Investigating the Affecting Factors of Speed Dispersion for Suburban Arterial Highways in Nanjing, China

Yongfeng Ma,1,2,3 Xin Gu,1,2,3 Jaeyoung Lee,4 and Qiaojun Xiang1,2,3

1 Jiangsu Key Laboratory of Urban ITS, Nanjing, China
2 Jiangsu Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing, China
3 School of Transportation, Southeast University, Nanjing, China
4 School of Traffic and Transportation Engineering, Central South University, Changsha, China

Correspondence should be addressed to Yongfeng Ma; mayf@seu.edu.cn

Received 31 August 2019; Revised 29 October 2019; Accepted 7 November 2019; Published 31 December 2019

Copyright © 2019 Yongfeng Ma 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.

Suburban arterial highways are usually characterized by mixed traffic environments, which is a major contributor to traffic crashes. It has been known that speed dispersion as a surrogate safety measure has a strong correlation with safety. The objective of this study is to identify the influencing factors of speed dispersion for suburban arterial highways. Two definitions of speed dispersion are proposed for comparison: (1) an individual vehicle speed variation along a highway segment and (2) a vehicle speed variation at a cross-section. Vehicle speeds, traffic composition, and driving interference data were obtained from high-resolution videos from 20 segments of the G205 highway in Nanjing, China. An exploratory factor analysis was used to detect initial relationships between latent influencing factors and 13 candidate variables selected based on traffic condition, road condition, and driving behavior. A multivariable regression model was applied to identify the impacts of latent influencing factors on speed dispersion. The results from the two models showed substantial differences. The road condition factor was not significant in the cross-sectional speed dispersion model, but was interpretive in the segmented speed dispersion model. Driving interferences and illegal driving behaviors had a greater effect on the segmented speed dispersion. Consequently, segmented speed dispersion showed a better performance for the analysis of suburban arterial highways. On the other hand, traffic disturbance caused by driving interferences and illegal driving behaviors is the greatest contributor to high speed dispersion on suburban arterial highways, which may be mitigated by effective traffic management measures. It is expected that this work will help traffic managers better understand speed dispersion in mixed traffic environments and to develop effective safety improvement strategies.

1. Introduction

Suburban arterial highways are located in the rural–urban fringe zone and they are often characterized by mixed traffic environment, resulting from a mix of user modes (e.g., pedestrians, motor vehicles, and nonmotor vehicles). Mixed traffic environment is one of the most important reasons that result in frequent traffic crashes of the suburban arterial highways. Consequently, the traffic safety issues of suburban arterial highways have been increasingly emphasized.

Crash-based safety analysis has been widely used to estimate safety situation of highways [1–5]. However, reliable crash data of suburban arterial highways in China are very rare. Surrogate safety measures based on roadway characteristics are often utilized in order to indirectly assess road safety management in case historical crash data are limited or unavailable [6]. Numerous studies have developed speed dispersion as a surrogate measure for traffic safety [6–14]. Chen et al. developed a speed-based model with appropriate speed measure to access roundabout safety performance [10]. Guo et al. proposed two new speed measures to estimate the safety level of freeway exits, in order to overcome the overestimation of safety level resulting from the conventional measure [11]. According to compare with historical crash data, the validation results affirmed the reasonability of the new speed measure to estimate the safety level. Qudus [13] found that speed variation is statistically and positively associated with accident rates. A 1% increase in speed variation is associated with a 0.3% increase in accident rates. On the other hand, some studies have also investigated the relationship between speed...
dispersion and safety, and the speed dispersion is found to be statistically and positively associated with crash rates [15–19]. Garber and Gadiraju [16] concluded that crash rates have a positive relationship with the speed variance; however, have no significant correlation with the average speed. Similar to Garber and Gadiraju [16], Liu and Popoff [17] also found that the most prevalent source of human error contributing to collisions may be speed-related and will increase with travel speed. According to previous studies, speed dispersion is an acceptable surrogate measure to conduct safety analysis. A comprehensive study on identifying factors affecting speed dispersion is beneficial for helping authorities to develop relevant effective countermeasures to reduce speed dispersion, which could lead to a reduction in crash rates.

Although previous studies have explored the probability of crash occurrence by speed dispersion and developed speed dispersion models [20–24], they still have some limitations. Most of the models focused on rural or urban expressways; however, no study has focused on suburban arterial highways until now. Suburban arterial highways are common in developing countries, especially in China, and they have to meet the complex traffic demands of both highways and urban roads. The suburban arterial highways have three main characteristics: (1) typical mixed traffic environment owing to complex traffic composition, especially after the popularity of electric vehicles [25, 26]; (2) high-density accesses directly linked to the main road, owing to land commercialization; (3) numerous traffic interferences resulting from illegal traffic behaviors. Consequently, suburban arterial highways have a more complex traffic environment, and previous speed dispersion models, therefore, might not be appropriate.

More importantly, the definition of speed dispersion varies by data resource and research method. Speed dispersion can be expressed by the speed variation at a certain place in a certain time interval, or the speed variation along a road section. The proposed approaches to define speed dispersion in previous studies are briefly summarized in Table 1. Most conventional definitions are based on cross-sectional aggregated data, and thus may not be the most appropriate for suburban arterial highways, owing to a lack of consideration for all traffic compositions in the mixed traffic environment and the impacts of traffic interference along the segment, in addition to the probability of ecological fallacy [11]. Ecological fallacy describes the phenomenon in which what seems true for a group may not be true for the individual, because some information gets lost during the aggregation process [11].

Along with the definition of speed dispersion, some studies have also been conducted to explore the influencing factors of speed dispersion [12, 27–29]. Moreno and García [12] proposed two speed dispersions, average individual speed and accumulated speed uniformity. And it concluded that speed limit, and traffic calming density are two factors influencing the average individual speed, but the average operating speed is the only one influencing factor of accumulated speed uniformity. Park et al. [27] developed hierarchical linear regression models to predict the speed dispersion. The results indicated that the speed dispersion is positively related to tangent speed and curvature. Chung and Recker [28] investigated the relationship between speed dispersion and traffic flow parameters. However, the methodological approaches used in previous studies rarely considered the complex relationships between the influencing factors, and thus may not produce precise estimation of the effects of various candidate variables on speed dispersion. For example, Moreno and García [12] explored speed distribution based on continuous GPS-speed data on cross-town roads in the European areas, and developed a speed variation model. However, the data of zero-speed vehicles or parking cars are removed, which could result in a limitation of influencing factors.

Hence, the current study aims to contribute to the literature by identifying more influencing factors of speed dispersion for suburban arterial highways. To this end, exploratory factor analysis (EFA) was used to explore the relationship between candidate variables. Then, a multivariable regression model is developed to identify the correlation between speed dispersion and influencing factors. Note that, for comparison purpose, two different definitions of speed dispersion are proposed to develop regression models. These methods could help us attain a more precise and comprehensive understanding of the impacts of different variables on speed dispersion on suburban arterial highways. It is expected that this work will help traffic managers better understand speed dispersion in mixed traffic environments and to develop effective safety improvement strategies.

2. Methodology

2.1. Definition of Speed Dispersion. An individual vehicle speed is the result of the combined effects from road and traffic conditions, driving behaviors, vehicle performance, and weather, and thus, its variation may represent the characteristics of speed dispersion. Boonsiripant et al. [6] studied the relationship between speed distribution and safety, and his results showed that acceleration noise as a surrogate safety measure had better correlation with accidents than aggregated speed-based surrogate safety measures. This study could provide supporting evidence because acceleration has a strong relationship with individual vehicle speed variation. Therefore, speed dispersion is defined by the individual speed variation along a road segment, which not only provides a more comprehensive picture of the characteristics of a road segment during a particular period of time, but also avoids the ecological fallacy [11] owing to information loss during the aggregation process. The expression of speed dispersion is as follows:

$$SP_{\text{segment}} = \sqrt{\frac{\sum_{m=1}^{m} (v_{i,j} - \bar{v}_j)^2}{m-1}}, \quad (1)$$

$$v_{i,j} = \frac{(1/2)i_{i-1,j} + (1/2)i_{i,j}}{t_{i,j}}, \quad (2)$$

$$\bar{v}_j = \frac{l_j}{t_j}, \quad (3)$$
### Table 1: Summary of speed dispersion definitions in previous studies.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Method</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solomon [18]</td>
<td>1964</td>
<td>Difference between individual vehicle speed and mean speed</td>
<td>Speed data along a segment</td>
</tr>
<tr>
<td>Garber &amp; Gadiraju [16]</td>
<td>1989</td>
<td>Speed variance</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Aarts and Van Schagen [20]</td>
<td>1989</td>
<td>Difference between 85th percentile speed and the average speed</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>McFadden and Elefteriadou [22]</td>
<td>2000</td>
<td>The 85th percentile maximum speed reduction</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Wang et al. [30]</td>
<td>2007</td>
<td>The standard deviation of the individual speed (SDIS); The average speed difference (ASD) of two neighbouring vehicles</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Guo et al. [11]</td>
<td>2011</td>
<td>The 85th percentile individual speed difference; The 85th percentile speed change rate</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Boonsiripant et al. [6]</td>
<td>2011</td>
<td>The variation of the 85th percentile speed; Mean of 85th percentile speed; Interquartile range of 85th percentile speed; The difference between the 95th percentile speed and the 5th percentile speed (speed band); Mean of speed band; Variation of speed band</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Moreno and García [12]</td>
<td>2012</td>
<td>Ra-relative area measure of uniformity; Ea-relative area measure of speeding</td>
<td>Continuous speed data along a segment</td>
</tr>
<tr>
<td>Chung and Recker [28]</td>
<td>2014</td>
<td>The coefficient of variation of speed</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Wang et al. [31]</td>
<td>2016</td>
<td>Standard speed deviation (SD)</td>
<td>Cross-sectional data</td>
</tr>
<tr>
<td>Qiao et al. [32]</td>
<td>2017</td>
<td>The average of the speed deviation in the statistical time; The difference between the 85% speed and the 15% speed; The ratio of the SD to the average speed of the traffic flow</td>
<td>Continuous speed data in the statistical time</td>
</tr>
<tr>
<td>Gao et al. [43]</td>
<td>2019</td>
<td>The speed standard deviation, the coefficient of variation (CV)</td>
<td>Cross-sectional data</td>
</tr>
</tbody>
</table>
where, $SP_{segment}$ is the speed dispersion based on individual vehicle data along a highway segment $p$, $v_{i,j}$ is the speed of vehicle $j$ at cross-section $i$, $l_{i−1,j}$ is the distance between cross-section $i − 1$ and cross-section $i$; $(1/2)l_{i−1,j} = 5$ m, if $i = 0$; $l_{i}$ is the distance between cross-section $i$ and cross-section $i + 1$, $t_{i,j}$ is the travel time of vehicle from location $(1/2)l_{i−1,j}$ to location $(1/2)l_{i+1,j}$, $\bar{v}$ is the mean speed of vehicle $j$ on the road segment $p$, $l_{p}$ is the length of road segment $p$, and $m$ is the number of cross-sections in the highway segment $p$.

For comparison, the conventional speed dispersion based on aggregated data at certain cross-sections was also considered. The expression is as follows:

$$SP_{cross-section} = \sqrt{\frac{\sum_{i=1}^{n} (v_i - \bar{v})^2}{n - 1}}.$$  

where, $SP_{cross-section}$ is the speed dispersion based on cross-sectional data, $v_i$ is the speed of vehicle $i$ at the observed cross-section, $\bar{v}$ is the mean speed of vehicles through the observed cross-section in specific time interval, and $n$ is the number of vehicles in the observed time interval.

2.2. Model Development. This study aims to propose a new analysis procedure for speed dispersion on suburban arterial highways. The overall modelling procedure that would achieve the objectives as follows:

1. Factor analysis to identify latent influencing factors: EFA is utilized to detect initial relationships between the latent influencing factors and the selected candidate variables in the study. This process can help the model development, by extracting a small number of latent factors to represent the large number of selected candidate variables with minimal loss of information.

2. Multivariable regression modelling to explore the impact of latent factors on speed dispersion: This step employs the multivariable regression method to develop a speed dispersion model. The latent factors are used as independent variables to estimate the relationships with speed dispersion. The effects of the selected candidate variables on speed dispersion can also be expressed.

3. Model comparison between $SP_{segment}$ and $SP_{cross-section}$: This step is conducted to explore the effects of different definitions on the analysis results of speed dispersion.

3. A Case Study

3.1. Site Selection. Suburban arterial highways with characteristics of typical mixed traffic environments were selected. This paper focused on the influencing factors of speed dispersion under continuous conditions, in order to ensure the reliability of the study, all research road segments need to keep away from the signalized intersections to avoid interruption. In addition, the alignment should avoid the curve and the longitudinal slope. Finally, 20 segments on ten different sites of the G205 highway in Nanjing, which crosses a rural–urban fringe area, were selected (Figure 1). The average number of cross-sections is 10.7 for each segment. The cross-section was selected at intervals of 20 m from the beginning of each segment. Note that when the detection point falls on the crosswalk or the access to surroundings etc., it will extend forward to 5 m after the end of the above facilities. The suburban arterial highway characteristics are summarized in Table 2.

3.2. Data Collection. Many collection methods have been applied to research speed and safety; for instance, earlier studies [38–41] have used GPS-based collection [11], loop detector-based collection [33, 34], video-based speed collection [9, 35], and video detection-based speed collection [36, 37]. To meet the research aims, a video-based collection method was adopted in this study, which can provide full and proper context for highway segments. High-resolution videos of suburban arterial highway segments were collected by unmanned aerial vehicle (UAV) (Figure 1). UAV photography allowed to shoot videos in the necessary surroundings, including the wind force below grade 4, sunny weather, good light, and no electromagnetic interference, which could output stable video pictures at a vertical angle. When processing the video data, the length of lane line was considered as a reference to calibrate the distance, so the error caused by the projection of the videos could be ignored. After the multiple tests, the final shooting location was determined to be at a height of approximately 200 m to ensure the ability to obtain high-resolution videos of highway segments with length of at least 350 m in good weather.

Video collection was conducted between 4 p.m. and 6 p.m. on weekdays except Friday, which also met UAV photography demands. The UAV could hover for 20 minutes at maximum and to obtain the stable data, a 15-minute useful highway segment video which takes out the periods of rise and fall are extracted each time. Finally, a total of ten-hour video data, 1 hour for a pair of highway segment, were collected in the survey.

As previously mentioned, earlier studies [38–41] have found that the speed dispersion was affected by the road conditions (e.g. median, roadside environment), traffic conditions (e.g. traffic volume, heavy vehicle mix rate), and road users’ behavior (e.g. cross and turn around rate, familiar or not familiar road users). In order to explore the influencing factors of speed dispersion for suburban arterial highways, key variables including road condition, traffic condition, and driving interference variables were selected considering the difficulty of data acquisition (Table 3). The road condition variables, which described the road situation, were collected directly from the field site. The traffic condition variables describing the traffic flow state and driving interference showing the turbulence into the vehicles along the segment could be observed from the videos. Vehicle speeds were computed with the distance and the through time between adjacent cross-sections by video processing with VideoStudio. Instead of selecting cross-sections at a fixed interval distance, the selected cross-sections (Figure 1) are determined by the location of accesses, to avoid the overlap between the cross-section and the location of the access. In addition, the data from sections that are totally covered by trees or erroneous data are removed.
Figure 1: Scheme of shooting location and data collection.

Table 2: Segments characteristics.

<table>
<thead>
<tr>
<th>Segments</th>
<th>Number of lanes</th>
<th>Median</th>
<th>Southbound/Northbound Land-use</th>
<th>Number of access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xishan bridge—DiKan Xinyuan</td>
<td>4</td>
<td>No</td>
<td>Commercial, residential/agriculture</td>
<td>4/0</td>
</tr>
<tr>
<td>Dikan—Jianning</td>
<td>4</td>
<td>No</td>
<td>Commercial, residential/agriculture</td>
<td>4/0</td>
</tr>
<tr>
<td>Jianning—Daishan interchange</td>
<td>4</td>
<td>No</td>
<td>Industrial, commercial/agriculture</td>
<td>6/0</td>
</tr>
<tr>
<td>Daishan interchange—Tianbao</td>
<td>4</td>
<td>No</td>
<td>Industrial, office, commercial/agriculture</td>
<td>7/0</td>
</tr>
<tr>
<td>Meishan hospital—Yongan</td>
<td>4</td>
<td>Yes</td>
<td>Commercial, office, residential/commercial, residential</td>
<td>2/3</td>
</tr>
<tr>
<td>Meishan iron—Qiaogao road</td>
<td>4</td>
<td>Yes</td>
<td>Agriculture, residential/industrial, residential</td>
<td>0/3</td>
</tr>
<tr>
<td>Qiaogao road—Dingxin road</td>
<td>4</td>
<td>Yes</td>
<td>Agriculture, residential/agriculture, residential</td>
<td>1/1</td>
</tr>
<tr>
<td>S337—Yefu road</td>
<td>4</td>
<td>Yes</td>
<td>Commercial, residential/industrial, office, residential</td>
<td>5/4</td>
</tr>
<tr>
<td>Yefu road—Jingtong</td>
<td>4</td>
<td>Yes</td>
<td>Commercial, residential/agriculture, residential, office</td>
<td>5/2</td>
</tr>
<tr>
<td>Shuguang road—Xingfu road</td>
<td>6</td>
<td>Yes</td>
<td>Industrial, office, residential/commercial, office, residential</td>
<td>1/2</td>
</tr>
</tbody>
</table>
4. Results and Analysis

4.1. Exploratory Factor Analysis (EFA). Correlation analysis was first conducted to explore the relationship between the 13 candidate variables. The results show that 75% of the correlation coefficients between variables are greater than 0.3. To avoid information loss due to clearing variables with multicollinearity and explore the interrelationship, the EFA was conducted to explore a small number of latent factors that could represent the observed candidate variables. The number of latent factors needed to describe the unique information in the observed candidate variables is classically determined based purely on the sufficiency of explanation of relationships among the candidate variables (i.e., the percentage of variance in the original variables accounted for by the set of factors selected).

EFA includes the following steps: (1) The Kaiser–Meyer–Olkin and Bartlett sphere test were conducted in SPSS to test if EFA was appropriate for the input data. The test result is between 0 and 1, and a high result indicates the EFA does well for the sample. In this paper, a sampling degree of 0.752 and 0.754 indicated the feasibility; (2) the number of the extracted latent factors was determined by Cattell’s Scree Plot method, only factors that have the entire confidence interval greater than 1.0 were retained. Four latent influencing factors were retained in this paper and they were found to account for 72% of the variance in the candidate variables. An indicator, the communality of the variable, could show the variance percentage of each of the original 13 variables explained by the four factors together (the $R^2$ of the regression of each of the 13 variables on all four factors). For the retained four latent factors, ten original variables’ communalities were in excess of 0.70, with the exception of “Illegal parking (parking on the road or roadside illegally) rate”, “Roadside environment”, “Cross and turn around rate”, which were 0.631, 0.653, and 0.646. (3) then the varimax rotation (i.e., a rotation that maximizes the sum of the variances of the factors loadings) was conducted to facilitate interpreting the extracted latent traffic variables. The factor loadings which express the correlations between the 13 candidate variables and the latent influencing factors are shown in Table 4. Scores on each of the eight rotated factors were calculated for each combination (weighted average) of the 13 candidate variables. The factor loadings of the original variables on the rotated factors (i.e., the correlations between the variables and factors) form the basis for the interpretation of the factors. To aid in interpretation, the original candidate variables with large factor loadings were then selected to represent each latent influencing factor, which are shown in boldface in Table 4. The most indicative parameter for each factor is identified by a box around its loadings [42]. According to the bold variables, each of the three types of candidate parameters (i.e., road condition, traffic condition, and driving interference) provides some unique information for these four indicative parameters.

4.1.1. EFA of $SP_{\text{segment}}$. Latent factor 1 is represented by the right and left turn rates at access, and the cross and turn around rate. This latent variable can be interpreted as the driving interference on the highway section, which can be denoted by the driving interference factor $F_{\text{interference}}$.

Latent factor 2 can be interpreted as the illegal driving behavior on the highway segment, which can be denoted by the illegal driving behavior factor $F_{\text{drive}}$. The selected three candidate variables with great factor loadings are the reverse driving rate, illegal parking rate, and nonmotor vehicle volume driving on the motor road.

Latent factor 3 is represented by the heavy vehicle mix rate, traffic volume, average speed, and the average lane changing rate. This latent variable can be interpreted as the traffic condition on the highway segment, for instance, traffic operation and traffic composition, which can be denoted by the traffic condition factor $F_{\text{traffic}}$.

Latent factor 4 can be interpreted as the road condition on the highway section, which can be denoted by the road condition factor $F_{\text{road}}$. The selected candidate variables with

---

**Table 3: Description of the candidate variables considered in this study.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Variables</th>
<th>$SP_{\text{segment}}$</th>
<th>$SP_{\text{cross-section}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Min</td>
</tr>
<tr>
<td>Road condition</td>
<td>Median (0 = if there is a median; 1 = otherwise)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Roadside environment (1 = agriculture; 2 = industrial, residential, office; 3 = commercial)</td>
<td>Percentage</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Heavy vehicle mix rate (%)</td>
<td>0.2</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Traffic volume (veh/h)</td>
<td>1336</td>
<td>432</td>
</tr>
<tr>
<td>Traffic condition</td>
<td>Average speed (km/h)</td>
<td>58.3</td>
<td>33.4</td>
</tr>
<tr>
<td></td>
<td>Headway (s)</td>
<td>30.1</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>Average lane changing rate (/veh)</td>
<td>1.4</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Right turn rate at access (/h)</td>
<td>982</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Left turn rate at access (/h)</td>
<td>582</td>
<td>0</td>
</tr>
<tr>
<td>Driving interference</td>
<td>Cross and turn around rate (/h)</td>
<td>848</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Reverse driving (driving in the wrong direction) rate (/h)</td>
<td>230</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Illegal parking (parking on the road or road side illegally) rate (/h)</td>
<td>835</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Nonmotor vehicle volume driving on the motor road (veh/h)</td>
<td>1272</td>
<td>44</td>
</tr>
</tbody>
</table>
large factor loadings are the median and roadside environment.

4.1.2. EFA of \( SP_{\text{cross-section}} \)  

Latent factor 1 is represented by the heavy vehicle mix rate, traffic volume, average speed, headway, and nonmotor vehicle volume driving on the motor road. This latent factor can be interpreted as the driving interference factor at the highway cross-section, which is similar with \( F_{\text{traffic}} \), and can be denoted by the traffic condition factor \( F_{\text{traffic}} \).

**Latent factor 2** can be interpreted as the driving interference at the highway cross-section, which can be denoted by the driving interference factor \( F_{\text{interference}} \). The selected three candidate variables with great factor loadings are the right and left turn rates at access, and the cross and turn around rate.

**Latent factor 3** is represented by the average lane changing rate, reverse driving rate, and illegal parking rate. This latent variable can be interpreted as the illegal driving behavior at the highway cross-section, which can be denoted by the illegal driving behavior factor \( F_{\text{r-drive}} \).

**Latent factor 4** can be interpreted as the road condition at the highway cross-section, which can be denoted by the road condition factor \( F_{\text{road}} \). The selected candidate variables with large factor loadings are the median and roadside environment.

4.2. Speed Dispersion Modelling. To explore the impact of latent factors on different speed dispersions, two multivariable regression models were developed. Thus, the effects of the selected candidate variables on speed dispersion can be also identified.

4.2.1. Model 1-for Segment Data-Based Speed Dispersion \( SP_{\text{segment}} \)  

A multivariable regression analysis was conducted to explore the effects of the latent factors on the speed dispersion on a suburban arterial highway segment. The regression result is as follows:

\[
SP_{\text{segment}} = 0.292 F_{\text{interference}} + 0.171 F_{\text{r-drive}} + 0.157 F_{\text{traffic}} + 0.123 F_{\text{road}},
\]  

where \( SP_{\text{segment}} \) is the speed dispersion based on individual vehicle data along a highway segment, \( F_{\text{interference}} \) is the driving interference factor, \( F_{\text{r-drive}} \) is the illegal driving behavior factor, \( F_{\text{traffic}} \) is the traffic condition factor, and \( F_{\text{road}} \) is the road condition factor, which could be calculated by candidate variables and EFA results.

\[
R^2 = 0.598
\]

which indicates that the latent factors well explain the speed dispersion.

According to the model results, the most important latent factor is \( F_{\text{interference}} \) which is positively related to the speed dispersion. \( F_{\text{interference}} \) is determined by “the right and left turn rates at access” and “the cross and turn around rate” with a positive relationship. Therefore, the speed dispersion will rise along with the increasing driving interference rate. This reflects the adverse impact of traffic disorder on the speed dispersion. The more vehicles turn into the surrounding areas or taking U-turns, the speed of the straight-through vehicles will be affected. On the contrary, the speed dispersion may effectively decrease if traffic management measures are taken, for instance, implementing “no left turn” and “no U-turn” controls.

The second most important factor is the illegal driving behavior \( F_{\text{r-drive}} \) which also has a positive relationship with the speed dispersion. The illegal driving behavior factor highly correlates with the rates of reverse driving, illegal parking, and nonmotor vehicle volume driving on the motor road. Reverse driving and illegal parking reduce the space that the vehicles can travel which have significant effects on the service level of the road. Moreover, these phenomena will also lead to the deceleration of the upstream fast-moving vehicles. Since the speed of nonmotor vehicle is small, higher volume of nonmotor vehicle tend to affect the speed of the vehicles. These candidate variables and \( F_{\text{r-drive}} \) belong to the traffic management category, which indicates that the reasons contributing to large speed dispersion on suburban arterial highways are a lack of traffic and safety consciousness and ineffective traffic management measures.

The other two influencing factors are the traffic condition \( F_{\text{traffic}} \) and the road condition \( F_{\text{road}} \), which have a positive...
correlation with the speed dispersion. $F_{traffic}$ is positively correlated with the average lane changing rate, heavy vehicle mix rate, and vehicle count, but negatively correlated with average speed. The reason for this is that a high average speed indicates good traffic conditions with less transverse interference, and in which the speed dispersion is low. This is consistent with the actual situation. $F_{road}$ mainly includes whether to set the inner median and the roadside environment. Setting the inner median can effectively reduce the occurrence of illegal U-turn and crossing behavior, and reduce the interference to the traffic flow, thus reducing the speed dispersion. Moreover, with the development of land use and the increase in the number of accesses, the interference will increase, which will intensify the speed dispersion.

4.2.2. Model 2 for Cross-Section Data-Based Speed Dispersion $SP_{cross-section}$. For comparison, the influence analysis for cross-sectional speed dispersion definition $SP_{cross-section}$ was also conducted. The expression for $SP_{cross-section}$ is as follows:

$$SP_{cross-section} = 0.432 F_{traffic}^i + 0.185 F_{interference}^i + 0.114 F_{drive}^i + 0.029 F_{road}^i$$

where $SP_{cross-section}$ is the speed dispersion based on i cross-sectional data, $F_{traffic}$ is the traffic condition factor, $F_{interference}$ is the driving interference factor, $F_{drive}$ is the illegal driving behavior factor, and $F_{road}$ is the road condition factor, which could be calculated by candidate variables and EFA results.

$R^2 = 0.526$ indicates that the latent factors interpret the speed dispersion well.

The latent factors can also be interpreted by the traffic condition, driving interference, illegal driving behavior, and road condition factors. As shown in Equation (4), the factor that have the greatest effect on $SP_{cross-section}$ is the traffic condition factor, which is positively correlated with $SP_{cross-section}$ and is mainly determined by the heavy vehicle mix rate, vehicle volume, average speed, headway, and nonmotor vehicle count driving on the motor road. Two other factors that affect $SP_{cross-section}$ are the driving interference and illegal driving behavior factors. However, the influence of the road condition factor on $SP_{cross-section}$ is insignificant.

4.2.3. Model Validation. A test dataset was used to calibrate the regression model. The observed speed dispersion was recorded as Field speed dispersion. The predicted value of speed dispersion was recorded as predicted speed dispersion. The numerical analysis between Field speed dispersion and predicted speed dispersion is shown in Table 5 and Figure 2. According to the results, the segment data-based speed dispersion model outperforms the model for cross-section data-based speed dispersion.

4.2.4. Model Comparison. The influences of the latent factors on the speed dispersion are different. For instance, the impact of the driving interference factor was the greatest in the $SP_{segment}$ model, but not in the $SP_{cross-section}$ model, in which the traffic condition factor had the greatest effect. The cross-sectional speed dispersion is affected by the vehicle speed distribution at a certain section through a certain time interval, which is determined by the number and types of vehicles through the cross-section. However, the segment-based speed dispersion analyzed the speed variation along a stretch of highway segment, which contains many factors, not only traffic condition, but also the road condition and driving interference. Therefore, the traffic condition factor plays a more important role in the $SP_{cross-section}$ model. Furthermore, the influence of the road condition factor in the $SP_{cross-section}$ model is not significant, in contrast to the $SP_{segment}$ model. This is because the segment-based data in the $SP_{segment}$ model can provide a comprehensive picture of the whole highway segment, which accumulates the effects of the road conditions.

Some of the candidate variables explained by the latent factors are different. For instance, nonmotor vehicle volume driving on the motor road, this variable is explained by the illegal driving behavior factor $F_{drive}$ in the segment-based factor analysis, but by the traffic condition factor $F_{road}$ in the cross-sectional factor analysis. This is because the variable has both traffic characteristics and driving characteristics. This behavior causes a decrease in the average speed at the observed cross-section in the cross-sectional analysis, which also contributes to a reduction in the speed dispersion. However, the same behaviors may disturb the adjacent motor vehicle and cause speed reduction and more lane changing behaviors, which are the main reasons for speed dispersion. Therefore, the variable nonmotor vehicle volume driving on the motor road shows more traffic characteristics in the $SP_{cross-section}$ model, but more driving characteristics in the $SP_{segment}$ model.

Similarly, the average lane changing rate has a high correlation with the traffic condition factor $F_{road}$ in the segment-based analysis, but more explained by the illegal driving behavior factor $F_{drive}$ in the cross-sectional analysis. Lane changing behaviors include free lane changing and compulsory lane changing. Lane changing behaviors observed at a certain cross-section are random, which cause a random characteristic in the cross-sectional analysis. However, free lane changing behaviors and compulsory lane changing behaviors are all observed and analyzed in the segment-based analysis, and thus the characteristics of the traffic condition factor are more obvious.

5. Conclusions

Suburban arterial highways are common in developing countries and have typical mixed traffic characteristics. For example, a mix of user modes, numerous direct-linking accesses to

<table>
<thead>
<tr>
<th>Validation Dataset</th>
<th>Max</th>
<th>Min</th>
<th>Mean</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field $SP_{segment}$</td>
<td>7.206</td>
<td>4.183</td>
<td>5.541</td>
<td>0.998</td>
<td>15.6%</td>
</tr>
<tr>
<td>Predicted $SP_{segment}$</td>
<td>8.529</td>
<td>3.372</td>
<td>5.810</td>
<td>0.998</td>
<td>15.6%</td>
</tr>
<tr>
<td>Field $SP_{cross-section}$</td>
<td>24.021</td>
<td>5.999</td>
<td>11.638</td>
<td>0.998</td>
<td>15.6%</td>
</tr>
<tr>
<td>Predicted $SP_{cross-section}$</td>
<td>23.742</td>
<td>5.838</td>
<td>12.085</td>
<td>0.998</td>
<td>15.6%</td>
</tr>
</tbody>
</table>

Note: RMSE denotes root mean square error; MAPE denotes absolute percentage error.
arterial highways, and disordered traffic resulting from illegal traffic behaviors, will cause an increased potential for traffic collisions. The safety issues of suburban arterial highways require more attention. This study sought to identify the influencing factors of speed dispersion for suburban arterial highways, which is a surrogate measure for safety risks. Two definitions of speed dispersion were presented and 20 different segments of the G205 highway in Nanjing were used for case study. In order to avoid the information loss, an exploratory factor analysis (EFA) was conducted to explore the relationship between the candidate variables. Finally, two multivariable regression models were developed to explore the impact of factors on different speed dispersions.

For exploring the influencing factors, thirteen variables which were selected from three different aspects, e.g. the traffic conditions, road conditions, and driving interference, were fully considered. The EFA results showed that the two different definitions of speed dispersion, which were calculated by individual vehicle data and aggregated data respectively would lead to the principal components results with different compositions and ranks. Another important finding is that the impacts of the same variable on \( SP_{\text{segment}} \) and \( SP_{\text{cross-section}} \) are different. For example, \( SP_{\text{cross-section}} \) is significantly affected by the traffic condition factor; however, for \( SP_{\text{segment}} \), the driving interference factor plays the most important role. One possible reason is that the cross-sectional vehicle data for the \( SP_{\text{cross-section}} \) model are significantly affected by the traffic volume and traffic composition passing through the cross-section. However, the individual speed variation along a highway segment, which was used for modelling the \( SP_{\text{segment}} \) model, shows comprehensive effects of all factors and reduces the impact of cross-sectional vehicle data on the \( SP_{\text{segment}} \). This is also consistent with the actual situation.

In addition, the model results showed that the speed dispersion for suburban arterial highways will increase under one of the following conditions: (1) there are many accesses along the arterial highways; (2) there is no inner median; (3) nonmotor vehicle volume is high; (4) there are many traffic violations, for example, illegal parking. This indicates that the speed dispersion of the suburban arterial highway is mainly caused by the lack of safety consciousness of traffic participants and the neglect of road management. Therefore, it is essential that safety countermeasures aimed at reducing speed dispersion should be promoted, which is also beneficial to road safety. One of such countermeasures would be to set the inner median which could reduce the reverse driving behavior and the cross and turn around rate, thus reducing the interference on the traffic stream. Another way to reduce the speed dispersion is setting roadside parking lots or strengthen the management of parking on the road in order to decrease the impact of parking cars. Moreover, the management of accesses and U-turn would also reduce speed dispersion.

This work could enrich the knowledge of speed dispersion on suburban arterial highways and can be applied to transportation networks in other countries. Moreover, further development of video-based moving target detection technology will be beneficial to obtain more accurate speed trajectory of traffic participants and calculate the speed dispersion, thereby avoiding the influence of random error due to cross-section selection depending on the land use and the layout of accesses. In addition, we will collect crash data of suburban arterial highways, then explore the relationship between speed dispersion and crashes, and identify the significant influencing factors to traffic crashes.

**Figure 2:** Comparison of field and predicted values of speed dispersion \( SP_{\text{segment}} \) and \( SP_{\text{cross-section}} \)

**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 regarding the publication of this paper.
Funding

This study was funded by the National Natural Science Foundation of China (No. 51208100, No. 71871059).

Acknowledgments

The authors acknowledge the assistance provided by the graduate research assistants at the School of Transportation, Southeast University, in field data collection.

References

[29] J. Li, W. H. K. Lam, and X. Li, "Modeling the effects of rainfall intensity on the heteroscedastic traffic speed dispersion on..."


