricing in discrete time stochastic day-to-day route choice models," *Transportation Research Part B: Methodological*, vol. 92, pp. 104–118, 2016.
- [30] Z. Liu, S. Wang, B. Zhou, and Q. Cheng, "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 58–72, 2017.
- [31] D. Henry and J. Emmanuel Ramirez-Marquez, "Generic metrics and quantitative approaches for system resilience as a function of time," *Reliability Engineering & System Safety*, vol. 99, pp. 114–122, 2012.
- [32] X. Yang, Z. Yajie, T. Jinjun, L. Jian, and I. Muhammad, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.
- [33] G. A. Davis and N. L. Nihan, "Large population approximations of a general stochastic traffic assignment model," *Operations Research*, vol. 41, no. 1, pp. 169–178, 1993.
- [34] Y. Hou, N. Wu, Z. Li et al., "Pareto-optimization for scheduling of crude oil operations in refinery via genetic algorithm," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 3, pp. 517–530, 2015.
- [35] M. Nemati, M. Braun, and S. Tenbohlen, "Optimization of unit commitment and economic dispatch in microgrids based on genetic algorithm and mixed integer linear programming," *Applied Energy*, vol. 210, pp. 944–963, 2018.
- [36] X. Li and L. Parrott, "An improved Genetic Algorithm for spatial optimization of multi-objective and multi-site land use allocation," *Computers, Environment and Urban Systems*, vol. 59, pp. 184–194, 2016.
- [37] S. M. Elsayed, R. A. Sarker, and D. L. Essam, "A new genetic algorithm for solving optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 27, pp. 57–69, 2014.
- [38] A. Jain, "Performance assessment of thirteen crossover operators using GA," in *Soft Computing for Problem Solving*, pp. 761–769, Springer, Berlin, Germany, 2019.
- [39] H. Yuan, J. Bi, and M. Zhou, "Spatiotemporal task scheduling for heterogeneous delay-tolerant applications in distributed green data centers," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1686–1697, 2019.
- [40] H. Yuan, J. Bi, and M. Zhou, "Multiqueue scheduling of heterogeneous tasks with bounded response time in hybrid green IaaS clouds," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 10, pp. 5404–5412, 2019.
- [41] K. Gao, Z. Cao, L. Zhang, Z. Chen, Y. Han, and Q. Pan, "A review on swarm intelligence and evolutionary algorithms for solving flexible job shop scheduling problems," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 4, pp. 904–916, 2019.
- [42] Z. Lei, J. Shen, Q. Wang, and Y. Chen, "Real-time weld geometry prediction based on multi-information using neural network optimized by PCA and GA during thin-plate laser welding," *Journal of Manufacturing Processes*, vol. 43, pp. 207–217, 2019.
- [43] A. Hoffmann, "EOS lumping optimization using a genetic algorithm and a tabu search," *Journal of Petroleum Science and Engineering*, vol. 174, pp. 495–513, 2019.
- [44] A. Hiassat, A. Diabat, and I. Rahwan, "A genetic algorithm approach for location-inventory-routing problem with perishable products," *Journal of Manufacturing Systems*, vol. 42, pp. 93–103, 2017.
- [45] H. Uğuz, "A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm," *Knowledge-Based Systems*, vol. 24, no. 7, pp. 1024–1032, 2011.

- [46] Q.-K. Pan, P. N. Suganthan, J. J. Liang, and M. F. Tasgetiren, "A local-best harmony search algorithm with dynamic sub-harmony memories for lot-streaming flow shop scheduling problem," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3252–3259, 2011.
- [47] S. Gao, M. Zhou, Y. Wang et al., "Dendritic neuron model with effective learning algorithms for classification, approximation, and prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 601–614, 2018.
- [48] J. Wang and T. Kumbasar, "Parameter optimization of interval Type-2 fuzzy neural networks based on PSO and BBBC methods," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 1, pp. 247–257, 2019.

Research Article

Developing Roadway Safety Models for Winter Weather Conditions Using a Feature Selection Algorithm

Bryce Hallmark ¹ and Jing Dong ²

¹HDR, 1917 S 67th Street, Omaha, NE 68106, USA

²Iowa State University, 2711 S Loop Dr, Ames, IA 50010, USA

Correspondence should be addressed to Jing Dong; jingdong@iastate.edu

Received 31 July 2020; Revised 17 November 2020; Accepted 15 December 2020; Published 29 December 2020

Academic Editor: Yajie Zou

Copyright © 2020 Bryce Hallmark and Jing Dong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inclement winter weather such as snow, sleet, and freezing rain significantly impacts roadway safety. To assess the safety implications of winter weather, maintenance operations, and traffic operations, various crash frequency models have been developed. In this study, several datasets, including for weather, snowplow operations, and traffic information, were combined to develop a robust crash frequency model for winter weather conditions. When developing statistical models using such large-scale multivariate datasets, one of the challenges is to determine which explanatory variables should be included in the model. This paper presents a feature selection framework using a machine-learning algorithm known as the Boruta algorithm and exhaustive search to select a list of variables to be included in a negative binomial crash frequency model. This paper's proposed feature selection framework generates consistent and intuitive results because the feature selection process reduces the complexity of interactions among different variables in the dataset. This enables our crash frequency model to better help agencies identify effective ways to improve roadway safety via winter maintenance operations. For example, increased plowing operations before the start of storms are associated with a decrease in crash rates. Thus, pretreatment operations can play a significant role in mitigating the impact of winter storms.

1. Introduction

Inclement winter weather such as snow, sleet, and freezing rain significantly impacts roadway safety. Every year, over 118,000 people in the United States are injured or killed due to winter-weather-related vehicle crashes [1]. Crash frequency models using Bayesian or negative binomial modeling [2–4] have been developed to investigate the safety implications of winter weather, maintenance operations, and traffic operations. However, due to inherent correlations among explanatory variables, conflicting results have been reported. For example, Qin et al. [5] used Wisconsin snowstorm and maintenance operation reports to develop a crash frequency model. Their estimated negative binomial model indicated that deploying more deicing material reduces the number of crashes, while deploying more salting material increases crash rates. This finding exemplifies the

complexity and difficulty of using multiple data sources when analyzing winter maintenance operations.

In this paper, winter weather crashes are aggregated by storm and city in order to model the impact of winter weather and maintenance operations on the expected crash rate for a particular winter event. A feature selection technique, called the Boruta algorithm [6, 7], is used to select the most impactful among highly correlated explanatory variables from a comprehensive dataset containing weather, maintenance operations, and traffic information.

As agencies continue to move towards data-driven decision-making, innovative data analytics are valuable for working with large datasets. In particular, when developing statistical models using large-scale multivariate datasets, one of the challenges is to determine which explanatory variables to include in the model. In this study, several datasets are combined to examine the occurrence of winter-weather-

related crashes, including weather (e.g., type, rate and duration of precipitation, temperature, and visibility), snowplow operations (e.g., plow frequency and material spreading rate), and traffic information (e.g., traffic volume). Some variables in these datasets are inherently correlated. For example, snowplow operations usually depend on the severity of winter storms and type of precipitation.

This paper proposes a framework for identifying which of highly correlated explanatory variables should be selected to develop a robust winter-weather-related crash frequency model. In particular, it adopts the Boruta algorithm to calculate the importance of each variable using a random forest wrapper. Based on relative importance scores, a set of variables is selected to be included in the negative binomial model. This feature selection process helps to create a robust crash frequency model for winter-weather-related crashes.

2. Literature Review

Past studies have shown that winter storm events generally increase the crash rate while winter maintenance operations tend to reduce the crash rate. Nixon and Qiu conducted a meta-analysis of studies that quantified the impact of weather on traffic crashes and found that the crash rate increases by 84% and the injury rate increases by 75% with snow conditions. Nixon and Qiu [8], on the other hand, showed an average reduction of 78% in crash rate on freeways and 87% on two-lane undivided highways due to salting. Earlier studies like these are usually based on hourly traffic count and weather data, as well as manually maintained winter storm records.

However, with the wide adoption of sensing, GPS tracking, and communication technologies in recent years, state agencies are now collecting more detailed and granular data. For example, today most state departments of transportation (DOTs) collect traffic volume and speed data using in-pavement or roadside sensors at 20-second to 5-minute aggregation intervals. Many agencies have also deployed cameras, sensors, and GPS tracking devices on snowplow trucks to collect location and operational data at subminute intervals [9, 10]. With these new data sources, researchers are again attempting to assess the effect of winter maintenance operations on mobility and safety in order to provide guidance for more effective maintenance operations [11–13].

As researchers begin to face larger and more complex datasets, the ability to identify meaningful relationships from such datasets becomes more important. Machine learning has been commonly applied in the transportation field when attempting to model and predict crashes, but usually with only minimal variable filtering [14–16]. While machine learning for feature selection has been done in other fields, its use in the transportation sector is nearly nonexistent. Feature selection, however, has been used with success when analyzing variables in evacuation behavior modeling and in determining gully erosion factors [17, 18]. In particular, the Boruta algorithm has been shown as an effective method for feature selection. For example, Prasad et al. [19] used the Boruta feature selection algorithm to provide a trimmed list for their extreme learning machine

(ELM) model to study weekly soil moisture. Similarly, to select optimal random forest predictive models for seabed hardness, Li et al. [20] compared five feature selection methods and recommended the averaged variable importance and Boruta algorithms as producing the most accurate predictive models. Various other studies have also found the Boruta method to produce accurate and stable results [6, 7, 21]. Based on the success of such past applications, this paper adopts the Boruta feature selection algorithm to select variables in a winter-weather-related crash frequency model.

3. Data Description

This paper analyzes winter weather events in eight major cities across Iowa. Geographic-based analysis is a commonly used method for aggregating winter weather data [2, 22]. For each city, a list of winter storm events was compiled. All relevant data were then associated with each storm based on the location and time stamp, as shown in Figure 1.

The scope of this study encompassed two winter seasons from November 2016 to May 2017 and from November 2017 to May 2018.

The Iowa DOT operates and maintains all of the Interstates, US highways, and state highways across Iowa. Only data pertaining to these Iowa DOT maintenance routes were analyzed for this study. A list of compiled variables from across the study's data sources can be seen in Table 1.

3.1. Roadway Data. The Iowa DOT maintains and provides roadway information via the Roadway Asset Management System (RAMS). The RAMS provides the geometric and operational features of the roadway such as the number of lanes, roadway surface type, and speed limit. In conjunction with the RAMS database, the Iowa DOT provides a Linear Reference System (LRS). An LRS is a mile reference system for each unique route. By selecting any location along the Iowa DOT maintenance routes; therefore, users can link their selected location's LRS mile marker reference to the RAMS database. Additionally, the LRS provides a means to combine any other data that can be linked to the roadway system, such as crashes and snowplow locations.

Each city contains some Iowa DOT maintenance roadways. Therefore, the current study's road miles variable is the length of the Iowa DOT roadway network within each city's boundary (Table 2). The lane miles variable accounts for each Iowa DOT roadway's number of lanes, thus providing an indicator of the scale of the Iowa DOT's roadway surface area in each city.

3.2. Traffic Data. The Iowa DOT has over 900 Wavetronix sensors placed throughout the state. These sensors collect traffic speed, occupancy, and volume data that are archived at 5-minute aggregation intervals. Most of these sensors are located in urban areas. Using Wavetronix sensor data in lieu of annual average daily traffic (AADT) provides more accurate traffic counts as roadway volumes can vary greatly during winter storm events [23].

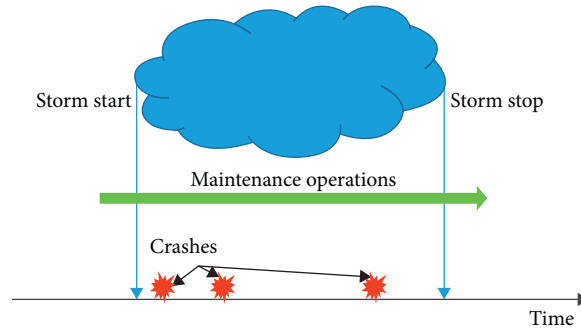


FIGURE 1: Associating multiple data sources related to a winter storm.

TABLE 1: Combined dataset variable descriptions.

Variable	Description (units)
AVLRecords	The number of 5-minute AVL records (count of records)
AVL_Hours	Hours of AVL operations (hours)
AfterRecords	The number of 5-minute AVL records from the end of the snow event to the end of the plowing event (count of records)
BeforeRecords	The number of 5-minute AVL records from the start of the plow event to the start of the storm event (count of records)
Exposure	Calculation for vehicle volume (exposure)
FreezingRain	The amount of freezing rain in the storm event (inches)
HourPrecipitation	The hourly rate of precipitation (inches/hour)
LaneLaps	The amount of truck distance traveled divided by the total roadway distance in the city (lane mile laps)
Laps	The amount of truck distance traveled divided by the length of roadways in the city (laps)
LnExposure	The natural log of the exposure variable
PCPN	Total precipitation (inches)
RecordsBefore	The number of 5-minute record counts from before the start of the storm event (count of records, before)
RoadCondition	The aggregate road condition for the storm event (1 to 5 categorical, 1 = least severe conditions, 5 = most severe conditions)
Snow	The total snow in the storm event (inches)
StormHours	The number of hours that the storms lasted (hours)
StormRecords	The number of 5-minute counts that the storm lasted (count of records)
TemperatureFah	The average temperature of the storm (Fahrenheit)
TemperatureFahASOS	The average temperature of the storm from the ASOS system (Fahrenheit)
TemperatureFahRWIS	The average temperature of the storm from the RWIS system (Fahrenheit)
TotalDistLiquid	The total distance of truck travel while spreading liquid material (miles)
TotalDistLiquidBefore	The total distance of truck travel while spreading liquid material between the start of the plow event and storm event (miles)
TotalDistPrewet	The total distance of truck travel while spreading prewet material (miles)
TotalDistPrewetBefore	The total distance of truck travel while spreading prewet material between the start of the plow event and storm event (miles)
TotalDistSolid	The total distance of truck travel while spreading solid material (miles)
TotalDistSolidBefore	The total distance of truck travel while spreading solid material between the start of the plow event and storm event (miles)
TruckDisPerLanMile	The total distance of truck travel divided by the amount of lane miles in each city (miles per lane mile)
TruckDisPerLanMileBefore	The total distance of truck travel divided by the amount of lane miles in each city before the start of plow operations (miles per lane mile before)
TruckDisPerLanMileMinute	The total distance of truck travel divided by the amount of lane miles in each city divided by the number of minutes of the storm (miles per lane mile per minute)
TruckDisPerLanMileMinuteBefore	The total distance of truck travel divided by the amount of lane miles in each city divided by the number of minutes of the storm before the start of the storm event
TruckDistMiles	The total truck distance driven (miles)
TruckDistMilesBefore	The total truck distance driven before the storm (miles before)
TruckLiquidPerLanMile	The total amount of liquid spread divided by the amount of lane miles (gal per mile per lane mile)

TABLE 1: Continued.

Variable	Description (units)
TruckLiquidPerLanMileBefore	The total amount of liquid spread divided by the amount of lane miles before the storm event (gal per lane mile before)
TruckLiquidPerLanMileMinute	The total amount of liquid spread divided by the amount of lane miles divided by the total minutes of the storm (gal per lane mile per minute)
TruckLiquidPerLanMileMinuteBefore	The total amount of liquid spread divided by the amount of lane miles divided by the total minutes of the storm before (gal per lane mile per minute before)
TruckMaterialLiquid	The total amount of liquid spread (gal)
TruckMaterialLiquid.Distance	The total amount of liquid spread divided by the distance traveled (gal/mile)
TruckMaterialLiquid.DistanceBefore	The total amount of liquid spread divided by the distance traveled before the storm event (gal/mile)
TruckMaterialLiquidBefore	The total amount of liquid spread before the storm (gal before)
TruckMaterialLiquidDistance.Distance	The total amount of liquid material spread divided by the distance traveled only while spreading occurred (gal/mile, of spreading miles traveled only)
TruckMaterialLiquidDistance.DistanceBefore	The total amount of liquid material spread divided by the distance traveled only while spreading occurred before (gal/mile, of spreading miles before traveled only)
TruckMaterialPreLiquid	The total amount of preliquid spread (gal)
TruckMaterialPreLiquid.Distance	The total amount of preliquid spread divided by the distance traveled (gal/mile)
TruckMaterialPreLiquid.DistanceBefore	The total amount of preliquid spread divided by the distance traveled before (gal/mile before)
TruckMaterialSolid	The total amount of material solid spread (lbs)
TruckMaterialSolid.Distance	The total amount of solid material spread divided by the total truck distance traveled (lbs/mile)
TruckMaterialSolid.DistanceBefore	The total amount of solid material spread divided by the total truck distance traveled before (lbs/mile, before)
TruckMaterialSolidBefore	The total amount of solid material spread before the storm (lbs, before)
TruckMaterialSolidDistance.Distance	The total amount of solid material spread divided by the truck distance that was traveled only while spreading occurred (lbs/mile, of mile traveled while spreading)
TruckMaterialSolidDistance.DistanceBefore	The total amount of solid material spread divided by the truck distance that was traveled only while spreading occurred before the storm (lbs/mile, of mile traveled while spreading, before)
TruckPreLiquidPerLanMile	The total amount of preliquid material spread divided by the length of lane miles (gal/lane mile)
TruckPreLiquidPerLanMileMinute	The total amount of preliquid material spread divided by the length of lane miles divided by the total minutes of the storm (gal/lane mile per minute)
TruckSolidPerLanMile	The total amount of solid material spread divided by the length of lane miles (lbs/lane mile)
TruckSolidPerLanMileBefore	The total amount of solid material spread divided by the length of lane miles before the storm (lbs/lane mile, before)
TruckSolidPerLanMileMinute	The total amount of solid material spread divided by the length of lane miles divided by the total minutes of the storm (lbs/lane mile per minute)
TruckSolidPerLanMileMinuteBefore	The total amount of solid material spread divided by the length of lane miles divided by the total minutes of the prestorm plow operations (lbs/lane mile per minute, before)
Visibility	The total visibility (miles)
WindGust	The max wind recorded (mph)
WindSpeed	The average wind speed (mph)
WindSpeedKnots	The average wind speed (knots)
WindSpeedKnotsASOS	The average wind speed obtained by the ASOS system (knots)
WindSpeedKnotsRWIS	The average wind speed obtained by the RWIS system (knots)
WorstFreezingRain	The worst intensity of freezing rain (score of 1–3 categorical variables, 1 = least severe, 3 = most severe)
WorstRoadCondition	The worst road condition (1 to 5 categorical variables, 1 = least severe conditions, 5 = most severe conditions)
WorstSnow	The worst snow intensity recorded (score of 1–3 categorical variables, 1 = least severe, 3 = most severe)
WorstVisibility	The lowest visibility recorded (miles)
City	The city location of the storm event (Ames, Des Moines, Council Bluffs, Sioux City, Waterloo, Iowa City, Quad cities, Cedar Falls)

Note: the acronym “PLM” stands for per lane mile.

TABLE 2: Roadway miles breakdown by city.

City	Road miles	Lane miles
Ames	142	285
Cedar Rapids	199	439
Council Bluffs	153	359
Des Moines	624	1380
Iowa City	227	563
Sioux City	218	517
Davenport	279	597
Waterloo	232	482

The average vehicle count (AVC) represents the average count of vehicles present at each Wavetronix sensor. AVC is calculated as follows:

$$AVC = \frac{\sum(\text{vehicle counts for all sensors})}{\text{number of sensors} \div \text{number of 5-minute intervals of the storm.}} \quad (1)$$

In essence, a city's total count of traffic volume is divided by its number of sensors (see Table 3) and then divided by the number of 5-minute intervals throughout the duration of a storm. Because the sensor data are aggregated over just 5 minutes, this essentially provides the expected count of vehicles at each portion of the roadway for the entirety of the storm.

The resulting average traffic volume based on Wavetronix sensor counts and the road miles variable together provide an "exposure" count:

$$\text{Exposure} = AVC * \text{Road Miles.} \quad (2)$$

3.3. Automatic Vehicle Location (AVL) Data. The Iowa DOT has over 900 snowplow trucks spread throughout 101 garages. Each snowplow's automatic vehicle location (AVL) system records the date and time, longitude and latitude, traveling speed, plow position (up vs. down), and material spreading rates at approximately a 10-second refresh rate. Three types of spreading rates are recorded, namely, solid rate, prewet rate, and liquid rate. Four types of plow wing records are available, namely, front plow, left wing, right wing, and underbelly plow.

A snowplow's capacity is 12,000 lbs for single-axle trucks and 24,000 lbs for tandem-axle trucks. Its spreading rate is approximately 200 lbs per lane mile for solid material and 60 gallons per lane mile for liquids. Its travel speed when plowing and spreading material is about 30 miles per hour. Its deadheading speed can be as high as the speed limit.

Earlier works have used ratios such as the total material spread normalized per precipitation event to examine the safety implications of snowplow operations [11]. As past works have had difficulties in showing conclusive results, we created an extensive list of ratios as candidate predictors for the impact of maintenance operations on crash rates. Additionally, our snowplow variables needed to be normalized per geographic region to facilitate appropriate comparison.

TABLE 3: Number of Wavetronix sensors in each city.

City	Number of Wavetronix sensors
Ames	82
Cedar Rapids	163
Council Bluffs	151
Des Moines	289
Iowa City	78
Sioux City	92
Davenport	46
Waterloo	61

This was done by dividing the material spread by the roadway surface area (i.e., lane miles). Again, Table 1 contains our full list of variables.

3.4. Weather Data. This study's weather data were obtained from the Iowa Environmental Mesonet system, which provides highly granular weather data across Iowa. This Multiradar/Multisensor (MRMS) project combines information from many sources and radar systems to provide precise weather information for 1-by-1-mile grid areas. The weather variables for each grid area include air temperature, wind speed, hourly and minute-based precipitation, daily snowfall, precipitation type, and so forth. Additionally, we rated the intensity of the precipitation and weather on a scale of 0 to 3, with 3 being the worst. When analyzing the data for winter storm events, such ratings are used to determine the intensity of a given storm.

4. Crash Data

This study's crash data were obtained from the Iowa DOT crash database. The crash data include information such as the location, time, crash severity, direction of travel, lighting conditions, and weather conditions that potentially contributed to the crash. In particular, the following data fields describe what weather conditions may have contributed to each crash as well as what the road surface conditions were like at the time of the crash:

- (i) Environmental contributing circumstances
- (ii) Weather1 (dominant weather condition)
- (iii) Weather2 (secondary weather condition)
- (iv) Surface conditions

Since this study focuses on winter weather crashes, a filter was used requiring a winter weather condition present for any of the four weather-related data fields for a crash to be considered as a winter weather crash. After filtering based on weather conditions, there were 5,089 winter-weather-related crashes along the Iowa DOT maintenance routes from 2016 to 2018. When counting only crashes that occurred within this study's geographic analysis regions and that could be linked to the LRS mile reference system, only 1,372 crashes remained.

5. Methodology

Different from previous safety analyses, this paper presents a model development process based on the feature selection method. Figure 2 summarizes this process. First, data from multiple sources are combined into a single dataset. Next, a ridge regression test is performed to determine the suitability of the data for analysis. If the test fails, the Boruta feature selection method is applied to trim the dataset before retesting. Once the data are determined to be suitable, an exhaustive search function is performed to produce a final trimmed list of explanatory variables. At this stage, the user can select the policy relevant variables to be included in the final model.

5.1. Data Integration. As the first step in our methodological framework, data from different sources are integrated based on the spatial and temporal information associated with each record. In particular, a geofence was created for each city to filter its weather, snowplow, crash, and traffic data. Then, the weather data were analyzed to identify a list of winter storm events for each city. In particular, any time period below 41 degrees with precipitation that lasted longer than 30 minutes was considered a winter storm event. Based on this geofence and storm event timing, relevant snowplow and crash data were extracted.

For each winter storm event, the snowplow data were separated into three time periods, that is, before, during, or after the storm. Making this distinction is important because many agencies, including the Iowa DOT, deploy operations before the start of a storm, and many are unable to properly clear their roadways within the timeframe of storms. Using our geofence, therefore, a list of snowplows in operation during each storm was created. For each snowplow, its continuous hours of operation were compiled. A continuous operation was defined as a snowplow in operation with a time gap of no more than 2 hours. Each snowplow's start- and end-of-continuous-operation timestamps were then used to distinguish its before-storm, during-storm, and after-storm plow data. Each storm event's before-storm aggregation thus constituted the time from the beginning of any of its snowplow operations to the start of the storm. Its during-storm aggregation was from the start to the end of the winter storm. Its after-storm aggregation was from the end of the storm to the end of all plow operations.

5.2. Ridge Regression Test. Ridge regression cross-validation is employed to determine dataset suitability for analyses such as exhaustive search. Ridge regression tunes the parameters of a model to minimize the ordinary least squares [24]. It outputs an expected coefficient for each variable via differing model tuning or lambda penalizations. Lambda penalizations tend to tune variables to a model coefficient value of 0. When lambda has been maximized, all possible variables will produce a coefficient of 0. That is, varying lambda between 0 and 1 will produce a model that removes only a portion of the input variables. By imposing lambda penalizations, ridge

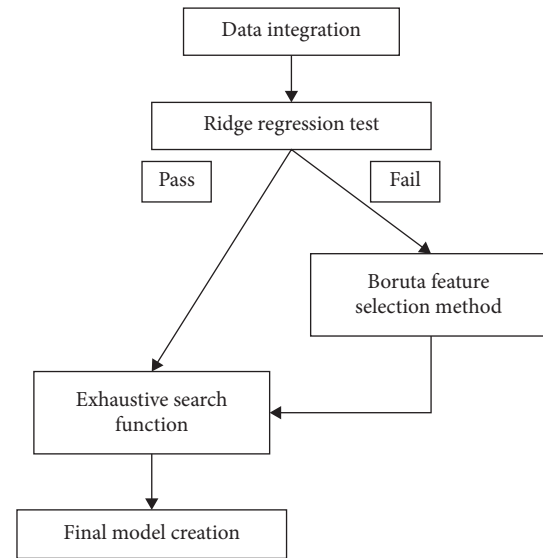


FIGURE 2: Methodological framework.

regression can provide a best-case tradeoff between bias (i.e., training set accuracy) and predictive variance (i.e., testing set accuracy). Ridge regression successively penalizes different variables' coefficients down to zero using lambda and compares the performance of its resulting model variants. This process effectively removes unimportant variables from a model as it helps identify the important variables in the dataset as well as the expected model outputs. Tracking expected model outputs as model complexity changes can provide insight into how data interact.

Normally, all variables in a dataset are presented in one ridge regression plot. However, to better show how the variables in our dataset interact, we split our many variables into groups based on their deviance away from lambda and plotted them accordingly in six subfigures (Figure 3). Each curve represents a variable from the dataset. The y -axis is the expected coefficient of each respective variable. Moving from left to right along the x -axis, the model becomes more complex as more variables move away from 0.

As more variables interact, it is expected that their estimated coefficients will vary. Most variables maintain a coefficient either above or below the x -axis, indicating a positive or negative relationship, respectively. Some variables, however, experience drastic and highly erratic behavior where they start with a positive coefficient and then drop to a negative coefficient as the model becomes more complex or vice versa. For example, the variable "Road-Condition" has a positive coefficient that increases as more variables are added. This is an acceptable change in an estimated coefficient. On the other hand, the variable "TruckDisPerLanMileMinute" is the total distance of snowplow trucks travel divided by the total lane miles in a given city divided by the number of minutes of a given storm. On the left-hand side, where the model is the least complex, the variable has a positive coefficient. As the model becomes more complex as more variables are added, this coefficient changes from positive to negative. In other words, depending on what other variables are included in the

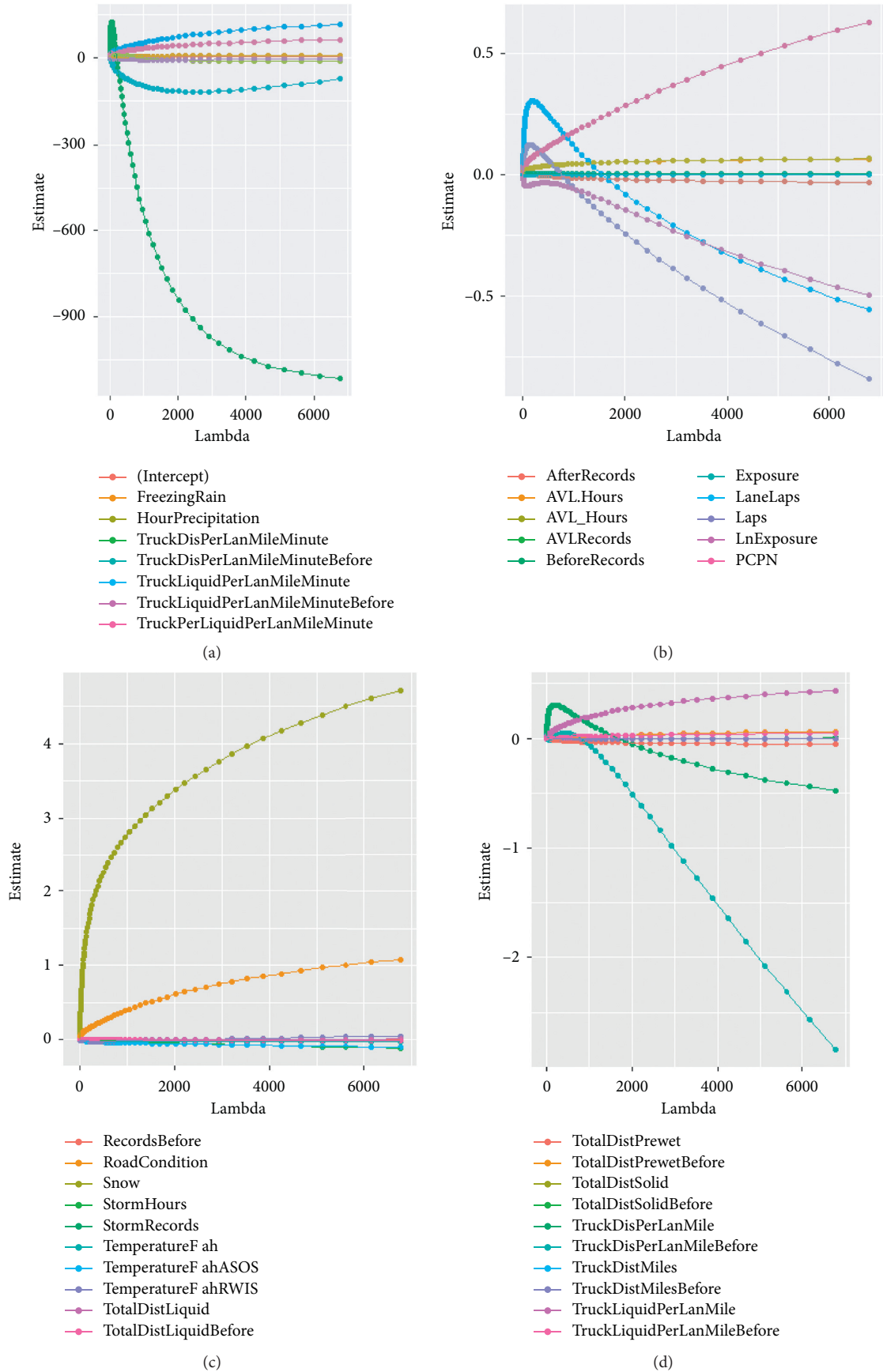


FIGURE 3: Continued.

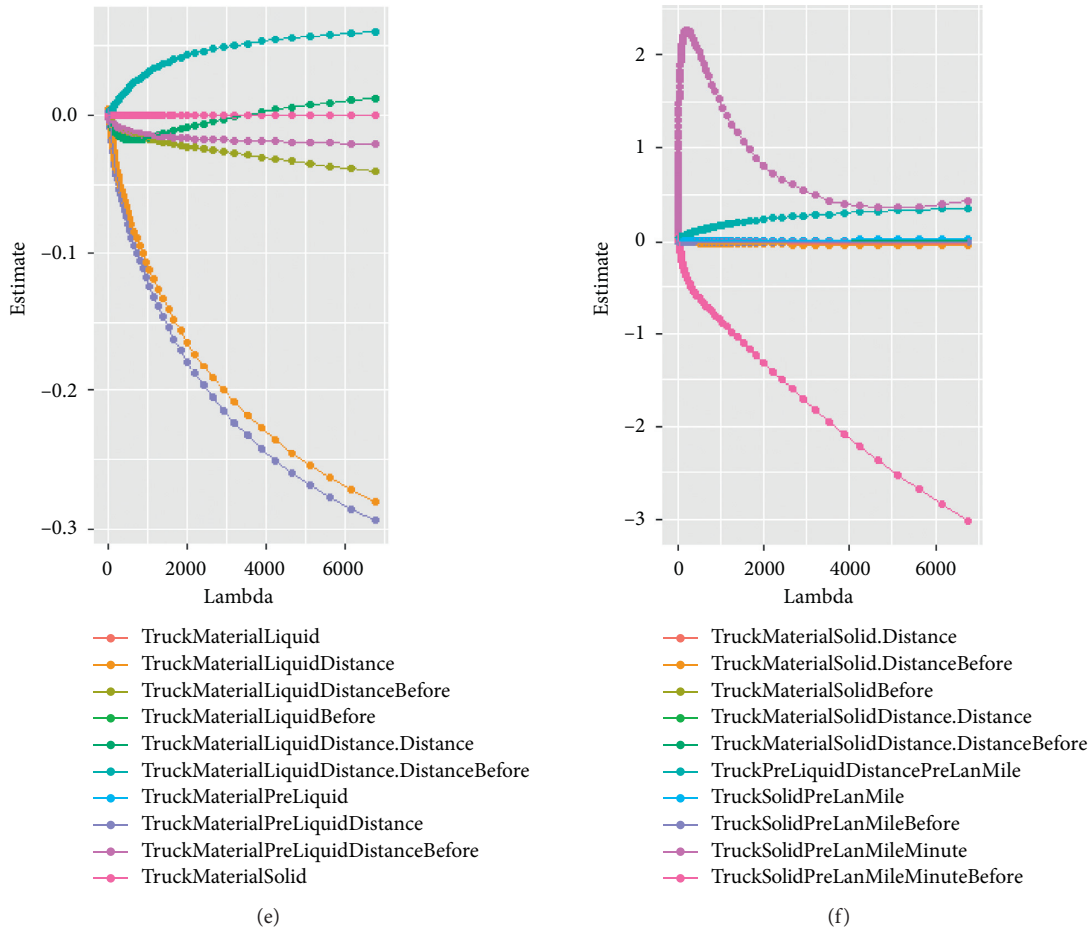


FIGURE 3: Ridge regression cross-validation.

model, the coefficient of this variable can be positive or negative. This explains why inconsistent results have been observed when analyzing similar datasets in the past. In the ridge regression test, however, having variables that cross over the x -axis constitutes a failure indicating that feature selection is required.

5.3. Boruta Feature Selection. If the ridge regression test fails, the Boruta feature selection method is applied to trim the dataset. The Boruta method creates an importance score for each variable [6, 7]. In particular, a random forest wrapper classification is employed on the variables. In the process, the variables are shuffled, while random shadow variables are created. The shadow variables are meant to identify ineffective input variables. To do so, each variable, as well as the random shadow variables, is assigned a relative importance score. Any variable that scores worse than the worst shadow variable will prove to be ineffective while variables that score higher than the highest shadow variable will be highly effective variables. These final importance scores are then used to create a set of noncorrelated data points.

The Boruta algorithm enables the development of a hierarchal list of variables ranked in terms of importance. These importance values can then be used to select which

explanatory variables should be included in statistical models. For better visualization, we split our variables into two groups, namely, the AVL variables shown in Figure 4 (and whose ranked IDs are matched in Table 4 to the respective variables they represent), and the traffic and weather variables, as shown in Figure 5. Figure 4 displays all the AVL variables with their respective ID numbers with variable 1 on the right moving in descending order to the left to variable 46 on the left.

The blue star variables represent the random shadow variables. Any variable that has a higher importance score resides to the right of these shadow variables and is colored green or yellow. Any variables between the two blue boxes are of negligible importance and are either red or yellow.

When creating statistical models, only one variable from a group of highly correlated variables should be included. Because large datasets are likely to contain many groups of highly correlated variables, manual feature selection is usually inefficient. The Boruta importance scores allow the highest-ranking variable from each correlated group to be selected, in order to create a trimmed list of noncorrelated variables. For example, in Figure 4, the highest-ranking variable was “AVLRecords.” Any variable in the AVL grouping that contained a high correlation with the “AVLRecords” variable (i.e., above 0.70) was therefore

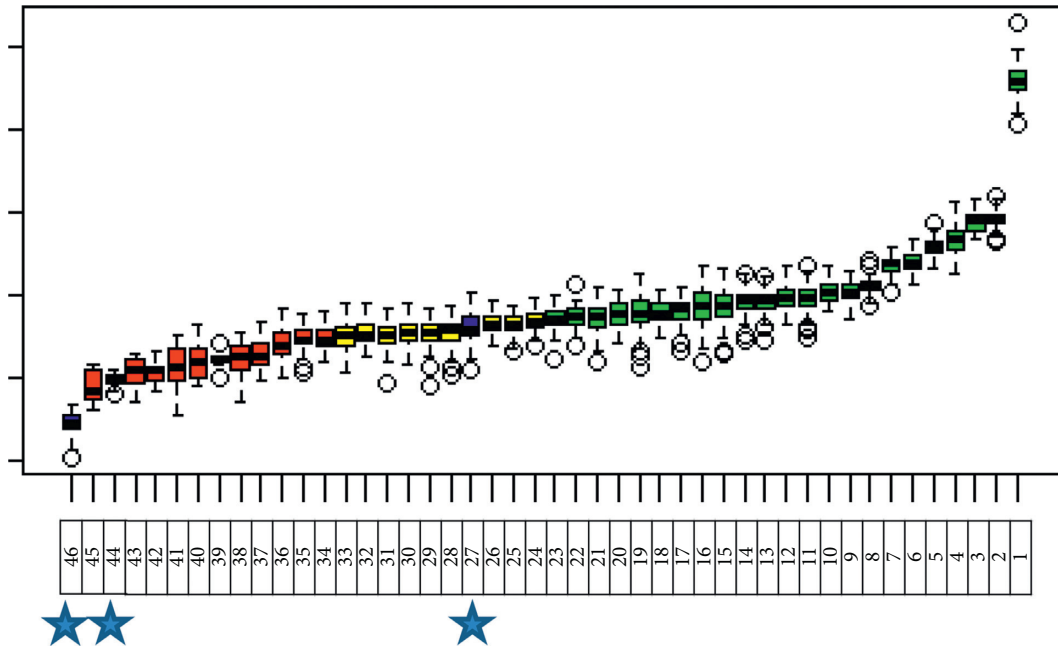


FIGURE 4: The importance scores of AVL variables.

TABLE 4: Boruta ranking of AVI variables.

ID	Variable
1	AvlRecords
2	Solid_PLM_Minute
3	Solid.Distance
4	DistMiles
5	Solid_PLM_
6	Solid
7	TotalDistSolid
8	SolidDistance.Distance
9	Dis_PLM_Minute
10	Solid_PLM_MinuteBefore
11	Laps
12	AfterRecords
13	Dis_PLM_
14	LaneLaps
15	BeforeRecords
16	RecordsBefore
17	PreLiquid_PLM_MinuteBefore
18	Liquid_PLM_Minute
19	Dis_PLM_MinuteBefore
20	PreLiquid_PLM_
21	Liquid_PLM_
22	DistMilesBefore
23	LiquidDistance.Distance
24	TotalDistLiquid
25	Solid.DistanceBefore
26	Liquid
27	Shadow Max
28	PreLiquid.Distance
29	TotalDistPrewet
30	Solid_PLM_Before
31	SolidBefore
32	Liquid.Distance
33	TotalDistSolidBefore
34	TotalDistPrewetBefore

TABLE 4: Continued.

ID	Variable
35	SolidDistance.DistanceBefore
36	Dis_PLM_Before
37	TotalDistSolidBefore
38	PreLiquid
39	LiquidBefore
40	Liquid_PLM_MinuteBefore
41	Liquid.DistanceBefore
42	PreLiquid.DistanceBefore
43	Liquid.DistanceBefore
44	Shadow Mean
45	LiquidDistance.DistanceBefore
46	Shadow Min

dropped from further consideration. The next variable was then considered until all highly correlated variables had been dropped. The resulting remaining list of variables was neither correlated nor displayed conflicting results from the ridge regression test. That is, as shown in Figure 6, none of these remaining variables display the erratic or unpredicted behavior as seen in Figure 3. Thus, our trimmed dataset listed in Table 5 is appropriate for estimating a negative binomial model. In other words, no matter which set of variables from the trimmed dataset are included in the final model, the results will be robust.

6. Exhaustive Search Function

To determine which of these trimmed variables should be included in our negative binomial model, an exhaustive search function was performed. The exhaustive search function calculates the expected outcome of each model

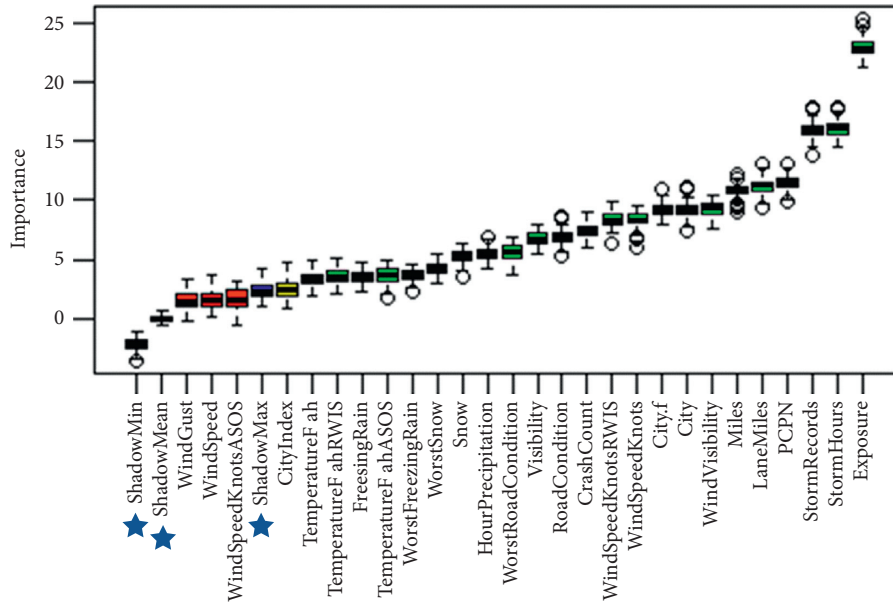


FIGURE 5: The importance scores of weather and traffic variables.

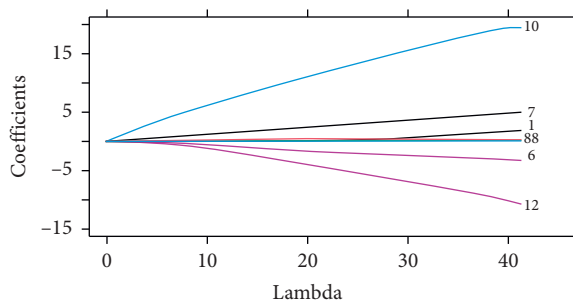


FIGURE 6: Ridge regression of trimmed variables.

TABLE 5: List of the trimmed variables.

ID	Variable
1	(Intercept)
2	LnExposure
3	RoadCondition
4	StormHours
5	TemperatureFahASOS
6	WindSpeedKnots
7	WorstFreezingRain
8	WorstSnow
9	WorstVisibility
10	AVLRecords
11	Solid_PLM_Minute
12	SolidDistance.Distance
13	Solid_PLM_MinuteBefore
14	AfterRecords
15	BeforeRecords
16	LiquidDistance.Distance
17	PreLiquid.Distance
18	SolidDistance.DistanceBefore

Note: see Table 1 for the definition of each variable.

based on the input variables and ranks it according to “R-squared” score or another specified criterion. For this analysis, our trimmed list of variables was used as the input for the exhaustive search function.

The highest criteria scores, that is, the highest adjusted R-squared scores, are reported in Figure 7. The maximum score is represented at the top of the Figure 7 graph along the y-axis. The x-axis resides the final set of variables selected from the trimmed dataset, or the variables displayed in Table 5. Each row represents a unique model’s combination of variable inputs, with the solid black cells indicating which variables are included for that specific model. When using the “regsubsets” package, as was used here, the default setting is that only the highest-ranking model for each varying number of input variables is displayed. For example, only 1 model with all 17 variables is displayed, along with only 1 model with 16 variables, and so on. This method shows how model accuracy changes as the input variables themselves change.

It can be seen in Figure 7 that the adjusted R-squared values of the top 7 rows are within 0.03 of each other. In other words, although the variables included in these models may differ, any of the top 7 rows of variables will provide accurate, reliable, and consistent results. This accords with the results from Figure 6 where none of these variables experienced a large swing in its coefficient based on complexity. Having multiple reliable model options allows researchers to include in their final model whichever set of variables are most policy relevant. Because this method only displays the highest scoring model for each number of variable inputs, there are many existing combinations of variables that will score approximately in the 0.4 range.

6.1. Final Model Creation. Once our final selection of variables had been made, a negative binomial model was used to

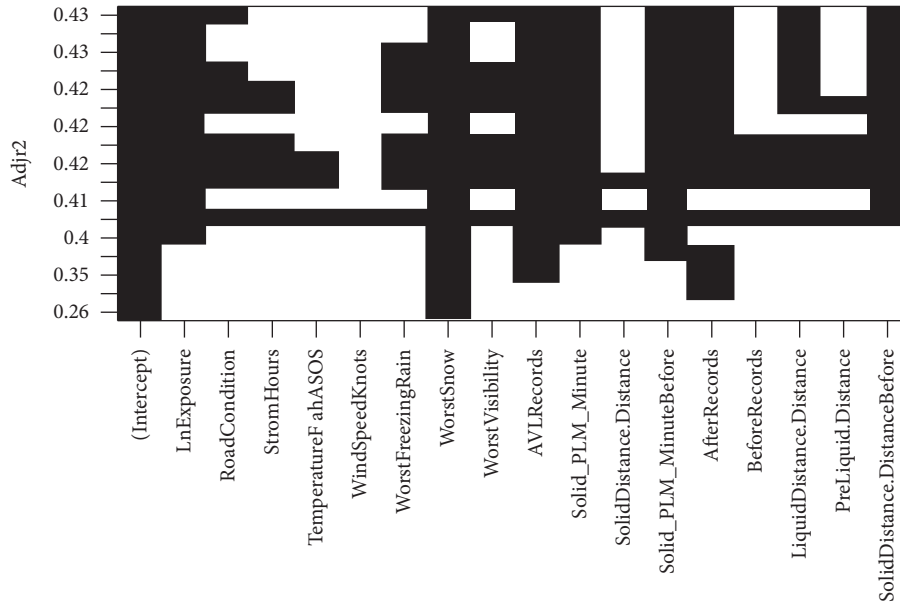


FIGURE 7: Adjusted R^2 exhaustive search output.

estimate crash frequency [25, 26]. The negative binomial model is written as in equation (1) using a fixed over-dispersion parameter [27]:

$$Y_i \sim NB(\mu_i, \alpha), \tag{3}$$

where Y_i is the number of crashes during a winter storm event i , ($i = 1, \dots, n$), μ_i stands for the mean crash frequency, and α is the overdispersion parameter.

It is assumed that μ_i is a function of explanatory variables such that

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \beta_{k+1} \ln(\text{Average Exposure}_i)), \tag{4}$$

where x_{ij} represents the j^{th} variable in event i . $\beta_0, \beta_1, \dots, \beta_{k+1}$ is a vector of regression parameters. As mentioned earlier, since the number of crashes is count data, to make it comparable across different events, the $\ln(\text{Exposure}_i)$ variable was devised as the offset variable in our negative binomial model.

7. Results and Discussion

Our final negative binomial model, presented in Table 6, includes variables from the top row of Figure 7. The “Pr ($>|t|$)” column provides a visual indicator of each variable’s significance (i.e., “ P value” below 0.05). Noted that “LiquidDistance.Distance” and “SolidDistance.DistanceBefore” variables were dropped in the final model. As seen from the R -squared score, no significant difference in performance was observed. Because of the selection process and validation, such adjustments based on researchers’ judgement will not impact model integrity as seen in Table 5 and Figure 6.

As expected, with “LnExposure” (i.e., vehicle count) and “Worst Snow,” the higher the number, meaning more traffic

TABLE 6: Crash frequency model.

Coefficients	Estimate	Std. Error	P value	Pr ($> t $)	
(Intercept)	-17.58	4.36	-4.03	0.00	***
LnExposure	1.42	0.60	2.35	0.02	*
RoadCondition	0.59	0.68	0.87	0.39	
WorstSnow	5.38	1.33	4.05	0.00	***
WorstVisibility	0.21	0.15	1.36	0.18	
AVLRecords	0.04	0.01	4.85	0.00	***
Solid_PLM_Minute	21.87	9.07	2.41	0.02	*
Solid_PLM_MinuteBefore	-11.90	3.81	-3.12	0.00	**
AfterRecords	0.03	0.01	1.96	0.05	.
R -squared					0.4359
F -statistic					18.35 on 8 and 190 DF

Significance codes: 0 “***”; 0.001 “**”; 0.01 “*”; 0.05 “.”.

or worse weather, the higher the expected crash frequency. The “AVLRecords” and “AfterRecords” variables signify that the more time snowplows are operating during and after a storm event, the higher the crash rate. This is probably because storms only mildly impacting roadway conditions do not require as much plowing effort as storms that severely impact roadway conditions. Note that the “RoadCondition” variable is not significant in the estimated model. This is because the limited number of Road Weather Information System (RWIS) sensors across the study area resulting in inadequate measurements of road weather conditions. Therefore, frequent snowplow operations can be considered as indicative of rapid degradation of roadway conditions.

The “Solid_PLM_Minute” variable represents the total amount of solid material spread divided by the total lane miles per minute of the storm event, i.e., how much material was spread on the roadway surface every minute during the storm. This variable has a positive coefficient, indicating that

the more solid material spread per lane mile per minute during snow events leads to a higher crash rate. This variable essentially represents the intensity of material spreading during a storm. In ideal conditions, agencies could plan for storm events and have the roadways treated with deicing material before the start of storms to limit degradation in roadway conditions. However, when storms are more severe, or they are flash storms that appear before proper pretreatment operations can take place, agencies might have to spread more materials during the storm, leading to a higher rate of spreading once the storm has begun. This may explain why a higher spreading rate per minute is associated with a higher crash rate.

The final variable, “Solid_PLM_MinuteBefore” represents the solid material spread per lane mile per minute before the storm starts. This variable has a negative coefficient, meaning that the more solid material spread before the start of the storm, the lower the crash rate. This suggests that proper planning can inhibit the degradation of conditions on the roadway. In turn, this will lower the amount of material needed during the storm event. In addition, past research has shown that a higher proportion of crashes occur at the beginning of a storm event [5]. Therefore, by mitigating the adverse conditions at the beginning of the storm event, a greater impact on the crash rate reduction can be achieved.

8. Conclusion

When working with large datasets including variables with complex interactions, agencies and researchers must find ways to perform effective analysis while also being able to present the results in a way that can be easily understood. This paper presents a framework for selecting variables from a complex and highly correlated dataset to develop a statistical model describing crash frequency in winter weather conditions. In particular, a machine-learning algorithm, known as the Boruta algorithm and exhaustive search are used to select a list of variables to be included in the final negative binomial crash frequency model. This method provides consistent and intuitive results because the process reduces the complexity of interactions amongst different variables in the dataset.

By following this process, the current paper developed a crash frequency model for winter-weather-related crashes. This model can help agencies identify effective ways to improve roadway safety via winter maintenance operations. For example, by increasing the plowing operations before the start of storms, a decrease in crash rates is observed. Previous works have shown that the beginning of a winter storm carries the highest proportion of crash events compared to any other point during the event [5, 13]. Thus, pretreatment operations could have a significant role in mitigating the impact of winter storms.

Data Availability

The data used in this research are available through Iowa DOT Open Data <https://data.iowadot.gov/>.

Conflicts of Interest

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

Acknowledgments

This research was funded by the Traffic Safety Improvement Program (TSIP) of Iowa Department of Transportation.

References

- [1] Federal Highway Administration (FHWA), *Road Weather Management Program*, Federal Highway Administration (FHWA), Montgomery, AL, USA, 2020, https://ops.fhwa.dot.gov/weather/weather_events/snow_ice.htm.
- [2] T. Usman, L. Fu, and L. F. Miranda-Moreno, “Quantifying safety benefit of winter road maintenance: accident frequency modeling,” *Accident Analysis & Prevention*, vol. 42, no. 6, pp. 1878–1887, 2010.
- [3] Z. Hans, N. Hawkins, K. Gkritza, and M. Shaheed, “Nlenanya, inya. “Safety and mobility impacts of winter weather—phase 3,”” *Intrans Project Reports*, vol. 82, 2014.
- [4] Federal Highway Administration, *Signalized Intersections Informations Guide*, Federal Highway Administration, Montgomery, AL, USA, 2013, <https://safety.fhwa.dot.gov/intersection/conventional/signalized/fhwasa13027/>, Second edition.
- [5] X. Qin, D. A. Noyce, C. Lee, and J. R. Kinar, “Snowstorm event-based crash analysis,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1948, no. 1, pp. 135–141, 2006.
- [6] M. B. Kurska and W. R. Rudnicki, “Feature selection with the boruta package,” *Journal of Statistical Software*, vol. 36, no. 11, 2010.
- [7] M. B. Kurska, A. Jankowski, and W. R. Rudnicki, “Boruta—a system for feature selection,” *Fundamenta Informaticae*, vol. 101, no. 4, pp. 271–285, 2010.
- [8] W. A. Nixon and L. Qiu, “Effects of adverse weather on traffic crashes: systematic review and meta-analysis,” in *Proceedings of the TRB 87th Annual Meeting 2008*, Washington, DC, USA, 2008.
- [9] A. S. Potter, M. R. Gallagher, and C. W. Bayer, *Synthesis on GPS/AVL Equipment Used for Winter Maintenance (No. CR 14-01)*, Department of Transportation Clear Roads Pooled Fund, Saint Paul, MN, USA, 2016.
- [10] W. Schneider, M. Crow, W. A. Holik et al., *Evaluation of the GPS/AVL Systems for Snow and Ice Operations Resource Management (No. FHWA/OH-2017-31)*, Department of Transportation Office of Statewide Planning and Research, Columbus, OH, USA, 2017.
- [11] Z. Hans, N. Hawkins, P. Savolainen, and E. Rista, *Operational Data to Assess Mobility and Crash Experience during Winter Conditions*, Center for Weather Impacts on Mobility and Safety, Cary, NC, USA, 2018.
- [12] C. Chen and X. Shi, *Modeling the Macroscopic Effects of Winter Maintenance Operations on Traffic Mobility on Washington Highways*, Center for Advanced Multimodal Mobility Solutions and Education, Charlotte, NC, USA, 2019.
- [13] B. Hallmark and J. Dong, “Examining the effects of winter road maintenance operations on traffic safety through visual analytics,” in *Proceedings of the 23rd International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Rhodes, Greece, September 2020.

- [14] S. Mafi, Y. AbdelRazig, and R. Doczy, "Machine learning methods to analyze injury severity of drivers from different age and gender groups," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 38, pp. 171–183, 2018.
- [15] O. A. Osman, M. Hajij, P. R. Bakhit, and S. Ishak, "Prediction of near-crashes from observed vehicle kinematics using machine learning," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 12, pp. 463–473, 2019.
- [16] A. Theofilatos, C. Chen, and C. Antoniou, "Comparing machine learning and deep learning methods for real-time crash prediction," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 8, pp. 169–178, 2019.
- [17] M. Amiri, H. R. Pourghasemi, G. A. Ghanbarian, and S. F. Afzali, "Assessment of the importance of gully erosion effective factors using boruta algorithm and its spatial modeling and mapping using three machine learning algorithms," *Geoderma*, vol. 340, pp. 55–69, 2019.
- [18] S. Demirogluk, M. A. Yazici, K. Ozbay, and J. A. Carnegie, "Feature selection for ranking of most influential variables for evacuation behavior modeling across disasters," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2599, no. 1, pp. 24–32, 2016.
- [19] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and boruta-random forest hybridizer algorithm approach," *Catena*, vol. 177, pp. 149–166, 2019.
- [20] J. Li, M. Tran, and J. Siwabessy, "Selecting optimal random forest predictive models: a case study on predicting the spatial distribution of seabed hardness," *PLoS One*, vol. 11, no. 2, Article ID e0149089, 2016.
- [21] M. Christ, "Distributed and parallel time series feature extraction for industrial big data applications," in *Proceedings of the ACML Workshop on Learning on Big Data*, Hong Kong, China, 2016.
- [22] B. Hallmark, "Analyzing winter weather impact on safety using snowplow automatic vehicle location," 2019.
- [23] M. Agarwal, T. H. Maze, and R. Souleyrette, *Impact of Weather on Urban Freeway Traffic Flow Characteristics and Facility Capacity*, Center for Transportation Research and Education, Ames, IA, USA, 2005.
- [24] K. Kim, "Ridge Regression for Better Usage: Medium, Towards Data Science," January 2019, <https://towardsdatascience.com/ridge-regression-for-better-usage-2f19b3a202db>.
- [25] D. Lord and F. Mannering, "The statistical analysis of crashfrequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.
- [26] A. Goswamy, S. Hallmark, and T. Litteral, *Impact of Destination Lighting and Other Factors on Driver's Injury Severity of Nighttime Crashes at Rural Stop-Controlled Cross-Intersections Using Proportional Odds Model*, Transportation Research Board, Washington, DC, USA, 2018.
- [27] X. Guo, L. Wu, Y. Zou, and L. Fawcett, "Comparative analysis of empirical bayes and bayesian hierarchical models in hot-spot identification," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 7, pp. 111–121, 2019.

Research Article

Origin-Destination-Based Travel Time Reliability under Different Rainfall Intensities: An Investigation Using Open-Source Data

Qi Zhang ¹, Hong Chen ¹, Hongchao Liu ², Wei Li ³, and Yibin Zhang ²

¹College of Transportation Engineering, Chang'an University, Xi'an 710000, China

²Department of Civil, Construction and Environmental Engineering, Texas Tech University, Lubbock, TX 79409, USA

³School of Civil Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

Correspondence should be addressed to Hong Chen; glch@chd.edu.cn

Received 19 July 2020; Revised 12 October 2020; Accepted 24 October 2020; Published 4 November 2020

Academic Editor: Yajie Zou

Copyright © 2020 Qi Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Origin-destination- (O-D-) based travel time reliability (TTR) is fundamental to next-generation navigation tools aiming to provide both travel time and reliability information. While previous works are mostly focused on route-based TTR and use either ad hoc data or simulation in the analyses, this study uses open-source Uber Movement and Weather Underground data to systematically analyze the impact of rainfall intensity on O-D-based travel time reliability. The authors classified three years of travel time data in downtown Boston into one hundred origin-destination pairs and integrated them with the weather data (rain). A lognormal mixture model was applied to fit travel time distributions and calculate the buffer index. The median, trimmed mean, interquartile range, and one-way analysis of variance were used for quantification of the characteristics. The study found some results that tended to agree with the previous findings in the literature, such that, in general, rain reduces the O-D-based travel time reliability, and some seemed to be unique and worthy of discussion: firstly, although in general the reduction in travel time reliability gets larger as the intensity of rainfall increases, it appears that the change is more significant when rainfall intensity changes from light to moderate but becomes fairly marginal when it changes from normal to light or from moderate to extremely intensive; secondly, regardless of normal or rainy weather, the O-D-based travel time reliability and its consistency in different O-D pairs with similar average travel time always tend to improve along with the increase of average travel time. In addition to the technical findings, this study also contributes to the state of the art by promoting the application of real-world and publicly available data in TTR analyses.

1. Introduction

Travel time reliability (TTR) plays a vital role in various applications such as evaluation of network performance [1], measuring the improvement of traffic operations and management strategies [2], quantification of service quality [3], enhancing the experience of traveler's route choice [4], and determining freeway bottlenecks [5].

Among the route level (microscopic), origin-destination (O-D) level (mesoscopic), and network level (macroscopic) studies, the route level TTR analyses have received much more attention in the past. Besides the demand from the practical side, route-based data which are relatively easier to obtain should be another reason. For instance, using the data from California State Route

91, one research found that traveler's route choice was more delicate to TTR than travel time [6]. Chepuri et al. assessed the performance of various TTR measures with bus route data collected in Chennai, India. They recommended using 95th percentile travel and buffer time as reliability indicators for bus routes [7]. Some recent route-based studies can be found in [8–12]. Because the focus of route-based studies is usually on one or a few specific routes, data were usually project-specific and most were discarded upon completion of a project, which makes continuous investigation difficult and sometimes impossible.

It is not uncommon that under certain circumstances route-based and O-D-based analyses may get similar or the same results because a route is associated with at least one

origin and one destination and thus can be viewed as a particular case of an O-D-based study. Comparison studies are limited, though. In [13], the authors concluded that there were no significant differences between O-D-based and route-based estimates in most part of the studied time periods. In [14], the researchers found that adding an alternative path tends to decrease the O-D-based TTR. Network level TTR studies are mostly simulation-based in a lack of real-world data. Some notable studies include but are not limited to the work of [15–18]. Studies based on traffic simulation are sometimes subject to serious errors caused by the underlying problems of the simulation model. A detailed discussion of simulation-based approaches is beyond the scope of the study.

Many factors, such as connected vehicles, traffic incidents, weather, work zones, special events, types of traffic control, and the dynamics of traffic flow, have impacts on TTR. Accordingly, the study of the impact of these factors has become one of the prominent topics in the TTR field [19]. For instance, in the literature [20], the authors attempted to quantify the contribution of various features on TTR and found that demand-capacity imbalance and accidents are the two factors that most affect TTR. In [21], researchers uncovered that deployment of connected vehicles improves TTR in the work zone environment, and higher benefits come along with higher market penetration levels. Additionally, the impact of rain on speed and travel time and the route level TTR have been well studied, and some results are conducive. For example, studies have discovered that speed reduction could vary from 10% to 25% in general rainy days [22] and an average increase of travel time by 11% might be expected in peak hours under the impact of a certain level of precipitation [23]. Adverse weather exacerbates TTR, especially during peak periods [24]. However, some findings are controversial and need further investigation, especially when it relates to TTR. Chien and Kolluri found that TTR would diminish when weather condition changes from dry to rain, as indicated by an expansion of 16% in the buffer index [10], while in another study [25], the authors suspected that rain and snow might have caused lower standard deviation and coefficient of variation of travel time and thus increased TTR. While it is understandable that different studies may produce controversial results, ad hoc data might have played a role. In a review of the literature, we found that studies are limited with respect to the impact of rain on the O-D level TTR; moreover, most of the data used in previous studies were project-specific and only covered a short period of time that was not even sound for a full-scale statistical analysis.

1.1. Research Objectives. The availability of publicly available open-source data in recent years has made a detailed investigation of O-D-based TTR possible. A major thrust of this study is to use Uber Movement data and Underground Weather data to systematically analyze the impact of rain on O-D-based TTR. Uber and Underground Weather data provide an ideal and probably the only opportunity for applying real-world data in such studies because Uber data

are O-D-specific and cover a lot longer time span while Underground Weather data provide very detailed weather data. A significant contribution of using publicly available data is that the results can be easily verified and compared to those that use ad hoc data or computer-aided simulation, studies based on real-world data always have a better value in the literature. Additionally, O-D-based TTR is fundamental to next-generation-navigation tools that are aimed at providing both travel time and reliability information. This paper only focuses on the impact of rain, but there are a lot more deserving further investigations along this line, such as the impact of other weather events and the combined effects of weather and work zones.

In this study, the authors investigated the impact of rain at various levels of intensity on O-D-based TTR, through the analyses of three-year travel time and weather data and a hundred O-D pairs collected from downtown Boston. A general lognormal mixture model was adopted to fit distributions and calculate the buffer index values. While a portion of the results was in proper alignment with previous studies, some turned out to be unique.

The rest of the paper is organized as follows: Section 2 introduces the data used in this research, which includes the O-D-based travel time data from Uber Movement and historical weather data collected from the Weather Underground website; Section 3 depicts the typical TTR measures and the analytical approach developed based on the Gaussian mixture model; Section 4 presents the results, and Section 5 summarizes the findings and conclusions and concludes the paper by discussions and future research.

2. Data

2.1. Uber Movement and Weather Data. The O-D-based travel time data used in this research are from Boston, the United States, retrieved from the Uber Movement website (<https://movement.uber.com>). The website provides detailed information on average travel time (ATT), classified by five-time intervals during a day, including early morning (00–07 h), AM peak (07–10 h), midday (10–16 h), PM peak (16–19 h), and evening (19–24 h). To make the results statistically sound, three-year data were used, which span from 1/1/2016 to 12/31/2018, and a hundred O-D pairs were selected. Uber already classified ATT ranges by 5-minute intervals. Considering that lots of data seem to be missing in the dataset with the ATT range of 25 minutes and beyond, we selected five categories in the analysis: (a) 0–5 mins, (b) 5–10 mins, (c) 10–15 mins, (d) 15–20 mins, and (e) 20–25 mins.

Figure 1 shows all the origin and destination nodes included in this study. Table 1 depicts the O-D pairs classified into five groups based on the ATT. Note that travel time in the table is directional and one-way (e.g., 2–16 indicates node 2 to node 16) because of the limited availability of two-way travel time in-between the O-D pairs. Twenty origin-destinations were selected in each ATT category. Table 2 presents descriptive statistics for ATT data from Uber, including mean and standard deviation (SD) in different periods. The studied area is in level terrain.

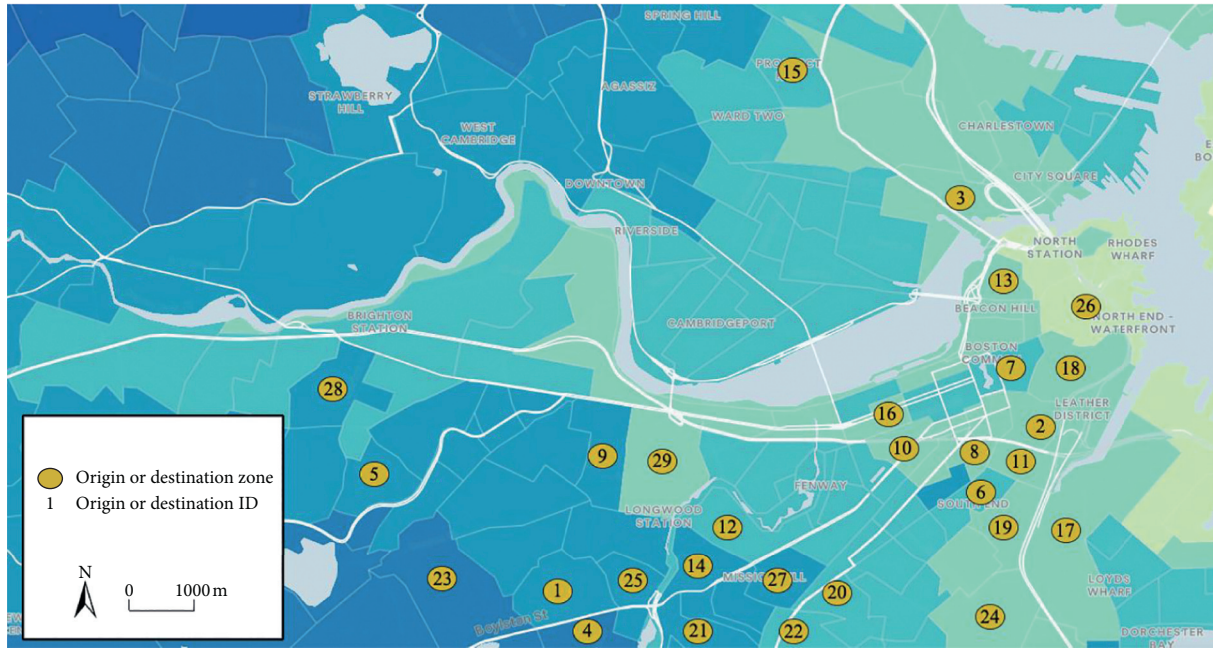


FIGURE 1: The origin and destination nodes.

TABLE 1: O-D pair information in each range of TTR.

ATT range (min)	O-D pairs (ID)
0-5	2-16, 6-10, 6-11, 7-18, 8-18, 8-29, 10-11, 11-6, 11-10, 11-26, 13-12, 17-18, 17-24, 18-2, 18-7, 18-26, 22-27, 26-13, 26-18, 27-22
5-10	2-13, 2-26, 6-7, 7-6, 8-24, 10-6, 12-16, 13-2, 13-16, 16-2, 16-13, 18-4, 18-8, 18-24, 24-8, 24-20, 26-2, 26-7, 26-11, 29-18
10-15	2-12, 6-26, 8-20, 10-26, 12-2, 12-13, 13-26, 15-26, 16-26, 18-20, 20-18, 20-29, 24-8, 24-29, 26-6, 26-10, 26-15, 26-16, 29-20, 29-24
15-20	2-9, 2-14, 9-2, 9-18, 9-26, 12-11, 12-17, 12-18, 12-19, 12-26, 14-2, 17-14, 18-12, 22-26, 26-9, 26-12, 26-14, 26-22, 26-27, 27-26
20-25	1-18, 1-26, 3-18, 5-26, 17-23, 18-1, 18-3, 18-21, 18-23, 18-25, 21-18, 23-17, 23-18, 23-26, 25-18, 26-2, 26-5, 26-23, 26-28, 28-26

TABLE 2: Descriptive statistics for ATT data from Uber (min).

ATT range (min)	O-D pair count	Early morning (00-07 h)		AM peak (07-10 h)		Midday (10-16 h)		PM peak (16-19 h)		Evening (19-24 h)	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0-5	20	4.11	1.39	4.53	2.01	4.35	1.35	4.95	2.15	3.88	1.15
5-10	20	7.27	1.70	7.82	1.92	7.33	1.66	8.39	2.62	6.31	1.42
10-15	20	11.96	2.38	13.36	3.06	13.01	2.50	14.83	4.01	10.87	1.95
15-20	20	16.62	3.15	17.90	3.64	18.61	2.83	19.82	4.47	15.81	2.07
20-25	20	23.88	4.24	24.54	4.57	23.86	3.28	25.55	5.49	21.07	3.06

The corresponding weather data in the subject area were collected from the Underground Weather data at <https://www.wunderground.com/history>. In this study, weather conditions with no precipitation, such as clear, cloudy, or overcast, are classified as normal weather. Meanwhile, fog or haze conditions were excluded so that the focus could be placed on the impact of rain. Rain condition was defined as the rainy weather that caused effective precipitations.

In alignment with the 24-hour travel time data, the sum of rainfall in each matched period (like 00-07 h) was calculated and converted into a 24-hour value. Table 3 summarizes the definition of the data used in this research.

2.2. Data Screening. The travel time data have already been preprocessed and filtered by Uber before uploading to the Internet. In general, the data were well prepared, and the work

TABLE 3: Summary of data attributions and interpretation.

Attribute	Interpretation
Origin-destination	See Table 1
Dates	From 1/1/2016 to 12/31/2018
Days of week	From Sunday to Saturday
Time periods	Early morning (00–07 h), AM peak (07–10 h), midday (10–16 h), PM peak (16–19 h), and evening (19–24 h)
Travel times	Average O-D-based travel times (s)
Rainfall	Sum of rainfall per time period (mm)
Weather conditions	Normal, light rain (0.0 and 10 mm/24 h), moderate rain (10.0 mm/24 h and 25.0 mm/24 h), heavy rain (25.0 mm/24 h and 50.0 mm/24 h), extreme rain (>50.0 mm/24 h), and others (e.g., snow and fog)

for outlier removal was rather simple. There are a couple of null cells without any data within the ATT range of 20–25 mins, such as from Lincoln Road (ID: 1) to Harborwark (ID: 26); these cells were removed. Additionally, Boston has considerable snowfalls in winter months, which may cause a negative impact for a longer period of time even after the snow. According to the analysis in [26], the impact of snow on travel time was associated with the severity of snow and road conditions, and it usually takes at least six hours after the snow for travel time to become stabilized. In this study, the data recorded one-day after regular snow (≤ 5.0 mm/24 h) and two-day after heavy snow (> 5.0 mm/24 h) were excluded. As a result, around 540,000 valid records, including nearly 68,000 light rain records, 18,000 moderate rain records, 11,000 heavy rain records, and 13,000 extreme rain records, were included in the study.

3. Methods

3.1. Measures of Travel Time Reliability. In addition to the conventional measures (mean, standard deviation, and coefficient of variation, e.g.), there are some other TTR measures, such as travel time variability (TTV), planning travel time index (PI), and buffer index (BI). Among these measures, BI has been widely utilized in existing literature [7, 8, 13, 27, 28] and as concluded in [29] has a high consistency with the coefficient of variation and thus is best suitable for the measure of TTR. The authors adopted the idea and took BI as the primary TTR measure. Then, we used interquartile range (IQR), the median, and trimmed mean of BI-values, as well as the analysis of variance for BI variation ratio, to quantify the impact from rainfall intensity.

The buffer index can be generally formulated as follows:

$$BI = \left(\frac{T_p - T_c}{T_c} \right), \quad (1)$$

where T_p is the percentile travel time and T_c is the contrastive travel time (e.g., mean travel time, median travel time, and free-flow travel time).

Obviously, the higher the BI-value is, the less reliable the travel time will be. In this study, the 95th percentile travel time T_p and the mean travel time T_c were adopted as the BI parameters. The interquartile range is the distance between the 75th and 25th percentiles, and the trimmed mean excludes the 5% highest and the 5% lowest data for reducing the error caused by the extreme data.

3.2. Lognormal Mixture Model. With respect to the calculation of the percentile value in BI, in earlier literature [30], the author directly calculated the percentile value according to the available data without considering the statistical regression, which was easily subject to statistical errors (e.g., regression to the mean). In later studies, various regression methods were applied, such as multiple linear regression [20], and continuous probability distribution functions, such as Weibull distribution [31], lognormal distribution [32], and generalized Pareto distribution [33]. Multiple regression was found to be more suitable for the multiparameter impact study, while for the single factor (e.g., rainfall intensity) analysis, the latter seems to be more desirable. However, the complexity of the problem makes it difficult for the real data to fit well with traditional prior distribution, such as the lognormal distribution. Recent studies attempted to use multistate models, such as the Gaussian mixture model, lognormal mixture model, and gamma mixture model for better results [8, 13, 14, 16, 17, 34–36]. Among these methods, the lognormal mixture model (LMM) is outperformed and was recommended by many researchers [8, 34, 36]. When LMM is applied, the best fitting usually occurs at a low K -value (e.g., $K = 2$ or 3), which may also help improve computational efficiency. Accordingly, LMM was selected for this research.

LMM is essentially a linear combination of multiple lognormal distributions with a weight sum value of 1. The general formula of LMM is as follows:

$$P_{LMM} \left(\frac{t}{\omega_k, \mu_k, \sigma_k} \right) = \sum_{k=1}^K \omega_k L \left(\frac{t}{\mu_k, \sigma_k} \right), \quad (2)$$

where t is the travel time; $\omega_k, \mu_k, \sigma_k$ are the weight, mean, and standard deviation of the k^{th} lognormal distribution, respectively; and L is the lognormal probability density. The equation is subject to $\sum_{k=1}^K \omega_k = 1$.

3.3. Expectation-Maximization Algorithm. Since mixture models (like LMM) involve latent variables, maximum likelihood estimate (MLE) cannot be used directly to estimate the parameters. Presently, the expectation-maximization (EM) algorithm is the most commonly used approach for multimodal parameter estimates, where an expectation (E) step calculates the expected log-likelihood by estimating the current parameters, and a maximization (M) step

maximizes the expectation of the log-likelihood in the E step. Algorithm 1 depicts the complete process of the EM algorithm.

3.4. Supplement Algorithms. Before the application of LMM and EM, two issues need to be addressed: the optimal K-value and the inverse function of the cumulative distribution function (CDF) of LMM. The former can be resolved by referring to the method in [34], where the K-value was determined by the minimum Akaike information criterion (AIC) estimation with the null hypothesis not rejected by the one-sample Kolmogorov-Smirnov (K-S) test. AIC is defined as

$$AIC = 2C - 2 \ln(Li), \quad (3)$$

where C is the number of parameters and Li is the likelihood function.

For the second issue, since there is no corresponding original form of the CDF of LMM, it is impossible to obtain the percentile value by solving the inverse function of the original function. For this reason, the bisection method was adopted, with a stop threshold of 0.00001. The complete TTR estimation framework is presented in Figure 2.

4. Results and Discussion

The authors calculated the BI-value of each O-D pair under six different weather conditions, that is, normal, light rain (0 and 10.0 mm/24 h), moderate rain (10.0 and 24.9 mm/24 h), heavy rain (25 and 50.0 mm/24 h), extreme rain (>50.0 mm/24 h), and rain (>0 mm/24 h). We summarized all calculations into five groups according to the TTR range, that is, 0–5 mins, 5–10 mins, 10–15 mins, 15–20 mins, and 20–25 mins, including three location measures (median, trimmed mean, and interquartile range) and the one-way analysis of variance. MATLAB was used to run the s, and the final buffer index is presented in the form of the average value calculated after 50 fittings.

4.1. Increasing TTR Reduction Impact by Rain. Figure 3 shows an example fitting of LMM from Harrison Ave (ID: 11) to Huntington Ave (ID: 10) in the ATT range of 0–5 mins, which depicts a higher TTR under light rain and lower TTR under the rest of the rainfall intensity. Moreover, the impact of the O-D-based TTR increased with the increase of rainfall intensity (the variation of BI-value from 0.3332 to 0.4379). Overall, it shows that rain reduced the O-D-based TTR (see in Figure 3(f)).

More analyses were conducted for further investigation. Figure 4 and Table 4 (median of BI-values) interpret the results in terms of the median value, which shows that rain reduced the O-D-based TTR in each ATT range. Additionally, four out of five subfigures in Figure 4 demonstrate an increasing trend in TTR reduction when the rainfall intensity increases, with only one exception when being within the ATT range of 20–25 mins (BI-value decrease under heavy rain (0.3489) compared to that under moderate rain (0.3588)). The global mean values of the median under

different rainfall intensity (the 8th row in Table 4) also revealed the trend.

Likewise, the trimmed mean of BI-values in Table 4 shows that rain has adverse effects on O-D-based TTR and the increasing reduction effect in terms of the global mean (the 14th row). In terms of time ranges, three out of five presented the increasing adverse effect in the range of 0–5 mins, 10–15 mins, and 20–25 mins.

Regarding the O-D-based results (see 8th column in Table 5), thirty-five out of a hundred O-Ds strictly met the regularity (an average of seven O-Ds in each range). Meanwhile, ninety-one out of a hundred O-D pairs (see 7th column in Table 5) show that the rain reduces the O-D-based TTR.

By far, not surprisingly, a dominating feature is that rain reduces O-D-based TTR, which aligns properly with people's perception as well as previous research at route level [10, 22, 23, 37, 38]. Notably, the low-probability anomalies (positive effect of rain), though not sufficient to negate the conclusion, may be a combination of multiple factors in a real environment, for example, a combination of rain, accidents, and work zones. This counterintuitive phenomenon will be discussed in a subsequent subsection specifically. More importantly, the results reveal that the negative impact grows with the increase in rainfall intensity. This trend seems doubtful, partly due to the exceptions in Table 4 and partly due to the fact that solely 35 percent of O-D pairs strictly conform to this trend. The reasons are generally twofold: (1) the interference of the complicated environment; (2) more seriously, the quite insignificant difference in impact between moderate, heavy, and extreme rain (this characteristic will be additionally discussed in subsequent subsection), resulting in extra challenging to achieve O-D pairs with satisfying the trend. Nonetheless, the current consequences can still expose the trend effectively.

4.2. Significant Impact from Light Rain to Moderate Rain.

A unique finding was revealed from the analysis; that is, the impact is a lot more significant when rainfall intensity changes from light to moderate, while other changes among the rainfall intensity categories seemed to cause only moderate impacts on the O-D-based TTR.

Comparing the variation ratio values in the parentheses in Table 4, we found that the average increase from light to moderate is up to 17.2% (median) and 14.6% (trimmed mean). On the contrary, the average BI variation ratio between the rest of the conditions is not significant, which is 2.7% (normal to light), 1.1% (moderate to heavy), and 5.8% (heavy to extreme), respectively, and 3.3%, 3.6%, and 4.0% in average trimmed mean values, respectively.

The one-way analysis of variance (ANOVA) was used to demonstrate the statistical significance of the BI variation ratio between different rainfall intensity categories. Considering the trimmed mean in the analysis, we trimmed 10% values for the test as well. The analysis was conducted for a significance level of 0.05. As presented in Table 6, the null hypothesis is rejected with the P value of $3.5874e-16$, far less than 0.05 in the 2-column source, which indicates significant

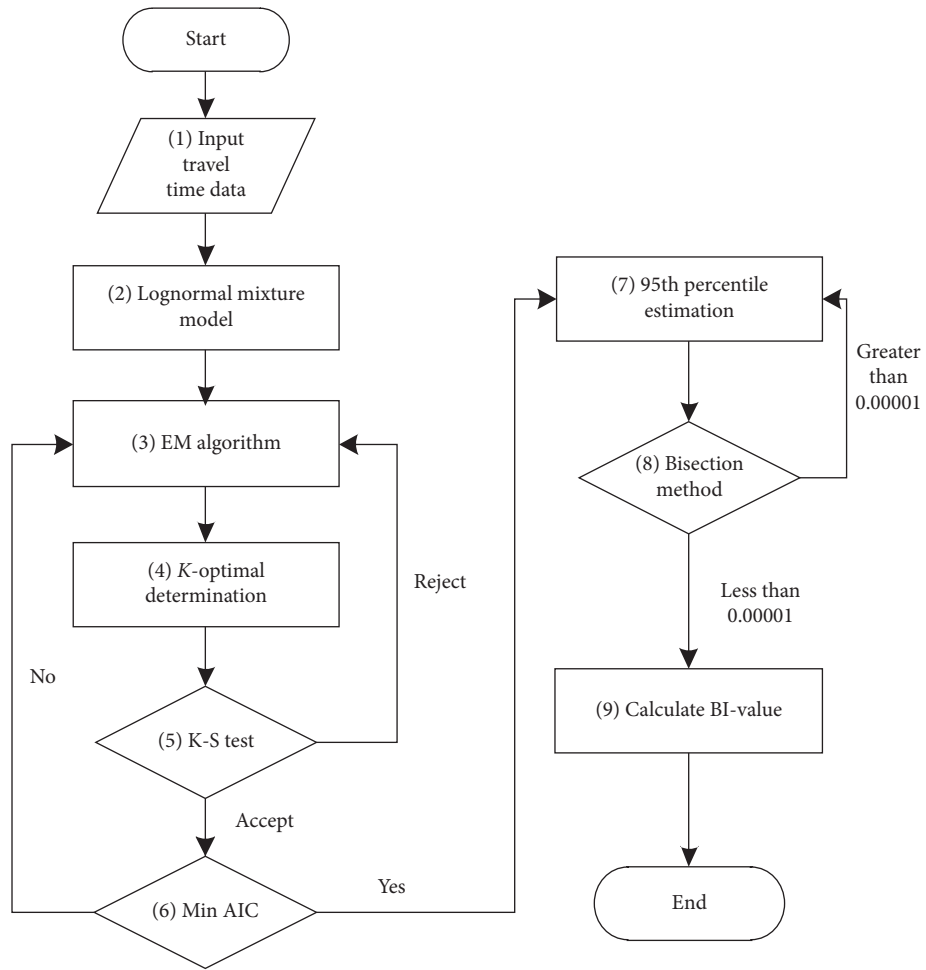


FIGURE 2: Framework for estimating O-D-based TTR.

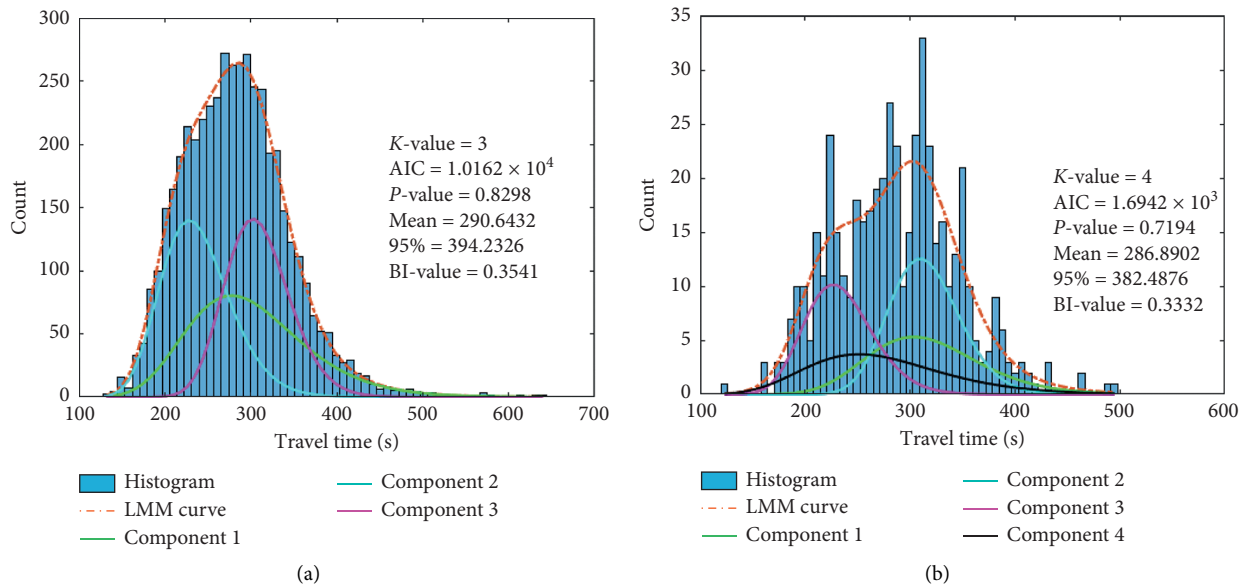


FIGURE 3: Continued.

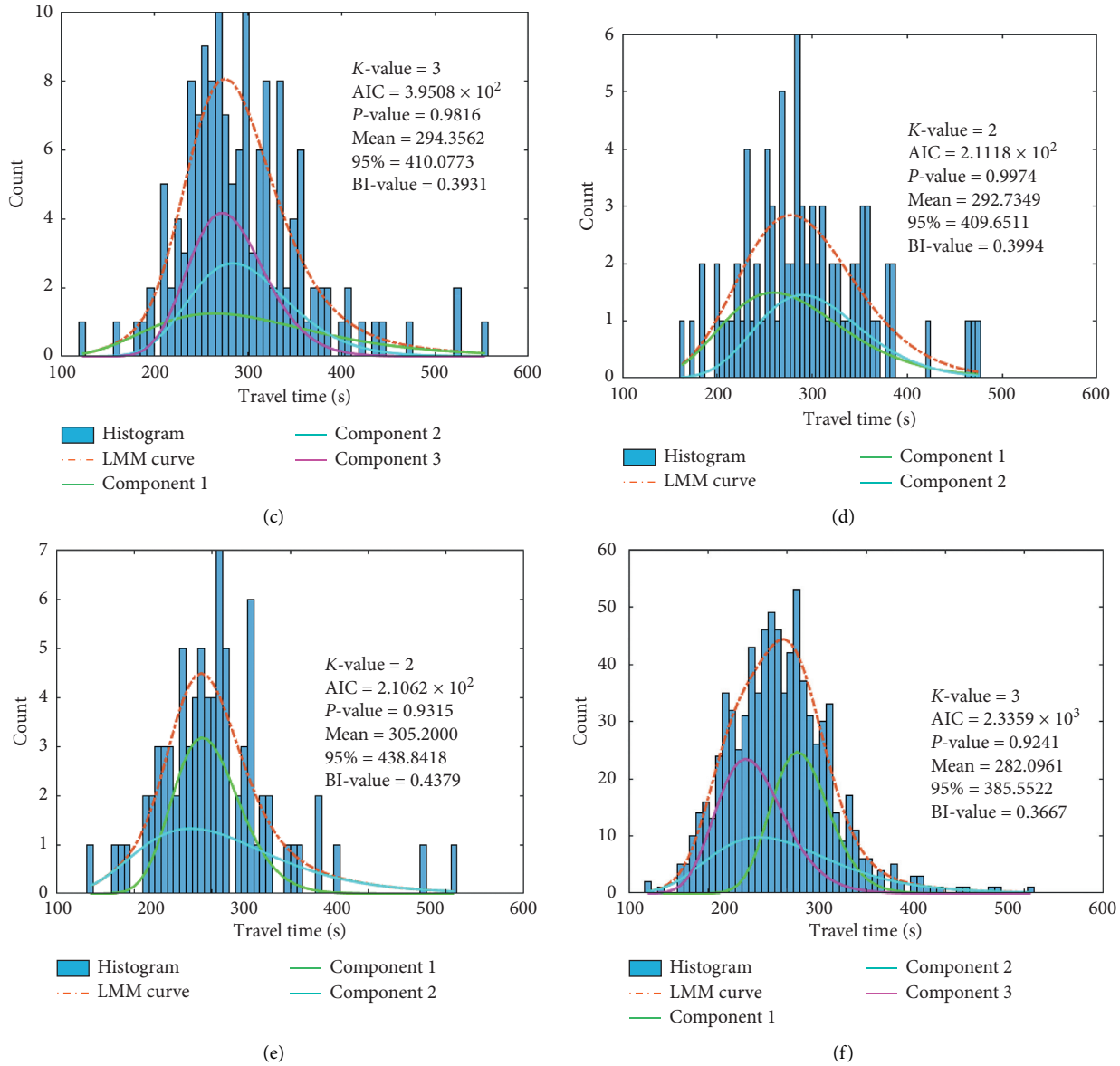


FIGURE 3: Fitting examples of LMM from Harrison Ave (ID: 11) to Huntington Ave (ID: 10): Part A: (a) normal and (b) light rain. Part B: (c) moderate rain and (d) heavy rain. Part C: (e) extreme rain and (f) rain.

variation between the light and the moderate rain condition. Notwithstanding, the null hypothesis is true when testing significance is in the 3-column source (P value > 0.05), which supports the previous analysis that there is no significant difference between moderate rain, heavy rain, and extreme rain.

The finding reveals that drivers are more sensitive to the change from light rain to moderate rain. While further investigation on driver behavior may be needed to fully explain this phenomenon fully, this finding is undoubtedly helpful in conducting more detailed and in-depth O-D-based TTR analysis.

4.3. Other Findings regarding O-D-Based TTR. Another notable finding from this research is that the O-D-based TTR tends to improve when ATT is longer, regardless of

the normal and the rainy weather. It was explicitly recognized from Table 4 (column 3 to column 8) that four out of six columns in the median and five out of six columns in the trimmed mean both illustrate that the O-D-based TTR increased along with the increase of average travel time.

Further investigation was conducted against the O-D pairs with the same ATT. The authors calculated the interquartile range of each travel time range, as illustrated in Table 7 and Figure 4 (the areas of the rectangle). It can be found that the areas of the rectangle are shrinking from Figures 4(a)–4(e). Referring to the columns in Table 7, two out of six columns follow this trend. For the rest of the columns, there is only one exception in each column (e.g., the last cell in column 3). The results demonstrate that the consistency of the O-D-based TTR in different

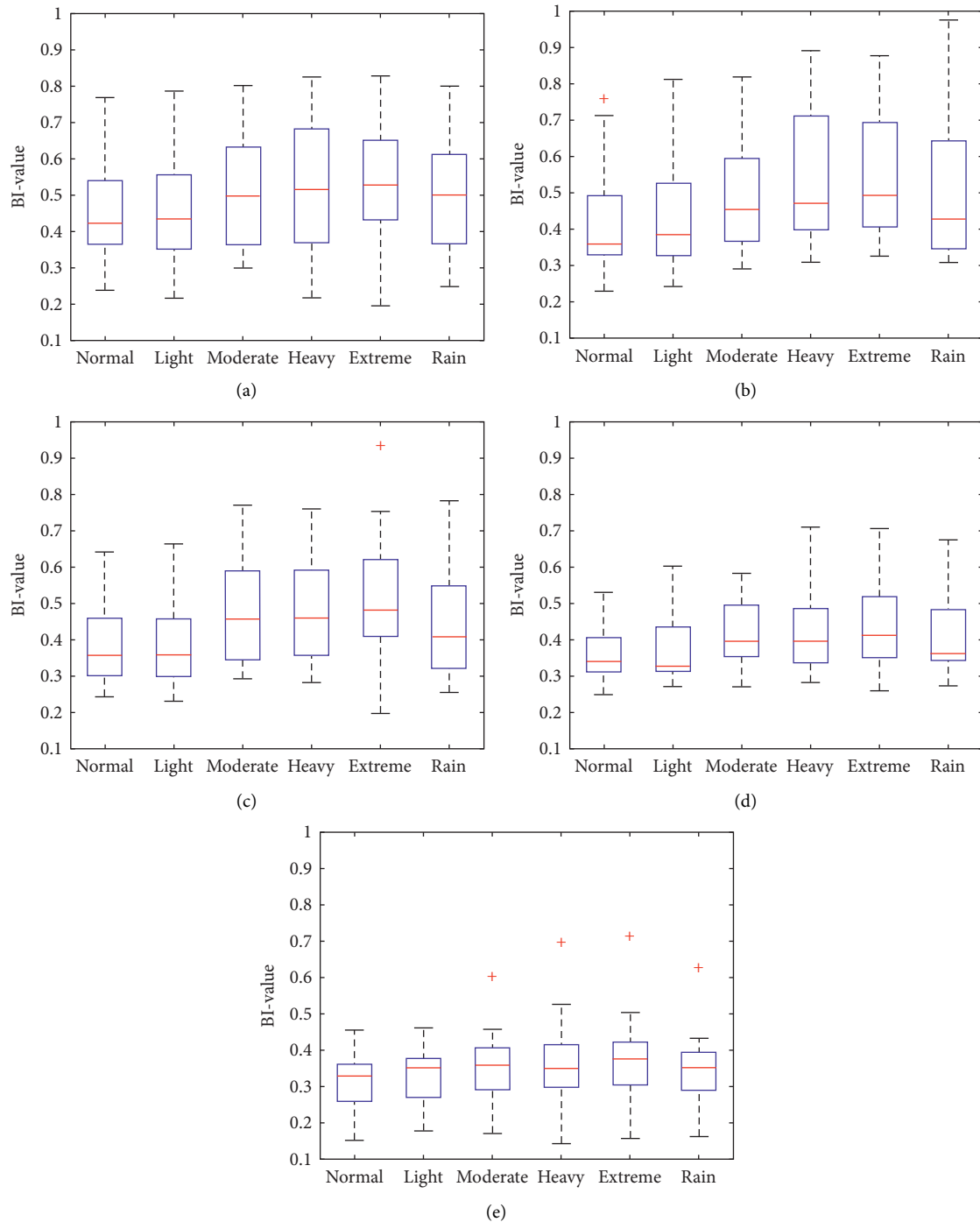


FIGURE 4: BI-value boxplots under six conditions within five travel time ranges. (a) Average O-D-based travel times within 0–5 mins. (b) Average O-D-based travel times within 5–10 mins. (c) Average O-D-based travel times within 10–15 mins. (d) Average O-D-based travel times within 15–20 mins. (e) Average O-D-based travel times within 20–25 mins.

O-D pairs with similar ATT range tends to improve as ATT gets longer.

Authors speculate that this may be attributable in large part to travel time fluctuations which have a decreasing effect on the longer ATT. For instance, the one-minute fluctuation exerts a greater influence on the O-D pair with a five-minute ATT than that with a ten-minute ATT. In practice, knowing this trend may significantly improve the accuracy of TTR prediction.

4.4. A Counterintuitive Phenomenon. Based on the statistics in Table 5, a so-called counterintuitive phenomenon (e.g., [10, 25]) was also found in this study, which is the O-D-based TTR which was improved under the rainy weather (the global probability is 16%). Although the negative effect of the rain on TTR is still the dominating conclusion considering the low probability of the positive effect, this phenomenon remains an issue to be investigated. According to [20], rain likely affects the effects of

Step 1:
Initialize K -value and LMM.

Step 2 (E step):
Calculate the probability γ_{ik} of each travel time t_i (sample size N) belonging to each lognormal distribution according to (i).
(i) $\gamma_{ik} = (\omega_k L(t_i/\mu_k, \sigma_k) / \sum_{k=1}^K \omega_k L(t_i/\mu_k, \sigma_k)), \quad i = 1, 2, \dots, N, k = 1, 2, \dots, K.$

Step 3 (M step):
Update all parameters in LMM according to (ii)–(iv).
(ii) $\hat{\omega}_k = (\sum_{i=1}^N \gamma_{ik} / N), \quad k = 1, 2, \dots, K.$
(iii) $\hat{\mu}_k = (\sum_{i=1}^N \gamma_{ik} \ln(t_i) / \sum_{i=1}^N \gamma_{ik}), \quad k = 1, 2, \dots, K.$
(iv) $\hat{\sigma}_k^2 = (\sum_{i=1}^N \gamma_{ik} (\ln(t_i) - \hat{\mu}_k) (\ln(t_i) - \hat{\mu}_k)^T / \sum_{i=1}^N \gamma_{ik}), \quad k = 1, 2, \dots, K.$

Step 4:
Repeat until convergence.

ALGORITHM 1: EM algorithm for estimating LMM.

TABLE 4: Summary of median and trimmed mean of BI-values and BI variation ratio under six weather conditions.

Location measure	ATT range (min)	Normal	Rainfall intensity (BI variation ratio)				Rain
			Light rain ((L-N)/ N * 100%)	Moderate rain ((M-N)/ N * 100%)	Heavy rain ((H-N)/ N * 100%)	Extreme rain ((E-N)/ N * 100%)	
Median of BI-values	0–5	0.4219	0.4342 (2.9%)	0.4985 (18.2%)	0.5157 (22.2%)	0.5274 (25.0%)	0.4997
	5–10	0.3586	0.3845 (7.2%)	0.4544 (26.7%)	0.4715 (31.5%)	0.4933 (37.6%)	0.4268
	10–15	0.3568	0.3586 (5.0%)	0.4573 (28.2%)	0.4598 (28.9%)	0.4823 (35.2%)	0.4080
	15–20	0.3399	0.3265 (–3.9%)	0.3956 (16.4%)	0.3962 (16.6%)	0.4113 (21.0%)	0.3604
	20–25	0.3287	0.3513 (6.9%)	0.3588 (9.2%)	0.3489 (6.1%)	0.3750 (14.1%)	0.3515
	Mean	0.3612	0.3710 (2.7%)	0.4329 (19.9%)	0.4384 (21.4%)	0.4579 (26.8%)	0.4092
Trimmed mean of BI-values	0–5	0.4413	0.4555 (3.2%)	0.5078 (15.1%)	0.5205 (17.9%)	0.5476 (24.1%)	0.4990
	5–10	0.4098	0.4351 (6.2%)	0.5047 (23.2%)	0.5399 (31.7%)	0.5376 (31.2%)	0.4809
	10–15	0.3754	0.3793 (1.0%)	0.4644 (23.7%)	0.4866 (29.6%)	0.5030 (34.0%)	0.4429
	15–20	0.3580	0.3675 (2.7%)	0.4151 (15.9%)	0.4121 (15.1%)	0.4304 (20.2%)	0.4007
	20–25	0.3190	0.3291 (3.2%)	0.3519 (10.3%)	0.3532 (10.7%)	0.3697 (15.9%)	0.3448
	Mean	0.3807	0.3933 (3.3%)	0.4488 (17.9%)	0.4625 (21.5%)	0.4777 (25.5%)	0.4337

N, L, M, H, and E indicate normal, light rain, moderate rain, heavy rain, and extreme rain, respectively.

TABLE 5: Summary of O-D number of higher reliability than normal and increasing reduction impact with rainfall increase.

ATT range (min)	O-D number of higher reliability than normal					Rain	O-D number of increasing negative impact with rainfall increase
	Light rain	Moderate rain	Heavy rain	Extreme rain	Sum/sample size		
0–5	8	1	4	2	15/80	3	7
5–10	3	1	1	1	6/80	2	8
10–15	10	1	1	2	14/80	1	6
15–20	7	0	3	2	12/80	1	7
20–25	8	2	5	2	17/80	2	7
Sum/sample size	36/100	5/100	14/100	9/100	64/400	9/100	35/100
Rate	36%	5%	14%	9%	16%	9%	35%

TABLE 6: One-way ANOVA test for BI variation ratio between varying rainfall intensities.

Source	SS	df	MS	F	P value
2 columns (light and moderate)	1.2280	1	1.2280	80.82	3.5874e-16
Error	2.7047	178			
Total	3.93263	179			
3 columns (from moderate to extreme)	0.2079	2	0.1040	2.9	0.0568
Error	9.5750	267	0.0359		
Total	9.7829	269			
4 columns (from light to extreme)	2.9091	3	0.9697	33.37	4.9919e-19
Error	10.3457	356	0.0291		
Total	13.2548	359			

TABLE 7: Summary of interquartile range of BI-values under six weather conditions.

Location measure	ATT range (min)	Normal (75th–25th)	Rainfall intensity (75th–25th)				Rain (75th–25th)
			Light rain	Moderate rain	Heavy rain	Extreme rain	
Interquartile range (IQR) of BI-values	0–5	0.1771 (0.5206–0.3635)	0.2043 (0.5543–0.3500)	0.2687 (0.6330–0.3643)	0.3138 (0.6829–0.3691)	0.2188 (0.6511–0.4323)	0.2450 (0.6119–0.3669)
	5–10	0.1645 (0.4929–0.3284)	0.1975 (0.5253–0.3278)	0.2278 (0.5847–0.3569)	0.3134 (0.7118–0.3984)	0.2881 (0.6939–0.4058)	0.2987 (0.6441–0.3454)
	10–15	0.1606 (0.4600–0.2994)	0.1612 (0.4586–0.2974)	0.2454 (0.5907–0.3453)	0.2342 (0.5910–0.3568)	0.2112 (0.6212–0.4100)	0.2292 (0.5487–0.3195)
	15–20	0.0944 (0.4065–0.3121)	0.1222 (0.4354–0.3132)	0.1432 (0.4952–0.3520)	0.1509 (0.4865–0.3356)	0.1698 (0.5195–0.3497)	0.1407 (0.4842–0.3435)
	20–25	0.1006 (0.3593–0.2587)	0.1083 (0.3786–0.2703)	0.1157 (0.4168–0.3011)	0.1173 (0.4202–0.3029)	0.1201 (0.4248–0.3047)	0.1056 (0.3941–0.2885)

some other substantial contributing factors for some specific O-D pairs (e.g., O-D with bottlenecks). Therefore, future studies should be focused on data with multiple contributing factors such as weather and bottlenecks.

5. Conclusions

This research uses open-source data to study the effects of varying rainfall intensity on O-D-based travel time reliability. The intensity covers light rain (0 and 10.0 mm/24 h), moderate rain (10.0 and 24.9 mm/24 h), heavy rain (25 and 50.0 mm/24 h), and extreme rain (>50.0 mm/24 h). An orithm based on the lognormal mixture model was adopted for analyzing the probability distribution functions of the O-D-based travel time data. Then the buffer index, the three location measures, and the one-way analysis of variance were used for detailed analysis.

In general, rain lowers O-D-based travel time reliability, and the negative impact grows with the increase in rainfall intensity. With respect to the abnormal phenomenon mentioned in [25], it is restricted in low probability from massive results in the study, which cannot be a general conclusion but deserve investigation profoundly. The study also confirmed the existence of the so-called counterintuitive phenomenon mentioned in previous work [25] that, in some cases, TTR may be enhanced in rainy weather. Particularly, we discovered that O-D-based TTR was more sensitive when rainfall intensity changes from light to moderate but less notable when changes are among other categories such as no rain to light rain and moderate rain to heavy rain. The study also demonstrates that the O-D-based TTR in different O-D pairs with a similar ATT range tends to improve as ATT gets longer.

This study contributed to disclose the characteristics of the O-D-based TTR under the varying rainfall with the open-access data. This study is helpful in enhancing current applications by providing more accurate O-D-based travel time information under rain conditions; for example, the trend that consistency of the O-D-based TTR tends to improve with ATT increase can help to improve the accuracy of TTR prediction, when missing enough travel time information. Meanwhile, the research is conducive to promote the use of publicly available data in such investigations so that the results are verifiable, and the studies are sustainable.

This paper only focuses on the impact of rain, but there are a lot of more deserving further investigations along the line of O-D-based TTR analysis. For future work, the impact of other weather events and the combined effects of weather and other factors such as work zones will be the focus.

Data Availability

The data used to support the findings of this study are available from the Uber Movement website (<https://movement.uber.com>) and Weather Underground website (<https://www.wunderground.com/history>).

Conflicts of Interest

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

Acknowledgments

The study was supported by the Technology Project of Shaanxi Transportation Department (Grant no. 15-39R), Special Fund for Basic Scientific Research of Central Colleges of Chang'an University (Grant no. 300102218409), and China Scholarship Council. The authors are also grateful for data retrieved from Uber Movement, (c) 2019 Uber Technologies, Inc., <https://movement.uber.com>, and Weather Underground website, <https://www.wunderground.com/history>.

References

- [1] Y. Lida, "Basic concepts and future directions of road network reliability analysis," *Journal of Advanced Transportation*, vol. 33, no. 2, pp. 125–134, 1999.
- [2] K. Lyman and R. L. Bertini, "Using travel time reliability measures to improve regional transportation planning and operations," *Transportation Research Record*, vol. 2046, pp. 1–10, 2008.
- [3] C. Chen, E. W. van Zwet, and P. Varaiya, "Travel-time reliability as a measure of service," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, no. 1, pp. 74–79, 2007.
- [4] N. Y. Tilahun and D. M. Levinson, "A moment of time: reliability in route choice using stated preference," *Journal of Intelligent Transportation Systems*, vol. 14, no. 3, pp. 179–187, 2010.
- [5] L. Gong and W. Fan, "Applying travel-time reliability measures in identifying and ranking recurrent freeway bottlenecks at the network level," *Journal of Transportation Engineering Part A: Systems*, vol. 143, no. 8, 2017.
- [6] H. X. Liu, W. Recker, and A. Chen, "Uncovering the contribution of travel time reliability to dynamic route choice using real-time loop data," *Transportation Research Part A: Policy and Practice*, vol. 38, no. 6, pp. 435–453, 2004.
- [7] A. Chepuri, J. Ramakrishnan, S. Arkatkar, G. Joshi, and S. S. Pulugurtha, "Examining travel time reliability-based performance indicators for bus routes using GPS-based bus trajectory data in India," *Journal of Transportation Engineering Part A: Systems*, vol. 144, no. 5, 2018.
- [8] C. Lu and J. Dong, "Estimating freeway travel time and its reliability using radar sensor data," *Transportmetrica B: Transport Dynamics*, vol. 6, no. 2, pp. 97–114, 2018.
- [9] F. Zheng, J. Li, H. van Zuylen, X. Liu, and H. Yang, "Urban travel time reliability at different traffic conditions," *Journal of Intelligent Transportation Systems*, vol. 22, no. 2, pp. 106–120, 2018.
- [10] S. I. Chien and K. K. Kolluri, "Evaluation of freeway travel time variability and reliability under adverse weather with TRANSMIT data," *Journal of Civil Engineering and Architecture*, vol. 6, no. 1, pp. 1–11, 2012.
- [11] Z. Chen and W. Fan, "Data analytics approach for travel time reliability pattern analysis and prediction," *Journal of Modern Transportation*, vol. 27, no. 4, pp. 250–265, 2019.
- [12] X. Zhang and M. Chen, "Quantifying the impact of weather events on travel time and reliability," *Journal of Advanced Transportation*, vol. 2019, Article ID 8203081, 9 pages, 2019.
- [13] S. Yang, C. An, Y.-J. Wu, and J. Xia, "Origin-destination-based travel time reliability," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2643, no. 1, pp. 139–159, 2017.
- [14] S. H. Kim and J.-H. Chung, "Exploration on origin-destination-based travel time variability: insights from Seoul metropolitan area," *Journal of Transport Geography*, vol. 70, pp. 104–113, 2018.
- [15] J. Dong, H. S. Mahmassani, and C. C. Lu, "How reliable is this route? predictive travel time and reliability for anticipatory traveler information systems," *Transportation Research Record*, vol. 1980, pp. 117–125, 2006.
- [16] J. Kim and H. S. Mahmassani, "Compound gamma representation for modeling travel time variability in a traffic network," *Transportation Research Part B: Methodological*, vol. 80, pp. 40–63, 2015.
- [17] J. Kim and H. S. Mahmassani, "A finite mixture model of vehicle-to-vehicle and day-to-day variability of traffic network travel times," *Transportation Research Part C: Emerging Technologies*, vol. 46, pp. 83–97, 2014.
- [18] H. S. Mahmassani, T. Hou, and J. Dong, "Characterizing travel time variability in vehicular traffic networks," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2315, no. 1, pp. 141–152, 2012.
- [19] M. A. P. Taylor, "Travel through time: the story of research on travel time reliability," *Transportmetrica B: Transport Dynamics*, vol. 1, no. 3, pp. 174–194, 2013.
- [20] J. Kwon, T. Barkley, R. Hranac, K. Petty, and N. Compin, "Decomposition of travel time reliability into various sources," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2229, no. 1, pp. 28–33, 2011.
- [21] H. Abdulsattar, A. Mostafizi, M. R. K. Siam, and H. Wang, "Measuring the impacts of connected vehicles on travel time reliability in a work zone environment: an agent-based approach," *Journal of Intelligent Transportation Systems*, vol. 24, no. 5, pp. 421–436, 2020.
- [22] L. C. Goodwin, "Weather impacts on arterial traffic flow," Report Prepared for the Road Weather Management Program, Mitretek Systems, Inc., Virginia, USA, 2002.
- [23] A. D. Stern, V. Shah, and L. Goodwin, *Analysis of Weather Impacts on Traffic Flow in Metropolitan Washington DC*, Institute of Transportation Engineers Meeting & Exhibit, Washington, DC, USA, 2003.
- [24] H. Caceres, H. Hwang, and Q. He, "Estimating freeway route travel time distributions with consideration to time-of-day, inclement weather, and traffic incidents," *Journal of Advanced Transportation*, vol. 50, no. 6, pp. 967–987, 2016.
- [25] C. Kanga and M. A. Yazıcı, "Temporal and weather related variation patterns of urban travel time: considerations and caveats for value of travel time, value of variability, and mode choice studies," *Transportation Research Part C: Emerging Technologies*, vol. 45, pp. 4–16, 2014.
- [26] M. Martchouk, F. Mannering, and D. Bullock, "Analysis of freeway travel time variability using bluetooth detection," *Journal of Transportation Engineering*, vol. 137, no. 10, pp. 697–704, 2011.
- [27] A. Higatani, T. Kitazawa, J. Tanabe, Y. Suga, R. Sekhar, and Y. Asakura, "Empirical analysis of travel time reliability measures in hanshin expressway network," *Journal of Intelligent Transportation Systems*, vol. 13, no. 1, pp. 28–38, 2009.
- [28] C. Lu, J. Dong, A. Houchin, and C. Liu, "Incorporating the standstill distance and time headway distributions into freeway car-following models and an application to estimating freeway travel time reliability," *Journal of Intelligent Transportation Systems*, pp. 1–20, 2019.
- [29] M. Saberi and R. L. Bertini, "Beyond corridor reliability measures: analysis of freeway travel time reliability at the segment level for hot spot identification," in *Proceedings of*

- 89th Annual Meeting of the Trans-Portation Research Board*, Washington, DC, USA, January 2010.
- [30] H. Tu, J. W. C. Van Lint, and H. J. Van Zuylen, "Impact of traffic flow on travel time variability of freeway corridors," *Transportation Research Record*, vol. 1993, no. 1, pp. 59–66, 2007.
 - [31] E. B. Emam and H. Al-Deek, "Evaluation method and influence factors of network travel time reliability," *Transportation Research Record: Journal of the Transportation Research Board of the National Academies*, vol. 1959, pp. 140–150, 2006.
 - [32] W. Pu, "Analytic relationships between travel time reliability measures," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2254, no. 1, pp. 122–130, 2012.
 - [33] K. Sohn and D. Kim, "Statistical model for forecasting link travel time variability," *Journal of Transportation Engineering*, vol. 135, no. 7, pp. 440–453, 2009.
 - [34] S. Yang and Y.-J. Wu, "Mixture models for fitting freeway travel time distributions and measuring travel time reliability," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2594, no. 1, pp. 95–106, 2016.
 - [35] F. Guo, H. Rakha, and S. Park, "Multistate model for travel time reliability," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2188, no. 1, pp. 46–54, 2011.
 - [36] F. Guo, Q. Li, and H. Rakha, "Multistate travel time reliability models with skewed component distributions," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2315, no. 1, pp. 47–53, 2013.
 - [37] M. Agarwal, T. H. Maze, and R. R. Souleyrette, "Impacts of weather on urban freeway traffic flow characteristics and facility capacity," in *Proceedings of the 2005 Mid-Continent Transportation Research Symposium*, Ames, Iowa, August 2005.
 - [38] N. E. Lownes and R. B. Machemehl, "Vissim: a multi-parameter sensitivity analysis," in *Proceedings of the 2006 Winter Simulation Conference*, pp. 1406–1413, California, December 2006.

Research Article

Predicting Wet-Road Crashes Using the Finite-Mixture Zero-Truncated Negative Binomial Model

Ying Chen¹ and Zhongxiang Huang²

¹*School of Traffic and Transportation Engineering, School of Architecture, Changsha University of Science & Technology, Changsha 410114, China*

²*School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China*

Correspondence should be addressed to Ying Chen; cycsust@163.com

Received 28 July 2020; Revised 23 September 2020; Accepted 28 September 2020; Published 21 October 2020

Academic Editor: Yajie Zou

Copyright © 2020 Ying Chen and Zhongxiang Huang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Inclement weather affects traffic safety in various ways. Crashes on rainy days not only cause fatalities and injuries but also significantly increase travel time. Accurately predicting crash risk under inclement weather conditions is helpful and informative to both roadway agencies and roadway users. Safety researchers have proposed various analytic methods to predict crashes. However, most of them require complete roadway inventory, traffic, and crash data. Data incompleteness is a challenge in many developing countries. It is common that safety researchers only have access to data on sites where a crash has occurred (i.e., zero-truncated data). The conventional crash models are not applicable to zero-truncated safety data. This paper proposes a finite-mixture zero-truncated negative binomial (FMZTNB) model structure. The model is applied to three-year wet-road crash data on 395 divided roadway segments (total 586 km), and the parameters are estimated using the Markov chain Monte Carlo (MCMC) method. Comparison indicates that the proposed FMZTNB model has better fitting performance and is more accurate in predicting the number of wet-road crashes. The model is capable of capturing the heterogeneity within the sample crash data. In addition, lane width showed mixed effects in different components on wet-road crashes, which are not observed in conventional modeling approaches. Practitioners are encouraged to consider the finite-mixture zero-truncated modeling approach when complete safety dataset is not available.

1. Introduction

According to the World Health Organization, more than 1.3 million roadway users died each year as a result of traffic crashes, and the cost of traffic crashes accounted for about 3% of the gross domestic product in most countries [1]. Road traffic injuries and deaths are a global problem, and traffic crashes are a leading cause for nonnatural death. Safety researchers and practitioners have made continuous efforts to reduce the number and severity of crashes.

A traffic crash is usually caused by one or several factors, including humans (i.e., vehicle driver, motorist, bicyclist, and pedestrians), vehicles, roadway facilities, and environment. Weather affects traffic safety, demand, selection of transportation mode, driving capabilities, vehicle

performance (i.e., stability, maneuverability, and traction), and roadway infrastructure (i.e., pavement friction) through visibility impairments, precipitation, and temperature. Inclement weather not only increases crash risk but also significantly affects users' travel time. According to crash statistics, more than 20 percent of crashes and more than 15 percent of traffic fatalities are weather-related [2]. It is necessary to accurately predict the occurrence of crashes under inclement weather conditions.

Statistical modeling approach has been extensively used in recent two to three decades to quantitatively predict number of crashes. Specifically, safety researchers have proposed various models for developing crash counts, e.g., Poisson, negative binomial (NB), Sichel [3, 4], Conway–Maxwell–Poisson [5, 6], zero-inflated Poisson [7],

Poisson–Tweedie [8], Tobit [9–12], machine learning techniques [13, 14], etc. For a detailed review of the crash regression techniques, readers can refer to the article by Lord and Mannering [15]. In all these models, crash counts are treated as response variables, and more importantly, all of these models require complete roadway inventory, traffic, and crash data. Safety data (e.g., roadway inventory, traffic, operation, and crash information) play a critical role in crash prediction model development, hotspot identification, and safety effectiveness evaluation. Inaccurate or incomplete crash records with the conventional crash prediction models may lead to various misleading results. These errors not only result in inefficient use of limited resources for safety improvements but also cause additional loss of lives. However, data incompleteness is a challenge in many developing countries. For example, only information on segments or intersections where a crash has occurred is collected. This type of data is known as zero-truncated data. Previous studies have shown that the conventional count models are not adequate to model zero-truncated crash data [16, 17]. How to develop reliable crash prediction models using zero-truncated data is an important topic for safety analysts.

This study is an extension of a recent study on modeling zero-truncated crash data [16]. The primary objective is to develop safety performance functions for wet-road crashes when zero's are truncated in the safety data considering the heterogeneity. Particularly, this paper proposes a finite-mixture zero-truncated negative binomial (FMZTNB) model structure and examines if the FMZTNB model provides better modeling results than the commonly used models.

The rest of this paper is organized as follows: Section 2 reviews the literature pertaining to the influence of weather on safety. Section 3 describes the details of the zero-truncated models. Section 4 briefly documents the zero-truncated data. Section 5 presents the modeling results, and Section 6 summarizes the study.

2. Literature Review

Because of the great influence of weather on roadway safety, transportation researchers have made continuous efforts in understanding the relationship between different weather conditions and traffic crashes.

Shankar et al. [18] conducted one of the earliest studies on the effect of weather conditions on roadway crashes. The researchers developed a negative binomial crash model with roadway geometrical and environmental factors. The modeling results suggest that both maximum rainfall and number of rainy days play significant and positive role in number of total crashes.

Maze et al. [19] studied how inclement weather affects traffic demand, traffic safety, and traffic flow relationships. The researchers pointed out that certain types of severe weather conditions (e.g., winter storms) bring a higher risk of being involved in a crash by 13 to 25 times. Weather conditions also impact the crash severity, but it varies depending on specific weather condition and crash location.

Qiu and Nixon [20] conducted a systemic review on the effect of adverse weather on the occurrence of roadway crashes. The researchers reviewed 112 studies conducted between 1967 and 2005 that had examined the association between weather and traffic crashes. Crash rates from each study were combined through a meta-analysis method. The researchers conclude that the crash rate usually increases during precipitation. Snow has a greater effect than rain does on crash occurrence. Specifically, snow can increase the crash rate by 84% and the injury rate by 75%.

Jung et al. [21] analyzed the influence of four weather factors (i.e., rainfall intensity, water film depth, temperature, and wind speed and direction) on the injury severity of rainy day multivehicle crashes. The study found that wind speed is associated with the outcome of crashes.

Recently, Das et al. [22] developed safety performance functions for two types of roadways (i.e., rural two-lane highway and rural multilane highway) in two states (i.e., Ohio and Washington). The researchers included speed measures and weather conditions in the models. Modeling results revealed that precipitation is negatively associated with number of crashes. This result is inconsistent with most previous studies, and the researchers noted that the vehicle speeds might reduce during the wet-weather conditions, hence resulting in fewer crashes.

To summarize, extensive studies have been conducted to analyze the relationship between weather and safety. Overall, crash rates increase significantly during inclement weather conditions. In the previous studies, almost all of them include weather data as factors in the regression models, and none of them have focused on developing a safety performance function for wet-weather crashes specifically. In addition, previous studies have used the common count models (e.g., negative binomial), which require complete safety data. Zero-truncated data are common in developing countries, and zero-truncated models have been proposed by researchers to analyze crash data in recent years [16, 17]. To the best of the authors' knowledge, no efforts have been made to analyze zero-truncated wet-road crashes. This study aims to fill this gap.

3. Methodology

This section discusses three crash modeling approaches: (1) the commonly used negative binomial model; (2) zero-truncated NB model; and (3) finite-mixture zero-truncated NB model.

3.1. Conventional NB Model. As has been mentioned in Section 2, various statistical methods have been developed by safety researchers to predict number of crashes. The NB model is still the most commonly used approach and is recommended by the first edition of *Highway Safety Manual (HSM)* [23]. This section briefly introduces the structure of the NB model.

The commonly used NB model assumes that the number of crashes occurred at a given site (a segment or an

intersection) during a certain period follows Poisson distribution as follows:

$$y_{i,t} \mid \lambda_{i,t} \sim \text{Poisson}(\lambda_{i,t}). \quad (1)$$

The probability mass function (PMF) of crash count is shown as follows:

$$P(y_{i,t} \mid \lambda_{i,t}) = \frac{\lambda_{i,t}^{y_{i,t}} \times e^{-\lambda_{i,t}}}{y_{i,t}!}, \quad y_{i,t} = 0, 1, 2, \dots, \quad (2)$$

where y denotes the crash count. The subscripts i and t represent site index and study period, respectively. $\lambda_{i,t}$ is the Poisson rate for the site during the period. For the ease of readers, the subscripts i and t are omitted in the rest of this paper.

Furthermore, assume that the Poisson rate λ follows gamma distribution:

$$\lambda \mid \mu, \alpha \sim \text{Gamma}(\mu, \alpha), \quad (3)$$

where μ is the mean for λ and α is the shape parameter (positive).

Assuming that the mean μ is associated with roadway features (e.g., traffic volume, segment length, and geometric characteristics),

$$\mu = f(\text{traffic volume, roadway features}). \quad (4)$$

Interpreting λ from equation (2), the PMF of the NB distribution can be obtained as

$$p(y \mid \mu, \alpha) = \int_0^\infty \frac{\lambda^y \times \exp(-\lambda)}{y!} \frac{1}{\mu^\alpha} \times \frac{\lambda^{\alpha-1} \times \exp(-\lambda/\mu)}{\Gamma(\alpha)} d\lambda. \quad (5)$$

The PMF of y is shown as follows:

$$p(y \mid \mu, \alpha) = \frac{\Gamma(y + \alpha)}{\Gamma(y + 1) \times \Gamma(\alpha)} \times \left(\frac{\mu/\alpha}{1 + (\mu/\alpha)} \right)^y \times \left(\frac{1}{1 + (\mu/\alpha)} \right)^\alpha, \quad (6)$$

where y is the response variable (i.e., crash count), μ indicates the mean response of the observation, and α is the

dispersion parameter (i.e., shape parameter in the Gamma distribution). For the detailed derivative of the NB model, readers can refer to [24]. It is important to note that, in the conventional NB model structure, the response variable y takes the values of all nonnegative counts (i.e., 0, 1, 2, 3, ...). In other words, all the observed crash counts should be included in the model development. Since the NB model has closed-form, the parameters can be easily estimated. Many software packages have been developed to estimate the unknown parameters, for example, the MASS package of R [25, 26].

3.2. Zero-Truncated NB Model. The NB model has been widely used in analyzing overdispersed count data; however, it requires completed observed data. When the zero's are truncated, the assumption of the NB model cannot be satisfied, and the estimated parameters are biased. Statisticians proposed truncated models [27]. In the truncated count model, the response variable, y , is also considered to follow Poisson distribution. But, it only takes positive numbers (i.e., conditional on that $y > 0$) as follows:

$$P(y \mid \lambda, y > 0) = \frac{P(y)}{P(y > 0)} = \frac{P(y)}{1 - P(y = 0)}, \quad y = 1, 2, 3, \dots \quad (7)$$

From equation (2), it can be derived that

$$P(y = 0 \mid \lambda) = \frac{\lambda^0 \times e^{-\lambda}}{0!} = e^{-\lambda}. \quad (8)$$

Substituting equation (8) into (7), the zero-truncated Poisson distribution can be obtained as follows:

$$P(y \mid \lambda, y > 0) = \frac{P(y)}{1 - e^{-\lambda}}, \quad y = 1, 2, 3, \dots, \quad (9)$$

where y is the response variable (truncated) and λ is the Poisson rate. Similarly, assuming that the Poisson rate λ follows Gamma distribution, the zero-truncated NB model can be obtained as follows:

$$P(y = 0 \mid \mu, \alpha) = \frac{\Gamma(y + \alpha)}{\Gamma(\alpha) \times \Gamma(y + 1)} \times \left(\frac{\mu/\alpha}{1 + (\mu/\alpha)} \right)^y \times \left(\frac{1}{1 + (\mu/\alpha)} \right)^\alpha = \left(\frac{1}{1 + (\mu/\alpha)} \right)^\alpha, \quad (10)$$

$$P(y \mid \mu, \alpha, y > 0) = \frac{\Gamma(y + \alpha)}{\Gamma(\alpha) \times \Gamma(y + 1)} \times \left(\frac{\mu/\alpha}{1 + (\mu/\alpha)} \right)^y \times \frac{(1/(1 + (\mu/\alpha)))^\alpha}{(1 - (1/(1 + (\mu/\alpha)))^\alpha)}, \quad (11)$$

where μ is the mean response of the observation and α is the dispersion parameter.

Compared to the conventional NB model, the zero-truncated NB model can be viewed as a conditional NB distribution that the response variable takes nonzero values. The conditional distribution (i.e., positive NB) brings complexity in estimating parameters. A few

software packages are available for estimating the ZTNB model, for example, VGAM with R [28].

3.3. Finite-Mixture Zero-Truncated NB Model. In both the conventional NB and zero-truncated NB models, the distribution of the response variable has only one component,

i.e., there is only one Poisson mean. The finite-mixture models assume that the response variables arise from two or more unobserved components with unknown proportions. This provides significant modeling flexibility than the conventional single component models [29]. As has been mentioned, statisticians have proposed the K -component finite mixture of negative binomial regression models (i.e., FMNB- K) as follows [29, 30]:

$$p(y) = \sum_{k=1}^K w_k \left[\frac{\Gamma(y + \alpha_k)}{\Gamma(y + 1)\Gamma(\alpha_k)} \left(\frac{\mu_k}{\mu_k + \alpha_k} \right)^y \left(\frac{\alpha_k}{\mu_k + \alpha_k} \right)^{\alpha_k} \right], \quad (12)$$

$$p(y) = \sum_{k=1}^K w_k \left[\frac{\Gamma(y + \alpha_k)}{\Gamma(\alpha_k) \times \Gamma(y + 1)} \times \left(\frac{\mu_k/\alpha_k}{1 + \mu_k/\alpha_k} \right)^y \times \frac{(1/(1 + (\mu_k/\alpha_k)))^{\alpha_k}}{(1 - (1/(1 + \mu/\alpha)))^\alpha} \right], \quad (13)$$

where y is the zero-truncated response variable (i.e., crash counts; $y=1, 2, 3, 4, \dots$); w_k is the weight factor of component k which sum to 1 ($\sum_{k=1}^K w_k = 1$); μ_k is the Poisson mean of component k ; and α_k is the dispersion parameter of component k .

In both the FMNB- K and FMZTNB- K models, a function is used to link the Poisson mean and roadway features; therefore,

$$\mu = \sum_{k=1}^K f_k(\text{traffic volume, roadway features}). \quad (14)$$

It can be seen that when $K=1$, the FMNB- K and FMZTNB- K models reduce to NB and ZTNB models, respectively. The FMNB models allow for additional heterogeneity within components not captured by the independent variables.

It is important to note that, as the number of components K increases, the FMNB model becomes more flexible. However, it also brings complexity in the parameter estimation. Previous studies have indicated that a two-component finite mixture of NB regression models (FMNB-2) was quite enough to characterize crash data [31–33]. Thus, this study considers the two-component finite mixture of zero-truncated NB model (FMZTNB-2) in the analyses.

In terms of parameter estimation, the commonly used maximum likelihood estimation (MLE) algorithm will not generate reliable results due to the complicated likelihood function in the FMZTNB-2 model. An alternative is the Gibbs sampling technique, also known as the Markov chain Monte Carlo (MCMC) method, which has been frequently used in estimating parameters of finite-mixture models [29, 34]. Package “rjags” is used to draw the samplings [35], and the FMNZTB-2 MCMC model is developed using JAGS (Just Another Gibbs Sampler) [36]. The truncation is represented using function T(.) in the JAGS.

where y is the response variable ($y=0, 1, 2, 3, 4, \dots$); w_k is the weight factor of component k which sum to 1 ($\sum_{k=1}^K w_k = 1$); μ_k is the Poisson mean of component k ; and α_k is the dispersion parameter of component k .

Analogous to equation (12), the K -component finite mixture of zero-truncated NB model (FMZTNB- K) can be constructed as

4. Data

This study collected data on 395 rural multilane-divided roadway segments, including traffic volume, lane width, average shoulder width, and median width. Three years of wet-road crash data were collected. A wet-road crash is defined as that the weather condition was rain, snow, or hail, or the surface condition was wet, snowy, ice, or standing water at the time of the crash occurred. In terms of independent variables, this paper mainly considered data availability and potential effects on the occurrence of crashes during rainy weather conditions from published literature [19–21]. Finally, the following six variables were selected from the dataset: segment length, traffic volume, lane width, average outside shoulder width, average inside shoulder width, and median width. Descriptive statistics of the roadway and crash data are illustrated in Table 1.

It is worth mentioning that the minimum crash count of the sample segments is 1 (see the last row in Table 1), rather than 0. This is because when collecting the roadway data, only information on segments where at least one crash had occurred is available to the authors. In other words, the safety data is zero-truncated.

5. Modeling Results

Previous studies have revealed that the commonly used NB model is not applicable for modeling zero-truncated crash data [16, 17]. The parameters can be heavily biased, and the results are not reliable. Thus, the conventional NB model is not used to the data collected in this study. This section presents the results of the ZTNB model and the FMZTNB-2 model, separately.

5.1. Modeling Result of ZTNB. The authors developed the ZTNB model with the data described in Section 4 with the following functional form.

TABLE 1: Descriptive statistics of data (sample size: 395).

Variable	Mean	Minimum	Maximum	Standard deviation
Segment length (km)	1.483	0.172	3.080	0.969
Traffic volume (veh/day)	11569.41	3192	26935	5505.763
Lane width (m)	3.826	2.9	5.0	0.513
Average outside shoulder width (m)	2.515	0.61	3.05	0.758
Average inside shoulder width (m)	1.536	0	3.05	0.705
Median width (m)	6.352	3.05	14.64	3.107
Wet-road crash count	2.42	1	15	2.439

$$\mu = \exp(\beta_0 + \beta_1 \log(\text{ADT}) + \beta_2 \text{LW} + \beta_3 \text{OSH} + \beta_4 \text{ISH} + \beta_5 \times \text{MW}), \quad (15)$$

where μ is the mean of the observed crash data; ADT is traffic volume; LW is lane width (m); OSH is average outside shoulder width; ISH is average inside shoulder width (m); MW is median width (m); $\beta_0, \beta_1, \dots, \beta_5$ are unknown parameters to be estimated. It is important to note that the length of a segment is considered as an offset variable, meaning that the number of crashes is proportional to the segment length. This assumption is consistent with the *HSM*.

Although studies have pointed out that varying dispersion parameter (i.e., α in equations (10) and (11)) benefits the crash prediction models [37–39], this study assumed that it is fixed among all the sites to make the computation easier and consistent with the FMZTNB-2 model in Section 5.2 [37, 40–42].

The modeling results of the ZTNB model is shown in Table 2. As can be seen, the parameters for traffic volume, average outside shoulder width, and average inside shoulder width are all statistically significant at the level of 90 percent or higher. Specifically, as the traffic volume increase, the predicted number of wet-road crashes also increases. The parameters for the other three roadway features are all negative, indicating that, with the increase of shoulder width or median width, the predicted number of wet-road crashes will decrease. For example, with one meter increase in average outside shoulder width, the predicted number of wet-roadway crashes will decrease by 14.6 percent (i.e., $1 - e^{-0.158}$). This is expected, as outside shoulders become wider, it provides additional recovery spaces for vehicles which slide away from the traveling lane due to the reduced skid number during rainy days. The results are in line with several previous studies [21, 43]. On the other hand, the parameter for lane width is -0.1 , and the result is not statistically significant. The dispersion parameter, α , is estimated as 1.615, which is also insignificant.

This study used four types of goodness-of-measure (GOF) to evaluate the model performance: Akaike

TABLE 2: Estimating results of the ZTNB model.

Variable	Estimate	Std. err.	P value	Significant level
Intercept, β_0	-3.638	1.698	—	Not significant
Log (ADT), β_1	0.770	0.169	<0.001	99.9%
Lane width, β_2	-0.100	0.159	0.530	Not significant
Ave. out. SHD, β_3	-0.158	0.082	0.054	90.0%
Ave. in. SHD, β_4	-0.310	0.080	<0.001	99.9%
Median width, β_5	-0.373	0.034	<0.001	99.9%
Disp. par., α	1.615	5.465	0.409	Not significant
AIC	1142.29	—	—	—
BIC	1170.14	—	—	—
MAE	0.64	—	—	—
RMSE	2.52	—	—	—

Note. Std. err. = standard error. Disp. par. = dispersion parameter. Bold and underline values indicate statistically significant at 90.0% or higher. — means not applicable.

information criterion (AIC), Bayesian information criterion (BIC), mean absolute error (MAE), and root mean square error (RMSE). The AIC, BIC, MAE, and RMSE for the ZTNB model are 1142.29, 1170.14, 0.64, and 2.52, respectively (see the last four rows in Table 2).

5.2. Modeling Result of FMZTNB-2. As has been mentioned in Section 3, this study utilized MCMC approach to estimate the parameters of the FMZTNB-2 model. Noninformative priors were used for hyperparameters. This study performed 1,000,000 MCMC iterations with two different chains, and the first 20,000 samples of each chain were discarded as burn-in samples from the MCMC outputs. Gelman–Rubin (G–R) convergence statistics and visual history plots were used to verify the MCMC process [44, 45]. The functional forms linking the Poisson mean and the roadway features are similar to those of the ZTNB model, except that there are two forms in the components, as shown in the following equations.

$$\mu_{c1} = \exp(\beta_{0,c1} + \beta_{1,c1} \log(\text{ADT}) + \beta_{2,c1} \text{LW} + \beta_{3,c1} \text{OSH} + \beta_{4,c1} \text{ISH} + \beta_{5,c1} \text{MW}), \quad (16)$$

$$\mu_{c2} = \exp(\beta_{0,c2} + \beta_{1,c2} \log(\text{ADT}) + \beta_{2,c2} \text{LW} + \beta_{3,c2} \text{OSH} + \beta_{4,c2} \text{ISH} + \beta_{5,c2} \text{MW}), \quad (17)$$

$$\mu = wt \times \mu_{c1} + (1 - wt) \times \mu_{c2}, \quad (18)$$

TABLE 3: Estimating results of the FMZTNB-2 model.

Variable	Estimate	Component 1			Component 2		
		Estimate	Std. err.	P value	Estimate	Std. err.	P value
Intercept	-4.047	2.819	0.151	-4.485	1.190	<0.001	
Log (ADT)	0.718	0.018	<0.001	0.993	0.195	<0.001	
Lane width	-0.080	0.010	<0.001	0.203	0.012	<0.001	
Ave. Out. SHD	-0.094	0.022	<0.001	-0.183	0.014	<0.001	
Ave. In. SHD	-0.241	0.177	0.174	-0.879	0.166	<0.001	
Median width	-0.416	0.010	<0.001	-0.494	0.251	0.049	
Disp. par.	1.499	0.142	<0.001	4.619	1.856	0.013	
Weight	0.712	0.082	<0.001	0.288	-	-	
AIC	1020.54	-	-	-	-	-	
BIC	1088.32	-	-	-	-	-	
MAE	0.22	-	-	-	-	-	
RMSE	2.18	-	-	-	-	-	

Note. Std. err. = standard error. Disp. Par. = dispersion parameter. Bold and italic values indicate statistically significant at level of 90.0 percent or higher. — means not applicable.

TABLE 4: Prediction comparison between ZTNB and FMZTNB models (three example sites).

Site number (level)	ZTNB			FMZTNB-2		
	Prediction	Std. err.	90% PI	Prediction	Std. err.	90% PI
138 (low)	0.0040	0.0636	[0.0036–0.0046]	0.0005	0.0007	[0.0005–0.0005]
90 (moderate)	0.0654	0.2608	[0.0392–0.109]	0.0627	0.0449	[0.0574–0.0684]
65 (high)	0.7350	1.0341	[0.0968–5.5788]	0.7625	0.5903	[0.2398–2.425]

Note. Std. err. = standard error. PI = prediction interval.

where μ is the mean of the observed crash data; μ_{c1} and μ_{c2} are the mean of observations in the two components, respectively; ADT is traffic volume; LW is lane width (m); OSH is average outside shoulder width; ISH is average inside shoulder width (m); MW is median width (m); and β' s are parameters to be estimated.

The modeling results of the FMZTNB-2 model are documented in Table 3. First, the estimated weight factor for component 1 is 0.712, with a standard error of 0.082. This result is statistically significant, indicating that the sample data include two components. Component 1 accounts for about 71.2 percent, and component 2 accounts for about 28.8 percent (i.e., $1 - 0.712$). Second, most of the parameters in both components are statistically significant (except average inside shoulder width in component 1 and median width in component 2). Overall, the signs of the parameters in the FMZTNB-2 model are the same as the corresponding parameters in the ZTNB model. For example, the parameter for average outside shoulder width is -0.158 in the ZTNB model. They are -0.094 and -0.183 in the two components, respectively, of the FMZTNB-2 model. All of them indicate that wider outside shoulders are associated with fewer wet-road crashes. The parameter for average inside shoulder in component 1 and that for median width in component 2 is not significant. Another interesting finding is that the estimated parameters for lane width in the two components have different signs, i.e., it is negative (-0.080) in component 1 and positive (0.203) in component 2. In other words, the lane width is negatively associated with wet-road crashes in the first group of

segments (i.e., component 1); however, it is positively associated with wet-road crashes in the second group of segments (i.e., component 2). This is in line with a few studies which have reported controversy effects of lane width on safety [46–48].

Finally, the AIC, BIC, MAE, and RMSE for the FMZTNB-2 model are 1020.54, 1088.32, 0.22 and 2.14, respectively (see the last four rows in Table 3). In addition to model goodness-of-fit, this paper also analyzed the prediction performance of the two models using three sites. The three sites represent relative low, moderate, and high crash levels, respectively. The crash mean prediction, standard deviation, as well as 90 percentile confidence intervals of the three sites by the two models are tabulated in Table 4. The results indicate that, for the three sites, the predicted crash mean (i.e., number of wet-weather crashes) between the two models are fairly close (except for the first site, which has a very small crash mean). For site 90, the predicted number of crashes of the ZTNB and FMZTNB-2 models are 0.0645 and 0.0627, respectively. Their standard deviation values are 0.2606 and 0.0449, respectively. The crash predictions with FMZTNB-2 model have significantly lower standard deviation values and narrower intervals, indicating that the model has higher prediction accuracy.

The FMZTNB-2 model shows superiority in modeling the wet-weather crash data. First, the FMZTNB-2 model fits the dataset better than the ZTNB model in terms of GOF measures (e.g., AIC, BIC, MAE, and RMSE). Second, the predictions using FMZTNB-2 model have lower standard deviations and narrower prediction intervals,

indicating that the predictions are more accurate. Finally, a few interesting relationships between variables and crashes are observed from the FMZTNB-2 model. For example, the parameters of lane width are opposite in the two components, indicating that this factor have mixed effects at different locations. These results indicate that the FMZTNB-2 model captures the heterogeneity of the crash data better than the ZTNB model.

6. Conclusions

Inclement weather increases both crash risk and travel time. Efforts have been made in the past decades to predict the occurrence of traffic crashes. However, very few of the previous studies have focused on predicting wet-road crashes. Most of the commonly used crash prediction models require complete roadway inventory, traffic, and crash data. Data missing is relative common in developing countries. How to analyze zero-truncated crash data and predict the number of wet-road crashes is the primary objective of this study. To better capture the heterogeneity of wet-road crash data, this study developed the two-component finite-mixture zero-truncated negative binomial model. The model is applied to three-year wet-road crash on 395 rural-divided roadways. The model results are compared with those based on zero-truncated negative binomial model. Comparison indicates that the proposed FMZTNB-2 model fits the wet-road crash data better than the ZTNB model. It is worth mentioning that, the wet-weather crash data were not modeled with the conventional NB model since previous studies have demonstrated that the application of NB model on truncated data is not recommended. There are trade-offs of using ZTNB or FMZTNB models in crash analyses. With zero-truncated data, the sample size is smaller than that of full data. The reduced sample size might increase uncertainty of parameter estimates.

There are some limitations with this study. First, only a number of roadway characteristics (i.e., segment length, lane width, shoulder width, and median width) and traffic data are available to the authors. There are other factors affecting the occurrence of wet-roadway crashes (e.g., precipitation, number of rainy days per year, and surface skid number). Unfortunately, they are not accessible to the authors. Second, previous studies have shown that the varying forms of dispersion parameter and weight factor for the components in the finite-mixture models improve both crash prediction and hotspot identification [33, 37, 49–51]. In this study, fixed dispersion parameter and weight factor were used to simplify the parameter estimation process. In the future, it is necessary to collect more data, especially those closed related to wet-road crashes, and to examine if varying forms of dispersion parameter and weight factor will further improve the model performance. Finally, the finite-mixture model provides better results than the previously proposed zero-truncated model (e.g., goodness-of-fit and prediction). However, parameter estimates with the FMZTNB-2 model require MCMC, and it increases the computational

time, which may be challenging for practitioners. The parameter estimating method in the FMZTNB-2 model needs to be further simplified 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 they have no conflicts of interest.

Acknowledgments

This research was sponsored jointly by the National Natural Science Foundation of China (project no. 51978082); the Outstanding Youth Foundation of Hunan Education Department (project no. 19B022); and the Young Teacher Development Foundation of Changsha University of Science & Technology (project no. 2019QJCZ056).

References

- [1] World Health Organization, *Global Status Report on Road Safety 2018: Summary*, World Health Organization, Geneva, Switzerland, 2018.
- [2] FHWA, *How Do Weather Events Impact Roads?*, FHWA, Washington, DC, USA, 2020, https://ops.fhwa.dot.gov/weather/q1_roadimpact.htm.
- [3] L. Wu, Y. Zou, and D. Lord, "Comparison of sichel and negative binomial models in hot spot identification," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2460, no. 1, pp. 107–116, 2014.
- [4] Y. Zou, D. Lord, Y. Zhang, and Y. Peng, "Comparison of sichel and negative binomial models in estimating empirical bayes estimates," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2392, no. 1, pp. 11–21, 2013.
- [5] D. Lord, S. R. Geedipally, and S. D. Guikema, "Extension of the application of conway-maxwell-Poisson models: analyzing traffic crash data exhibiting underdispersion," *Risk Analysis*, vol. 30, no. 8, pp. 1268–1276, 2010.
- [6] D. Lord, S. D. Guikema, and S. R. Geedipally, "Application of the conway-maxwell-Poisson generalized linear model for analyzing motor vehicle crashes," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1123–1134, 2008.
- [7] D. Lord, S. P. Washington, and J. N. Ivan, "Poisson, poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory," *Accident Analysis and Prevention*, vol. 37, no. 1, pp. 35–46, 2005.
- [8] D. Saha, P. Alluri, E. Dumbaugh, and A. Gan, "Application of the Poisson-tweedie distribution in analyzing crash frequency data," *Accident Analysis & Prevention*, vol. 137, Article ID 105456, 2020.
- [9] Y. Guo, Z. Li, P. Liu, and Y. Wu, "Modeling correlation and heterogeneity in crash rates by collision types using full bayesian random parameters multivariate tobit model," *Accident Analysis & Prevention*, vol. 128, pp. 164–174, 2019.
- [10] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident

- rates using the random-parameters Tobit model,” *Accident Analysis & Prevention*, vol. 45, pp. 628–633, 2012.
- [11] Y. Guo, T. Sayed, and M. Essa, “Real-time conflict-based bayesian Tobit models for safety evaluation of signalized intersections,” *Accident Analysis & Prevention*, vol. 144, Article ID 105660, 2020.
- [12] Y. Guo, Z. Li, and T. Sayed, “Analysis of crash rates at freeway diverge areas using bayesian tobit modeling framework,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 4, pp. 652–662, 2019.
- [13] S. Das, L. Minjares-Kyle, L. Wu, and R. H. Henk, “Understanding crash potential associated with teen driving: survey analysis using multivariate graphical method,” *Journal of Safety Research*, vol. 70, pp. 213–222, 2019.
- [14] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, “Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models,” *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.
- [15] D. Lord and F. Mannering, “The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives,” *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.
- [16] Y. Chen and Z. Huang, “Modeling intersection traffic crashes using a zero-truncated negative binomial model,” *China Journal of Highway and Transport*, vol. 36, no. 15, pp. 138–145, 2020.
- [17] D.-G. Kim and Y. Lee, “Modelling crash frequencies at signalized intersections with a truncated count data model,” *International Journal of Urban Sciences*, vol. 17, no. 1, pp. 85–94, 2013.
- [18] V. Shankar, F. Mannering, and W. Barfield, “Effect of roadway geometrics and environmental factors on rural freeway accident frequencies,” *Accident Analysis & Prevention*, vol. 27, no. 3, pp. 371–389, 1995.
- [19] T. H. Maze, M. Agarwal, and G. Burchett, “Whether weather matters to traffic demand, traffic safety, and traffic operations and flow,” *Transportation Research Record: Journal of Transportation Research Board*, vol. 1948, no. 1, pp. 170–176, 2006.
- [20] L. Qiu and W. A. Nixon, “Effects of adverse weather on traffic crashes: systematic review and meta-analysis,” *Transportation Research Record*, vol. 2055, no. 1, pp. 139–146, 2008.
- [21] S. Jung, X. Qin, and D. A. Noyce, “Injury severity of multi-vehicle crash in rainy weather,” *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 50–59, 2012.
- [22] S. Das, S. R. Geedipally, and K. Fitzpatrick, “Inclusion of speed and weather measures in safety performance functions for rural roadways,” *IATSS Research*, 2020.
- [23] AASHTO, *Highway Safety Manual*, American Association of State Highway and Transportation Officials, Washington, DC, USA, 2010.
- [24] J. M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, Cambridge, UK, 2nd edition, 2011.
- [25] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, and D. Firth, “Package “mass,”” 2014, <http://cran.r-project.org/web/packages/MASS/MASS.pdf>.
- [26] R Core Team, “R: a language and environment for statistical computing,” 2019, <http://www.R-project.org/>.
- [27] A. C. Cohen, “Estimating the parameter in a conditional Poisson distribution,” *Biometrics*, vol. 16, no. 2, pp. 203–211, 1960.
- [28] T. Yee, *Vgam: Vector Generalized Linear and Additive Models*, Springer, Berlin, Germany, 2017.
- [29] B.-J. Park, D. Lord, and L. Wu, “Finite mixture modeling approach for developing crash Modification factors in highway safety analysis,” *Accident Analysis & Prevention*, vol. 97, pp. 274–287, 2016.
- [30] X. Zhong, Y. Zou, S. Yuan, Z. Dong, and M. Ijaz, “Finite mixture survival model for examining the variability of urban arterial travel time for buses, passenger cars and taxis,” *IET Intelligent Transport Systems*, vol. 45, 2020, In press.
- [31] B.-J. Park and D. Lord, “Application of finite mixture models for vehicle crash data analysis,” *Accident Analysis and Prevention*, vol. 41, no. 4, pp. 683–691, 2009.
- [32] B.-J. Park, D. Lord, and J. D. Hart, “Bias properties of bayesian statistics in finite mixture of negative binomial regression models in crash data analysis,” *Accident Analysis and Prevention*, vol. 42, no. 2, pp. 741–749, 2010.
- [33] Y. J. Zou, J. E. Ash, B. J. Park, D. Lord, and L. T. Wu, “Empirical bayes estimates of finite mixture of negative binomial regression models and its application to highway safety,” *Journal of Applied Statistics*, vol. 45, no. 9, pp. 1652–1669, 2018.
- [34] B.-J. Park, D. Lord, and C. Lee, “Finite mixture modeling for vehicle crash data with application to hotspot identification,” *Accident Analysis & Prevention*, vol. 71, pp. 319–326, 2014.
- [35] M. Plummer, A. Stukalov, and M. Denwood, “Bayesian graphical models using MCMC,” *R Package Version*, vol. 4, no. 6, 2016.
- [36] A. Inglis, A. Ahmed, B. Wundervald, and E. Prado, “Jags: just another gibbs sampler,” 2018, <http://mcmc-jags.sourceforge.net/2004>.
- [37] Y. Meng, L. Wu, C. Ma, X. Guo, and X. Wang, “A comparative analysis of intersection hotspot identification: fixed vs. varying dispersion parameters in negative binomial models,” *Journal of Transportation Safety & Security*, pp. 1–18, 2020.
- [38] E. Hauer, “Statistical road safety modeling,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1897, no. 1, pp. 81–87, 2004.
- [39] E. Hauer, “Overdispersion in modelling accidents on road sections and in empirical bayes estimation,” *Accident Analysis and Prevention*, vol. 33, no. 6, pp. 799–808, 2001.
- [40] D. Lord and L. F. Miranda-Moreno, “Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a bayesian perspective,” *Safety Science*, vol. 46, no. 5, pp. 751–770, 2008.
- [41] D. Lord and P. Y.-J. Park, “Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on Empirical Bayes estimates,” *Accident Analysis & Prevention*, vol. 40, no. 4, pp. 1441–1457, 2008.
- [42] S.-P. Miaou and D. Lord, “Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and bayes versus empirical bayes methods,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1840, no. 1, pp. 31–40, 2003.
- [43] K. Wang, S. Zhao, and E. Jackson, “Multivariate Poisson lognormal modeling of weather-related crashes on freeways,” *Transportation Research Record*, vol. 2672, no. 38, pp. 184–198, 2018.
- [44] P. D. Hoff, *A First Course in Bayesian Statistical Methods*, Springer, New York, NY, USA, 2009.
- [45] S. P. Brooks and A. Gelman, “General methods for monitoring convergence of iterative simulations,” *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 434–455, 1998.

- [46] D. M. Belmont, "Effect of shoulder width on accidents on two-lane tangents," 2014, <http://www.ktc.uky.edu/files/2012/09/1980-The-Effect-of-Lane-and-Shoulder-Widths-on-Accident-Reductions-on-Rural-Two-Lane-Roads-Report-No.-561.pdf>.
- [47] E. Hauer, "Lane width and safety," 2014, <http://www.roadsafetyresearch.com/>.
- [48] I. B. Potts, D. W. Harwood, and K. R. Richard, "Relationship of lane width to safety on urban and suburban arterials," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2023, no. 1, pp. 63–82, 2007.
- [49] Y. Zou, K. Henrickson, L. Wu, Y. Wang, and Z. Zhang, "Application of the empirical bayes method with the finite mixture model for identifying accident-prone spots," *Mathematical Problems in Engineering*, vol. 2015, p. 10, Article ID 958206, 2015.
- [50] Y. Zou, Y. Zhang, and D. Lord, "Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models," *Analytic Methods in Accident Research*, vol. 1, pp. 39–52, 2014.
- [51] Y. Zou, Y. Zhang, and D. Lord, "Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis," *Accident Analysis & Prevention*, vol. 50, pp. 1042–1051, 2013.

Research Article

The Effect of Regret-Based Risky Route Choice on the Traffic Equilibrium for Emergency Evacuation

Ze Wang ^{1,2}, Haiqiang Yang,³ and Linglin Ni⁴

¹*Institute of Intelligent Transportation Systems, Zhejiang University, Hangzhou 310058, China*

²*Department of Traffic Management Engineering, Zhejiang Police College, Hangzhou 310053, China*

³*Institute for Future (IFF), Qingdao University, Qingdao 266071, China*

⁴*Department of Business Management, Zhejiang University of Finance and Economics Dongfang College, Hangzhou 310012, China*

Correspondence should be addressed to Ze Wang; wangze2@zjcx.cn

Received 26 May 2020; Revised 13 September 2020; Accepted 27 September 2020; Published 12 October 2020

Academic Editor: Yajie Zou

Copyright © 2020 Ze Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Following the research on human decision-making under risk and uncertainty, the purpose of this paper is to analyze evacuees' risky route decision behavior and its effect on traffic equilibrium. It examines the possibility of applying regret theory to model travellers' regret-taking behavior and network equilibrium in emergency context. By means of modifying the utility function in expected utility theory, a regret-based evacuation traffic equilibrium model is established, accounting for the evacuee's psychological behavior of regret aversion and risk aversion. Facing two parallel evacuation routes choice situation, the effect of evacuees' risk aversion and regret aversion on traffic equilibrium is numerically investigated as well as the road capacity reduction from natural disaster. The findings reveal that evacuees prefer the riskless route with the lower travel time as the increase of the regret aversion degree. The equilibrium tends to be achieved when more evacuees choose the safer route jointly affected by risk aversion and regret aversion. Moreover, an optimization model for disaster occurring possibility is formulated to assess the traffic system performance for evacuation management. These findings are helpful for understanding how the regret aversion and risk aversion influence traffic equilibrium.

1. Introduction

Traffic equilibrium is a key process of transportation demand analysis and planning. The assumption of the presence of traffic equilibrium can help to predict route flow patterns in the network and evaluate the associated measures of system performance. It is critical to efficiently identify the potential periodic patterns from massive time-series data and provide accurate predictions for travel time and freeway traffic speed based on statistical, Markov chains and machine learning prediction models, especially for real evacuation scenario [1–3]. Previous studies have shown expected utility theory's popularity on depicting travel behaviors and solving traffic equilibrium problem [4–8]. It postulates that each individual traveler attempts to maximize the utility of the chosen route as the principle rule.

Travel time is an important indicator for traveller decision-making and traffic assignment. Besides travel time, other route spatiotemporal factors including intersection delay, path distance, and path size are also considered in the route choice process [9]. Current research on route choice turns to model travellers' responses to uncertainty. However, the principle assumption of expected utility maximization lacks behavioral realism in some risky decision-making occasions, especially the emergency evacuation context [10–12].

It is widely acknowledged that the notion of regret is highly relating to individual's decision-making [13]. The regret theory relaxes the traditional behavioral assumption and provides an opportunity to account for the regret aversion psychology, especially when people face risky choice decision-making [14]. A recent effort in modelling

individual's travel choices using the regret theory has been made, such as route choice, mode choice, and departure times [15, 16]. A small but growing body of studies has been performed to analyze traffic equilibrium based on the regret theory. The regret theory has great potential in modelling traffic equilibrium compared with expected utility theory and prospect theory [16–18]. The outstanding advantage lies in its parsimonious form to explain the actual behaviors and being consistent with empirical studies. A regret-based stochastic user equilibrium model has been proposed in more general choice sets that can depict the traveller's route choice behaviors more flexibly [19]. Considering the travellers' way of thinking, a noncompensatory multiobjective framework has been developed, and conflicts among multiple objectives can be solved through the model [20]. Since the monetary cost was also an important factor when making a decision between alternative routes, a biattribute user equilibrium model (i.e., travel time and monetary cost) has been established in which travellers aim to minimise their regret [21].

In recent years, although the prospect theory is increasingly applied to explain traveler's risky route choice behaviors, it is difficult to select a scientific reference point [22–24]. A unique pure nash equilibrium point has been achieved under the principles of prospect theory considering users' diverse behavioral patterns [25]. Cloud computing environments, flexible 5G access technology, and clustering mechanisms may help enhance evacuees' mobility characteristics [26–28]. The regret aversion psychology was conformed to affect the traffic equilibrium in the studies conducted by Chorus [17]. The impacts of regret and risk aversions increase and appear to reinforce one another, both implying equilibrium shifts towards safer routes. Regret is weaker in the environment with weaker risk degree, but it is stronger in the risky environment [29].

The review of the existing literature reveals that few studies focus on the traffic equilibrium with regret aversion under evacuation risky decision-making process. Evacuees' decision behaviors when facing the emergency evacuation are different from that in the regular conditions; thus their regret and risk aversion should be considered [30, 31]. Therefore, the application of the regret-based model in the field of emergency transportation needs to be addressed urgently. One motivation of this study is to analyze the traffic equilibrium under evacuation condition based on the regret theory. Meanwhile, the global optimization model for disaster state is constructed to assess the performance of traffic system in order to make feasible evacuation guidance. The analysis of the traffic equilibrium is the premise of assessing the performance of traffic system. The contribution of this study highlights to capture unobserved risk and regret aversions psychology in the evacuation decision-making process.

The remainder of this paper is organized as follows. Section 2 presents the regret-based traffic equilibrium models for emergency evacuation. Section 3 explores how the risk and regret aversions affect the evacuation traffic equilibrium, and Section 4 formulates the system performance assessment model. The Section 5 concludes the study and discusses future research directions.

2. Regret-Based Traffic Equilibrium Models

Travel behavior has a potential significant impact on the traffic equilibrium. In contrast to previous studies based on utility framework, a regret-based traffic equilibrium model was formulated accounting for risk and regret aversion decisions. The psychological regret/rejoice may occur when the alternative is worse/better than other alternatives. The decision-makers try to avoid the higher regret, specifically regret aversion. Facing a risky decision-making situation, more attentions were paid on regret theory rather than traditional utility theory.

2.1. Regret-Based Utility Functions. Assume a situation that evacuees should leave home for a safe destination to avoid the damage from natural disaster (e.g., flood and storms) through main evacuation routes a or b with risky travel time as shown in Figure 1. The route a is closer to the disaster than route b , but the route distance is shorter. The perceived travel time is uncertain, and its distribution is known to evacuees. More specifically, the evacuee knows the occurrence probability of random travel time p_r for every disaster state. Three different states s of the disaster including "good state," "medium state," and "bad state" may occur with the occurrence probability p_s .

According to the regret theory, the perceived utility function consists of two parts: basic utility function and regret-rejoice function [10, 12]. The regret-based utilities for the two routes are shown as follows:

$$RU(a) = U(t_{sa}) + R[\Delta U(a)] = U(t_{sa}) + R[U(t_{sa}) - U(t_{sb})], \quad (1)$$

$$RU(b) = U(t_{sb}) + R[\Delta U(b)] = U(t_{sb}) + R[U(t_{sb}) - U(t_{sa})], \quad (2)$$

where $U(t_{sa})$ and $U(t_{sb})$ represent the basic utility of route a and b ; t_{sa} and t_{sb} are the travel time of routes a and b in the disaster state s ; and $R[\Delta U(a)]$ and $R[\Delta U(b)]$ represent the regret-rejoice value of route a and b . When $R[\Delta U(a)] > 0$, it becomes a rejoice value, indicating that the evacuee perceives the rejoice psychology from choosing route a . When $R[\Delta U(b)] < 0$, it becomes a regret value, indicating that the evacuee perceives the regret psychology from choosing route b . Notably, the part $R(\Delta U)$ is a function of the utility difference between route a and b .

There are different forms depending on the risk aversion level for the utility function U . For a cost-attribute (e.g., travel time), U is a monotonically decreasing concave function, and it satisfies the conditions $U' < 0$ and $U''(\Delta U) > 0$ [32, 33]. For a benefit-attribute (e.g., service), U is a monotonically increasing concave function, and it satisfies the conditions $U' > 0$ and $U''(\Delta U) < 0$. The exponential expression was applied as the utility function for travel time attribute as follows [10, 12, 17]:

$$U(t_s) = \frac{1 - \exp(\alpha \cdot t_s)}{\alpha}, \quad (3)$$

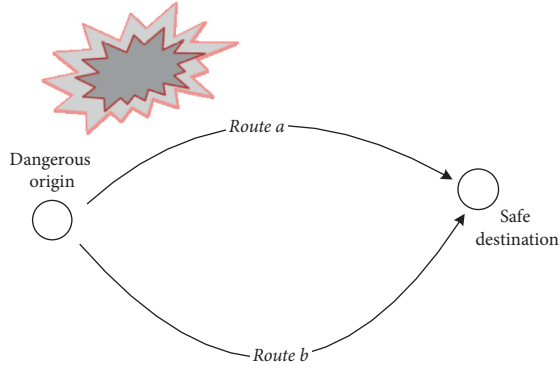


FIGURE 1: A schematic description of the route choice decision.

where α represents a nonnegative risk aversion parameter and it ranges from 0 to 1. Generally, the decision-makers are risk averse when facing risky choices. The influence of risk aversion on the utility with respect to travel times is shown in Figure 2. It is found that the risk aversion level of the evacuee is increasing with risk aversion parameter.

The regret-rejoice function is highly dependent on the regret-aversion level. In most situations, the evacuees are risk averse when facing disaster. The function $R(\Delta U)$ is a monotonically increasing concave function, and it satisfies the conditions $R'(\Delta U) > 0$, $R''(\Delta U) < 0$ and $R(0) = 0$ [33, 34]. The regret-rejoice utility with regret aversion parameter is shown as follows [10, 12, 17]:

$$R(\Delta U) = 1 - \exp(-\beta \cdot \Delta U), \quad (4)$$

where β represents a nonnegative regret aversion parameter, and it ranges from 0 to 1. The influence of regret aversion on the utility difference is shown in Figure 3. When $\Delta U > 0$, $|R(-\Delta U)| > R(\Delta U)$ is found with increase in β . It reveals that the perceived psychology to $-\Delta U$ is more sensitive than that to ΔU for an evacuee. In other words, the evacuees are regret averse, and the regret aversion level increases with the increases in the risk aversion parameter.

Consequently, considering the risky travel time, the regret-based utility for route a and b can be written as follows:

$$ERU(a) = \sum_s \sum_r p_s \cdot p_r \cdot \left(\left[\frac{1 - \exp(\alpha \cdot t_{sa})}{\alpha} \right] - \left[1 - \exp\left(-\beta \cdot \left[\frac{\exp(\alpha \cdot t_{sb}) - \exp(\alpha \cdot t_{sa})}{\alpha} \right]\right) \right] \right), \quad (5)$$

$$ERU(b) = \sum_s \sum_r p_s \cdot p_r \cdot \left(\left[\frac{1 - \exp(\alpha \cdot t_{sb})}{\alpha} \right] - \left[1 - \exp\left(-\beta \cdot \left[\frac{\exp(\alpha \cdot t_{sa}) - \exp(\alpha \cdot t_{sb})}{\alpha} \right]\right) \right] \right). \quad (6)$$

2.2. Traffic Equilibrium Conditions. From the perspective of behavior science, individual route choice decisions lead to the transportation network flow pattern evolving to traffic equilibrium [29]. More specifically, no user can decrease his utility by unilaterally switching routes. In this study, traffic equilibrium condition can be extended into a regret-based decision framework. The equilibrium is achieved when no user can

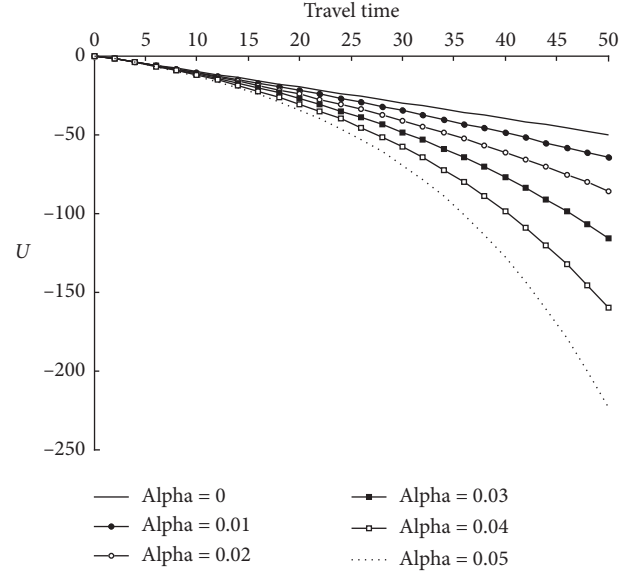


FIGURE 2: Influence of risk aversion on the utility with respect to travel times.

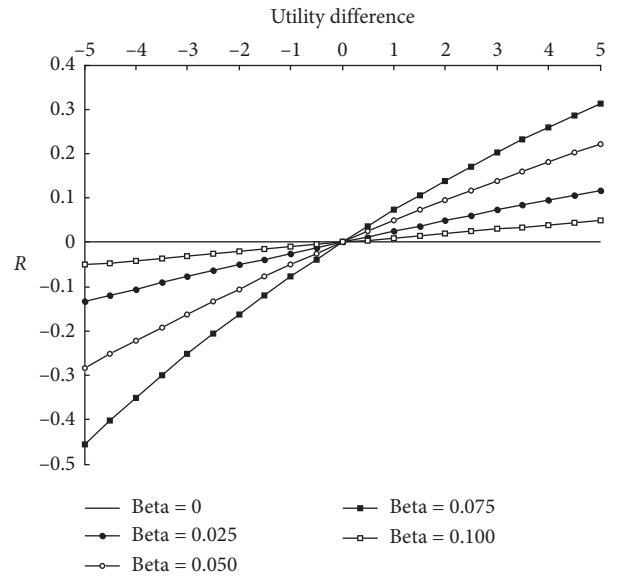


FIGURE 3: Influence of regret aversion on the utility difference.

increase his regret-based utility by unilaterally switching routes. According to user equilibrium principle proposed by Wardrop [35], the network is considered in equilibrium state when all traffic patterns stabilize and no user has any incentive to change its current route, as shown as follows:

$$\begin{cases} f(a) = 0, & \text{if } ERU(a) < ERU(b), \\ f(b) = 0, & \text{if } ERU(a) > ERU(b), \\ f(a) \cdot f(b) \neq 0, & \text{if } ERU(a) = ERU(b). \end{cases} \quad (7)$$

The disaster has a destructive effect on the road capacity. In order to capture this influence on network equilibrium in the real evacuation context, a capacity reduction parameter is introduced to BPR (Bureau of Public Road, BPR) function form adopted in Avineri [36] as follows:

$$t_{sa} = t_{saf} \cdot \left[1 + k \cdot \left(\frac{Q_{sa}}{\gamma_{sa} \cdot C_{sa}} \right)^\lambda \right] + t_{sar}, \quad (8)$$

where t_{saf} is free-flow travel time for route a under the state s ; k and λ are the parameters; Q_{sa} is the traffic flow on the route a under the state s ; C_{sa} is the basic capacity of the route a under the state s ; γ_{sa} is the road capacity reduction parameter under the state s ; the smaller the γ value, the greater the capacity loss; and t_{sar} is the random travel time of the route a under the state s .

Here, the BPR function is used as a route travel cost function when the links on one route are homogeneous (i.e., the capacity and the service level are the same) like Avineri [36]. Substituting equation (8) in equations (5) and (6), the network travel flow and user utility can be obtained in traffic equilibrium state.

3. Risk and Regret Aversion Parameter Analysis

One numerical example is conducted in this study to recognize the different effects of risk aversion level and regret aversion level on traffic equilibrium. The values of assumed parameters are partly taken from Chorus [17]. In contrast to the study done by Chorus, the regret-based traffic equilibrium is analyzed under emergency evacuation context, considering the effect of disaster on road capacity. Route a is closer to the disaster than route b but the route distance is shorter in Figure 1, it means route a is more likely to be affected by the disaster and suffer a greater capacity reduction in the emergency situation. For detailed information on the numerical example, the following settings are assumed: $t_{saf} = 10$, $t_{sbf} = 12$, $k = \lambda = 2$, and $C_a = C_b = 200$. There are 200 people that should be evacuated. With regards to the three states, the random parts of travel time and the capacity reduction parameters are different due to the damage degree of the disaster, as shown in Table 1. For the three states, the occurrence probability of the random parts of travel time is assumed to 1/3.

3.1. Different Risk Aversion Scenarios. Take the good state as the example, the parameter settings are assumed: $\beta = 0.05$, $\gamma_a = 0.85$, and $\gamma_b = 1$. When Q_a denotes the traffic flows on the route a , the expected regret-based utility of route a is captured when α varies from 0 to 0.05, as shown in Figure 4. It is found that the expected regret-based utility of route a decreases with the increase in the risk aversion parameter.

When $\beta = 0.05$, $\gamma_a = 0.85$, and $\gamma_b = 1$, the traffic flow achieves different equilibrium when α varies from 0 to 0.05, as shown in Figure 5. The traffic flows on the route a at the

traffic equilibrium state are 104, 101, 99, 96, 92, and 88 when α equals to 0, 0.01, 0.02, 0.03, 0.04, and 0.05, respectively. Take the case of $\alpha = 0.02$ as an example, the traffic flow on the route a is 99 cars, and the expected travel time is 31.8, compared with 101 cars on the route b and expected travel time of 33.1. As a result, the travel flow shifts from route a towards route b , when the risk aversion parameter increases.

3.2. Different Regret Aversion Scenarios. Take the good state as the example, the parameter settings are assumed: $\alpha = 0.03$, $\gamma_a = 0.85$, and $\gamma_b = 1$. When Q_a denotes the traffic flows on the route a , the expected regret-based utility of route a is captured when β varies from 0 to 0.1, as shown in Figure 6. It is found that the expected regret-based utility of route a decreases as the increase of regret aversion parameter.

When $\alpha = 0.03$, $\gamma_a = 0.85$, and $\gamma_b = 1$, the traffic flow achieves different equilibrium when β varies from 0 to 0.1, as shown in Figure 7. The traffic flows on the route a at the traffic equilibrium state are 96, 96, 95, 95, and 94 when β equals to 0, 0.025, 0.050, 0.075, and 0.100, respectively. Take the case of $\beta = 0.075$ as an example, the traffic flow on the route a is 95 cars, and the expected travel time is 31.2, compared with 105 cars on the route b and expected travel time of 33.6. As a result, the travel flow shifts from route a towards route b , when the risk aversion parameter increases.

3.3. Different Risk-Regret Aversion Scenarios. The traffic equilibrium without risk and regret aversions is shown in Figure 8. It can be seen that there are 104 cars on the route a and 96 cars on the route b at traffic equilibrium state. Expected travel times on the two routes equal 32.5. These findings are certain in line with the study conducted by Chorus [16]. The value 104 is smaller than 114 provided by Chorus; it is due to the fact that the capacity reduction is considered in this study.

The traffic equilibrium with different risk and regret aversions in each state is shown in Table 2. It can be seen that the traffic flows on the route a at the equilibrium state are the same when there is no risk aversion (i.e., $\alpha = 0$) whatever β varies. The traffic flows on route a are 104, 97, and 91 in the good, medium, and bad states, respectively. As shown in Table 2, the traffic flow on the route a shifts seriously with the regret aversion parameter increase when the risk aversion parameter increases simultaneously. This finding is consistent with that of the work done by Chorus [16]. In other words, the influence of increase in regret aversion on traffic equilibrium is larger for the traveler with more risk aversion. The impacts of increases in regret aversion and risk aversion appear to reinforce one another, both implying traffic equilibrium shifts towards safer routes (i.e., route b).

With respect to the capacity reduction parameter, the disaster damage degree reinforces the impacts from the regret aversion and risk aversion further. The traffic equilibrium for the route a is 101 when $\alpha = 0.04$ without damage. However, the traffic equilibrium is 93 when $\alpha = 0.04$ with a capacity reduction on route a (i.e., $\gamma_a = 0.85$), indicating that evacuees are more alike to choose the less damage routes.

TABLE 1: Random travel time and capacity reduction parameters for the three states.

Route	Parameters	Good state	Medium state	Bad state
Route <i>a</i>	Random travel time	0	5	10
		15	20	25
		30	35	40
	Capacity reduction parameter	0.85	0.75	0.65
Route <i>b</i>	Random travel time	12.5	15	17.5
		15	17.5	20
		17.5	20	22.5
	Capacity reduction parameter	1.00	0.85	0.75

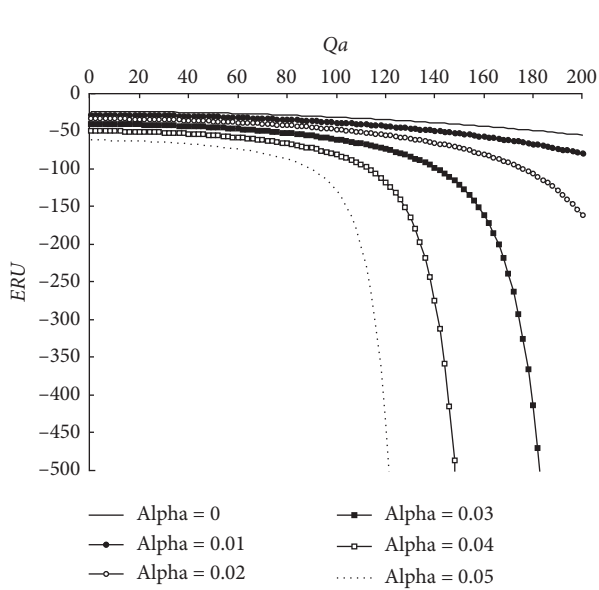


FIGURE 4: Expected regret-based utility of route *a* based on different risk aversion scenarios.

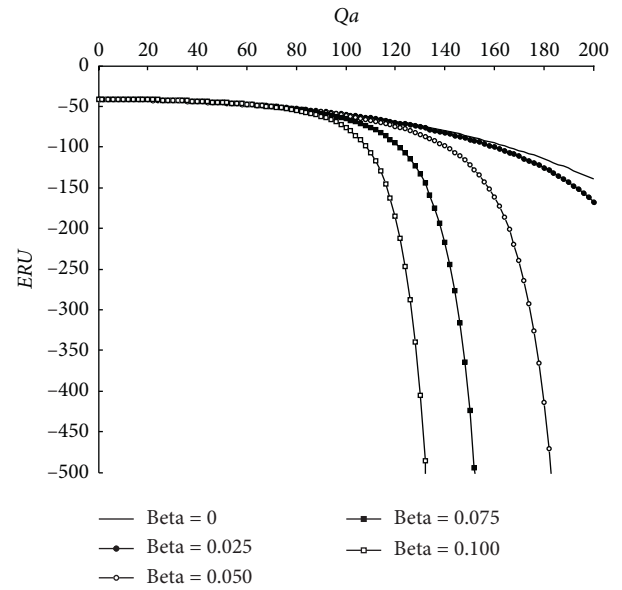


FIGURE 6: Expected regret-based utility of route *a* based on different regret aversion scenarios.

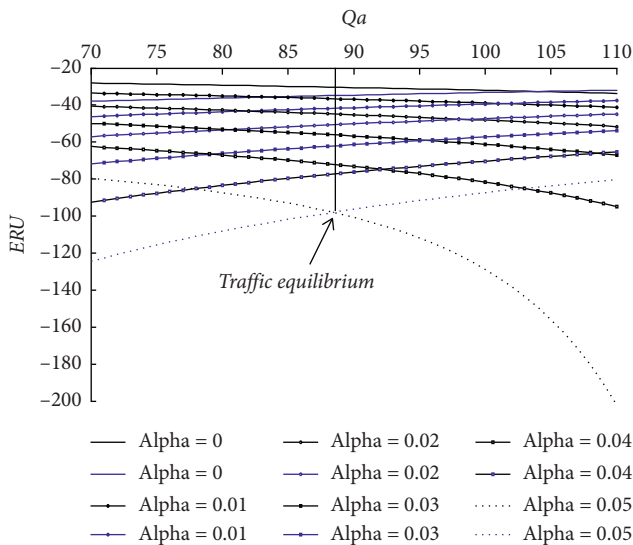


FIGURE 5: Influence of risk aversion parameter on traffic equilibrium. Note: the black line denotes route *a*, and the blue line denotes route *b*.

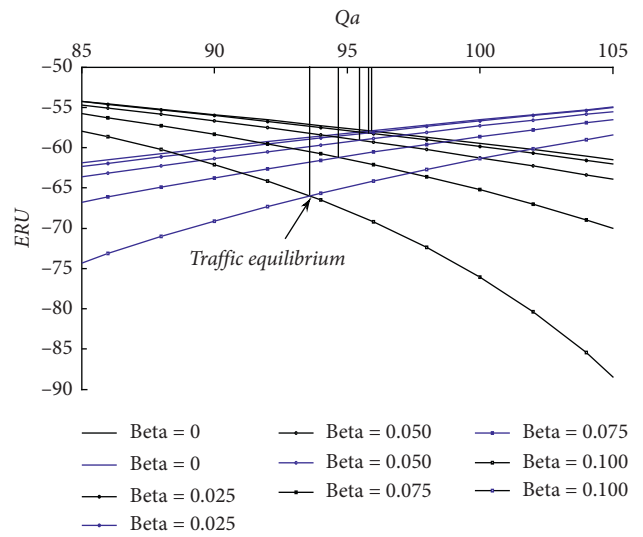


FIGURE 7: Influence of regret aversion parameters on traffic equilibrium. Note: the black line denotes route *a*, and the blue line denotes route *b*; $Q_b = 200 - Q_a$.

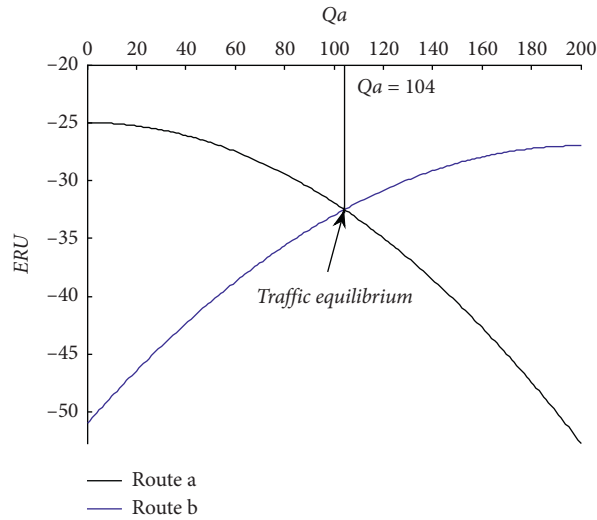


FIGURE 8: Expected regret-based utility without risk and regret aversion. Note. the black line denotes route a , and the blue line denotes route b ; $Q_b = 200 - Q_a$.

TABLE 2: Influence of risk aversion and regret aversion on traffic equilibrium.

β	γ_a and γ_b	$\alpha = 0$	$\alpha = 0.01$	$\alpha = 0.02$	$\alpha = 0.03$	$\alpha = 0.04$	$\alpha = 0.05$
$\beta = 0$	$\gamma_a = \gamma_b = 1.00$	114	110	107	104	101	98
	$\gamma_a = 0.85, \gamma_b = 1.00$	104	101	99	96	93	91
	$\gamma_a = 0.75, \gamma_b = 0.85$	97	95	93	91	89	87
	$\gamma_a = 0.65, \gamma_b = 0.75$	91	89	88	86	84	83
$\beta = 0.025$	$\gamma_a = \gamma_b = 1.00$	114	110	107	104	101	98
	$\gamma_a = 0.85, \gamma_b = 1.00$	104	101	99	96	93	90
	$\gamma_a = 0.75, \gamma_b = 0.85$	97	95	93	91	89	86
	$\gamma_a = 0.65, \gamma_b = 0.75$	91	89	88	86	84	82
$\beta = 0.05$	$\gamma_a = \gamma_b = 1.00$	114	110	107	104	100	95
	$\gamma_a = 0.85, \gamma_b = 1.00$	104	101	99	96	92	88
	$\gamma_a = 0.75, \gamma_b = 0.85$	97	95	93	90	87	84
	$\gamma_a = 0.65, \gamma_b = 0.75$	91	89	87	85	83	80
$\beta = 0.075$	$\gamma_a = \gamma_b = 1.00$	114	110	107	103	98	93
	$\gamma_a = 0.85, \gamma_b = 1.00$	104	101	98	95	91	86
	$\gamma_a = 0.75, \gamma_b = 0.85$	97	95	92	89	86	83
	$\gamma_a = 0.65, \gamma_b = 0.75$	91	89	87	85	82	80
$\beta = 0.100$	$\gamma_a = \gamma_b = 1.00$	114	110	106	101	96	91
	$\gamma_a = 0.85, \gamma_b = 1.00$	104	101	98	94	89	85
	$\gamma_a = 0.75, \gamma_b = 0.85$	97	94	92	88	85	82
	$\gamma_a = 0.65, \gamma_b = 0.75$	91	88	87	84	82	80

4. Traffic System Performance Assessment

From the viewpoint of decision-making department, it is helpful to know the global traffic system performance. Because the occurrence probability of the disaster at different states is uncertain, the traffic system performance is not a determined value. The global utility of traffic system at equilibrium can be considered as a performance indicator. To recognize this point, the interval value is used to present the occurrence probability. In detail, the interval values for occurrence probabilities of the good state, medium state, and bad state are assumed to be [0.2,

0.5], [0.4, 0.7], and [0.1, 0.3], respectively. The best state and the worst state of the traffic system can be estimated using the global optimal state model. For every disaster state, based on the perceived regret-based utility of the routes, the global utility of the traffic system using an optimization model as follows:

$$f(s) = \sum_s \sum_r p_s \cdot ERU_{sr} \tag{9}$$

In this study, the utilities for the three states when $\alpha = 0.04$ and $\beta = 0.1$ are used. The global optimal occurrence probability can be obtained using the following model:

$$f(s) = -161.12p_1 - 169.28p_2 - 240.84p_3, \quad (10)$$

$$\begin{cases} 0.2 \leq p_1 \leq 0.5, \\ 0.4 \leq p_2 \leq 0.7, \\ 0.1 \leq p_3 \leq 0.3, \\ p_1 + p_2 + p_3 = 1. \end{cases}$$

At last, the global optimal state probabilities can be obtained through solving equation (10). The best utility of traffic system at the network equilibrium is -172.356 that mainly depends on the good state and medium state, that is, $p_1 = 0.5$, $p_2 = 0.4$, and $p_3 = 0.1$. The increase of global utility due to the action on bad state is relatively limited. The worst utility of traffic system at the network equilibrium is -189.116 that mainly depends on the bad state and medium state, that is, $p_1 = 0.2$, $p_2 = 0.5$, and $p_3 = 0.3$. The interval value of traffic system utility is between -189.116 and -172.356 . The performance of traffic system can obtain a minimum regret-based disutility at 172.356 . According to the model results, the findings provide useful insights for the emergency agency that the management department should pay more attentions on the slight and medium disaster damage.

5. Conclusions

A regret-based utility function was used to account for the risk aversion and regret aversion psychology under emergency context. Meanwhile, the traffic equilibrium conditions were constructed considering the road capacity reduction. It is found that risk aversion, regret aversion, and capacity reduction have important effects on traffic equilibrium.

How risk aversion and regret aversion parameters influence the traffic equilibrium was explored. It is found that the traffic equilibrium tends to be achieved when safer routes are the main choice. It reflects that most evacuees are willing to choose the safer routes in the evacuation text. They are highly risk and regret aversion, and the outcome of the traffic equilibrium is dependent on the risk aversion and regret aversion levels.

The traffic equilibrium model under regret theory is an appealing approach compared with traditional utility theory, especially in the evacuation traffic context. It is reasonable that aversion psychology is taken into the analysis of evacuation issues. Different parameters produce different degree of effects on the travel decision. With the increase of aversion levels, evacuees prefer the stable routes. Future research can be found in a few directions. The application of the regret-based traffic equilibrium model can be conducted in complicated network. For example, in a network with three evacuation routes a , b , and c , $RU(a)$ can be modified to equal $U(t_{sa}) + R[U(t_{sa}) - \max\{U(t_{sa}), U(t_{sb}), U(t_{sc})\}]$. Meanwhile, facing a risk situation, evacuees' psychological behavior of regret aversion should not be all the same. The heterogeneity of regret-taking behavior should be taken into consideration in the future study. Real field data also should be utilized to validate the conclusion, and implementation cost can be well evaluated for evacuation practice.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was mainly supported by the National Natural Science Foundation of China (no. 71704161), Zhejiang Provincial Natural Science Foundation of China (no. LGF19G030002), China Postdoctoral Science Foundation (no. 2017M621935), National Natural Science Foundation of China (no. 71801188), and Zhejiang Provincial Philosophical and Social Science Program (no. 19NDJC167YB).

References

- [1] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.
- [2] J. Chen, K. Li, H. Rong, K. Bilal, K. Li, and P. S. Yu, "A periodicity-based parallel time series prediction algorithm in cloud computing environments," *Information Sciences*, vol. 496, pp. 506–537, 2019.
- [3] J. Tang, J. Hu, W. Hao, X. Chen, and Y. Qi, "Markov chains based route travel time estimation considering link spatio-temporal correlation," *Physica A: Statistical Mechanics and its Applications*, vol. 545, 2020.
- [4] T. A. Arentze and H. J. P. Timmermans, "Information gain, novelty seeking and travel: a model of dynamic activity-travel behavior under conditions of uncertainty," *Transportation Research Part A*, vol. 39, no. 2-3, pp. 125–145, 2005.
- [5] A. de Palma and N. Picard, "Route choice decision under travel time uncertainty," *Transportation Research Part A: Policy and Practice*, vol. 39, no. 4, pp. 295–324, 2005.
- [6] J. W. Polak, S. Hess, and X. Liu, "Characterizing heterogeneity in attitudes to risk in expected utility models of mode and departure time choice," in *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, January 2008.
- [7] R. D. Connors and A. Sumalee, "A network equilibrium model with travellers' perception of stochastic travel times," *Transportation Research Part B: Methodological*, vol. 43, no. 6, pp. 614–624, 2009.
- [8] A. de Palma and N. Picard, "Equilibria and information provision in risky networks with risk-averse drivers," *Transportation Science*, vol. 40, no. 4, pp. 393–408, 2006.
- [9] J. Tang, Y. Wang, W. Hao, F. Liu, H. Huang, and Y. Wang, "A mixed path size logit-based taxi customer-search model considering spatio-temporal factors in route choice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1347–1358, 2020.
- [10] D. E. Bell, "Regret in decision making under uncertainty," *Operations Research*, vol. 30, no. 5, pp. 961–981, 1982.
- [11] C. G. Chorus, T. A. Arentze, and H. J. P. Timmermans, "A Random Regret-Minimization model of travel choice,"

- Transportation Research Part B: Methodological*, vol. 42, no. 1, pp. 1–18, 2008.
- [12] G. Loomes and R. Sugden, “Regret theory: an alternative theory of rational choice under uncertainty,” *The Economic Journal*, vol. 92, no. 368, pp. 805–824, 1982.
- [13] T. Starver, “Anticipating regret: why fewer options may be better,” *Econometrica*, vol. 76, no. 2, pp. 263–305, 2008.
- [14] M. Boeri and L. Masiero, “Regret minimisation and utility maximisation in a freight transport context,” *Transportmetrica A: Transport Science*, vol. 10, no. 6, pp. 548–560, 2014.
- [15] C. G. Chorus, “A new model of random regret minimization,” *European Journal of Transport and Infrastructure Research*, vol. 10, no. 2, pp. 181–196, 2010.
- [16] C. Chorus, “Random regret minimization: an overview of model properties and empirical evidence,” *Transport Reviews*, vol. 32, no. 1, pp. 75–92, 2012.
- [17] C. G. Chorus, “Regret theory-based route choices and traffic equilibria,” *Transportmetrica*, vol. 8, no. 4, pp. 291–305, 2012.
- [18] G. M. Ramos, W. Daamen, and S. P. Hoogendoorn, “Expected utility theory, prospect theory, and regret theory compared for prediction of route choice behavior,” *Journal of Transportation Research Record*, vol. 2230, no. 1, pp. 19–28, 2011.
- [19] M. Li and H.-J. Huang, “A regret theory-based route choice model,” *Transportmetrica A: Transport Science*, vol. 13, no. 3, pp. 250–272, 2017.
- [20] S. Ye, S. Ma, and N. Jia, “A bi-attribute user equilibrium model considering travellers’ regret aversion,” *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1440–1458, 2019.
- [21] Y. Yu, J. Zhou, and W. Xu, “Regret-based multi-objective route choice models and stochastic user equilibrium: a non-compensatory approach,” *Transportmetrica A: Transport Science*, vol. 16, no. 3, pp. 473–500, 2020.
- [22] S. Gao, E. Frejinger, and M. Ben-akiva, “Adaptive route choices in risky traffic networks: a prospect theory approach,” *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 727–740, 2010.
- [23] Z. Li and D. Hensher, “Prospect theoretic contributions in understanding traveller behaviour: a review and some comments,” *Transport Reviews*, vol. 31, no. 1, pp. 97–115, 2011.
- [24] H. Xu, Y. Lou, Y. Yin, and J. Zhou, “A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing,” *Transportation Research Part B: Methodological*, vol. 45, no. 2, pp. 311–328, 2011.
- [25] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, “Risk-aware resource control with flexible 5G access technology interfaces,” in *Proceedings of the 2019 IEEE 20th International Symposium on “A World of Wireless, Mobile and Multimedia Networks” (WoWMoM)*, IEEE, Washington, DC, USA, June 2019.
- [26] P. Vamvakas, E. E. Tsiropoulou, and S. Papavassiliou, “On the prospect of UAV-assisted communications paradigm in public safety networks,” in *Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops*, IEEE, Paris, France, April 2019.
- [27] E. E. Tsiropoulou, K. Koukas, and S. Papavassiliou, “A socio-physical and mobility-aware coalition formation mechanism in public safety networks,” *EAI Endorsed Transactions on Future Internet*, vol. 18, no. 13, 2018.
- [28] M. Shin, S. T. Shah, M. Y. Chung et al., “Moving small cells in public safety networks,” in *Proceedings of the 2017 International Conference on Information Networking (ICOIN)*, Da Nang, Vietnam, January 2017.
- [29] P. Yuan and Z. Juan, “Risk route choice analysis and the equilibrium model under anticipated regret theory,” *PROMET—Traffic & Transportation*, vol. 26, no. 1, pp. 33–43, 2014.
- [30] C. G. Chorus, “Risk aversion, regret aversion and travel choice inertia: an experimental study,” *Transportation Planning and Technology*, vol. 37, no. 4, pp. 321–332, 2014.
- [31] P. Murray-tuite and B. Wolshon, “Evacuation transportation modeling: an overview of research, development, and practice,” *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 25–45, 2013.
- [32] C. E. Llaciana and E. U. Weber, “Correcting expected utility for comparisons between alternative outcomes: a unified parameterization of regret and disappointment,” *Journal of Risk and Uncertainty*, vol. 36, no. 1, pp. 1–17, 2008.
- [33] H. Levy, “Absolute and relative risk aversion: an experimental study,” *Journal of Risk and Uncertainty*, vol. 8, no. 3, pp. 289–307, 1994.
- [34] J. Quiggin, “Regret theory with general choice sets,” *Journal of Risk and Uncertainty*, vol. 8, no. 2, pp. 153–165, 1994.
- [35] J. G. Wardrop, “Road paper. Some theoretical aspects of road traffic research,” *Proceedings of the Institution of Civil Engineers*, vol. 1, no. 3, pp. 325–362, 1952.
- [36] E. Avineri, “The effect of reference point on stochastic network equilibrium,” *Transportation Science*, vol. 40, no. 4, pp. 409–420, 2006.

Research Article

Exploring the Application of the Linear Poisson Autoregressive Model for Analyzing the Dynamic Impact of Traffic Laws on Fatal Traffic Accident Frequency

Yue Zhang,¹ Yajie Zou ,¹ Lingtao Wu,² Jinjun Tang ,³ and Malik Muneeb Abid^{1,4}

¹Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, No. 4800 Cao'an Road, Shanghai 201804, China

²Texas A&M Transportation Institute, Texas A&M University System, 3135 TAMU, College Station 77843, TX, USA

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

⁴Department of Civil Engineering, College of Engineering and Technology, University of Sargodha, Sargodha, Pakistan

Correspondence should be addressed to Yajie Zou; yajiezou@hotmail.com and Jinjun Tang; jinjuntang@csu.edu.cn

Received 3 July 2020; Revised 10 September 2020; Accepted 20 September 2020; Published 9 October 2020

Academic Editor: Hailong Liu

Copyright © 2020 Yue Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Annual fatal traffic accident data often demonstrate time series characteristics. The existing traffic safety analysis approaches (e.g., negative binomial (NB) model) often cannot accommodate the dynamic impact of factors in fatal traffic accident data and may result in biased parameter estimation results. Thus, a linear Poisson autoregressive (PAR) model is proposed in this study. The objective of this study is to apply the PAR model to analyze the dynamic impact of traffic laws and climate on the frequency of fatal traffic accidents occurred in a large time span (from 1975 to 2016) in Illinois. Besides, the NB model, NB with a time trend, and autoregressive integrated moving average model with exogenous input variables (ARIMAX) are also developed to compare their performances. The important conclusions from the modelling results can be summarized as follows. (1) The PAR model is more appropriate for analyzing the dynamic impacts of traffic laws on annual fatal traffic accidents, especially the instantaneous impacts. (2) The law that allows motorcycles and bicycles to proceed on a red light following the rules applicable after a “reasonable period of time” leads to an increase in the frequency of annual fatal traffic accidents by 14.98% in the short term and 30.69% in the long term. The climate factors such as average temperature and precipitation concentration period have insignificant impacts on annual fatal traffic accidents in Illinois. Thus, the modelling results suggest that the PAR model is more appropriate for annual fatal traffic accident data and has an advantage in estimating the dynamic impact of traffic laws.

1. Introduction

A report of National Highway Traffic Safety Administration (NHTSA) reveals that 37,461 people were killed in 34,439 motor vehicle crashes, an average of 102 deaths per day in year 2016. To reduce the number of people killed in traffic accidents, it is important to analyze the influential factors affecting the frequency of fatal traffic accidents. Among many factors, the traffic law is considered an effective measure to reduce the severity of injuries and the number of fatalities as a means of macroeconomic regulation and control. Some existing studies have analyzed the impact of

certain traffic laws on the number of traffic accidents, such as the seat belt law [1], driving under the influence (DUI) law [2, 3], and alcohol control law [4]. However, the dynamic impacts of these traffic laws on traffic accidents have not been adequately studied.

So far, numerous traffic safety analysis models have been developed. Since the frequency of traffic accident is non-negative and integer, many studies assumed such events follow a Poisson distribution and modelled the frequency of traffic accidents using a Poisson regression model [5, 6]. However, the Poisson model cannot handle overdispersed or underdispersed data and may result in biased estimation. In

order to analyze the overdispersed data, many studies proposed different mixed Poisson models, such as the Poisson-gamma model (the negative binomial (NB) model) [7–13], Poisson-lognormal model [14–16], and Poisson-inverse gamma model [17]. For the data with many zeros (i.e., excess zero-count data), the zero-inflated models were applied, including the zero-inflated Poisson model [18, 19], zero-inflated negative binomial model [20–22], and their extension models (i.e., multiple random parameter zero-inflated negative binomial regression model [20] and zero expansion Poisson regression model with random parameter effect [23]). Although rare, crash data can sometimes be characterized by underdispersion. The Conway–Maxwell–Poisson model [24] and diagonal inflated bivariate Poisson regression model [25] were appropriate.

Several recent works about regression models cannot properly address the time series characteristics of the traffic accident count data. Noland et al. [26] proposed a time trend variable as an explanatory variable in the count regression model to consider the series correlation. However, this model may not clearly account for the effects of serial correlation. An alternative approach was modelling possible dynamics in the traffic accident count data with a lagged dependent variable in the Poisson or NB models. These models failed to represent adequately the dynamics in persistent time series because they implied that the growth rate of the process was the exponentiated coefficient on the lagged dependent variable. Such a process may potentially generate time series data rather than dynamic data [27]. These two kinds of models were dynamic models with a trend, but not necessarily a cyclical or dynamic component. Another approach to handle the time series was the autoregressive integrated moving average (ARIMA) model and its extensions including the seasonality autoregressive integrated moving average (SARIMA) [28] and nonlinear autoregressive exogenous (NARX) [29]. These time series models may not be applicable to discrete time series variables (e.g., traffic accident count data). To consider both time series and discrete characteristics of the response variables, an integer-valued autoregressive (INAR) Poisson model [30–34] was developed. However, the dynamic characteristics of the influential factors were not adequately described in the INAR model. Few approaches can adequately model the dynamics and distribution of annual fatal traffic accident data.

To address the above issues, this study proposed a linear Poisson autoregressive (PAR) model. The objective of this study is to apply the PAR model to analyze the dynamic impact of traffic laws on annual fatal traffic accident frequency from the macroscopic point of view using the data collected in Illinois from year 1975 to 2016. The contribution is to demonstrate the performance of the PAR model in the analysis of the dynamic influence of factors on traffic accident frequency and quantitatively analyze the impacts of traffic laws.

The rest of the paper is organized as follows. Section 2 introduces the specification, estimation, and interpretation of the PAR (p) model. Section 3 describes the dataset used in

this study and the source of our data. In Section 4, the results of statistical modelling are shown to understand the contribution of different factors to the annual fatal traffic accidents in Illinois and compare the performances of various models. A conclusion and future recommendations are provided in Section 5.

2. Methodology

2.1. The PAR Model. Before presenting the model, the linear autoregressive (AR) process is firstly introduced. The AR model describes the random variables at some time by using the linear combination of random variables at earlier time as equation (1). It is a common form of time series.

$$y_t = \sum_{i=1}^p \rho_i y_{t-i} + \lambda, \quad (1)$$

where y_t is the traffic accident count value at time t , y_{t-i} is the past traffic accident count value at the i moments before time t , $\rho_i \in [-1, 1]$ is autocorrelation coefficient, and λ is a random error term.

Because there are explanatory variables in the PAR model, it is necessary to redefine the variables in the AR process. The conditional data Y_{t-i} replace y_{t-i} in the AR (p) model, which is a vector that included all the observed values of the dependent and independent variables at time t .

$$Y_{t-1} = (y_0, y_1, \dots, y_{t-1}; X_0, X_1, \dots, X_{t-1}), \quad (2)$$

where X_{t-i} are the past explanatory variables (factors affecting traffic accidents) at the i moments before time t . Y_{t-1} can be regarded as all the prior information about the series of interest at time t . Assume that y_t is a realization from a Markov process with the conditional transition probability $\Pr(y_t | Y_{t-1})$ and $E[Y_0] = \mu$. Let the conditional expectation $E[y_t | Y_{t-1}] = m_t$ at time t have a finite mean. Then, y_t is a p th order linear autoregressive process as shown in the following equation:

$$E[y_t | Y_{t-1}] = \sum_{i=1}^p \rho_i Y_{t-i} + \lambda. \quad (3)$$

Then, we can obtain equations (4) and (5) by using iterated expectations [35] of equation (3):

$$E[E[y_t | Y_{t-1}]] = E\left[\sum_{i=1}^p \rho_i Y_{t-i} + \lambda\right], \quad (4)$$

$$E[Y_t] = \sum_{i=1}^p \rho_i E[Y_{t-i}] + \lambda, \quad (5)$$

where equation (5) is a geometric series for ρ_i ; then,

$$\lim_{t \rightarrow \infty} E[Y_t] = \frac{\lambda}{(1 - \sum_{i=1}^p \rho_i)} \equiv \mu. \quad (6)$$

Since $[Y_0] = \mu$, equation (3) can be written as

$$E[y_t | Y_{t-1}] = \sum_{i=1}^p \rho_i Y_{t-i} + \left(1 - \sum_{i=1}^p \rho_i\right) \mu. \quad (7)$$

This is a linear AR(p) process, where the distribution of y is not used. The only role of distribution of y_t is to define the possible value of ρ_i .

The PAR (p) model can be defined as follows. The assumptions of the model are that the observed traffic accident counts y_t ($t = 1, 2, \dots, T$) are generated from a Poisson distribution on the condition of m_t . Then, the measurement equation for the observation value is obtained in the following equation:

$$\Pr(y_t | m_t) = \frac{m_t^{y_t} e^{-m_t}}{y_t!}. \quad (8)$$

Assume that m_t is the conditional mean of the linear AR process of $E[y_t | Y_{t-1}]$, which defines the state variable for the model. According to the measurement equation (8), this state density with its mean m_t and variance σ_t is in the exponential family. The mean or state variable of the marginal Poisson distribution evolves according to a stationary AR (p) process with autocorrelation parameters ρ_i , and its transition equation is expressed as follows:

$$m_t = \sum_{i=1}^p \rho_i y_{t-i} + \left(1 - \sum_{i=1}^p \rho_i\right) \exp(X_t \delta), \quad (9)$$

where X_t are $T \times k$ -matrices of explanatory variables composed of various factors affecting traffic accidents, δ is a $k \times 1$ -vector of regression parameters, and k is the number of factors.

Finally, assume that the density of the state variable has a gamma-distributed conjugate prior with gamma's location parameter a_{t-1} and scale parameter b_{t-1} , so

$$\Pr(m_t | Y_{t-1}) = \Gamma(\sigma_{t-1} m_{t-1}, \sigma_{t-1}), \quad m_{t-1} > 0, \sigma_{t-1} > 0, \quad (10)$$

where $m_t = E[y_t | Y_{t-1}]$ and $\sigma_{t-1} = \text{Var}[y_t | Y_{t-1}]$. The prior is constructed using the observed traffic accidents data. The prior can help to find the conditional mean and variance at time t based on the previous $t - 1$ observations.

Since the prior is gamma, using an extended Kalman filter, the conditional distribution at time t given $t - 1$ is also gamma, that is, $m_{t|t-1} \sim \Gamma((m_{t|t-1} \sigma_{t|t-1}), \sigma_{t|t-1})$. Since the measurement equation is Poisson and the state equation is gamma, an estimate of the posterior at time t is equation (11). This is a negative binomial distribution.

$$\begin{aligned} \Pr(y_t | Y_{t-1}) &= \int_{\theta} \Pr(y_t | \theta_t) \Pr(\theta_t | Y_{t-1}) d\theta \\ &= \int_{\theta} \frac{\theta_t^{y_t} e^{-\theta_t}}{y_t!} \cdot \frac{e^{-\sigma_{t|t-1} \theta} \theta^{\sigma_{t|t-1} m_{t|t-1} - 1} \sigma_t^{\sigma_{t|t-1} m_{t|t-1}}}{\Gamma(\sigma_{t|t-1} m_{t|t-1})} \\ &= \frac{\Gamma(\sigma_{t|t-1} m_{t|t-1} + y_t)}{\Gamma(y_t + 1) \Gamma(\sigma_{t|t-1} m_{t|t-1})} (\sigma_{t|t-1})^{\sigma_{t|t-1} m_{t|t-1}} \\ &\quad \times (1 + \sigma_{t|t-1})^{-(\sigma_{t|t-1} m_{t|t-1} + y_t)}. \end{aligned} \quad (11)$$

Replace the AR (1) process with m_t and obtain the PAR (1) model with a negative binomial predictive distribution. The one-step ahead conditional forecast function for the PAR (p) model is expressed as follows:

$$\begin{aligned} E[y_{t+1} | Y_t] &= m_{t+1|t} \\ &= \sum_{i=1}^p \rho_i m_{t|t-1} + \left(1 - \sum_{i=1}^p \rho_i\right) \mu. \end{aligned} \quad (12)$$

2.2. The Impact Multipliers. Because the PAR (p) model considers an explanatory variable matrix X_t and $\mu = \exp(X_t \delta)$, the interpretation differs from the Poisson and NB models significantly [36]. There is a concept of impact multiplier as in a Gaussian linear autoregressive model, which is the effect of a change in explanatory variable X_t . The instantaneous impact multiplier can be obtained by calculating the first derivative of the mean function for this change. The calculation process is shown in the following equation:

$$\begin{aligned} \frac{\partial m_t}{\partial X_t} &= \frac{\partial (\sum_{i=1}^p \rho_i Y_{t-i} + (1 - \sum_{i=1}^p \rho_i) \exp(X_t \delta))}{\partial X_t} \\ &= \left(1 - \sum_{i=1}^p \rho_i\right) \exp(X_t \delta) \cdot \delta, \end{aligned} \quad (13)$$

where δ is the coefficient of the explanatory variable X_t . This is the instantaneous effect of a shock in factors affecting traffic accidents X_t on the mean of traffic accidents m_t . The long-run impact multiplier which means the total effect of a shock to X_t can be calculated by equation (14), as in the Gaussian time series analysis. The long-run multiplier can be compared with the parameter estimation of other count regression models, which measures the impact of a shock on the conditional mean number of events.

$$\begin{aligned} \frac{\partial m_t / \partial X_t}{(1 - \sum_{i=1}^p \rho_i)} &= \frac{(1 - \sum_{i=1}^p \rho_i) \exp(X_t \delta) \cdot \delta}{(1 - \sum_{i=1}^p \rho_i)} \\ &= \exp(X_t \delta) \cdot \delta. \end{aligned} \quad (14)$$

The long-run impact multiplier and the instantaneous impact multiplier correspond to the concepts of average impact and marginal impact in economics, respectively. In economics, the average impact corresponds to the whole time, while the marginal impact corresponds to the "present" in time. The long-run impact is compared with the whole of the past, which should consider the impact of the past. However, the instantaneous impact multiplier or the marginal impact (the marginal effect is obtained by the partial derivation of X_t as shown in equation (13)) focuses on the impact on the future without considering the past.

In the Poisson and NB models, the long-run and instantaneous impact multipliers are the same and they can also be calculated by equation (14). The reason for this difference is

that the PAR (p) model accounts for the influence of explanatory variables on traffic accidents and the dynamic responses to the changes in explanatory variables over time.

3. Data Description

The annual fatal traffic accident frequency data for Illinois from year 1975 to 2016 were obtained from NHTSA's Fatality Analysis Reporting System (FARS). The annual fatal traffic accident was taken as the dependent variable to avoid underreporting due to various definitions of traffic accidents. Some traffic laws were considered to evaluate the instantaneous and long-run impact on traffic accidents from a quantitative perspective. According to WHO [37], five road safety risk factors (i.e., speeding, drunk-driving, the use of helmets, seatbelts, and child restraint systems) played an important role in traffic injuries and deaths. And Senna et al. [38] concluded that driving under the influence of alcohol is always a dominant problem. Due to the far-reaching influence of law on traffic accidents, the research data of law are basically based on the year [1, 39]. Thus, the traffic laws in Table 1 were selected for analysis of annual fatal traffic accidents. The variables related to law were set as binary variables here. For example, the initial effective date of safety belt law in Illinois was January 1, 1988, and then this traffic law variable equalled 0 for the first 13 periods and 1 thereafter (every traffic law used in this study and their effective date are shown in Table 1). If the date of implementation of the law was in the latter half of the year, the law was considered to work from the second year.

Besides, various factors identified as related to traffic safety are composed of our dataset to analyze how the explanatory variables affect the annual fatal traffic accident. In order to be consistent with legal factors, we also select some macroscopic indicators ranging from 1975 to 2016 with an interval of year in Illinois. The dataset is assembled from a variety of sources including the U.S. Energy Information Administration, Federal Highway Administration, and National Institute on Alcohol Abuse and Alcoholism. The dataset covers economic, social, driver, climate, and law factors. Summary statistics of the dataset are shown in Table 1. Note that the climate factors mentioned are different from the microscopic weather factor, which indicates the wet and dry conditions over a large space-time range. In the dataset, gross domestic product (GDP), total vehicle miles of travel (VMT), rural VMT as proportion of total VMT, per capita beer consumption, gasoline price, and safety belt law are used to analyze the influence of various factors on traffic accidents [3]. Geedipally et al. [3] demonstrated that DUI laws, beer consumption, the proportion of rural VMT, and shocks in the economy had a significant effect in traffic fatalities. Note that all the economic indicators are converted to 2016 dollars using the consumer price index (CPI) calculator at the Bureau of Labor Statistics.

4. Modelling Results

Four models (i.e., the ARIMAX, PAR, NB, and NB with a time trend model) are developed using the Illinois data, where the ARIMAX, NB, and NB with a time trend model

are considered as benchmark models. All the models constructed in this paper are implemented with R. The main purpose is to identify the influential traffic laws affecting the fatal traffic accidents. Figure 1 shows the trend of the fatal traffic frequency accident and VMT over time. As can be seen from Figure 1, the frequency of fatal traffic accident decreases significantly with time and shows sequence correlation, while VMT increases with time.

For the PAR(p) model mentioned in the methodology, the order p of the PAR (p) model is determined firstly (Table 2). Based on the PAR (1) model, stepwise regression is used to select all significant variables as the combination of explanatory variables of each model. Thus, the final models include only a subset of the original explanatory variables, which is shown in Tables 2 and 3. The Akaike information criterion (AIC) which is an estimator of the relative quality of statistical models for a given set of data provided another means for order selection. The smaller its value is, the better the fitting effect of the model is. When $p > 4$, the model complexity increases but AIC does not decrease significantly. As discussed by Eluru et al. [40], different from the AIC, the Bayesian information criterion (BIC) imposes a larger penalty on model overfitting with excess parameters. As can be seen from Table 2, the BIC value for the PAR (4) and PAR (5) models differ slightly because the PAR (5) model has more parameters than the PAR (4) model. Besides, the estimated insignificant parameters are shown in bold in Table 2. There are insignificant parameters from $p > 2$, which will affect the analysis of the impact multiplier of explanatory variables. Based on the modelling results in Table 2, the PAR (2) model is chosen as the final model.

In addition to the PAR (2) model, the ARIMAX, NB, and NB with a time trend models are also compared as alternative models. Because of the time series in the data, the ARIMAX model considering the explanatory variables is selected as one of the alternative models. Based on AIC values, the final model is determined as ARIMAX (1, 1, 0). Besides, the NB model, which is most commonly used in traffic accident frequency analysis, is considered as one of the alternative models. Furthermore, the NB with a time trend model which can consider the time series and discrete characteristic of the traffic accident frequency by using a simple solution is also compared. The parameter values of these models are estimated using the maximum likelihood estimation method.

The results of these models are shown in Table 3. According to the AIC and BIC values in Table 3, the PAR (2) model fits this dataset best, followed by the NB with a time trend model and ARIMAX (1, 1, 0) model. The NB model provides the least fitting performance. The models with a time trend structure seem to fit best since the traffic accidents appear to be serially correlated. Taking AIC as example, the fitting performance of the PAR model increases by 12% compared to the NB model, 6% compared to the ARIMAX model, and 5% compared to the NB with a time trend model. However, the coefficients of explanatory variables estimated by the ARIMAX model are not significant; in other words, the ARIMAX model is not able to explain how these factors affect annual fatal traffic accidents. For the

TABLE 1: Summary statistics of annual fatal traffic accidents and factors in Illinois.

Variables	Min	Max	Mean	St. error
Fatal traffic accidents	832	1877	1316	299.31
Total VMT (million)	60943	109135	89696	16914.83
GDP (in \$100,000)	405739	807458	593554	134797
Rural VMT proportion (%)	0.24	0.34	0.30	0.03
Gasoline price (\$ per gallon)	1.54	3.71	2.41	0.64
Per capita beer consumption (gallons)	1.12	1.45	1.30	0.10
Average temperature (°F)	49.38	55.78	52.16	1.52
Precipitation concentration period	28.54	51.18	39.68	5.68
Safety belt law	January 1988		Before: 0; after: 1 (year 1988: 1)	
DUI toughened penalties	January 1982		Before: 0; after: 1 (year 1982: 1)	
Drivers under age 18 may not use a wireless phone while driving	July 2005		Before: 0; after: 1 (year 2005: 0)	
Persons under age 18 who obtain a graduated driver’s license may not drive during the first 6 months of the license, or until the person reaches age 18, with more than one person under age 20 in the vehicle	January 2004		Before: 0; after: 1 (year 2004: 1)	
Disallow parents/guardians to knowingly permit the consumption of alcohol by underage invitees at their residence	August 2007		Before: 0; after: 1 (year 2007: 0)	
Motorcycles and bicycles may proceed on a red light following the rules applicable after a “reasonable period of time”	January 2012		Before: 0; after: 1 (year 2012: 1)	
Graduated driver’s license provisions strengthened	January 2008		Before: 0; after: 1 (year 2008: 1)	
Limit the use of mobile phones while operating a motor vehicle	January 2010		Before: 0; after: 1 (year 2010: 1)	

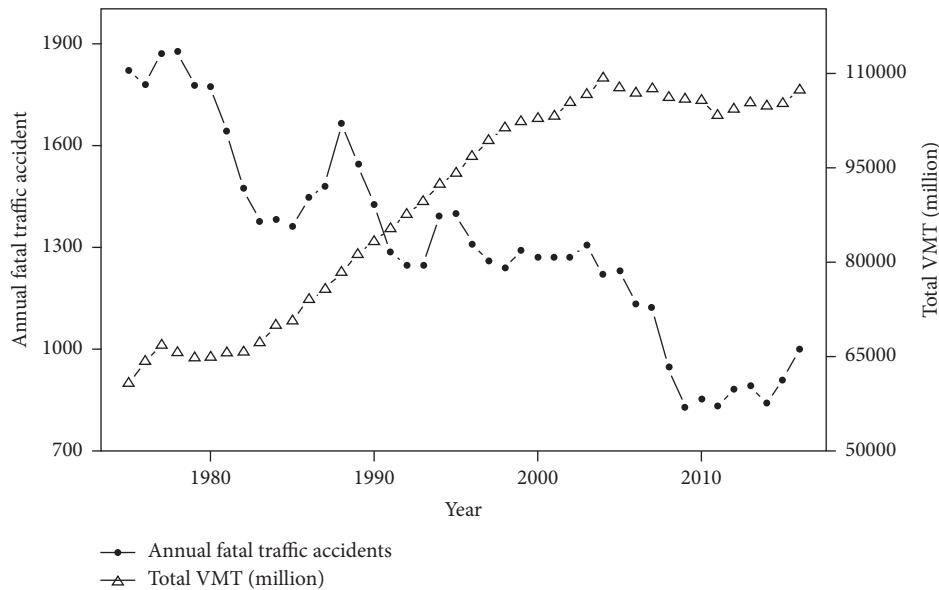


FIGURE 1: Annual fatal traffic accidents and VMT in Illinois.

NB with a time trend model, most of the explanatory variables are significant and the values of parameter estimates are similar to those of the NB and PAR (2) models. However, the belt law and beer consumption variables are insignificant. This phenomenon shows that this model may not help us to explain the impact of various legal factors on the frequency of traffic accidents completely. Since the PAR (2) model and NB model have all statistically significant variables, their fitting performance and parameter estimates are compared in the following paragraphs.

Except the ARIMAX model, other models belong to the regression model. During modelling the fatal traffic accidents, total VMT is considered as an offset term because there is a linear relationship between total VMT and fatal traffic accidents [3]. From a qualitative point of view, both the coefficients estimated by the NB and PAR models show that beer consumption has the greatest impact among these factors on the frequency of annual fatal traffic accidents (Table 3). However, the AIC and BIC values of the PAR (2) model are much smaller than those of the NB model. Note

TABLE 2: Modelling results for different orders of PAR models.

Model	PAR (1)	PAR (2)	PAR (3)	PAR (4)	PAR (5)
Intercept	-6.216 (0.375)	-6.821 (0.328)	-6.811 (0.322)	-7.028 (0.341)	-6.787 (NA)
Gasoline price	-0.152 (0.038)	-0.074 (0.028)	-0.074 (0.029)	-0.101 (0.037)	-0.124 (0.048)
Beer consumption	2.12 (0.244)	2.37 (0.204)	2.363 (0.197)	2.476 (0.209)	2.491 (0.225)
Belt law	-0.196 (0.046)	-0.157 (0.037)	-0.159 (0.038)	-0.118 (0.042)	-0.093 (0.049)
DUI toughened penalties	-0.379 (0.05)	-0.279 (0.043)	-0.278 (0.05)	-0.206 (0.076)	-0.451 (NA)
Red running	0.393 (0.083)	0.301 (0.059)	0.296 (0.065)	0.383 (0.089)	0.434 (0.113)
Alcohol law	-0.237 (0.067)	-0.219 (0.049)	-0.216 (0.052)	-0.247 (0.062)	-0.276 (0.075)
ρ_1	0.628 (0.057)	0.706 (0.093)	0.704 (0.095)	0.754 (0.095)	0.718 (0.098)
ρ_2	—	-0.194 (0.079)	-0.179 (0.119)	-0.218 (0.12)	-0.157 (0.127)
ρ_3	—	—	-0.015 (0.086)	-0.265 (0.118)	-0.262 (0.12)
ρ_4	—	—	—	0.287 (0.082)	0.144 (0.12)
ρ_5	—	—	—	—	0.147 (0.09)
N^1	42	42	42	42	42
Log-likelihood	-234.125	-222.116	-217.543	-200.139	-197.746
AIC	482.249	460.232	453.069	420.277	417.492
BIC	494.414	474.133	468.725	437.655	436.606

¹ N is the number of samples.

TABLE 3: Fatal traffic accident analysis (standard errors in parentheses).

Model	ARIMAX (1, 1, 0)	PAR (2)	NB	NB with a time trend
Intercept	—	-6.821 (0.328)	-5.835 (0.347)	7.118 (0.29)
Gasoline price	3.836 (32.897)	-0.074 (0.028)	-0.077 (0.028)	-0.05 (0.02)
Beer consumption	575.665 (514.852)	2.37 (0.204)	1.805 (0.242)	0.393 (0.211)
Belt law	191.826 (66.103)	-0.157 (0.037)	-0.225 (0.048)	0.029 (0.039)
DUI toughened penalties	-98.515 (77.537)	-0.279 (0.043)	-0.445 (0.048)	-0.227 (0.042)
Red running	58.631 (63.765)	0.301 (0.059)	0.162 (0.059)	0.096 (0.039)
Alcohol law	-144.166 (67.999)	-0.219 (0.049)	-0.221 (0.06)	-0.228 (0.041)
Linear trend	—	—	—	-0.007 (0.003)
ρ_1	—	0.706 (0.093)	—	—
ρ_2	—	-0.194 (0.079)	—	—
α^1	—	—	0.006	0.002
N^2	42	42	42	42
Log-likelihood	-237.85	-222.116	-254.633	-232.543
AIC	491.71	460.232	525.27	483.087
BIC	505.61	474.133	539.168	498.726

¹The dispersion parameter in the NB model. ² N is the number of samples.

that the PAR model can capture dynamics in fatal traffic accident data and provide better fitting performance. In addition to the goodness-of-fit statistics, we further compare the modelling results of the PAR (2) model and the NB model.

For dynamic models, the results cannot be fully obtained by displaying coefficients in Table 3, which represented the average effect of explanatory variables [41]. Besides, the generalized linear model includes a link function, which makes it difficult to explain the original coefficients of the model independently [41]. In other words, the estimated coefficients in Table 3 cannot directly quantify the impact of factors per unit change on fatal traffic accident frequency. Thus, in order to compare the effects of explanatory variables on fatal traffic accidents in different models, the long-run and instantaneous multipliers are calculated by using equations (13) and (14). Since this study focuses on the impact of traffic laws on annual fatal traffic accidents, only the impact multipliers of law are presented in the first

column of Tables 4 and 5. Because of the dynamic structure of the PAR (2) model, the value of long-run impact multiplier is larger than that of instantaneous impact multiplier ($0 < \sum_{i=1}^2 \rho_i < 1$). The three laws (e.g., belt law, DUI toughened penalties, and alcohol law) lead to a decrease in fatal traffic accident, among which the DUI toughened penalties law has the greatest influence. After the implementation of the DUI toughened penalties, the frequency of fatal traffic accident decreases by 91 in the short run and 186 in the long run. However, implementation of the law that allows motorcycles and bicycles to proceed on a red light following the rules applicable after a “reasonable period of time” (red running) leads to an increase in the frequency of fatal traffic accidents. Before this study, Pai and Jou [42] have revealed the high association between bicyclist red-running and accidents in Taiwan. The effect of red running is a total increase of 201 fatal traffic accidents in the long run and 72 fatal traffic accidents in the short run, which indicates that this law is not conducive to traffic safety. This law may be

TABLE 4: Fatal traffic accident analysis (long-run effect).

Variable	PAR (2)		NB	
	Long-run multiplier	Percentage change in mean (%)	Long-run multiplier	Percentage change in mean (%)
Belt law	-105.137	-15.00	-155.447	-20.15
DUI toughened penalties	-186.371	-25.65	-309.795	-35.92
Red running	201.172	30.69	112.659	17.59
Alcohol law	-146.495	-20.63	-153.839	-19.83

TABLE 5: Fatal traffic accident analysis (instantaneous effect).

Variable	PAR (2)		NB	
	Instantaneous multiplier	Percentage change in mean (%)	Instantaneous multiplier	Percentage change in mean (%)
Belt law	-51.322	-7.32	-155.447	-20.15
DUI toughened penalties	-90.975	-12.52	-309.795	-35.92
Red running	98.2	14.98	112.659	17.59
Alcohol law	-71.51	-10.07	-153.839	-19.83

designed to increase the traffic efficiency of nonmotorized vehicles and reduce travel time, but it is not desirable to improve traffic efficiency at the expense of traffic safety.

Finally, the impact multipliers of the NB and PAR (2) models are compared (Tables 4 and 5). All signs of parameters estimated by the NB and PAR models are the same. It can be observed that the frequency of fatal traffic accidents decreases with the increase of gasoline prices, the implementation of the belt law, and the enforcement of the DUI penalty law. With beer consumption rising and red running allowed, the frequency of fatal traffic accident increases. However, the values of impact multipliers estimated by the NB and PAR (2) models differ significantly. Taking the implementation of red running law as an example, which is the only law variable leading to an increase in the frequency of traffic accidents in the dataset, the instantaneous impact of the red running law is about 98 for the PAR (2) model, which means that the implementation of red running law has increased the number of accidents by 98 at this point. The long-run multiplier of red running law is 201, which means the frequency of accidents increased by 201 in the long run. The NB model estimates the impact of enforcing the red running law by a multiplier of 113. These percentage changes are shown in Tables 4 and 5. For the PAR (2) model, the total percentage change due to this intervention in the number of fatal traffic accidents is an increase of 30.69% while the instantaneous percentage change is 14.98%. For the NB model, the total change in the number of fatal traffic accidents is 17.59%. The instantaneous effect of the PAR (2) model is smaller than the instantaneous effect of the NB model, and long-run effects of the PAR (2) model are larger than the long-run effect of the NB model. The reason for this phenomenon is that the estimated coefficients of the PAR (2) model include dynamic characteristics. The long-run multiplier takes into account the impact of the previous stage on the present, while the instantaneous effect only takes into account the current impact. Since δ describes an average effect, the multipliers calculated from $\exp(X_t \delta) \cdot \delta$ can only describe the average effect. Thus, the instantaneous effect of

the NB model actually reflects the average effect, which leads to overestimation of the short-run effect. The NB model cannot properly consider the dynamics of fatal traffic accident data, which leads to overestimation of the instantaneous impact of explanatory variables on fatal traffic accidents.

For the remaining three variables, both the long-run and instantaneous impact multipliers estimated by the PAR model are smaller than those estimated by the NB model. Taking the belt law as an example, the instantaneous and long-run percentage changes due to the intervention estimated by the PAR model are -7.32% and -15%, respectively, and the estimated multiplier of the NB model is -20.15%. This phenomenon indicates again that the NB model overestimates the impact of explanatory variables, especially for the instantaneous impacts. The dynamic nature of the PAR (p) model makes it more suitable for estimating the dynamic impact of traffic laws on annual fatal traffic accidents. The instantaneous impact of a safety intervention strategy can inform the transportation management agencies to design more appropriate traffic laws, while the NB model cannot provide such information.

5. Conclusions

Annual fatal traffic accidents are count data with time series characteristics. The existing traffic accident analysis models cannot fully model their dynamic characteristics and analyze the dynamic influence of explanatory variables on annual fatal traffic accidents. Among many explanatory variables of traffic accident analysis, the dynamic effect of the enforcement of traffic laws has not been widely concerned. In this study, a linear Poisson autoregressive model is proposed to analyze the long-run and instantaneous impact of traffic laws on annual fatal traffic accidents. Then, the modelling results of PAR (p), ARIMAX, NB, and NB with a time trend models are compared. Several major conclusions are summarized as follows:

- (1) The PAR model can outperform the ARIMAX, NB, and NB with a time trend models in terms of fitting performance and estimation of dynamic effects. The PAR (p) model is more suitable for analyzing the dynamic impact of traffic laws on annual fatal traffic accidents. Compared with the ARIMAX model, the PAR (p) model can consider discrete characteristics in the accident data and analyze the influence of factors. Compared with the NB with a time trend model, the PAR (p) model can accurately analyze the influence of more explanatory variables on the frequency of fatal traffic accidents. Compared with the NB model, the PAR (p) model can capture the time series in annual fatal traffic accident frequency and calculate the dynamic effect of traffic laws and other explanatory variables. The omission of the dynamics from the NB model leads to biased parameter estimates, especially the inability to estimate the instantaneous multipliers of factors. However, instantaneous multipliers can indicate the immediate effects of traffic law interventions for traffic safety management agencies and help to make new laws.
- (2) Some climate and traffic law factors are considered to quantitatively evaluate their impact on annual fatal traffic accidents in Illinois. The average temperature and precipitation concentration period have insignificant impacts. The law of DUI toughened penalties results in a decrease of annual fatal traffic accidents by 12.52% in the short run and 25.65% in the long run, which has the greatest inhibitory effect on fatal traffic accidents among the analyzed laws. However, the law allowing red running leads to an increase of annual fatal traffic accidents by 14.98% in the short term and 30.69% in the long term. Therefore, controlling the DUI behaviors and modifying the red running law may significantly reduce the frequency of annual fatal traffic accidents, which provide guidance for future traffic law development.

The PAR (p) model can be widely applied to analyze the time series count data. Besides the traffic laws mentioned in this paper, the applicable explanatory variables exhibiting a sudden change can be extended to the factors such as the emergence of policies and regulations, and the dynamic impact of these kinds of variables can be well explained by the PAR (p) model. For future research, the PAR (p) model can be applied to investigate traffic accident data collected from other sites. Furthermore, with the development of data acquisition technology, multisource datasets [43–46] can be used to analyze traffic accidents.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This research was sponsored jointly by the National Key Research and Development Program of China (grant no. 2018YFE0102800), the Shanghai Science and Technology Committee (grant no. 19210745700), and the Fundamental Research Funds for the Central Universities (grant no. 22120200035).

References

- [1] J. Lv, D. Lord, Y. Zhang, and Z. Chen, "Investigating Peltzman effects in adopting mandatory seat belt laws in the US: evidence from non-occupant fatalities," *Transport Policy*, vol. 44, pp. 58–64, 2015.
- [2] V. R. Campos, R. d. S. E. Silva, S. Duailibi, J. F. D. Santos, R. Laranjeira, and I. Pinsky, "The effect of the new traffic law on drinking and driving in São Paulo, Brazil," *Accident Analysis & Prevention*, vol. 50, pp. 622–627, 2013.
- [3] S. R. Geedipally, "In-depth investigation of factors that contributed to the decline in fatalities from 2008 to 2012 in the United States," in *Proceedings of the Transportation Research Board 97th Annual Meeting*, Washington, DC, USA, 2018.
- [4] K. Chang, C.-C. Wu, and Y.-H. Ying, "The effectiveness of alcohol control policies on alcohol-related traffic fatalities in the United States," *Accident Analysis & Prevention*, vol. 45, pp. 406–415, 2012.
- [5] M. Hosseinpour, S. Sahebi, Z. H. Zamzuri, A. S. Yahaya, and N. Ismail, "Predicting crash frequency for multi-vehicle collision types using multivariate Poisson-lognormal spatial model: a comparative analysis," *Accident Analysis & Prevention*, vol. 118, pp. 277–288, 2018.
- [6] X. Ye, "A semi-nonparametric Poisson regression model for analyzing motor vehicle crash data," *PLoS One*, vol. 13, 2018.
- [7] C. Dong, "An improved deep learning model for traffic crash prediction," *Journal of Advanced Transportation*, vol. 2018, p. 13, 2018.
- [8] C. Ma, W. Hao, A. Wang, and H. Zhao, "Developing a coordinated signal control system for urban ring road under the vehicle-infrastructure connected environment," *Ieee Access*, vol. 6, pp. 52471–52478, 2018.
- [9] W. Ma and Z. Yuan, "Analysis and comparison of traffic accident regression prediction model," in *Proceedings 3rd International Conference on Electromechanical Control Technology and Transportation (ICECTT 2020)*, New York, NY, USA, 2018.
- [10] M. R. R. Shaon, X. Qin, M. Shirazi, D. Lord, and S. R. Geedipally, "Developing a random parameters negative binomial-lindley model to analyze highly over-dispersed crash count data," *Analytic Methods in Accident Research*, vol. 18, pp. 33–44, 2018.
- [11] K. Wang and X. Ye, *Development of Alternative Stochastic Frontier Models for Estimating Time-Space Prism Vertices*, Transportation, New York, NY, USA, 2020.
- [12] Y. Zou, "Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety," *Journal of Applied Statistics*, vol. 14, pp. 1–18, 2017.

- [13] Y. Zou, L. Wu, and D. Lord, "Modeling over-dispersed crash data with a long tail: examining the accuracy of the dispersion parameter in Negative Binomial models," *Analytic Methods in Accident Research*, vol. 5-6, pp. 1-16, 2015.
- [14] M. Shirazi and D. Lord, "Characteristics based heuristics to select a logical distribution between the Poisson-gamma and the Poisson-lognormal for crash data modelling," *Transportmetrica A: Transport Science*, vol. 14, pp. 1-22, 2019.
- [15] K. Wang, S. Zhao, and E. Jackson, "Multivariate Poisson lognormal modeling of weather-related crashes on freeways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 38, pp. 184-198, 2018.
- [16] M. Zhao, "Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach," *Journal of Transportation Safety & Security*, vol. 10, no. 3, pp. 251-265, 2018.
- [17] S. H. Khazraee, V. Johnson, and D. Lord, "Bayesian Poisson hierarchical models for crash data analysis: investigating the impact of model choice on site-specific predictions," *Accident Analysis & Prevention*, vol. 117, pp. 181-195, 2018.
- [18] V. N. Shankar, G. F. Ulfarsson, R. M. Pendyala, and M. B. Nebergall, "Modeling crashes involving pedestrians and motorized traffic," *Safety Science*, vol. 41, no. 7, pp. 627-640, 2003.
- [19] S. Unhapipat, M. Tiensuwan, and N. Pal, "Bayesian predictive inference for zero-inflated Poisson (ZIP) distribution with applications," *American Journal of Mathematical and Management Sciences*, vol. 37, no. 1, pp. 66-79, 2018.
- [20] C. Liu, M. Zhao, W. Li, and A. Sharma, "Multivariate random parameters zero-inflated negative binomial regression for analyzing urban midblock crashes," *Analytic Methods in Accident Research*, vol. 17, pp. 32-46, 2018.
- [21] M. A. Raihan, P. Alluri, W. Wu, and A. Gan, "Estimation of bicycle crash modification factors (CMFs) on urban facilities using zero inflated negative binomial models," *Accident Analysis & Prevention*, vol. 123, pp. 303-313, 2019.
- [22] H. Wang, Y. Yang, and Y. Ma, "Research on the Yangtze River accident casualties using zero-inflated negative binomial regression technique," in *Proceedings of the 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, Berlin, Germany, 2016.
- [23] M. Haque, H. C. Chin, and H. Huang, "Modeling random effect and excess zeros in road traffic accident prediction," in *Proceedings of the 19th KKCNN Symposium on Civil Engineering*, Berlin, Germany, 2006.
- [24] D. Lord, S. R. Geedipally, and S. D. Guikema, "Extension of the application of conway-maxwell-Poisson models: analyzing traffic crash data exhibiting underdispersion," *Risk Analysis*, vol. 30, no. 8, pp. 1268-1276, 2010.
- [25] Y. Lao, Y.-J. Wu, J. Corey, and Y. Wang, "Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 220-227, 2011.
- [26] R. Noland, M. Quddus, and W. Ochieng, "The effect of the congestion charge on traffic casualties in London: an intervention analysis," in *Proceedings of the Transportation Research Board (TRB) Annual Meeting*, Washington, DC, USA, January 2006.
- [27] P. T. Brandt, J. T. Williams, B. O. Fordham, and B. Pollins, "Dynamic modeling for persistent event-count time series," *American Journal of Political Science*, vol. 44, no. 4, pp. 823-843, 2000.
- [28] S. Yousefzadeh-Chabok, "A time series model for assessing the trend and forecasting the road traffic accident mortality," *Archives of Trauma Research*, vol. 5, no. 3, 2016.
- [29] A. Sukhai, A. P. Jones, B. S. Love, and R. Haynes, "Temporal variations in road traffic fatalities in South Africa," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 421-428, 2011.
- [30] T. Brijs, D. Karlis, and G. Wets, "Studying the effect of weather conditions on daily crash counts using a discrete time-series model," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1180-1190, 2008.
- [31] Q. Duan, X. Ye, J. Li, and K. Wang, "Empirical modeling analysis of potential commute demand for carsharing in shanghai, China," *Sustainability*, vol. 12, no. 2, p. 620, 2020.
- [32] C. Ma, R. He, and W. Zhang, "Path optimization of taxi carpooling," *PLoS One*, vol. 13, no. 8, 2018.
- [33] M. Quddus, "Time-eries regression models for analysing transport safety data," in *Safe Mobility: Challenges, Methodology and Solutions*, L. Dominique and W. Simon, Eds., pp. 279-296, Emerald Publishing Limited, Berlin, Germany, 2018.
- [34] Y. Sunecher, N. M. Khan, and V. Jowaheer, "Estimation methods for a flexible INAR (1) COM-Poisson time series model," *Journal of Applied Mathematics, Statistics and Informatics*, vol. 14, no. 1, pp. 57-82, 2018.
- [35] G. K. Grunwald, R. J. Hyndman, and L. M. Tedesco, "A unified view of linear AR (1) Models," 1995.
- [36] P. T. Brandt and J. T. Williams, "A linear Poisson autoregressive model: the Poisson AR(p) model," *Political Analysis*, vol. 9, no. 2, pp. 164-184, 2001.
- [37] W. H. O. D. O. Violence, *Global Status Report on Road Safety: Time for Action*, World Health Organization, Geneva, Switzerland, 2009.
- [38] M.-C. Senna, M. Augsburg, B. Aebi et al., "First nationwide study on driving under the influence of drugs in Switzerland," *Forensic Science International*, vol. 198, no. 1-3, pp. 11-16, 2010.
- [39] D. Mark Anderson, B. Hansen, and D. I. Rees, "Medical marijuana laws, traffic fatalities, and alcohol consumption," *The Journal of Law and Economics*, vol. 56, no. 2, pp. 333-369, 2013.
- [40] N. Eluru, M. Bagheri, L. F. Miranda-Moreno, and L. Fu, "A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings," *Accident Analysis & Prevention*, vol. 47, pp. 119-127, 2012.
- [41] B. J. Fogarty and J. E. Monogan Iii., "Modeling time-series count data: the unique challenges facing political communication studies," *Social Science Research*, vol. 45, pp. 73-88, 2014.
- [42] C.-W. Pai and R.-C. Jou, "Cyclists' red-light running behaviours: an examination of risk-taking, opportunistic, and law-obeying behaviours," *Accident; Analysis and Prevention*, vol. 62C, pp. 191-198, 2013.
- [43] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *Journal of Advanced Transportation*, vol. 2020, pp. 1-16, 2020.
- [44] X. Zhong, Y. Zou, Z. Dong, S. Yuan, and M. Ijaz, "Finite mixture survival model for examining the variability of urban arterial travel time for buses, passenger cars and taxis," *IET Intelligent Transport Systems*, vol. 45, 2020 In Press.
- [45] X. Chen, X. Xu, Y. Yang, H. Wu, J. Tang, and J. Zhao, "Augmented ship tracking under occlusion conditions from maritime surveillance videos," *IEEE Access*, vol. 8, pp. 42884-42897, 2020.
- [46] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, "Robust ship tracking via multi-view learning and sparse representation," *Journal of Navigation*, vol. 72, no. 1, pp. 176-192, 2018.