

## Research Article

# Modeling Lane-Keeping Behavior of Bicyclists Using Survival Analysis Approach

**Hongwei Guo, Wuhong Wang, Weiwei Guo, and Facheng Zhao**

*Department of Transportation Engineering, Beijing Institute of Technology, Beijing 100081, China*

Correspondence should be addressed to Wuhong Wang; [wangwuhong@bit.edu.cn](mailto:wangwuhong@bit.edu.cn)

Received 27 September 2012; Accepted 9 January 2013

Academic Editor: Geert Wets

Copyright © 2013 Hongwei Guo 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.

Bicyclists may cross the bicycle lane and occupy the adjacent motor lanes for some reason. The mixed traffic consisting of cars and bicycles shows very complicated dynamic patterns and higher accident risk. To investigate the reason behind such phenomenon, the lifetime analysis method is adopted to examine the observed data for the behavior that bicycles cross the bicycle lane and occupy the adjacent motor lanes. The concepts named valid volume and probability of lane-keeping behavior are introduced to evaluate the influence of various external factors such as lane width and curb parking, and a semiparametric method is used to estimate the model with censored data. Six variables are used to accommodate the effects of traffic conditions. After the model estimation, the effects of the selected variables on the lane-keeping behavior are discussed. The results are expected to give a better understanding of the bicyclist behavior.

## 1. Introduction

Our daily life and work are closely related to traffic and mobility. Nowadays, in consequence of dramatically increasing traffic demand, traffic congestion has an immense negative impact on daily life and modern society [1, 2]. In many developing countries, like China, a typical traffic phenomenon is called “mixed traffic” which consists of cars (representing motorized vehicles) and bicycles (representing nonmotorized vehicles). Such inhomogeneous traffic flow has been considered as an important cause of traffic congestions and accidents. Additionally, bicycles are usually used as a kind of green traffic tool and it is important to improve traffic condition. Therefore, researchers turned their attention to the traffic characteristic of bicycles and mixed traffic composed of bicycles and cars. They did research from theoretical and practical points of view. Of the research on bicycles and mixed traffic, traffic simulation is popular method. Jiang et al. proposed a stochastic multivalued cellular automata model for bicycle flow [3]. Faghri and Egyházióvá developed a microscopic simulation model of mixed motor vehicle and bicycle traffic over an entire urban network [4]. Zhao et al. described mixed traffic flow by combining the NaSch model and the Burger cellular automata (BCA) model and investigated the

mixed traffic system near a bus stop [5, 6]. Mallikarjuna and Rao extended the cellular automata (CA) model to study the heterogeneous traffic [7]. On the other hand, theoretical models derived from empirical data were proposed. Oketch proposed a microscopic model to describe the mixed traffic flow by using the combination of car-following model and lateral movement [8]. Yang et al. presented a road capacity model for mixed traffic flow at the curbside stop based on queuing theory and gap acceptance theory [9]. Guo et al. used PLM model and Weibull's distribution to analyze the lane-crossing behavior of nonmotorized vehicles under the influence of curb parking [10].

In an urban street without segregated facility, the bicyclists may drive in the motor lane because of the blockage in the bicycle lane. Once the bicyclists do not satisfy the traffic condition, they will arbitrarily change travel route and even occupy the motor lane. Particularly for the position near bus stop or parking area, the occupancy of motor lane has a strong impact on traffic performance and safety [9]. The above-mentioned literature mainly focused on analyzing the influence of mixed traffic flow on traffic performance, and most research adopted the microsimulation method. However, the research on the traffic behavior of bicyclist is

TABLE 1: Analogy between lifetime data analysis and bicyclist behavior analysis.

	Lifetime analysis	Lane-crossing behavior analysis
Parameter	Time $t$	bicycle volume $q$
Failure event	Death at time $t$	Lane-crossing behavior at volume $q$
Variable	Lifetime $T$	Valid volume $q_c$
Censoring	Lifetime $T$ is longer than the duration of the observation	Valid volume $q_c$ is greater than travel demand
Survival function	$S(t) = 1 - F(t)$	$F_{lk}(q)$
Probability distribution function	$F(t)$	$F_{lc}(q)$

limited. In this paper, the traffic behavior that bicyclists keep in bicycle lane (called by lane-keeping behavior) is considered. The question why bicyclists cross the bicycle lane is discussed. With this aim, a survival-analysis-based approach is used to model the lane-keeping behavior of bicyclists under various external factors. To give a quantitative analysis of the lane-keeping behavior, the probability of lane-keeping behavior is studied using field data. A concept named valid volume is proposed and a semiparametric method is used to perform the analysis. It is hoped to provide reference frame for evaluating the influence of traffic conditions on the traffic behavior of bicyclist.

## 2. Method

**2.1. Analysis of the Bicyclist Behavior.** Some bicyclists are apt to travel in the adjacent motor lanes in order to get their desired driving conditions, for example, speed and space. When the bicycle lane is blocked due to some reason, the probability of the lane-crossing behavior would increase obviously. Assume that the lane-crossing behavior occurs if the bicycle volume  $q$  is higher than a critical value  $q_c$ , yielding

$$F_{lc}(q) = \text{Pro}(q \geq q_c), \quad (1)$$

where  $F_{lc}(q)$  is the distribution function of the lane-crossing behavior.

In this paper, the critical volume  $q_c$  is defined as valid volume in order to represent the maximum volume that bicycles would travel in the bicycle lane. That is, the lane-crossing behavior would not occur if the bicycle volume is lower than the valid volume ( $q < q_c$ ). Here, another definition that the probability of lane-keeping behavior is used:

$$F_{lk}(q) = 1 - F_{lc}(q) = \text{Pro}(q < q_c). \quad (2)$$

In terms of  $F_{lk}(q)$ , the influence of external factors on the behavior of bicyclist can be reflected by the distribution function of a lane-crossing behavior or lane-keeping behavior. Therefore, the influential factors (e.g., narrowing lane and obstructing lane) can be analyzed from a macroscopic perspective. Considering the volume as the input variable, the data acquisition and the data processing can be simplified because some variables involving the traffic behavior of a bicyclist are difficult to be quantized.

**2.2. Modeling Bicyclist Behavior Based on Lifetime Analysis.** Survival analysis models (also called lifetime analysis) have been used extensively for several decades in biometrics and industrial engineering as a means of determining causality in lifetime data [11, 12]. In recent years, they have been applied in the field of transportation [13, 14], including the analysis of activity participation and scheduling, vehicle transactions analysis, and incident-duration analysis. These models concern the distribution of lifetime  $T$ :

$$F(t) = 1 - S(t) = \text{Pr}(t \geq T), \quad (3)$$

where  $F(t)$  is the distribution function of lifetime data representing the probability that an individual fails before  $t$ ;  $S(t)$  is the survival function representing the probability that an individual survives longer than  $t$ .  $S(t)$  is also called reliability function.

The traffic behavior that bicyclists travel in bicycle lanes can be considered as a valid state under particular conditions (e.g., lane widths, traffic volume, and curb parking). Such a valid state continues with an increasing bicycle volume. If the volume is greater than the valid volume, the lane-crossing behavior will occur. It means that the particular conditions are hard to satisfy the travel demands of bicyclists. The continual process of valid state is similar to the continued life. If the lane-crossing behavior is regarded as the termination of life, the methods for lifetime data analysis can be used to estimate the valid volume  $q_c$ , which is the analogy of the lifetime  $T$ . In addition, due to the randomness of traffic behavior and the influence of curb parking, the same volume may correspond with two contrary behaviors, that is, crossing the lane or not. In this case, the lifetime analysis models are appropriate to solve such problem by censor analysis though the general statistical methods are no longer applicable [10]. The whole analogy between the bicyclist behavior analysis and lifetime data analysis is given in Table 1.

Firstly, an important concept, hazard function, is introduced. A hazard function at specified volume  $q$  in mathematical definition is

$$h(q) = \lim_{\Delta q \rightarrow 0} \frac{\text{Pro}(q \leq q_c \leq q + \Delta q \mid q_c \geq q)}{\Delta q}. \quad (4)$$

The result in the hazard function is hazard rate (or hazard), which is the instantaneous probability that the lane-crossing will occur in an infinitesimally small volume  $\Delta q$  after

TABLE 2: External factors and explanation.

Variable	Name	Type	Explanation
$X_1$	Effective width	Continuous variable	The effective width of bicycle lane
$X_2$	Travel speed	Continuous variable	The average travel speed for the survey interval
$X_3$	Car volume	Continuous variable	Car volume in the adjacent lane (veh/30 s) <sup>a</sup>
$X_4$	Curb parking	Binary indicator	1 if there are curb parking cars along the bicycle lane, 0 otherwise
$X_5$	Retrograde motion	Binary indicator	1 if there are retrograde bicycles in the lane at the observed interval, 0 otherwise
$X_6$	Safe gap	Binary indicator	1 if the adjacent lane is clear of moving car, 0 otherwise

<sup>a</sup>“veh” is the abbreviation of vehicle.

time  $q$ .  $h(q)\Delta q$  is the approximate probability of the lane-crossing behavior in  $[q, q + \Delta q]$ .

According to the mathematical relation between the hazard function and survival function, the probability of lane-keeping can be obtained:

$$F_{lk}(q) = \exp \left[ - \int_0^q h(s) ds \right]. \quad (5)$$

To accommodate the effects of external factors, the hazard function can be written as

$$h(q) = h_0(q) g(\mathbf{x}, \boldsymbol{\beta}), \quad (6)$$

where  $h_0(q)$  is the baseline hazard function,  $g(\cdot)$  is a known function to represent the effects of covariates,  $\mathbf{x}$  is a column vector of covariates and it is independent of duration time,  $\boldsymbol{\beta}$  is a row vector of unknown parameters. The form of (4) is one of the popular mathematical models used for duration analysis and its name is proportional hazard (PH) model.

In this study, a framework of nonparametric baseline hazard, which was proposed by Cox using  $g(\mathbf{x}, \boldsymbol{\beta}) = \exp(\boldsymbol{\beta}\mathbf{x})$ , is adopted [15]. With this parameterization, the hazard function is

$$h(q) = h_0(q) \exp(\boldsymbol{\beta}\mathbf{x}). \quad (7)$$

The endurance probability function combining (5) and (7) can be written as

$$S(q) = [S_0(q)]^{\exp(\boldsymbol{\beta}\mathbf{x})}, \quad (8)$$

where  $S_0(q)$  is the baseline probability function for the lane-keeping behavior. It represents the probability without any external influence.

The shape of  $h_0(q)$  in the PH model has important implications for data analysis. Also a parametric shape could be chosen according to data distribution. In this paper, the nonparametric baseline hazard is used to avoid the error when the assumed parametric form is incorrect. The parameter estimation can use the partial likelihood method. Other methods can be referred to in [11, 12].

**2.3. External Factors Selection.** The selection of external factors takes into account the previous researches and arguments regarding the effects of the exogenous variables and human

factors on bicyclist behavior. Three broad sets of variables may influence the bicyclist behavior: personal characteristics, traffic conditions, and trip characteristics. In this paper, the traffic conditions are considered. The following factors, as shown in Table 2, are adopted to construct the model.

### 3. Survey and Data

The field survey is conducted in the urban roads with no isolation facilities. The selected survey sites are monitored by video cameras. Then, the bicycle volumes in the lanes with different effective widths can be acquired. According to [10], the effective width of bicycle lane is defined as the physical width minus the margin of safety (0.5 m). Such safety margin indicates the influence of curb parking so that bicycles would keep a safe distance from the parked cars. On the other hand, the data related to the external factors are also derived from the video survey. The data acquisition is performed by manual counting and recording and the assistance of video processing tool.

The length of the observed section is 25 m and there is no influence of bus stop and pedestrian crosswalk. In consideration of the discrete arrival of bicycles and the nonuniform volume, short observed interval may not include enough samples while long interval may influence the definition of data status. Therefore, the observed interval is 30 s. The status of each interval is defined as (a) censored data if there is no bicycle entering the motor lane in the interval and (b) distinct data if the lane-crossing behavior occurs in the interval [10]. The field surveys are conducted in four typical urban roads in Beijing, including East Jiaoda Road, North Yufang Road, West Tucheng Road, and Da Liushu Road. The basic features of observed sections are shown in Table 3.

### 4. Model Estimation and Discussion

**4.1. Model Estimation.** Table 4 shows the model estimation of the lane-keeping behavior of bicyclists. The LR statistic of the estimated model clearly indicates the overall goodness-of-fit (the LR statistic is 98.4, which is greater than the chi-squared statistic with 6 degrees of freedom at any reasonable level of significance). The statistical significance of each variable is examined by the Z test, which has an asymptotically normal distribution with mean zero and variance one. The significant

TABLE 3: Basic features of observed sections.

Effective width (m)	Sample size	Volume distribution (veh/30 s)			Lane-crossing ratio (%)
		Min	Max	Mean	
1.5	118	2	28	12	47
2.5	378	4	35	17	33
3	381	2	34	26	28
3.5	430	3	39	19	29

TABLE 4: Model estimation.

Variable	Estimate	Standard error	Z test	P value
$X_1$	-0.271	0.12	-2.264	0.024
$X_2$	0.002	0.152	2.092	0.036
$X_3$	-0.201	0.192	1.047	0.295
$X_4$	0.317	0.001	-2.238	0.025
$X_5$	0.414	0.001	-3.777	<0.001
$X_6$	0.148	0.148	-1.014	0.303

level corresponding to each covariate is given by  $P$  value in Table 4. From the results, most of the included variables are statistically significant at the 0.05 level of significance. It means that these covariates are significantly related to the violation behavior. Two covariates ( $X_3$  car volume and  $X_6$  safe gap) have a relatively low significant level. The effects of the variables on lane-crossing behavior will be discussed in the next section.

Figure 1 shows the lane-keeping probability by the proposed model with the average of all variables. It means that the curve in Figure 1 can reflect the average lane-keeping probability of the typical bicycle flow which has an average value for every external factor. The curve is monotonely decreasing. The median of the distribution is 25 vehicle/30 s, indicating that over a half of the observed interval would result in lane-crossing behavior if the bicycle volume is greater than 25 veh/30 s in average condition (2.5 m lane, bicycle speed is 12 km/h, car volume is 450 pcu/h/lane, and about 30% of the bicyclists are affected by curb parking). If the variable changes (representing the change of external factor), the probability of lane keeping would change correspondingly.

**4.2. Effects of External Factors.** In the proportional hazard model, the effects of variables are multiplicative on the baseline hazard function. A negative coefficient on a variable implies that an increase in the corresponding variables decreases the hazard rate, or equivalently increases the valid volume. The greater valid volume means that the occurrence of lane-crossing behavior decreases. The effects of external factors are analyzed in the following.

(1) Effective width of bicycle lane shows a significant negative on hazard. It means the wider the lane width for bicycles, the less the lane-crossing behavior would occur. Figure 2 shows various distributions of the lane-keeping behavior with different lane widths. According to the empirical data and the estimated model, if the effective width of a bicycle

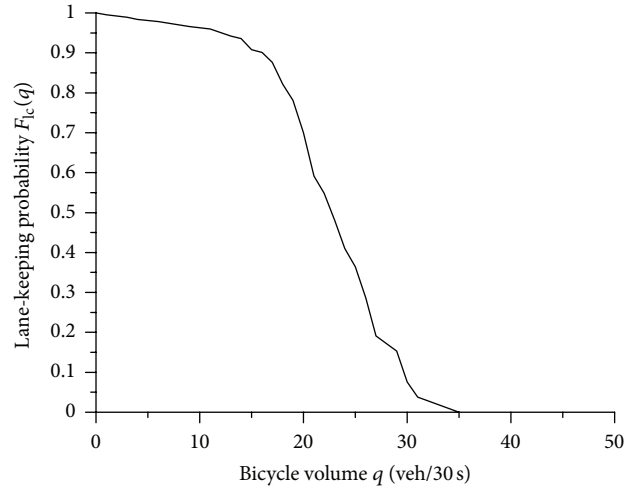


FIGURE 1: Distribution of the lane-keeping probability in average condition.

lane decreases from 3.5 m to 2.0 m, the valid volume will decline from 28 veh/30 s to 20 veh/30 s (the valid probability is 0.8). Assume that the average volume is 15 veh/30 s, the probability that all bicycles travel in the bicycle lane would be 0.6 under the condition of the narrow lane. Meanwhile, the valid volume would decrease by 40%.

(2) Curb parking also shows a positive effect on the hazard while it means that the curb parking can increase the hazard or decrease the probability of lane-keeping behavior. The effect of curb parking on the probability of the lane-keeping behavior is shown in Figure 3. The road sections for a comparative analysis show differences in the distribution of valid volume. Namely, probabilities of the lane-keeping behavior in the sections with curb parking are lower than those without influence of curb parking. It should be noted that the probability that the lane-crossing behavior occurs is low when the volume of the bicycle is low. Meanwhile, the influence of curb parking is insignificant. When the volume of the bicycle is high, the relation between the occurrence of the lane-crossing behavior and curb parking is also insignificant. From the results, the influence of curb parking on the valid probability is significant when the volume of bicycle distributes in median range. Taking the width of 3.0 m as an example, the influence of curb parking on the lane-keeping behavior is significant when the volume ranges between 22 and 32 veh/30 s.

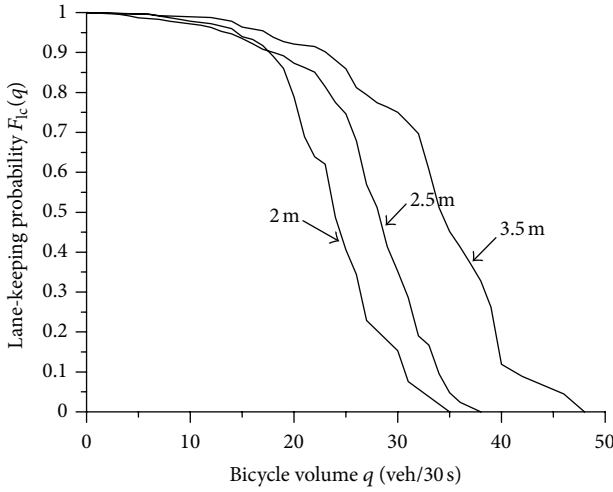


FIGURE 2: Distributions of the lane-keeping probability with various lane widths.

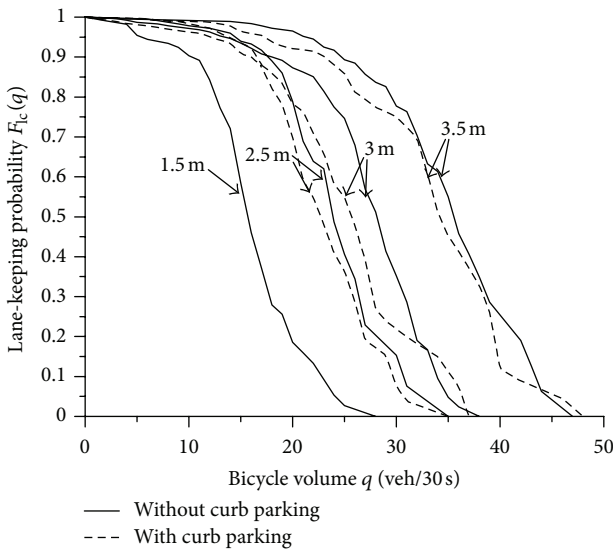


FIGURE 3: Distributions of lane-keeping probability under influence of curb parking.

(3) The travel speed of a bicycle can have a positive effect on the hazard function. The faster the bicycles travel, the higher the possibility they cross the bicycle lane. In this paper, the electric bicycle is considered as bicycle. According to the field survey, there is a certain number of electric bicycles that cross the bicycle lane and travel in the motor lane. The electric bicycles travel faster than other bicycles; thus, they want to seek ideal travel space. Particularly, when there are bicycles travelling in low speed or curb parking car in front of the faster electric bicycles, they change the travel direction and overtake the blockage via the motor lane without the least hesitation.

(4) Retrograde motion has a positive effect on the hazard function, like travel speed of bicycle. As shown in Figure 4,

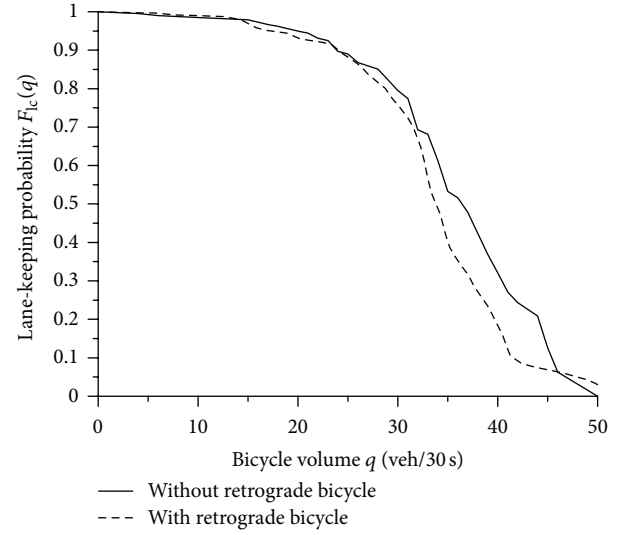


FIGURE 4: Effect of retrograde motion on lane-keeping behavior.

the existence of retrograde bicycle can decrease the probability of lane-keeping behavior. Such effect is more significant in the condition of higher bicycle volume. The retrograde motion of bicycles can hinder the travel routes of other bicycles; it is easy to provide a motivation for changing travel route, even changing the lane.

(5) According to the estimation results, the effect of car volume and safe gap is not significant from the perspective of statistic. However, the car volume still can influence the bicyclist behavior. If the car volume is high, the bicyclist may have little chance to travel in the motor lane. Additionally, the variable of safe gap can also reflect the chance and the safety for a bicyclist to travel in the motor lane. From the field survey and the estimated results, a certain number of bicycles travel in the motor lane when the car volume is very high. These bicyclists neglect the accident risk caused by the lane-crossing behavior. It is dangerous for the bicyclists to travel in the motor lane in heavy car flow. And the lane-crossing behavior can enforce a blockade against the moving car so that the traffic performance reduces obviously.

## 5. Conclusion

This paper proposed a model to describe the lane-keeping behavior of a bicyclist in urban street by using survival analysis. A concept of valid volume is also proposed to describe the relation between the lane-crossing behavior and the bicycle volume. The volume data are defined as censored data and uncensored data. Proportional hazard method is used to estimate the field data with censored data. In order to capture the effect of external factors involving traffic conditions, six variables are selected to construct the PH model. The results show that the effective width of bicycle lane, travel speed, curbs parking, and retrograde motion have significant effect on the lane-keeping behavior. Two variables (car volume and safe gap) show relatively low significance. It is concluded that the lane-keeping behavior results from



various related factors such as personal features, traffic conditions, and environmental factors, and any change of the influential factors can modify the lane-keeping behavior. Therefore, the planning and designing of urban street should consider these influential factors apprehensively.

The future work will focus on the influential factors. More factors will be introduced into the model and the field surveys of sites will be increased to obtain more empirical data. For example, the average speed of bicycle travelling in the car lane could be an important influential factor on the lane-keeping behavior of cyclists. Also the significances of variables and their effects on bicyclist behavior will be discussed deeply.

## Acknowledgments

The authors would like to thank the anonymous editor and referees for their valuable comments. This research is supported by the Programme of Introducing Talents of Discipline to Universities under Grant no. B12022 and the Programme of International Science and Technology Cooperation in the Beijing Institute of Technology.

## References

- [1] W. H. Wang, W. Zhang, H. W. Guo, H. Bubb, and K. Ikeuchi, "A safety-based behavioural approaching model with various driving characteristics," *Transportation Research C*, vol. 19, no. 6, pp. 1202–1214, 2011.
- [2] W. Wang, Y. Mao, J. Jin et al., "Driver's various information process and multi-ruled decision-making mechanism: a fundamental of intelligent driving shaping model," *International Journal of Computational Intelligence Systems*, vol. 4, no. 3, pp. 297–305, 2011.
- [3] R. Jiang, B. Jia, and Q. S. Wu, "Stochastic multi-value cellular automata models for bicycle flow," *Journal of Physics A*, vol. 37, no. 6, pp. 2063–2072, 2004.
- [4] A. Faghri and E. Egyházióvá, "Development of a computer simulation model of mixed motor vehicle and bicycle traffic on an urban road network," *Transportation Research Record*, no. 1674, pp. 86–93, 1999.
- [5] X. M. Zhao, B. Jia, Z. Y. Gao, and R. Jiang, "Traffic interactions between motorized vehicles and nonmotorized vehicles near a bus stop," *Journal of Transportation Engineering*, vol. 135, no. 11, pp. 894–906, 2009.
- [6] X. M. Zhao, B. Jia, Z. Y. Gao, and R. Jiang, "Congestions and spatiotemporal patterns in a cellular automaton model for mixed traffic flow," in *Proceedings of the 4th International Conference on Natural Computation (ICNC '08)*, pp. 425–429, October 2008.
- [7] C. Mallikarjuna and K. R. Rao, "Cellular Automata model for heterogeneous traffic," *Journal of Advanced Transportation*, vol. 43, no. 3, pp. 321–345, 2009.
- [8] T. G. Oketch, "New modeling approach for mixed-traffic streams with nonmotorized vehicles," *Transportation Research Record*, no. 1705, pp. 61–69, 2000.
- [9] X. Yang, Z. Gao, X. Zhao, and B. Si, "Road capacity at bus stops with mixed traffic flow in China," *Transportation Research Record*, no. 2111, pp. 18–23, 2009.
- [10] H. W. Guo, Z. Y. Gao, X. M. Zhao, and X. B. Yang, "Traffic behavior analysis of non-motorized vehicle under influence of curb parking," *Journal of Transportation Systems Engineering and Information Technology*, vol. 11, no. 1, pp. 79–84, 2011.
- [11] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New Jersey, NJ, USA, 2002.
- [12] C. R. Bhat, "Duration modeling," in *Handbook of Transport Modelling*, K. J. Button and D. A. Hensher, Eds., pp. 91–111, Elsevier Science, Amsterdam, The Netherlands, 2000.
- [13] D. A. Hensher and F. L. Mannering, "Hazard-based duration models and their application to transport analysis," *Transport Reviews*, vol. 14, no. 1, pp. 63–82, 1994.
- [14] D. Nam and F. Mannering, "An exploratory hazard-based analysis of highway incident duration," *Transportation Research Part A*, vol. 34, no. 2, pp. 85–102, 2000.
- [15] D. R. Cox, "Regression Models and Life Tables," *Journal of the Royal Statistical Society B*, vol. 34, no. 2, pp. 187–220, 1972.

