Speed, speed variation and crash relationships for urban arterials

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: WANG, X. ... et al, 2018. Speed, speed variation and crash relationships for urban arterials. Accident Analysis and Prevention, 113, pp.236-243.

Additional Information:

- This paper was published in the journal Accident Analysis and Prevention and the definitive published version is available at https://doi.org/10.1016/j.aap.2018.01.032.

Metadata Record: https://dspace.lboro.ac.uk/2134/33317

Version: Accepted for publication

Publisher: © Elsevier

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Speed, Speed Variation and Crash Relationships for Urban Arterials

Xuesong Wang\(^{a}\), Qingya Zhou\(^{a}\), Mohammed Quddus\(^{b}\), Tianxiang Fan\(^{a}\), Shou’en Fang\(^{a}\)

\(^{a}\)School of Transportation Engineering, Tongji University, Shanghai, 201804, China
\(^{b}\)Transport Studies Group, School of Architecture, Building and Civil Engineering, Loughborough University, Leicestershire, LE11 3TU, UK

*Corresponding author. Tel: +862169583946. E-mail address: wangxs@tongji.edu.cn.

ABSTRACT

Speed and speed variation are closely associated with traffic safety. There is, however, a dearth of research on this subject for the case of urban arterials in general, and in the context of developing nations. In downtown Shanghai, the traffic conditions in each direction are very different by time of day, and speed characteristics during peak hours are also greatly different from those during off-peak hours. Considering that traffic demand changes with time and in different directions, arterials in this study were divided into one-way segments by the direction of flow, and time of day was differentiated and controlled for. In terms of data collection, traditional fixed-based methods have been widely used in previous studies, but they fail to capture the spatio-temporal distributions of speed along a road. A new approach is introduced to estimate speed variation by integrating spatio-temporal speed fluctuation of a single vehicle with speed differences between vehicles using taxi-based high frequency GPS data. With this approach, this paper aims to comprehensively establish a relationship between mean speed, speed variation and traffic crashes for the purpose of formulating effective speed management measures, specifically using an urban dataset. From a total of 234 one-way road segments from eight arterials in Shanghai, mean speed, speed variation, geometric design features, traffic volume, and crash data were collected. Because the safety effects of mean speed and speed variation may vary at different segment lengths, arterials with similar signal spacing density were grouped together. To account for potential correlations among these segments, a hierarchical Poisson log-normal model with random effects was developed. Results show that a 1% increase in mean speed on urban arterials was associated with a 0.7% increase in total crashes, and larger speed variation was also associated with increased crash frequency.

Keywords: Urban arterial; Mean speed; Speed variation; Crash frequency; GPS Data; Hierarchical Poisson log normal models with random effects
1. Introduction

Most traffic on urban road networks is carried by their arterials. These roads are, however, the least safe. In downtown Shanghai, urban arterials have several unusual characteristics including uneven and short signal spacing, and the absence of posted speed limits. Operational conditions are also complex on those arterials due to a large variation in speed by time of day; for example, speed might be lower during peak hours. Both higher and lower mean speeds have been found to be associated with increased crash frequency, and large speed variation has consistently been found to increase crash counts (Garber and Gadiraju, 1989; Solomon, 1964; Oh et al., 2001). Although speed management measures are generally implemented on suburban arterials and highways in Shanghai, speeding is currently not strictly enforced on urban arterials. Quantification of the relationships of mean speed and speed variation to crash frequency can be a guideline for authorities to improve traffic safety on urban arterials (Wang et al., 2015).

Speed data employed in developing speed-crash relationships have typically been collected with fixed-based sensors such as radar and loop detectors (Figueroa and Tarko, 2005; Poe et al., 1998). These methods only capture speeds at fixed points, failing to represent spatio-temporal distributions of speed along a road. Spatio-temporal speed data can be collected by probe vehicles, a method known as floating car data (FCD), in which vehicles (e.g., taxis) are equipped with global positioning systems (GPS) that can continuously record vehicle position, speed and direction at a high sampling frequency (normally 1 Hz). Speed measurements from sample vehicles are then employed to calculate both mean speed and speed variation by road segment at given intervals (Wang et al., 2014; Xie et al., 2013; Wang et al., 2016). There is a dearth of research, however, on the use of spatio-temporal GPS speed measurements in developing crash-speed relationships while controlling for other mediating factors (e.g., traffic volume). In addition, most of the previous studies on crash-speed relationships have been developed using data either from freeways or rural roads.

A recent study by Wang et al. investigated how road geometry and other traffic-related factors affect traffic safety in Shanghai (Wang et al., 2015). Their results showed that higher average speed was associated with higher crash frequency. However, there are some limitations in their study and others, which are addressed in this paper. First, although speed variation is widely reported to have an effect on safety, and existing studies indicate that the inclusion of speed variation in a model has a significant influence on the impact of speed on crashes (Quddus, 2013; Gargoum and El-Basyouny, 2016), speed variation was not considered by Wang et al. (2015). Second, it is apparent that traffic conditions on a road segment vary by the direction of traffic flow for a given time of day. Most existing studies aggregate traffic conditions (e.g., speed, volume, geometry) from both sides into a single measurement/observation and assume the crash generation process is the same in both directions at a given time period (Wang et al., 2015; Wang et al., 2016). Third, traffic conditions also vary by time of day. For example, the mean speed during peak hours tends to be lower than during off-peak hours because of congestion. Therefore, in modeling, the relationship between speed, speed variation and crashes should differentiate and control for time of day. Finally, the heterogeneity of segments with different signal spacing density was not accounted for in the mentioned research while, in this study, arterials were assigned into homogeneous groups by signal spacing density so as to obtain more rational modeling results.

The objective of this study is therefore to examine a relationship between mean speed, speed variation and crash frequency on urban arterials by overcoming the above mentioned limitations. While existing studies employed segment-based aggregated speed measurements in calculating speed variation, this study aims to use a more representative measure of speed variation from vehicle-level GPS speed data. In this study, 8 arterials consisting of 234 one-way segments were selected, representing a total length of about 77 km. Other mediating and
exogenous factors were also considered, and their relationships with crash frequency were
developed using a hierarchical Poisson log-normal (HPLN) model with random effects.

2. Literature review

There are currently two schools of thought in developing speed, speed variation and
relationships: (1) using the temporal aggregation of speed measurements (e.g., mean
speed) of an entity (e.g., segment, link or area) over a given time period and then calculating
the number of crashes occurring on the same entity over the same time period (Quddus, 2013);
this is normally known as a segment-based analysis; (2) employing speed measurements just
before the crash occurrence and then synthesizing the speed data into bins and calculating the
number of crashes by speed bin (Imprialou et al., 2015); this is known as a scenario-based
analysis. Both approaches have advantages and disadvantages (Imprialou et al., 2015); this
paper, however, employs the former approach as the objective is to develop segment-based
safety management measures.

The balance of this review is devoted to speed, speed variation and crash relationships, as
well as the statistical methods employed in developing these relationships.

2.1. Mean speed and crashes

In a meta-analysis performed by Finch et al. (1994) of rural roads in Finland, Denmark,
Switzerland and the United States, a linear function was constructed and it was reported that a
1 mph increase in speed would lead to a 4.9% increase in crash rate. Similarly, in a study of
Swedish rural roads with speed limits from 90km/h to 110 km/h, Nilsson (1982) found that a
reduction in mean speed was accompanied by a reduction in crashes. The author explained
that from a kinetic point of view, stopping distance was greater at higher speeds, thus more
dangerous.

Inconsistent findings were reported in the literature. When looking into the speed-crash
relationship on rural European single carriageways with speed limits between 70 km/h and
100 km/h, Baruya (1998) found that when mean speed was low, crash frequency was high,
and inferred that poor road environment and road design caused serious congestion. Another
cross-sectional study on Hong Kong freeway segments conducted by Pei et al. (2012) reached
the same conclusion. Using GPS data from 480 taxis, they inferred that higher mean speed
was associated with shorter time exposure; thus crash frequencies decreased.

In a recent study, Cameron and Elvik (2010) suggested that the safety effect of mean
speed would vary with road types. They inferred that Nilsson’s model appeared satisfactory
for rural highways and freeways, but was not applicable on urban arterials because speed
limit changes were fewer on arterials. In order to study a variety of road types, Garber and
Gadiraju (1989) developed generalized linear models using spot-collected vehicle speeds on
interstate roads, arterial roads and major rural collector roads. They found that the crash rate
was higher at lower mean speeds, which may be a result of different geometric characteristics.
More in accord with relationships found on rural roads, a cross-sectional study focusing on
urban arterials conducted by Wang et al. (2015) found that an increase of 10km/h in mean
speed was associated with a 3% rise in crash frequency during peak periods.

Although mean speed is widely adopted as an important indicator of crash frequency, its
safety effects are not consistent. This inconsistency may result from the use of different
research approaches, different data sources and low data quality. According to a study by
Garber and Gadiraju (1989), the total accident rate and fatal accident rate on urban arterials
were both much higher than on other road types studied, indicating that the safety conditions
on urban arterials require more attention.
In summary, most research on speed and crash frequency has been based on rural roads or freeways. Urban arterials however are different, characterized by small and uneven signal spacing, heavier traffic load, higher speed variation throughout the day and significant difference between peak and off-peak periods due to the large proportion of commuter traffic. This paper, therefore, aims to investigate the relationship between mean speed and crash frequency on urban arterials.

2.2. Speed variation and crashes

Speed variation is used to represent the inconsistency of vehicle speed along a segment. Despite a variety of research methods and speed collection methods, crash frequency has consistently been found to rise as speed variation increases (Garber and Gadiraju, 1989; Solomon, 1964; Oh et al., 2001). However, this conclusion is to a great extent based on research on rural roads and freeways, while the relationship between speed variation and safety on urban arterials remains unclear.

Based on freeway spot speed data, Oh et al. (2001) found that speed variation was the most significant indicator for potential crash frequency. Solomon (1964) studied 600 miles of rural highways, measuring single vehicle speeds of 10,000 drivers. The author found a U-shaped curve relationship between speed variation and crash frequency. That is, crash counts were the lowest for travel speeds near the mean speed, and increased with greater differences as they moved away from the mean speed in either direction. In a cross-sectional study of highways and arterials, Garber and Gadiraju (1989) found that speed variation on a segment was greater when the difference between design speed and speed limit was greater. Another study on highway safety conducted by Pei et al. (2012) used FCD to calculate speed variation as the standard deviation of speed for vehicles travelling through a highway segment. In their study, speed variation was not significantly associated with crash occurrence. They argued that their results conflicted with the usual findings because the speed variation calculation could not accurately represent speed differences in a mixed traffic flow with various types of vehicles.

Although the speed variation-safety relationship has been demonstrated as generally consistent, previous studies have mainly investigated freeways and rural roads. The limited evidence available in the literature about the relationship between speed variation and crash counts on urban arterials demands more investigation.

2.3. Models for mean speed, speed variation and safety

Commonly used Poisson and negative binomial regression models in crash data analysis assume that each observation within a sample is independent, which may not be true in this study. Crash observations aggregated by segment within an urban arterial are likely to be correlated, as adjacent segments share similar road geometric design and commuting population, resulting in closely correlated traffic characteristics.

Hierarchical models are suggested for handling the data structure in which the lower level units are nested or grouped in the next higher level units. Yanmaz-Tuzel and Ozbay (2010) found that a hierarchical structure improved the predictability of Poisson log-normal (PLN) models because of its ability to adjust for spatial and temporal correlations, such as this study warrants due to its hierarchy of arterial and segment.

Apart from potential spatial correlations, the impact of segment length should also be taken into account in modeling. Existing research (Dinh and Kubota, 2013) has found that mean speed and speed variation can be affected by segment length. A longer length might result in higher values for both mean speed and speed variation because there is more space for acceleration and interaction with other vehicles. Considering the uneven arterial signal spacing in Shanghai that results in differing segment lengths, speed characteristics may vary.
across arterials. If the impact of segment length is not controlled in modeling, estimation results are likely to be biased.

To address the above limitations and account for the effects of signal spacing and segment length, arterials were divided into homogeneous groups that share similar signal spacing density. Further, random effect terms were introduced into the model to consider the potential correlation between segments within an arterial, and similarities in speed characteristics within a group. Consequently, a hierarchical Poisson log-normal (HPLN) model with random effects was employed in this study.

3. Data preparation

Geometric design features, traffic volume, speed and crash data were obtained from a sample of eight urban arterials with a total length of 76.59km.

3.1. Urban arterial segments

Eight arterials were selected from downtown Shanghai (see Figure 1). The selected arterials were divided into 117 segments, each defined by the two signalized intersections that terminate them. The divided segments were then separated by the direction of traffic flow. This resulted in a total of 234 one-way road segments that were used as sample units in developing mean speed, speed variation and crash relationships.

The eight arterials were selected from several randomly sampled urban arterials for the following reasons: (i) tunnel and elevated segments were excluded because their geometric design features are much different from surface segments, (ii) to ensure retrieving traffic volume, segments without loop detectors were also excluded and (iii) as some arterials have loop detectors in one direction but not the other, segment counts of an arterial could be different in each direction.
3.2. Geometric design features and traffic volume

Geometric design features consist of segment length, number of lanes, presence of medians and presence of non-motorized lanes. These data were obtained from a base map in geographic information system (GIS) and Google Earth®. Field surveys were also conducted as supplements.

Traffic volume data were obtained from Sydney Coordinated Adaptive Traffic System (SCATS) loop detectors during three time periods in 2011: 7:00-10:00 (morning), 11:00-14:00 (midday) and 16:00-19:00 (afternoon). Since this study focuses on road segments, only through-lane traffic volume was used.

Density of signal spacing (DSS), an arterial-level variable, was calculated as a reciprocal of average segment length of an arterial. For ease of the interpretation of the modeling results, the exposure variable was calculated as the logarithm of vehicle kilometers travelled (VKT); VKT was calculated by the product of traffic volume and segment length. Table 1 summarizes statistics of variables related to geometrics and traffic volume.

<table>
<thead>
<tr>
<th>TABLE 1: Summary Statistics of Variables Related to Geometrics and Traffic Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent Variable</strong></td>
</tr>
<tr>
<td>Arterial-level</td>
</tr>
<tr>
<td>DSS</td>
</tr>
<tr>
<td>Segment-level</td>
</tr>
<tr>
<td>Segment length</td>
</tr>
<tr>
<td>Number of lanes</td>
</tr>
<tr>
<td>Presence of medians</td>
</tr>
<tr>
<td>Presence of non-motorized lanes</td>
</tr>
<tr>
<td>Traffic volume</td>
</tr>
<tr>
<td>Exposure</td>
</tr>
</tbody>
</table>

Note: No. of observations is 702 (i.e., 234 segments ×3 time periods)

3.3. Mean speed and speed variation

Unlike aggregated speed data from loop detectors employed in many existing studies, vehicle-level speed data obtained using the FCD method from taxi-based GPS were employed in this study. This is advantageous in the sense that calculating mean speed and speed variation from vehicle-level data provides more realistic measurements than aggregated segment-based data. To maintain data consistency, it should be noted that taxi-based GPS data were obtained during the same time periods of retrieving data on the other traffic characteristics.

Three efforts were made to improve the taxis’ representativeness of general traffic. First, taxis without passengers were excluded because they travel more slowly than general traffic...
(Liu et al., 2009). Second, data with sampling intervals of more than 30s were excluded because there were not enough GPS records located in one segment. Third, it was ensured that extracted sample vehicles in each segment accounted for more than 5% of traffic volume. The sample size of taxis with passengers was 21,762; 22,000; and 24,363 in the morning, midday and afternoon periods respectively. Chen and Chien (2000) found that a sample of 3% of floating cars was sufficient for a 95% confidence level in travel time estimates.

Mean speed of a segment was calculated by employing all vehicle-based GPS speed data that were matched to the segment. This was achieved through the use of a space-mean speed formula. If a vehicle $j$ left $n$ GPS records during its trip through a segment, the single vehicle speed $V_j$ was computed as space-mean speed, as shown below:

$$V_j = \frac{nL}{\sum_{i=1}^{n} t_i} = \frac{nL}{\sum_{i=1}^{n} \left( \frac{L}{v_i} \right)}$$  \hspace{1cm} (1)

where $n$ is the total number of GPS records for vehicle $j$ on the segment with length $L$; and $t_i$ is the time required to travel the segment with the individual speed record of the vehicle $v_i$.

Mean speed of a segment $k$ (i.e. $\overline{MS}_k$) was then computed by

$$\overline{MS}_k = \frac{\sum_{j=1}^{m} V_j}{m}$$  \hspace{1cm} (2)

where $m$ is the number of vehicles that travelled on segment $k$.

Speed variation of a segment $k$ (i.e. $SV_k$) needs to consider both speed differences among vehicles and speed changes of individual vehicles along a segment. This is similar to the estimation of standard deviation. Using all GPS records in FCD, speed variation was calculated by

$$SV_k = \sqrt{\sum_{j=1}^{m} \sum_{i=1}^{l_j} (\bar{v}_{ij} - \bar{v})^2} / (n - m - 1) \text{ where } l_j = n_j - 1 \text{ and } \bar{v}_{ij} = \frac{\sum_{i=1}^{n_j} \left( \frac{v_i}{2} \right)}{n_j}$$  \hspace{1cm} (3)

where $\bar{v}_{ij}$ is the average speed between two consecutive GPS records from a vehicle, $j$; $\bar{v}$ is the grand mean value of all $v_i$ from all vehicles; and $n_j$ is the total number of GPS records for vehicle $j$.

Segment-based mean speed ($\overline{MS}_k$) and speed variation ($SV_k$) were then calculated for each of the segments by equations (1) and (3) for the different time periods (i.e., morning, midday and afternoon). Mean speed was found to be 36.0 km/h with a standard deviation of 8.2, a maximum value of 60.4 km/h and a minimum value of 15.6 km/h. The average speed variation was calculated as 15.9 km/h with a maximum value of 26.4 km/h and a minimum value of 3.7 km/h. Significant differences were observed in mean speeds and speed variations by segment and time of day.

### 3.4. Crash data

Since 2009, Shanghai traffic police have recorded crash locations according to a linear referencing method. For a specific crash that occurs in an arterial segment, the road name, reference point (usually the nearest intersection), direction and distance of the reference point to the crash are recorded. To date, more than 99% of police-reported crashes have been precisely located at a specific side of a road according to whether the crash is located in an approach lane or an exit lane. In this study, directional crash data were extracted from the same time periods (morning, midday, afternoon) of retrieval of the traffic volume and the FCD.
Intersection crashes located within boundaries of stop lines were excluded due to potentially different crash causation. Major contributory factors of intersection crashes include signal phase design, red light violations, issues with give way and the heterogeneous design features of the intersections. Since the objective of this paper is to analyze the relationship between speed, speed variance and crashes and given the fact that it is not possible to have a number of GPS records from a single vehicle in order to measure speed and speed variation within an intersection since GPS sampling frequency is 1s to 30s, we decided to exclude intersection crashes.

Crashes occurring on the main road segments were therefore extracted and employed in the analysis. The number of crashes during the three studied periods (morning, midday, and afternoon) was 1,567, accounting for 57% of the year’s total crashes. Of the 1,567 crashes used in the modeling, only 6 were fatal crashes, 12 were severe injuries and the rest, 1,549, were slight injury or property damage only crashes. Since crash data is already divided by traffic direction and by time period, the mean crash count is quite low (i.e. 6.7 crashes per segment in 2009), and hence crashes are not further disaggregated by their severity category meaning that the dependent variable is the total crashes by segment in 2009.

4. Methodology
4.1. Grouping of arterials
The eight full arterials were divided into sixteen one-way segments in order to separate them by the direction of traffic flow. Signal spacing of each arterial was uneven, which produced segments of uneven lengths. Considering potential effects of signal spacing on mean speed and speed variation (Dinh and Kubota, 2013), a Chi-squared automatic interaction detection (CHAID) decision tree was used in SPSS® to group the arterials, with crash frequency in each segment treated as a response variable, and DSS as an independent variable. The decision tree helped to reveal non-linear relationships among factors (Elmitiny et al. 2010). Based on this algorithm, the sixteen arterials were divided into six homogeneous groups, and the differences between groups were further accounted for by adding a group-level random effect into the model, which improved the model performance.

4.2. HPLN model with random effects
The sample data exhibit a hierarchical structure (i.e., segments are nested within arterials and arterials are then nested within groups). Consequently, crashes among the segments from an arterial may be correlated, and likewise, crashes among the arterials from a group may also be correlated. Therefore, a three-level hierarchical model is needed to control for within-segment, between-arterial and between-group differences. In this hierarchical modeling, crashes on a segment are assumed to be influenced by variables from all three levels. The data hierarchy is schematically shown in Figure 2.

![FIGURE 2 Framework of the HPLN model with random effects.](image)

The crash frequency was assumed to follow a Poisson distribution, which was given by
\[ Y_{ijk} \mid \theta_{ijk} \sim \text{Poisson}(\theta_{ijk}) \] (4)

where \( Y_{ijk} \) is the crash frequency for arterial segment \( i \), arterial \( j \) and group \( k \), and \( \theta_{ijk} \) is the expected crash frequency for the same segment.

A three-level Poisson log-normal model can be developed for a single explanatory variable \((x)\) as:

Segment-level (level 1):

\[ \ln(\theta_{ijk}) = \beta_{0jk} + \beta_{1jk}x_{ijk} + e_{ijk} \] (5)

Arterial-level (level 2):

\[ \beta_{0jk} = \delta_{00k} + u_{0jk}; \quad \beta_{1jk} = \delta_{10k} + u_{1jk}; \] (6)

Group-level (level 3):

\[ \delta_{00k} = \gamma_{000} + \vartheta_{00k}; \quad \delta_{10k} = \gamma_{100} + \vartheta_{10k} \] (7)

The composite equation can be expressed as:

\[ \ln(\theta_{ijk}) = \gamma_{000} + (\gamma_{100} + u_{1jk} + \vartheta_{10k})x_{ijk} + \vartheta_{00k} + u_{0jk} + e_{ijk} \] (8)

in which \( \gamma_{000} \) is the final model intercept, \( u_{0jk} \) is the random arterial-level intercept, \( \vartheta_{00k} \) is the group-level random intercept, \( e_{ijk} \) is the segment-level residual, Level-1 (segment) variance of \( e_{ijk} \) is \( \sigma_e^2 \), Level-2 (arterial) variance of \( u_{0jk} \) is \( \sigma_{u_0}^2 \) and Level-3 (group) variance of \( \vartheta_{00k} \) is \( \sigma_{\vartheta_{00}}^2 \). \( \gamma_{100} \) is the fixed slope coefficient for explanatory variable \( x \), \( u_{1jk} \) is the random arterial-level slope coefficient for \( x \), and \( \vartheta_{10k} \) is the random group-level slope coefficient for \( x \). All random components are assumed to follow a normal distribution with a mean of zero and a constant standard deviation. Equations (4) - (8) can collectively be termed as a three-level random-intercept and random-slope (i.e., hierarchical) Poisson log-normal model for a single explanatory variable. If \( u_{1jk} \sim 0 \) and \( \vartheta_{10k} \sim 0 \) are assumed (i.e., no variation in the slope coefficient in Level-2 and Level-3) then the model becomes a three-level random-intercept Poisson log-normal model. Model parameters can be estimated in a Bayesian modeling framework using the Markov Chain Monte Carlo (MCMC) method.

5. Modeling results

A Bayesian approach was used to estimate the model parameters. This approach considers parameters as random variables, which makes it possible to incorporate empirical knowledge about independent variables into the model (Ntzoufras, 2011). Posterior distributions were obtained using the Markov Chain Monte Carlo (MCMC) method. Two chains of 20,000 iterations were set in WinBUGS® (Spiegelhalter et al., 2003), and the chains that converged after the first 5,000 iterations were treated as burn-in. Deviance information criterion (DIC) was used as a means to compare the adequacy and complexity of the models (Spiegelhalter et al., 2003). Smaller DIC values indicate better-fitting models. When comparing models, differences of more than 10 in DIC values are sufficient to determine better fit.

To select the best fitted model in examining the relationships between mean speed, speed variation and crash frequency, a PLN model and a random-intercept HPLN model were
developed. Compared to the 2571 DIC of the HPLN model with random effects, the DIC of the PLN model was much greater (3188.87), while standard errors of coefficients in the HPLN model were all larger. Different specifications of mean speed were examined but it was found that the linear form provided the best goodness-of-fit.

Informative priors can make use of the known information about data, and thus make the model more robust. A maximum likelihood estimation was used in this paper to obtain relevant priors by running a simple negative binomial model for each variable separately, and the corresponding mean and variance of each parameter were estimated and then employed to formulate a normal distribution (Wang et al., 2015). Table 2 shows the posterior summaries of the HPLN model with random effects. Only variables significantly associated with crash frequency were included in the model.

Intra-class correlation (ICC1) was calculated to determine if the sample exhibited a cluster structure (Raudenbush and Bryk, 2002). Larger ICC values indicate a more demanding need to develop a hierarchical model (Huang et al., 2008). The overall ICC of the model (i.e., Level 2 and Level 3 relative to Level 1) was calculated at 0.473, indicating that up to 47.3% of the variation in segment-level crash frequency is explained by the three-level hierarchical structure of the data, which justifies the employment of hierarchical model.

### TABLE 2 Posterior Estimation of the Parameters of the HPLN Model with Random Effects

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>SE</th>
<th>95% BCIs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.688</td>
<td>0.172</td>
<td>(-3.970, -3.403)</td>
</tr>
<tr>
<td>Signal spacing density (1/km)</td>
<td>0.252</td>
<td>0.012</td>
<td>(0.232, 0.271)</td>
</tr>
<tr>
<td>Number of lanes:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 lane vs. 2 lane</td>
<td>0.558</td>
<td>0.070</td>
<td>(0.443, 0.673)</td>
</tr>
<tr>
<td>4 lane vs. 2 lane</td>
<td>0.739</td>
<td>0.058</td>
<td>(0.645, 0.835)</td>
</tr>
<tr>
<td>5 or 6 lane vs. 2 lane</td>
<td>1.075</td>
<td>0.089</td>
<td>(0.930, 1.221)</td>
</tr>
<tr>
<td>Exposure (lnVKT)</td>
<td>0.170</td>
<td>0.008</td>
<td>(0.157, 0.183)</td>
</tr>
<tr>
<td>Mean speed (km/h)</td>
<td>0.019</td>
<td>0.001</td>
<td>(0.018, 0.021)</td>
</tr>
<tr>
<td>Speed variation (km/h)</td>
<td>0.046</td>
<td>0.002</td>
<td>(0.042, 0.049)</td>
</tr>
<tr>
<td>Time period:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midday vs. Morning</td>
<td>0.483</td>
<td>0.062</td>
<td>(0.381, 0.563)</td>
</tr>
<tr>
<td>Afternoon vs. Morning</td>
<td>0.792</td>
<td>0.059</td>
<td>(0.695, 0.890)</td>
</tr>
<tr>
<td>Random effects (variance at group-level)</td>
<td>0.032</td>
<td>0.075</td>
<td>(0.001, 0.131)</td>
</tr>
<tr>
<td>Random effects (variance at arterial-level)</td>
<td>0.198</td>
<td>0.109</td>
<td>(0.079, 0.401)</td>
</tr>
<tr>
<td>Random effects (variance at segment-level)</td>
<td>0.606</td>
<td>0.069</td>
<td>(0.499, 0.724)</td>
</tr>
<tr>
<td>DIC</td>
<td>2,571</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Variables were statistically significant with 95% certainty. The abbreviation of Coef. stands for Coefficients, SE for Standard Error, and BCI for Bayesian Credible Intervals.

#### 5.1. Safety effect of mean speed

Exponential relationship between speed and crash frequency, as reported in this paper, is supported by a previous study (Elvik 2013). The findings of this paper are also consistent with those of most other studies, that higher mean speed is associated with increased crash frequency.

When other variables were controlled at their average values (i.e., VKT at 143 km·pcu/h/lane, speed variation at 16 km/h, and DSS at 3.15/km) on a two-lane segment

\[ ICC = \frac{\sigma^2_{\theta} + \sigma^2_{\theta 0}}{\sigma^2 + \sigma^2_{\theta} + \sigma^2_{\theta 0}} \]
during the morning (two-lane and morning as base case), the expected crash frequency at the
maximum mean speed was 140% higher than that of the minimum mean speed. When the
mean speed increased by 1%, then expected crash frequency increased by 0.7%. It should be
noted that over 99% of crashes observed in this study were property damage only (PDO) and
slight injury. A previous study (Elvik, 2009) showed that, on freeways and rural roads, a 1%
increase in mean speed led to a 1.1% increase in crash frequency for slight injury, and a 1.5%
increase for PDO crashes; the same study found, on urban roads, that these values were
slightly lower: 1% and 0.8% for slight injury crashes and PDO crashes, respectively. That is,
this study, with its even lower value of 0.7% on average, confirms that elasticity\(^1\) of crashes
with respect to mean speed is lower on urban roads than on rural roads and freeways.
However, on urban arterials with speed limit signs, the impact of mean speed on crash
frequency is even lower. Urban roads with posted speed limits in Canada were studied by
Gargoum and El-Basyouny (2016) to explore their speed-safety relationship. The researchers
reported that a 1% higher average speed was associated with a 0.018% higher collision
frequency, a much lower value than the Chinese urban arterials studied in this paper.

5.2. Safety effect of speed variation

When other variables were controlled at the average level (i.e., VKT at 143
km·pcu/h/lane, mean speed at 36 km/h, and DSS at 3.15 km\(^1\)), crash frequency at the
maximum speed variation on two-lane segments during the morning was 184% higher than at
minimum speed variation. An increase of 1% in speed variation was associated with a 0.74%
increase in crash frequency. In comparison, when speed variation increased by 1% in a study
by Garber and Gadiraju (1989), the researchers reported only a 0.057% increase in crash rate
(i.e., accidents per 100 million vehicle miles travelled) of all crash types on urban arterials;
while Quddus (2013) found a 0.313% increase in the slight injury accident rate (i.e., average
accidents per km) on freeways and A-class roads (i.e., major roads apart from freeways).
These findings are generally consistent with previous studies (Garber and Gadiraju, 1989;
Solomon, 1964; Oh et al., 2001; Quddus, 2013), that larger speed variation is associated with
increased crash frequency. However, the extent of the impact is significantly different for
urban arterials in Shanghai.

5.3. Safety effects of other variables

DSS, introduced as an arterial-level variable, was found to be positively associated with
.crash frequency. Higher DSS indicates that arterial traffic is more frequently interrupted by
intersecting roads.

Exposure is a log-transformed product of traffic volume and segment length. Crash
frequency increased in this study with increases in exposure, consistent with previous
findings (Wang et al., 2015; Greibe, 2003) that both traffic volume and segment length were
positively associated with crash frequency. Since exposure is the logarithm of VKT, the
coefficient value shown in Table 2 indicates its elasticity. That is, a 1% increase in VKT was
associated with a 0.17% increase in crash frequency.

Number of lanes was a categorical variable. Road segments with three or more lanes had
more crashes than those with two lanes, a finding consistent with an earlier study by Greibe
(2003). Time period was another categorical variable. Compared to the morning period,
crashes increased in both midday and the afternoon. This indicates that crash mechanisms
differ at different time periods in a day.

\(^1\) Elasticity is calculated as: $\frac{\partial y}{\partial x} = \gamma_{100} \cdot \bar{x}$; $\gamma_{100}$ is the fixed slope coefficient associated with
respective explanatory variable and $\bar{x}$ is the average value of the same explanatory variable for the
sample data.
6. Discussion

In Shanghai, operational conditions are complex on urban arterials, and quantifications of their relationships to crash frequency are necessary for implementing speed management measures. With sufficient FCD and qualified directional crash data, this study used a cross-sectional method to examine the relationships of mean speed and speed variation to crash frequency. Crash frequency was found to rise with the increase of mean speed and speed variation on urban arterials in Shanghai.

Figure 3 illustrates the relationship, based on the model estimation results, between mean speed, speed variation and expected crash frequency for a two-lane urban arterial in the morning. It is noticeable that the expected segment-based crash frequency increases with the increase of both mean speed and speed variation.

![Figure 3 Expected crash counts based on mean speed and speed variation.](image)

It is generally argued that because the available time for high-speed drivers to react is shorter, the likelihood of crashes is therefore higher at high driving speeds. The findings of this paper further confirm the validity of this position for urban arterials by showing that a 1% increase in mean speed resulted in a 0.7% increase in crash frequency. This result of 0.7 elasticity is generally in line with the findings of Wang et al. (2015) in which the elasticity of mean speed was found to be 0.51, resulting in a difference of just 37% in the elasticity value. It should be noted that Wang et al. (2015) did not control for speed variation and inherent data hierarchy in the modelling. This finding is, however, significantly different from the results of a study conducted by Xie et al. (2013), who found that a 1% increase in mean speed would lead to a 1.36% higher crash frequency. The inconsistency in research findings may be attributed to difference in data and research method. In Xie et al. (2013), the observation...
units from which crash data were collected were signalized intersections; and the road
operational speed was put in the model as a corridor-level variable.

In previous studies, speed variation was traditionally calculated with segment-based
aggregated (e.g. 5-minute or 15-minute) speed data (Figueroa and Tarko, 2005; Poe et al.,
1998). Such calculations cannot accurately reflect segment-level speed variation, as fixed
based methods cannot gather sufficient data for speed changes for a single vehicle travelling
through a segment. The speed variation proposed in this study was, instead, calculated from
spatio-temporal GPS-based speed data by taking into account speed differences among
vehicles and speed changes for individual vehicles; therefore, this calculation better reflects
the real traffic conditions and provides the consistency with previous results as expected.

A potential concern with the analysis carried out in this paper relates to the fact that all
crashes were combined into the total crash frequency and a single statistical model was
estimated. This is because the mean crashes per segment is quite low (i.e. 6.7 crashes per
segment in 2009). However, literature indicates that the impact of speed on crashes may
differ by crash severity (Imprialou et al., 2015). Therefore, any future study shall consider in
combining crash data from multiple years so as to facilitate the development of prediction
models by crash severity.

The key findings from this paper can be utilized to formulate policy relevant to speed
management in two ways: (i) speed management and (ii) speed variation control. Lower
mean speed is observed to be associated with lower crash frequency; it is therefore important
to establish appropriate speed on urban arterials, for example, by setting posted speed limits
(PSL). Research findings (Gargoum and El-Basyouny, 2016) of urban roads with PSL
indicate that higher PSL may lead to more collisions through its positive effect on speed.
Therefore, it is important to set and post appropriate speed limits along with matched
enforcement (e.g. fining drivers who exceed the speed limit) in order to limit the travel
speeds to a reasonable range, thereby reducing the number of those crashes resulting from
speeding.

Another important policy implication relates to the reduction in speed variation, and can
be termed proactive arterial safety management, similar to active traffic management for
freeways in several countries including the UK and USA. Even when drivers observe the
speed limit, there can still be an unsafe degree of speed variation. Reducing this variation can
be achieved by estimating optimal speed for given traffic conditions in real-time (a number of
algorithms are already available in the area of active traffic management) and displaying this
information through variable message signs, similar to a scheme already implemented on UK
freeways (Quddus, 2013). This, too would require enforcement, so the success of this
scheme would depend on the implementation of automatic enforcement devices such as
cameras.

7. Conclusion

This paper constructed an HPLN model with random effects to investigate the
relationship between mean speed, speed variation and crashes on urban arterials in Shanghai.
A total of 234 one-way segments were studied, and the time of day was controlled by
considering traffic heterogeneity in two directions and the variability of speed in different
time periods. Speed data from GPS was used to calculate mean speed and speed variation, as
it can capture the speed at any location of a segment. The arterials were divided into
homogeneous groups where average segment lengths were similar, which considered the
potential effects of segment length on speed characteristics.

The results show that crashes occurred more frequently at higher mean speeds. A 1%
increase in mean speed was found to be related to a 0.7% higher crash frequency on the
studied urban arterials, which have no posted speed limits. It should be noted that this value is
higher than that of urban roads with posted speed limits. However, it is lower than that of freeways and rural roads as reported by an existing study (Elvik, 2009). Although the impact of mean speed on crash frequency is smaller on urban arterials than on freeways and rural roads, this study of urban arterials offers a different perspective that could shed light on the conflicting results in previous studies on speed-crash relationships. However, most previous findings have been consistent regarding speed variation. This study further confirms that speed variation is observed to have a significant positive effect on crashes, with a 1% increase in speed variation associated with a 0.74% higher crash frequency. The elasticity values of crashes with respect to both speed and speed variation are significantly different on urban arterials compared with those on freeways and rural roads.

This study addresses an inherent knowledge gap in current studies of the relationships between speed, speed variation and crash frequency for urban arterials, particularly in China. Findings from this paper reveal the importance of speed management in improving traffic safety on urban arterials, especially the reduction in speed variation that could be addressed through a proactive arterial traffic management system.

Acknowledgements
This study was jointly sponsored by the Science and Technology Commission of Shanghai Municipality (15DZ1204800) and the Chinese National Science Foundation (51522810).

References
Elvik, R. A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. Accident Analysis and Prevention, Vol. 50, 2013, pp. 854-860.
Garber, N.J., and R. Gadiraju. Factors Affecting Speed Variance and its Influence on Accidents. In Transportation Research Record: Journal of the Transportation Research


