
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 1. Introduction

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

14 Speed data employed in developing speed-crash relationships have typically been
15 collected with fixed-based sensors such as radar and loop detectors (Figuroa and Tarko,
16 2005; Poe et al., 1998). These methods only capture speeds at fixed points, failing to
17 represent spatio-temporal distributions of speed along a road. Spatio-temporal speed data can
18 be collected by probe vehicles, a method known as floating car data (FCD), in which vehicles
19 (e.g., taxis) are equipped with global positioning systems (GPS) that can continuously record
20 vehicle position, speed and direction at a high sampling frequency (normally 1 Hz). Speed
21 measurements from sample vehicles are then employed to calculate both mean speed and
22 speed variation by road segment at given intervals (Wang et al., 2014; Xie et al., 2013; Wang
23 et al., 2016). There is a dearth of research, however, on the use of spatio-temporal GPS speed
24 measurements in developing crash-speed relationships while controlling for other mediating
25 factors (e.g., traffic volume). In addition, most of the previous studies on crash-speed
26 relationships have been developed using data either from freeways or rural roads.

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

45 The objective of this study is therefore to examine a relationship between mean speed,
46 speed variation and crash frequency on urban arterials by overcoming the above mentioned
47 limitations. While existing studies employed segment-based aggregated speed measurements
48 in calculating speed variation, this study aims to use a more representative measure of speed
49 variation from vehicle-level GPS speed data. In this study, 8 arterials consisting of 234 one-
50 way segments were selected, representing a total length of about 77 km. Other mediating and

1 exogenous factors were also considered, and their relationships with crash frequency were
2 developed using a hierarchical Poisson log-normal (HPLN) model with random effects.
3
4

5 **2. Literature review**

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

16 The balance of this review is devoted to speed, speed variation and crash relationships, as
17 well as the statistical methods employed in developing these relationships.
18

19 **2.1. Mean speed and crashes**

20 In a meta-analysis performed by Finch et al. (1994) of rural roads in Finland, Denmark,
21 Switzerland and the United States, a linear function was constructed and it was reported that a
22 1 mph increase in speed would lead to a 4.9% increase in crash rate. Similarly, in a study of
23 Swedish rural roads with speed limits from 90km/h to 110 km/h, Nilsson (1982) found that a
24 reduction in mean speed was accompanied by a reduction in crashes. The author explained
25 that from a kinetic point of view, stopping distance was greater at higher speeds, thus more
26 dangerous.

27 Inconsistent findings were reported in the literature. When looking into the speed-crash
28 relationship on rural European single carriageways with speed limits between 70 km/h and
29 100 km/h, Baruya (1998) found that when mean speed was low, crash frequency was high,
30 and inferred that poor road environment and road design caused serious congestion. Another
31 cross-sectional study on Hong Kong freeway segments conducted by Pei et al. (2012) reached
32 the same conclusion. Using GPS data from 480 taxis, they inferred that higher mean speed
33 was associated with shorter time exposure; thus crash frequencies decreased.

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

44 Although mean speed is widely adopted as an important indicator of crash frequency, its
45 safety effects are not consistent. This inconsistency may result from the use of different
46 research approaches, different data sources and low data quality. According to a study by
47 Garber and Gadiraju (1989), the total accident rate and fatal accident rate on urban arterials
48 were both much higher than on other road types studied, indicating that the safety conditions
49 on urban arterials require more attention.

1 In summary, most research on speed and crash frequency has been based on rural roads
2 or freeways. Urban arterials however are different, characterized by small and uneven signal
3 spacing, heavier traffic load, higher speed variation throughout the day and significant
4 difference between peak and off-peak periods due to the large proportion of commuter traffic.
5 This paper, therefore, aims to investigate the relationship between mean speed and crash
6 frequency on urban arterials.

7 8 **2.2. Speed variation and crashes**

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

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

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

33 34 **2.3. Models for mean speed, speed variation and safety**

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

40 Hierarchical models are suggested for handling the data structure in which the lower
41 level units are nested or grouped in the next higher level units. Yanmaz-Tuzel and Ozbay
42 (2010) found that a hierarchical structure improved the predictability of Poisson log-normal
43 (PLN) models because of its ability to adjust for spatial and temporal correlations, such as
44 this study warrants due to its hierarchy of arterial and segment.

45 Apart from potential spatial correlations, the impact of segment length should also be
46 taken into account in modeling. Existing research (Dinh and Kubota, 2013) has found that
47 mean speed and speed variation can be affected by segment length. A longer length might
48 result in higher values for both mean speed and speed variation because there is more space
49 for acceleration and interaction with other vehicles. Considering the uneven arterial signal
50 spacing in Shanghai that results in differing segment lengths, speed characteristics may vary

1 across arterials. If the impact of segment length is not controlled in modeling, estimation
2 results are likely to be biased.

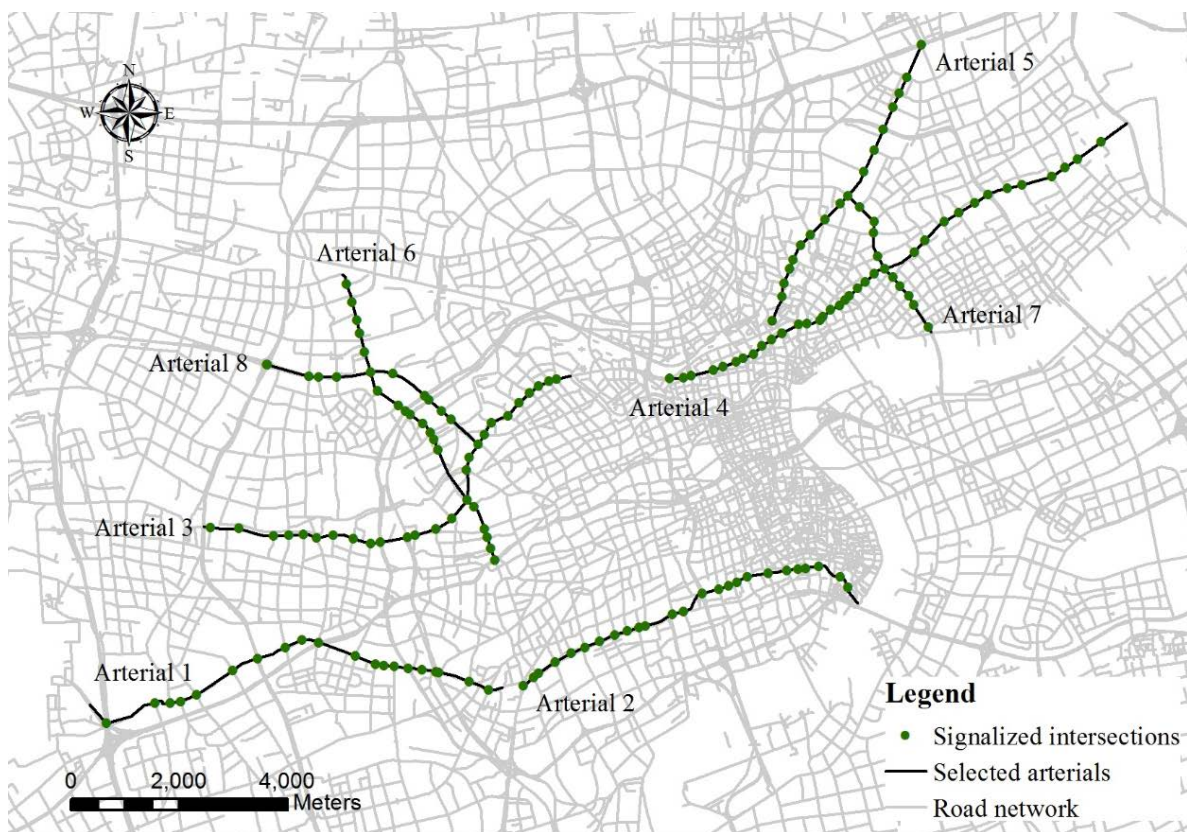
3 To address the above limitations and account for the effects of signal spacing and
4 segment length, arterials were divided into homogeneous groups that share similar signal
5 spacing density. Further, random effect terms were introduced into the model to consider the
6 potential correlation between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

9 **3. Data preparation**

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

14 **3.1. Urban arterial segments**

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



21 **FIGURE 1 Selected urban arterials in downtown Shanghai.**

22
23
24 The eight arterials were selected from several randomly sampled urban arterials for the
25 following reasons: (i) tunnel and elevated segments were excluded because their geometric
26 design features are much different from surface segments, (ii) to ensure retrieving traffic
27 volume, segments without loop detectors were also excluded and (iii) as some arterials have
28 loop detectors in one direction but not the other, segment counts of an arterial could be
29 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 1 Summary Statistics of Variables Related to Geometrics and Traffic Volume

Independent Variable	Description	Summary Statistics
Arterial-level		
DSS	Density of signal spacing (1/km)	Mean=3.15; SD=0.54
Segment-level		
Segment length	Length of a one-way segment (km)	Mean=0.33; SD=0.17
Number of lanes	Number of lanes of a one-way segment	Two lanes: 72 Three lanes: 231 Four lanes: 303 More than five lanes: 96
Presence of medians	A one-way segment with median or not	Yes:96; No:606
Presence of non-motorized lanes	A one-way segment with non-motorized lane or not	Yes:540 No:162
Traffic volume	Average through-lane volume per hour per lane in a one-way segment (pcu/h/lane)	Mean=452.82; SD=140.93
Exposure	Log-transformed vehicles' kilometers travelled in a one-way segment (km·pcu/h/lane)	Mean=142.76; SD=78.85

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:

$$V_j = \frac{nL}{\sum_{i=1}^n t_i} = \frac{nL}{\sum_{i=1}^n (\frac{L}{v_i})} \quad (1)$$

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

$$\overline{MS}_k = \sum_{j=1}^m V_j / m \quad (2)$$

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

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

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.

1 Intersection crashes located within boundaries of stop lines were excluded due to
 2 potentially different crash causation. Major contributory factors of intersection crashes
 3 include signal phase design, red light violations, issues with give way and the heterogeneous
 4 design features of the intersections. Since the objective of this paper is to analyze the
 5 relationship between speed, speed variance and crashes and given the fact that it is not
 6 possible to have a number of GPS records from a single vehicle in order to measure speed
 7 and speed variation within an intersection since GPS sampling frequency is 1s to 30s, we
 8 decided to exclude intersection crashes.

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

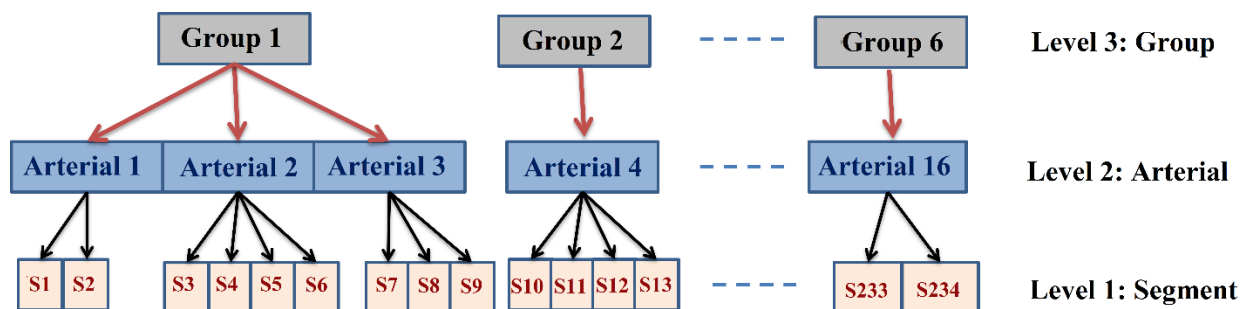
17 4. Methodology

18 4.1. Grouping of arterials

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

29 4.2. HPLN model with random effects

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



40
41 **FIGURE 2 Framework of the HPLN model with random effects.**
42

43 The crash frequency was assumed to follow a Poisson distribution, which was given by

$$1 \quad Y_{ijk} | \theta_{ijk} \sim \text{Poisson}(\theta_{ijk}) \quad (4)$$

2 where Y_{ijk} is the crash frequency for arterial segment i , arterial j and group k , and θ_{ijk} is the
3 expected crash frequency for the same segment.

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

6
7 Segment-level (level 1):

$$8 \quad \ln(\theta_{ijk}) = \beta_{0jk} + \beta_{1jk}x_{ijk} + e_{ijk} \quad (5)$$

9
10 Arterial-level (level 2):

$$11 \quad \beta_{0jk} = \delta_{00k} + u_{0jk}; \quad \beta_{1jk} = \delta_{10k} + u_{1jk}; \quad (6)$$

12
13 Group-level (level 3):

$$14 \quad \delta_{00k} = \gamma_{000} + \vartheta_{00k}; \quad \delta_{10k} = \gamma_{100} + \vartheta_{10k} \quad (7)$$

15
16 The composite equation can be expressed as:

$$17 \quad \ln(\theta_{ijk}) = \gamma_{000} + (\gamma_{100} + u_{1jk} + \vartheta_{10k})x_{ijk} + \vartheta_{00k} + u_{0jk} + e_{ijk} \quad (8)$$

18
19 in which γ_{000} is the final model intercept, u_{0jk} is the random arterial-level intercept, ϑ_{00k} is
20 the group-level random intercept, e_{ijk} is the segment-level residual, Level-1 (segment)
21 variance of e_{ijk} is σ_e^2 , Level-2 (arterial) variance of u_{0jk} is $\sigma_{u_0}^2$ and Level-3 (group) variance
22 of ϑ_{00k} is $\sigma_{\vartheta_0}^2$. γ_{100} is the fixed slope coefficient for explanatory variable x , u_{1jk} is the
23 random arterial-level slope coefficient for x , and ϑ_{10k} is the random group-level slope
24 coefficient for x . All random components are assumed to follow a normal distribution with a
25 mean of zero and a constant standard deviation. Equations (4) - (8) can collectively be termed
26 as a three-level random-intercept and random-slope (i.e., hierarchical) Poisson log-normal
27 model for a single explanatory variable. If $u_{1jk} \sim 0$ and $\vartheta_{10k} \sim 0$ are assumed (i.e., no variation
28 in the slope coefficient in Level-2 and Level-3) then the model becomes a three-level
29 random-intercept Poisson log-normal model. Model parameters can be estimated in a
30 Bayesian modeling framework using the Markov Chain Monte Carlo (MCMC) method.
31
32

33 5. Modeling results

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

44 To select the best fitted model in examining the relationships between mean speed, speed
45 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

Variable	Coef.	SE	95% BCIs
Intercept	-3.688	0.1722	(-3.970, -3.403)
Signal spacing density (1/km)	0.252	0.012	(0.232, 0.271)
Number of lanes:			
3 lane vs. 2 lane	0.558	0.070	(0.443, 0.673)
4 lane vs. 2 lane	0.739	0.058	(0.645, 0.835)
5 or 6 lane vs. 2 lane	1.075	0.089	(0.930, 1.221)
Exposure (lnVKT)	0.170	0.008	(0.157, 0.183)
Mean speed (km/h)	0.019	0.001	(0.018, 0.021)
Speed variation (km/h)	0.046	0.002	(0.042, 0.049)
Time period:			
Midday vs. Morning	0.483	0.062	(0.381, 0.563)
Afternoon vs. Morning	0.792	0.059	(0.695, 0.890)
Random effects (variance at group-level)	0.032	0.075	(0.001, 0.131)
Random effects (variance at arterial-level)	0.198	0.109	(0.079, 0.401)
Random effects (variance at segment-level)	0.606	0.069	(0.499, 0.724)
DIC	2,571		

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

$${}^1 ICC = \frac{\sigma_{u_0}^2 + \sigma_{\theta_{00}}^2}{\sigma_{\theta}^2 + \sigma_{u_0}^2 + \sigma_{\theta_{00}}^2}$$

1 during the morning (two-lane and morning as base case), the expected crash frequency at the
2 maximum mean speed was 140% higher than that of the minimum mean speed. When the
3 mean speed increased by 1%, then expected crash frequency increased by 0.7%. It should be
4 noted that over 99% of crashes observed in this study were property damage only (PDO) and
5 slight injury. A previous study (Elvik, 2009) showed that, on freeways and rural roads, a 1%
6 increase in mean speed led to a 1.1% increase in crash frequency for slight injury, and a 1.5%
7 increase for PDO crashes; the same study found, on urban roads, that these values were
8 slightly lower: 1% and 0.8% for slight injury crashes and PDO crashes, respectively. That is,
9 this study, with its even lower value of 0.7% on average, confirms that elasticity¹ of crashes
10 with respect to mean speed is lower on urban roads than on rural roads and freeways.

11 However, on urban arterials with speed limit signs, the impact of mean speed on crash
12 frequency is even lower. Urban roads with posted speed limits in Canada were studied by
13 Gargoum and El-Basyouny (2016) to explore their speed-safety relationship. The researchers
14 reported that a 1% higher average speed was associated with a 0.018% higher collision
15 frequency, a much lower value than the Chinese urban arterials studied in this paper.
16

17 5.2. Safety effect of speed variation

18 When other variables were controlled at the average level (i.e., VKT at 143
19 km-pcu/h/lane, mean speed at 36 km/h, and DSS at 3.15 km⁻¹), crash frequency at the
20 maximum speed variation on two-lane segments during the morning was 184% higher than at
21 minimum speed variation. An increase of 1% in speed variation was associated with a 0.74%
22 increase in crash frequency. In comparison, when speed variation increased by 1% in a study
23 by Garber and Gadiraju (1989), the researchers reported only a 0.057% increase in crash rate
24 (i.e., accidents per 100 million vehicle miles travelled) of all crash types on urban arterials;
25 while Quddus (2013) found a 0.313% increase in the slight injury accident rate (i.e., average
26 accidents per km) on freeways and A-class roads (i.e., major roads apart from freeways).

27 These findings are generally consistent with previous studies (Garber and Gadiraju, 1989;
28 Solomon, 1964; Oh et al., 2001; Quddus, 2013), that larger speed variation is associated with
29 increased crash frequency. However, the extent of the impact is significantly different for
30 urban arterials in Shanghai.
31

32 5.3. Safety effects of other variables

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

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

42 Number of lanes was a categorical variable. Road segments with three or more lanes had
43 more crashes than those with two lanes, a finding consistent with an earlier study by Greibe
44 (2003). Time period was another categorical variable. Compared to the morning period,
45 crashes increased in both midday and the afternoon. This indicates that crash mechanisms
46 differ at different time periods in a day.

¹ Elasticity is calculated as: $\frac{\partial y}{\partial x} \frac{x}{y} = \gamma_{100} * \bar{x}$; γ_{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.

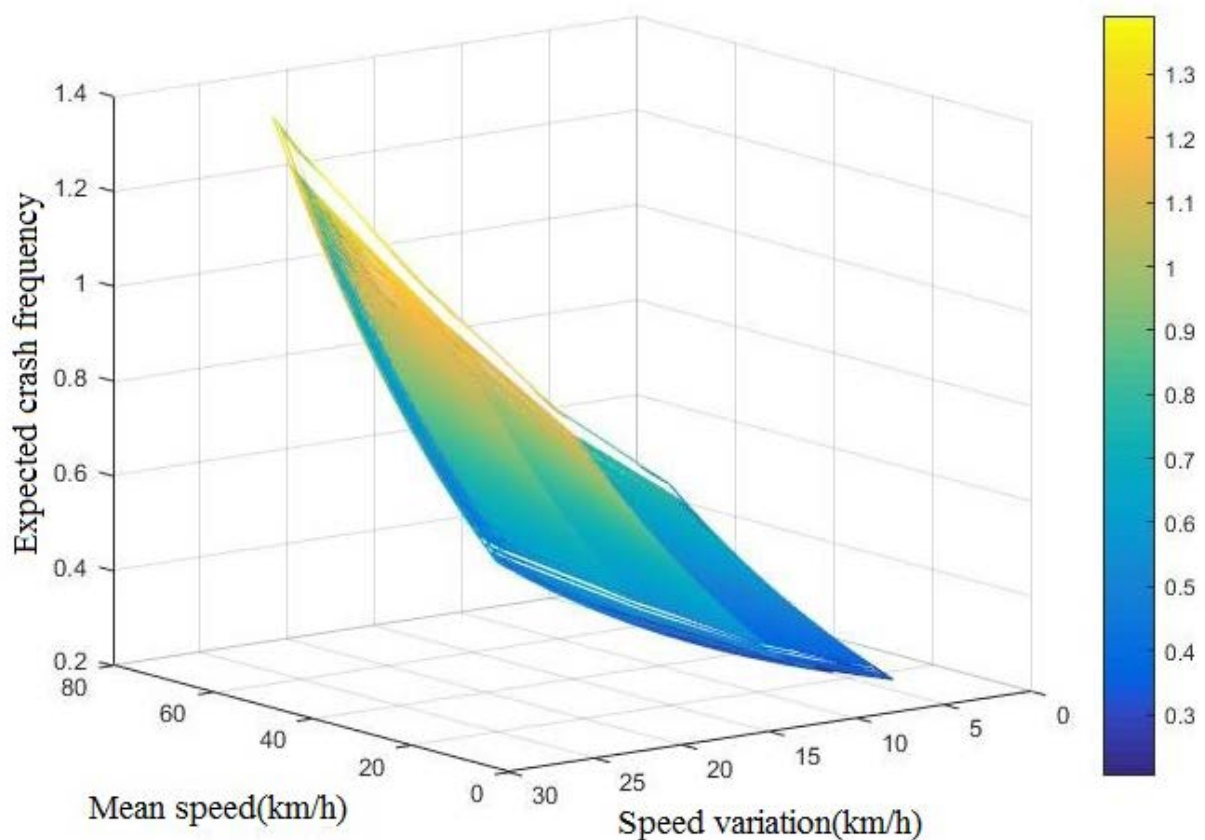


FIGURE 3 Expected crash counts based on mean speed and speed variation.

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

1 units from which crash data were collected were signalized intersections; and the road
2 operational speed was put in the model as a corridor-level variable.

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

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

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

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

38 39 **7. Conclusion**

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

48 The results show that crashes occurred more frequently at higher mean speeds. A 1%
49 increase in mean speed was found to be related to a 0.7% higher crash frequency on the
50 studied urban arterials, which have no posted speed limits. It should be noted that this value is

1 higher than that of urban roads with posted speed limits. However, it is lower than that of
2 freeways and rural roads as reported by an existing study (Elvik, 2009). Although the impact
3 of mean speed on crash frequency is smaller on urban arterials than on freeways and rural
4 roads, this study of urban arterials offers a different perspective that could shed light on the
5 conflicting results in previous studies on speed-crash relationships. However, most previous
6 findings have been consistent regarding speed variation. This study further confirms that
7 speed variation is observed to have a significant positive effect on crashes, with a 1%
8 increase in speed variation associated with a 0.74% higher crash frequency. The elasticity
9 values of crashes with respect to both speed and speed variation are significantly different on
10 urban arterials compared with those on freeways and rural roads.

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

17 Acknowledgements

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

22 References

- 23 Baruya, B. Speed-accident Relationships on European roads. *Proceedings of the Conference*
24 *'Road Safety in Europe'*, Bergisch Gladbach, Germany, September 21–13 23, VTI
25 Konferens No.10A, Part 10, 1998, pp.1–17.
- 26 Cameron, M. H., and R. Elvik. Nilsson's Power Model Connecting Speed and Road Trauma:
27 Applicability by Road Type and Alternative Models for Urban Roads. *Accident Analysis*
28 *and Prevention*, Vol. 42, No. 6, 2010, pp. 1908-1915.
- 29 Chen, M., and S. Chien. Determining the Number of Probe Vehicles for Freeway Travel
30 Time Estimation using Microscopic Simulation. Presented at 79th Annual Meeting of the
31 Transportation Research Board, Washington, D.C., 2000.
- 32 Dinh, D. D., and H. Kubota. Profile-speed Data-based Models to Estimate Operating Speeds
33 for Urban Residential Streets with a 30 km/h Speed Limit. *Accident Analysis and*
34 *Prevention*, Vol. 36, No. 2, 2013, pp. 115-122.
- 35 Elmitiny N., X. Yan, E. Radwan, C. Russo, and D. Nashar. Classification Analysis of
36 Driver's Stop/Go Decision and Red-light Running Violation. *Accident Analysis and*
37 *Prevention*, Vol. 42, 2010, pp. 101-111.
- 38 Elvik, R. The Power Model of the relationship between speed and road safety: update and
39 new analyses, Technical report, Institute of Transport Economics, 2009.
- 40 Elvik, R. A re-parameterisation of the Power Model of the relationship between the speed of
41 traffic and the number of accidents and accident victims. *Accident Analysis and*
42 *Prevention*, Vol. 50, 2013, pp. 854-860.
- 43 Figueroa, A., and A. Tarko. Speed Factors on Two-lane Rural Highways in Free-flow
44 Conditions. Presented at 85th Annual Meeting of the Transportation Research Board,
45 Washington, D.C., 2005.
- 46 Finch, D.J., Kompfner, P., Lockwood, C.R., Maycock, G. Speed, speed limits and crashes.
47 Project Record S211G/RB/Project Report PR 58. Transport Research Laboratory TRL,
48 Crowthorne, Berkshire, 1994.
- 49 Garber, N.J., and R. Gadiraju. Factors Affecting Speed Variance and its Influence on
50 Accidents. In *Transportation Research Record: Journal of the Transportation Research*

-
- 1 *Board, No. 1213*, Transportation Research Board of the National Academies,
2 Washington, D.C., 1989, pp. 64-71.
- 3 Gargoum S A and El-Basyouny K. Exploring the association between speed and safety: A
4 path analysis approach. *Accident analysis and prevention*, Vol. 93, 2016, pp. 32-40.
- 5 Greibe, P. Accident Prediction Models for Urban Roads. *Accident Analysis and Prevention*,
6 Vol. 35, No. 2, 2003, pp. 273-285.
- 7 Huang H, Chin H C, Haque M M. Severity of driver injury and vehicle damage in traffic
8 crashes at intersections: A Bayesian hierarchical analysis. *Accident Analysis &*
9 *Prevention*, Vol. 40, No. 1, 2008, pp. 45-54.
- 10 Imprialou M, Quddus M, Pitfield D E, et al. Re-visiting crash–speed relationships: A new
11 perspective in crash modelling. *Accident analysis and prevention*, Vol 86, 2015, pp. 173-
12 185.
- 13 Liu, B., K. Liu, and H. Mizuta. Study on Impact of Taxi Status to Probe Car System
14 Performance. In *Intelligent Transport Systems Telecommunications (ITST), 9th*
15 *International Conference on IEEE*, 2009, pp. 136-137.
- 16 Nilsson, G. *The Effects of Speed Limits on Traffic Crashes in Sweden*. Swedish National
17 Road and Transport Research Institute, VTI särtryck, 1982.
- 18 Ntzoufras, I. *Bayesian Modeling using WinBUGS (Vol. 698)*. John Wiley & Sons, Inc., 2011.
- 19 Oh, C, J. S. Oh, S. Ritchie, and M. Chang. Real-time Estimation of Freeway Accident
20 Likelihood. Presented at 80th Annual Meeting of the Transportation Research Board,
21 Washington, D.C., 2001.
- 22 Pei, X., S. C. Wong, and N. N. Sze. The Roles of Exposure and Speed in Road Safety
23 Analysis. *Accident Analysis and Prevention*, Vol. 48, 2012, pp. 464-471.
- 24 Poe, C., J. Tarris, and J. Mason. Operating Speed Approach to Geometric Design of Low-
25 Speed Urban Streets. Presented at 78th Annual Meeting of the Transportation Research
26 Board, Washington, D.C., 1998.
- 27 Quddus, M. Exploring the relationship between average speed, speed variation, and accident
28 rates using spatial statistical models and GIS. *Journal of Transportation Safety &*
29 *Security*. Vol. 5, No. 1, 2013, pp.27–45.
- 30 Raudenbush, S. W., and A. S. Bryk. *Hierarchical Linear Models: Applications and Data*
31 *Analysis Methods (Second edition)*. Thousand Oaks, CA: Sage Publications, 2002.
- 32 Solomon, D. *Accidents on Main Rural Highways Related to Speed, Driver, and Vehicle*. US
33 Department of Commerce, Federal Bureau of Highway, Washington D.C., 1964.
- 34 Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. *WinBUGS User Manual*. MRC
35 Biostatistics Unit, Cambridge, 2003.
- 36 Wang, X., Fan, T., Li, W., Yu, R., et al. Speed variation during peak and off-peak hours on
37 urban arterials in shanghai. *Transportation Research Part C Emerging Technologies*, Vol.
38 67, 2016, pp. 84-94.
- 39 Wang, X., H. Liu, R. Yu, B. Deng, X. Chen, and B. Wu. Exploring Operating Speeds on
40 Urban Arterials using Floating Car Data: Case Study in Shanghai. *Journal of*
41 *Transportation Engineering*, Vol. 140, No. 9, 2014.
- 42 Wang, X., T. Fan, M. Chen, B. Deng, B. Wu., and P. Tremont. Safety Modeling of Urban
43 Arterials in Shanghai, China. *Accident Analysis and Prevention*, Vol. 83, 2015, pp. 57-66.
- 44 Xie, K., X. Wang, H. Huang, and X. Chen. Corridor-level Signalized Intersection Safety
45 Analysis in Shanghai, China using Bayesian Hierarchical Models. *Accident Analysis and*
46 *Prevention*, Vol. 50, 2013, pp. 25-33.
- 47 Yanmaz-Tuzel, O., and K. Ozbay. A Comparative Full Bayesian Before-and-after Analysis
48 and Application to Urban Road Safety Countermeasures in New Jersey. *Accident*
49 *Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 2099-2107.
- 50