Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

A Novel Measure of speed Variation for Crash Frequency Prediction

Roshan Jose[a]*, Sudeshna Mitra[b]

[a]*Ranbir and Chitra Gupta School of Inrastructure Design and Mangement, IIT Kharagpur, West Bengal, 721302, India*
[b]*Civil Engineering Department, IIT Kharagpur, West Bengal, 721302, India*

Abstract

Among the different reasons for traffic crashes, excessive speed variation is accepted as one of the major contributors to high frequency of crashes. Traditionally, data to measure speed variation had restricted capability as data capturing locations were limited. However, with the use of Global Positioning System (GPS) device data from fleet vehicles, it is now possible to obtain extensive, yet fine, spatiotemporal speed data. This paper uses GPS data from a fleet of trucks to explore a novel approach in measuring speed variation to estimate the frequency of crashes on highways. This measure is based on the observed pattern of speed distribution of highway segments. The measure is developed with the assumption that a location will show a higher frequency of crashes when there exist at least two groups of vehicles traveling at statistically different speeds. It is calculated based on the ratio of area between the different speed regimes and the difference in mean speed between the different regimes. Findings of this study show that this calculated measure is capable of estimating the frequency of crashes along a highway.

Keywords: speed regime; speed variation; fleet GPS data; temporal variation of speed

1. Introduction

Speed variation increases the frequency of interaction between vehicles and hence is accepted as one of the significant contributors to traffic accidents (Solomon, 1964; Cirillo, 1968). Studies by Lave (1985), Garber and Gadiraju (1989), Fides et al. (1991), Kockelman and Ma (2010) and Quddus (2013) (to list a few) have identified and evaluated the importance of speed variation as a traffic crash indicator. Despite the current large body of research, there are some limitations in the current studies on excessive speed variation as a contributor to traffic accidents; with one such limitation being the current methods of speed data collection. The current methods of speed data collection

* Corresponding author. Tel.: +91-8116569101.
E-mail address: roshan.jose@iitkgp.ac.in

(loop detectors, videographic survey, floating car survey, and on-site radar-based methods) have practical limitation in simultaneous data collection from a large number of locations. This limitation affects the speed variation research as the speed of vehicle varies based on environment, road characteristics, situational factors, vehicle factors and individual characteristics of drives (Sadia et al., 2017). Hence, there is a need to examine speed variation along the entire highway at a finer level (continuous small segments of highway) to improve our understanding of the relationship between speed variation and frequency of crashes.

Global Positioning System (GPS) equipped vehicles provides this data as it is capable of transmitting its location, time and speed at a fixed interval to a central server. Further, according to Wang et al. (2016) the use of a GPS equipped fleet could provide sufficient data along a highway. This GPS data are used in a variety of applications such as map matching (Quddus and Washington, 2015; Chen et al., 2014; Yang and Meng, 2015), link speed and travel time estimation (Li et al., 2016; Lee et al. 2006; Li and McDonald, 2002), monitoring vehicle performance (Holt and Sarder, 2017; Chowdhury and Arsenault, 2015; Ma et al. 2011; Zhao et al., 2012; Zheng et al., 2008), detecting and classifying traffic states (D'Andrea and Marcelloni, 2017; Zeroual et al., 2017; D'Andrea and Marcelloni, 2016; Yong-chuan et al., 2011) and estimating crash frequency (Stipancic et al., 2017; Pande et al., 2017).

A few studies have assessed the relationship between speed variation and crashes as discussed in the following section, however such studies have limitations which merits additional research to further our understanding. The motivation of this paper is to add to the current research by examining speed variation along a 362 km long highway corridor and comparing the crashes along the highway over a two-and-a-half-year period. The study tries to develop a new measure to quantify speed variation based on the observed pattern of speed variation along the highway using which frequency of crashes can be predicted.

2. Literature Study

Solomon (1964) conducted one of the early studies on the effect of speed variation on crashes and suggested that the relationship between speed deviation from mean speed and frequency of crashes is a u-shaped parabolic curve, with high crash involvement for vehicles travelling at speeds far greater or far lesser than the mean speed. This research implies that crashes are triggered at extremely low or high variations from mean speed. The idea was also supported by other researches like Munden (1967) and Cirillo (1968). Subsequent studies (Baruya and Finch, 1994; Fildes et al., 1991; Kloeden et al.,1997, 2002; and Taylor et al., 2000) suggested that speed is linearly or exponentially related to crashed. Additionally, studies (Baruya, 1998; Stuster, 2004) have also been able to show that this relationship between speed and crash frequency is negative; while other studies (Garber and Gadiraju, 1989; Lave, 1985, Kockelman and Ma, 2010; Quddus, 2013) shows that such relationships are statistically insignificant. The inconsistency in the relationship between speed and crash frequency is explained by Pei et al. (2012) based on the 'measure of exposure'. Pie et al. (2012) explains that a negative relationship exists when distance-based exposure is used and a positive relationship exists when time-based exposure is used.

In addition to speed, studies have identified other factors which impact the frequency of crashes such as traffic flow (Aarts and Van Schagen, 2006), geometric design of road (Anastasopoulos and Mannering, 2009; Chang, 2005; Milton and Mannering, 1998), and vehicle occupancy (Garber and Subramanyan, 2001; Lord et al., 2005). Additionally higher frequency of crashes are observed in steep vertical grades (Anastasopoulos and Mannering, 2009; Chang, 2005; Milton and Mannering, 1998; Shankar et al., 1995) and on sharp horizontal curves (Abdel-Aty and Radwan, 2000; Anastasopoulos and Mannering, 2009). In such studies, the impact of higher speed variation is not explored due to the lack of fine vehicle-level data which could lead to a high reliability in estimates. The studies employ relatively highly aggregated data through current data collection techniques which leads to inconclusive results (Garber and Ehrhart, 2000; Kockelman and Ma, 2010).

Viewing the speed variation and crash estimation from a methodological perspective, crash counts are generated from a homogeneous links or road segments along the study network. Although this approach is logical and effective, it is undeniable that traffic conditions will vary significantly even within a homogeneous link or road segment, unless a very small segment of road network is observed. Hence, the assumption of homogeneity of a road segment which includes up to several miles of roadway and sometime bidirectional traffic is inaccurate. Further, the values of central tendency used for examining the factors may not be representative of actual condition at the time and site of the crash. The use of aggregate measures such as hourly/annual averages is not suitable for suddenly developed extreme traffic conditions (Abdel-Aty and Pande, 2005; Hossain and Muromachi, 2013). Hence, using such measures will lead to loss of information which is crucial in crash occurrences and may lead to erroneous and inconsistent conclusions.

It is interesting to note that one common limitation in all the studies is the method adopted in calculating speed. The studies rely on data from loop detectors and video surveillance from critical limited locations along the study segments. However, as observed by Poe et al. (1998), the speed of vehicles changes even along a homogeneous road segment. Hence, speed acquired using this method may not be representative of speeds over the entire segment. To mitigate the limitations of current spot speed studies, Wang et al. (2016) suggests the use of GPS enabled fleet data.

GPS data from probe and fleet vehicles are utilized for a variety of researches such as detecting and classifying traffic states (D'Andrea and Marcelloni, 2017; Zeroual et al., 2017), link speed and travel time estimation (Li et al., 2016; Lee et al. 2006; Li and McDonald, 2002), map matching (Quddus and Washington, 2015; Chen et al., 2014) and monitoring vehicle performance (Holt and Sarder, 2017; Chowdhury and Arsenault, 2015). Studies related to traffic crashes using GPS data are limited such as Stipancic et al. (2017) and Pande et al. (2017). A brief outline of the two studies are mentioned below.

Stopancic et al. (2017) corelates traffic flow measures such as congestion, speed and variation of speed to crash frequency and severity using data of approximately 4000 drivers in the Quebec City, Canada. The relationship between crash frequency and traffic flow was measured using Spearman’s correlation while pairwise Kolmogorov-Smirnov tests were used to determine the severity. The study showed a positive correlation of speed and crash frequency and a non-comparable relation between speed and crash severity. Additionally, average speed was found to be negatively correlated with crash frequency while no relation was found between average speed and severity. One of the limitation of the study was the lack of overlap in temporal coverage of traffic flow data and crash data.

Pande et al. (2017) used GPS data from 33 volunteer members travelling along US Highway 101, San Louis Obispo, California to assess crash prone locations. The study hypothesized that a road segment will be unsafe when the frequency of hard braking is higher compared to other locations. A negative binomial regression showed that the proportion of sharp deceleration along the freeway segments was significantly related to crash frequency of those segments. The study further observes that curvature and auxiliary lanes were not significantly related to crash frequency for these highway segments.

In this study, GPS data from truck fleet is used to measure the spatio-temporal variation of speed along a highway corridor. The speed data reveals the presence of speed regimes which is then measured. Crash database is linked with the highway corridor to examine the relation between crash frequency and the measured speed regimes. The contribution of this research is different from previous efforts in three critical ways. Firstly, studies have focused on spot speed from limited locations along a road segment to evaluate speed character, however, this study used GPS data to assess the speed character at every location of the highway. Secondly, while speed variation is accepted as one major cause of crashes, a loss of information due to aggregate data under-represents the extreme road conditions. To overcome this loss of information, this paper estimates the frequency at every 100 m segment of the highway. Thirdly, a new method of measuring speed variation is proposed which shows promise in predicting frequency of accidents at a location.

3. Data Collection and Processing

3.1. GPS Data

The study uses GPS data obtained from a fleet of 6600 trucks. The GPS devices transmits location, time, speed unique vehicle ID and trip ID through cellular technology to a central server at a 10-minute frequency. The data transmits uninterrupted with a 2.5 Circular Error Probability in location. The study observed data from January 2013 to May 2015 (30 months) of trucks plying through the selected corridors as mentioned in the following section.

3.2. Segmenting Road Network

The study area consists of a 362 km highway corridor divided into three distinct but connected highways. Corridor 1 stretches from Panagarh to Dankuni and is a part of National Highway 2, Corridor 2 begins at Dankuni and ends at

Kharagpur and is a part of National Highway 6, and Corridor 3 lies between Kharagpur and Balasore and is a part of National Highway 6 as illustrated in Fig. 1. The highway passes through minor towns and rural settlements while bypassing major towns and cities.

The study divides the highway into smaller 100m segments to capture the speed character at every location of the highway. 100 m is considered small enough to capture the speed variation accurately and is assumed to be small enough that speed variation due to other factors such as geometry will be negligible. It is important to note that the direction of travel was not considered in the subsequent analysis.

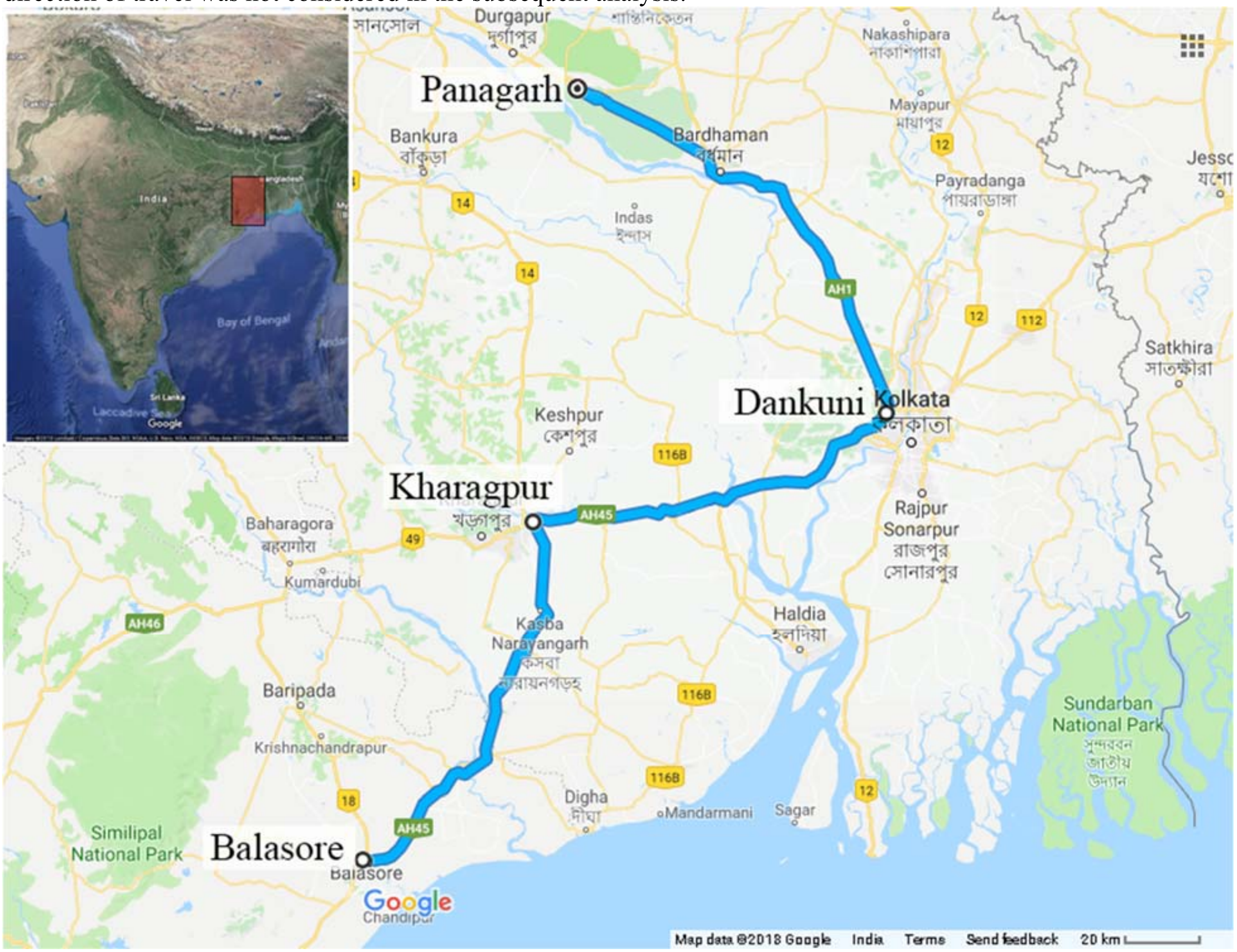

Fig. 1. Map of Corridor

3.3. GPS Data Processing

The GPS data obtained contain cross country movement, hence a GIS software was used to extract the GPS records which lie on the study corridor. A 25 m buffer from the road centerline of the study corridor was used to identify the GPS locations within the corridor. This buffer accounts for the tota width of the road including the shoulder. The extracted GPS data contains some undesirable data such as i) data of trucks stopped along the shoulder for extended period of time and ii) vehicle movement captured when trucks move across the highway instead of along the highway, i.e., cross traffic. The reason for removing undesirable data is that the trucks stopped for an extended period along the shoulder generates a significant amount of zero speed data, which does not reflect the actual vehicle speed along the highway. The study explicitly removed such data which shows extended stopped data. Additionally, trucks which do not travel along the highway are also filtered out. A total of more than 9 million GPS records were available for analysis.

3.4. Crash Data

The highways in this study are maintained by private concessionaries and as a part of the contract of the operation and maintenance of highways, the private concessionaires need to maintain an accident database. The databased contains the type of accident, time, weather, vehicle details, location in chainage and additional remarks stating the reason for accidents along with crashes involving property damage, injury, insurance claim or towing facility. The study excludes accidents such as i) accidents which arise due to collision with animals and ii) accidents arising due to poor vehicle maintenance as such accidents are not due to speed variation. A total of 2617 accidents were observed and Table 1 shows the yearly accident frequency count at the different corridors. It is important to mention that crash data for the year 2013 was not available on Corridor 2 and Corridor 3. Crash data is collected from the site in person and does not contain geographic coordinates. Hence, the location of crash data was first identified based on chainage and the location description, and then converted to GPS coordinates.

Table 1. Crash count along the highway

Year	Corridor 1	Corridor 2	Corridor 3
2013 (1st Jan – 31st Dec)	No records	No records	385
2014 (1st Jan – 31st Dec)	574	872	379
2015 (1st Jan – 31st May)	279	101	27

4. Methodology

The study initially investigated the relation between speed statistics and crash frequency. Insignificant statistical correlation between i) different speed statistics such as mean speed, 50th percentile, 85th percentile, 95th percentile speed, speed deviation, and range, ii) geometric features such as curvature of the highway, median openings, access and egress, and construction zone, and iii) additional features such as of truck parking along the shoulder, the presence of pedestrian activity on the highway, the presence of fuel stations and bus stops/bays, and presence of bridges/ flyovers suggested that a more in-depth investigation with more attributes were required. The authors assumed that such changes in any attribute could affect speed and observations in speed variation could offer some insights into the frequency of crashes. While studies (Hossain and Iqbal, 1999; Donell et al., 2001) suggest that speed data along a highway follows a normal distribution, or binomial distribution (Ko and Guensler, 2005; Zhao et al., 2013), these distributions are observed for low volume traffic flow with near free flow conditions (normal distribution) or high-volume traffic flows with different peak hours (binomial distribution). However, the data obtained in this study is of a homogeneous vehicle type in a heterogeneous environment. Hence, the authors investigated the speed distribution using Silhouette analysis on k-means clustering. K-means clustering is used as i) the speed data is one-dimensional, ii) mean values of clusters are considered important and iii) due to its low linear complexity.

4.1. Silhouette analysis

The Silhouette analysis of speed data reveals that the speed of vehicles may be grouped into three, as illustrated in Fig. 2 (top). It is to be noted that, since Silhouette analysis cannot be applied to evaluate a single cluster, the scores of two regimes are evaluated. A score of less than 0.75 for a two-speed regime segment is taken to be a single speed regime. The different speed groups or regimes are not present in all the segments of the highway. It may be noted that some segments have only 2 regimes while others have just one Fig. 2 (middle and bottom respectively). Hence, it may be stated that even with homogeneous group of vehicles, there exists external factors which influence the speed and creates different regimes of travel speed.

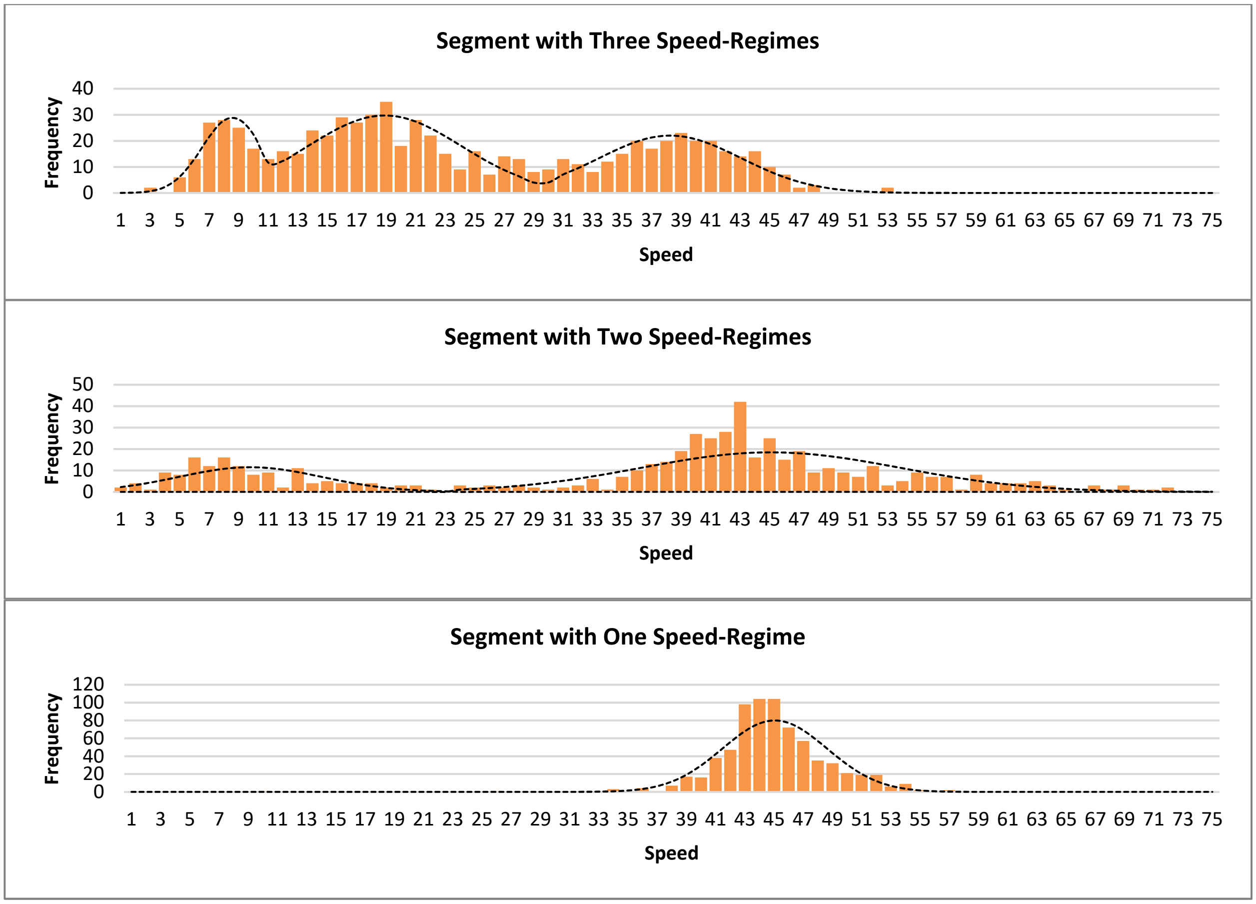

Fig. 2. Different Speed-regimes present on a highway corridor

The Silhouette analysis conducted for bihourly duration of each highway segment reveals that the number of speed regimes of that segment. It is possible that the number of regimes is different for each segment at different time intervals, depending on external conditions such as traffic, geometry, and weather. Fig. 3 illustrates the changes in a speed regime even within a highway segment due to temporal variation. It may be observed that in 'Road Segment 1401' there is a two-speed regime between 12:00pm and 02:00pm which slowly changes to a three-speed regime from 02:00pm to 04:00 pm and 04:00pm to 06:00pm. The three-speed regime then changes back to a two-speed regime from 006:00pm to 08:00 pm. Similarly, in 'Road Segment 2078' the single speed regime changes to a two-speed regime. It is also to be noted that there are segments which do not change the number of speed regimes with the temporal variation. The authors assume that the frequency of crashes increases based on the relation between i) the relative area of the different speed-regimes and ii) the difference in mean speed between the different regimes.

The reason behind these assumptions are that, firstly, the relative area of the different speed-regimes shows the ratio of the volume of vehicles which travel at different speed cluster. More the volume of vehicles which travel at different speed cluster, should mean a more the likelihood of traffic crash. Secondly, the difference in mean speed between the different regimes shows the extent of the difference in the faster-moving regime and slower-moving regime. It is hence assumed that the more the difference between the regimes, the more the likelihood of a crash.

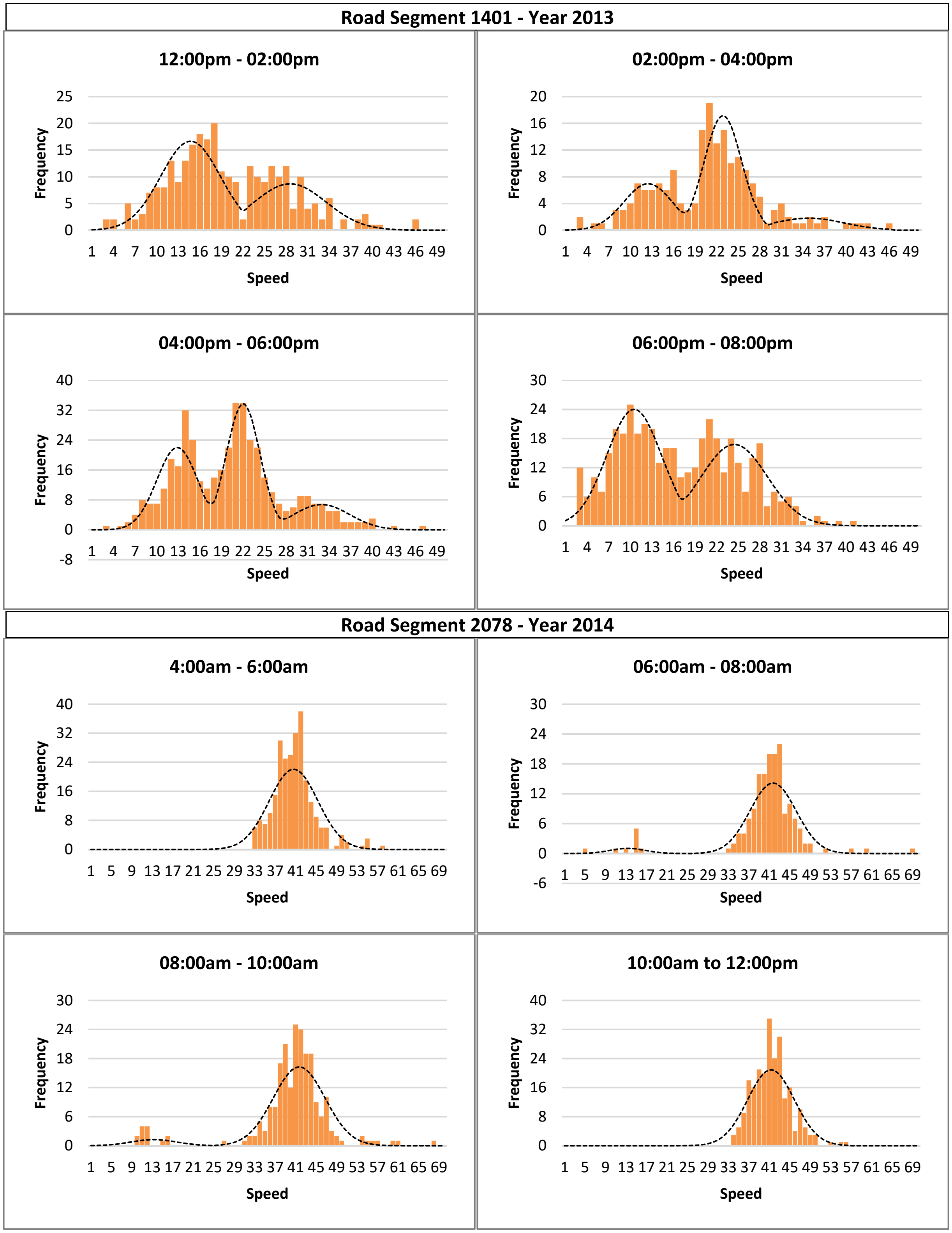

Fig. 3. Change in Speed-regimes across different years and different segments

4.2. Quantifying the Speed Regimes

From the observations on travel speed (illustrated in Fig. 3), it may be concluded that the presence of three distinct speed regimes shows that homogeneous groups of vehicles travel at statistically different speeds in heterogeneous conditions. The authors tried to quantify the speed regimes so that the different road segments may be compared for accident prediction.

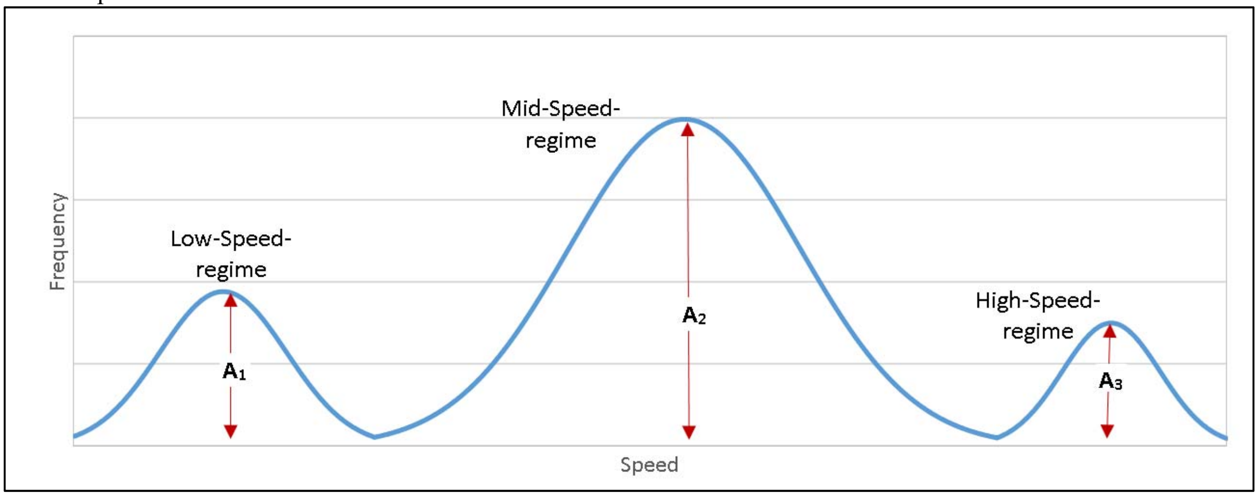

Fig. 4. Illustration of a Three-speed regime

Fig. 4 illustrates hypothetical three speed regime road segment. The authors assume that accidents happen only between vehicles of different regimes and that there is an equal probability of crash between any two vehicles from different speed-regimes. Hence, the value of Interaction is based on the number of vehicle-to-vehicle encounters between the different speed-regimes which is calculated based on the area under the curve. In Fig. 4, consider the low-speed-regime, the interaction of vehicles of low-speed-regime to the mid-speed-regime is the ratio of area between the two speed-regimes multiplied by the distance between the mean of speed-regimes as shown in Eq. 1. Similarly, the interaction of vehicles between low-speed-regime to the high-speed-regime may be calculates as shown in Eq. 2.

$$\frac{\sigma_2 * A_2}{\sigma_1 * A_1}\,(\mu_2 - \mu_1) \tag{1}$$

$$\frac{\sigma_3 * A_3}{\sigma_1 * A_1}\,(\mu_3 - \mu_1) \tag{2}$$

Where, σ is the standard deviation, A is the amplitude of curve under the speed-regime, μ is the mean of speed-regime, and 1 and 2 are low-speed and mid-speed regimes respectively.

While computing interaction of vehicles, it is logical to consider the ratio of the faster regime to the slower regime, as explained below.

Case 1: Consider two speed-regimes in which the (σ * A) of the slower speed is 100 and the faster is 1000. This low value of slow-speed regime means that fewer vehicles are traveling in the slow speed-regime. In such a case, there is a limited number of vehicle-to-vehicle encounters based on the time the slow-moving vehicles spend on a segment. Since there are a limited number of slow-moving vehicles, the majority of the vehicles in the faster speed regime will have no encounters with any slow-moving vehicle.

Case 2: In the exact opposite situation, where the (σ * A) of low speed-regime is 1000, and the faster regime is 100, means that the majority of the vehicles are in the slow-speed regime. Since there are a high number of slow-moving vehicles, every fast-moving vehicle will have a higher number of encounters with slow-moving vehicles thereby increasing the value.

The summation of all the three speed-regimes is expressed as shown in Eq. 3. In case of segments with only one speed-regime, the value is considered to be 'zero.'

$$\frac{\sigma_2 * A_2}{\sigma_1 * A_1}(\mu_2 - \mu_1) + \frac{\sigma_3 * A_3}{\sigma_2 * A_2}(\mu_3 - \mu_2) + \frac{\sigma_3 * A_3}{\sigma_1 * A_1}(\mu_3 - \mu_1) \tag{3}$$

Where, σ is the standard deviation, A is the amplitude of curve under the speed-regime, μ is the mean of speed-regime, and 1,2, and 3 are low-speed, mid-speed, and high-speed regimes respectively.

4.3. Crash correlation

The crash data and the calculated value of speed deviation are correlated to examine the relationship between the two variables. It is important to note that there are 130,320 segments of highway (i.e. 362 km x 10 segments per km x 12 bihourly divisions per segment x 3 years) and only 2,617 crashes. Since majority of the segments did not experience a crash, only those road segments which have crashes were considered for the comparison. Fig. 5 shows the three-year crash frequency vs the calculated value. It may be observed that the frequency of crashes increases with the increase in value across the years.

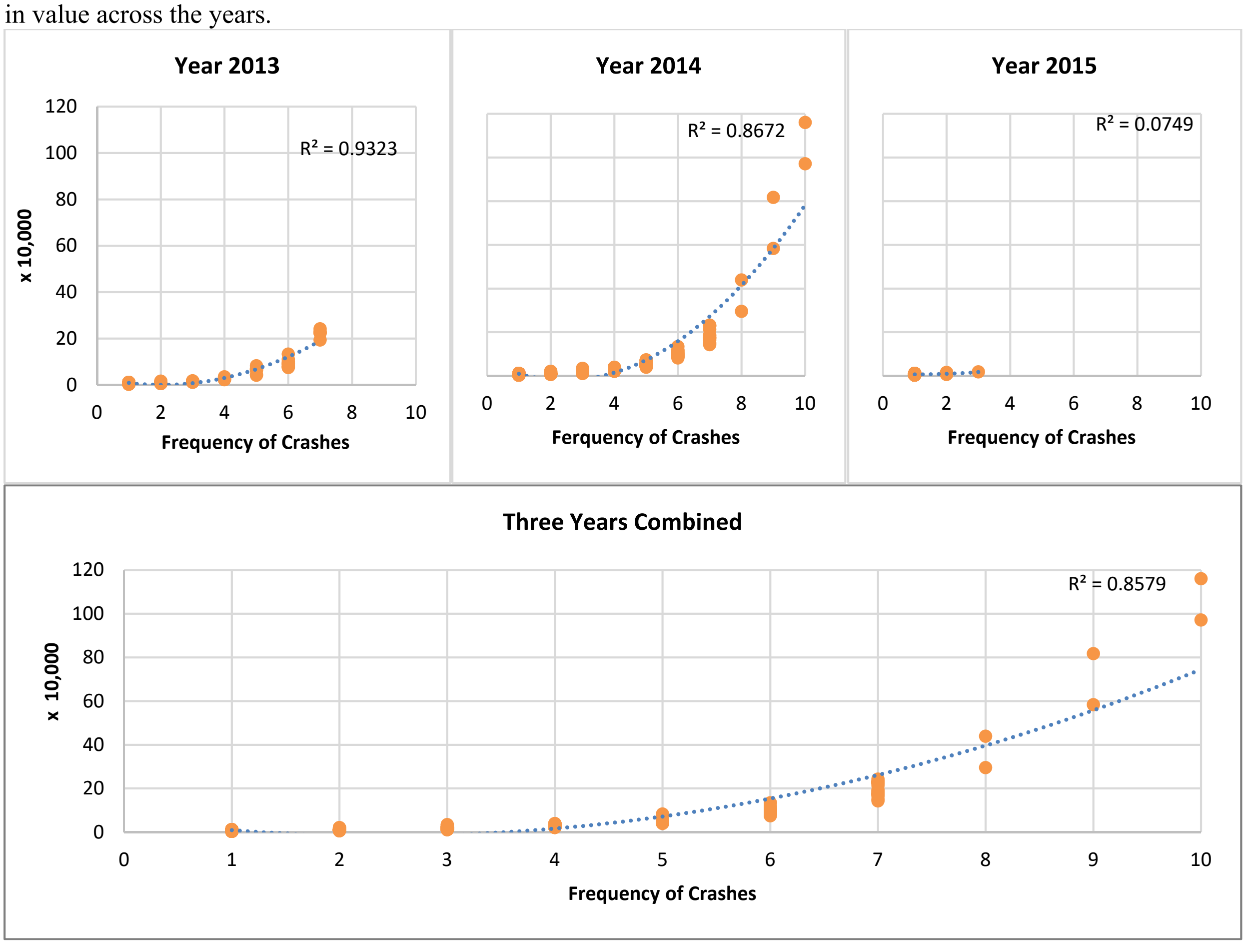

Fig. 5. Calculated value vs Frequency of Accidents

5. Results and Discussion

The study clearly show that frequency of accidents increases with the increase in interaction between vehicles of different speed regimes. Table 2 shows the total number of segments, total number of accidents and maximum frequency of accidents per segments for a turbulence range and it is evident that the segments with a single speed regime (i.e. segments where vehicles tend to travel at near similar speeds) have no occurrence of crashes. This is an important result as 30% (~39500 segments) of the segments fall in this category, this means that crashes happen (on a highway with uniform geometry) when there is a statistically significant difference in speed. In addition, only three crashes are observed for an additional 17% (22,759 segments) of the locations with two speed regimes. It is important to mention that some segments (~2% of the total segments) showed more than three speed-regimes as the optimal number during Silhouette analysis. All such locations are clustered into three speed-regimes.

Table 2. Frequency of crashes in different range of measure

Bins	Range of Measure	Total Number of Crashes	Maximum Frequency of Crash per segment	Number of segments
0	0	0	0	39337
1	0 - 58000	2227	5	81954
2	58000- 116000	195	6	6521
3	116000 - 174000	53	7	1446
4	174000 - 232000	77	7	645
5	232000 - 290000	7	7	172
6	290000 - 348000	8	8	100
7	348000 - 406000	0	0	46
8	406000 - 464000	8	8	38
9	464000 - 522000	0	0	21
10	522000 - 580000	0	0	12
11	580000 - 638000	9	9	4
12	638000 - 696000	0	0	7
13	696000 - 754000	0	0	3
14	754000 - 812000	0	0	3
15	812000 - 870000	9	9	3
16	870000 - 928000	0	0	2
17	928000 - 986000	10	10	1
18	986000 - 1044000	0	0	0
19	1044000 - 1102000	0	0	1
20	1102000 - 1160000	10	10	2

While it is observed that the frequency of crashes is related to the interaction of vehicles, a high value does not necessarily mean a high frequency of crashes at the location. It appears that other external factors such as geometric condition, site features, and weather also influences on the number of crashes at a given location.

6. Conclusion and Future Research

This study examines the relation between speed variation and frequency of crashes by proposing a new measure. By examining the speed pattern using Silhouette analysis, it was shown that speeds exist in a dynamic three-speed regime. The authors assumed that the relation between the speed regimes to be a gauge for predicting frequency of traffic crashes. The study evaluates interaction between vehicles as the summation of relationship between i) ratio of area under the regimes and ii) difference in mean speed of regimes. This method of evaluating speed regimes to estimate traffic crashes is unique as it not only estimates the potential frequency of crashes on a very small segment of a highway but also shows those segments which are potentially crash free. This is important as the results of the study can be utilized to identify critical locations which needs immediate improvement and also shows those locations which require no infrastructure changes.

The significant contribution of this study is in terms of proposing and testing a new measure, using GPS data from trucks, to interpret the speed variation along a highway with uniform geometry in order to predict the frequency of crashes. With many ongoing research initiatives on naturalistic driving data, a thorough understanding of speed variation at a micro-segment level becomes more relevant, and this study shows one such promising direction.

Limitation on this work is that temporal variation is limited by GPS data and the study does not account for accident severity. However, the assumption underlying the relationship used in this study should remain fairly stable, though more research is needed in this area. The authors also foresee that this method works best for segments of highways with high traffic flow, as speed regimes are dependent on volume. Future work will focus on the features of road segments which shows similar character at a wider network level. Not only will a network demonstrate the practical application, but also it will control for factors ignored in this study (relation between previous and following segment, varied geometry, etc.). Such a focus will contribute to a better understanding of relationship between speed variation within and between segments, and crash frequency and severity.

References

Aarts, Letty, and Ingrid Van Schagen. "Driving speed and the risk of road crashes: A review." Accident Analysis & Prevention 38.2 (2006): 215-224.

Abdel-Aty, M., & Pande, A. (2005). Identifying crash propensity using specific traffic speed conditions. Journal of safety Research, 36(1), 97-108.

Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. Accident Analysis & Prevention, 32(5), 633-642.

Anastasopoulos, P. C., & Mannering, F. L. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis & Prevention, 41(1), 153-159.

Baruya, A., & Finch, D. J. (1994, September). Investigation of traffic speeds and accidents on urban roads. In Traffic Management and Road Safety. Proceedings of Seminar held at the 22nd PTRC European Transport Forum, University of Warwick, England, September 12-16, 1994. VOLUME P381.

Baruya, A. (1998). MASTER: Speed-accident relationship on European roads. In Working Paper R 1.1. 3, Deliverable D7. Technical Research Centre of Finland VTT Espoo.

Chang, L. Y. (2005). Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network. Safety science, 43(8), 541-557.

Chen, B. Y., Yuan, H., Li, Q., Lam, W. H., Shaw, S. L., & Yan, K. (2014). Map-matching algorithm for large-scale low-frequency floating car data. International Journal of Geographical Information Science, 28(1), 22-38.

Chowdhury, T., & Arsenault, J. F. (2015). Freight Performance Micro Analysis Using Truck GPS data. In Canadian Transportation Research Forum 50th Annual Conference-Another 50 Years: Where to From Here?//Un autre 50 ans: qu'en est-il à partir de maintenant? Montreal, Quebec, May 24-26, 2015.

Cirillo, J.A., 1968. Interstate system crash research; study II, interim report II. Public Roads 35 (3), 71–76.

D'Andrea, E., & Marcelloni, F. (2016, May). Incident Detection by Spatiotemporal Analysis of GPS Data. In Smart Computing (SMARTCOMP), 2016 IEEE International Conference on (pp. 1-5). IEEE.

D'Andrea, E., & Marcelloni, F. (2017). Detection of traffic congestion and incidents from GPS trace analysis. Expert Systems with Applications, 73, 43-56.

Fildes, B. N., Rumbold, G., & Leening, A. (1991). Speed behaviour and drivers' attitude to speeding. Monash University Accident Research Centre, Report, 16, 186.

Garber, N., & Ehrhart, A. (2000). Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. Transportation Research Record: Journal of the Transportation Research Board, (1717), 76-83.

Garber, N. J., and R. Gadiraju. Factors Affecting Speed Variance and Its Influence on Accidents. In Transportation Research Record 1213, TRB, National Research Council, Washington, D.C., 1989, pp. 64–71.

Garber, N., & Subramanyan, S. (2001). Incorporating crash risk in selecting congestion-mitigation strategies: Hampton Roads area (Virginia) case study. Transportation Research Record: Journal of the Transportation Research Board, (1746), 1-5.

Holt, D. H., & Sarder, M. D. (2017). Analyzing Truck Traffic in Mississippi via GPS Transponders(No. 17-05943).

Hossain, M., & Muromachi, Y. (2013). Understanding crash mechanism on urban expressways using high-resolution traffic data. Accident Analysis & Prevention, 57, 17-29.

Kloeden, C.N., Mclean, A.J., Moore, V.M., Ponte, G., (1997). Travelling Speed and theRisk of Crash Involvement Volume 1 – Findings, No CR 172. NHMRC RoadAccident Research Unit, The University of Adelaide, South Australia.

Kloeden, C. N., McLean, J., & Glonek, G. F. V. (2002). Reanalysis of travelling speed and the risk of crash involvement in Adelaide South Australia. Australian Transport Safety Bureau.

Kockelman, K. K., & Ma, J. (2010, October). Freeway speeds and speed variations preceding crashes, within and across lanes. In Journal of the Transportation Research Forum (Vol. 46, No. 1).

Lave, C. A. (1985). Speeding, coordination, and the 55 mph limit. The American Economic Review, 75(5), 1159-1164.

Lee, S. H., Lee, B. W., & Yang, Y. K. (2006, May). Estimation of link speed using pattern classification of GPS probe car data. In International Conference on Computational Science and Its Applications (pp. 495-504). Springer, Berlin, Heidelberg.

Li, Y., & McDonald, M. (2002). Link travel time estimation using single GPS equipped probe vehicle. In Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference on (pp. 932-937). IEEE.

Li, Z., Cai, C., Menon, A. K., Xu, Y., & Chen, F. (2016). Estimation of Link Speed Distribution from Probe Vehicle Data. Transportation Research Record: Journal of the Transportation Research Board, (2595), 98-107.

Lord, D., Manar, A., & Vizioli, A. (2005). Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments. Accident Analysis & Prevention, 37(1), 185-199.

Ma, X., McCormack, E., & Wang, Y. (2011). Processing commercial global positioning system data to develop a web-based truck performance measures program. Transportation Research Record: Journal of the Transportation Research Board, (2246), 92-100.

Milton, J., & Mannering, F. (1998). The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation, 25(4), 395-413.

Munden, J.M., (1967). The Relation Between a Driver's Speed and His Accident Rate.Road Research Laboratory, Ministry of Transport, Crowthorne, England.

Pande, A., Chand, S., Saxena, N., Dixit, V., Loy, J., Wolshon, B., & Kent, J. D. (2017). A preliminary investigation of the relationships between historical crash and naturalistic driving. Accident Analysis & Prevention, 101, 107-116.

Pei, X., Wong, S. C., & Sze, N. N. (2012). The roles of exposure and speed in road safety analysis. Accident Analysis & Prevention, 48, 464-471.

Poe, C., Tarris, J., & Mason, J. (1998). Operating speed approach to geometric design of low-speed urban streets. Transportation research circular, (E-C003), 10-1.

Quddus, M., & Washington, S. (2015). Shortest path and vehicle trajectory aided map-matching for low frequency GPS data. Transportation Research Part C: Emerging Technologies, 55, 328-339.

Quddus, M. (2013). Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. Journal of Transportation Safety & Security, 5(1), 27-45.

Sadia, R., Bekhor, S., & Polus, A. (2017). Structural equations modelling of drivers' speed selection using environmental, driver, and risk factors. Accident Analysis & Prevention.

Shankar, V., Mannering, F., & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accident Analysis & Prevention, 27(3), 371-389.

Solomon, D., 1964. Crashes on main rural highways related to speed, driver and vehicle. In: Bureau of Public Roads. U.S. Department of Commerce. United States Government Printing Office, Washington, D.C.

Stipancic, J., Miranda-Moreno, L., & Saunier, N. (2017). Impact of Congestion and Traffic Flow on Crash Frequency and Severity: Application of Smartphone-Collected GPS Travel Data. Transportation Research Record: Journal of the Transportation Research Board, (2659), 43-54.

Stuster, J. (2004). Aggressive driving enforcement: evaluations of two demonstration programs. US Department of Transportation, National Highway Traffic Safety Administration.

Taylor, M. C., Lynam, D. A., & Baruya, A. (2000). The effects of drivers' speed on the frequency of road accidents. Crowthorne: Transport Research Laboratory.

Wang, X., Fan, T., Li, W., Yu, R., Bullock, D., Wu, B., & Tremont, P. (2016). Speed variation during peak and off-peak hours on urban arterials in Shanghai. Transportation Research Part C: Emerging Technologies, 67, 84-94.

Yang, J., & Meng, L. (2015). Feature selection in conditional random fields for map matching of GPS trajectories. In Progress in Location-Based Services 2014 (pp. 121-135). Springer, Cham.

Yong-chuan, Z., Xiao-qing, Z., & Zhen-ting, C. (2011). Traffic congestion detection based on GPS floating-car data. Procedia Engineering, 15, 5541-5546.

Zeroual, A., Harrou, F., Sun, Y., & Messai, N. (2017). Monitoring road traffic congestion using a macroscopic traffic model and a statistical monitoring scheme. Sustainable cities and society, 35, 494-510.

Zhao, W., McCormack, E., Dailey, D. J., & Scharnhorst, E. (2012). Using truck probe GPS data to identify and rank roadway bottlenecks. Journal of Transportation Engineering, 139(1), 1-7.

Zheng, Y., Liu, L., Wang, L., & Xie, X. (2008, April). Learning transportation mode from raw gps data for geographic applications on the web. In Proceedings of the 17th international conference on World Wide Web (pp. 247-256). ACM.