This is a draft version; please cite it as: Zheng, Z., 2012. Empirical analysis on relationship between traffic conditions and crash occurrences. Procedia - Social and Behavioral Sciences *43, 302-312.*

Empirical analysis on relationship between traffic conditions and crash occurrences

Zuduo Zheng[a*]

[a]*Civil Engineering & Built Environment School, Science and Engineering Faculty, Queensland University of Technology, 2 George St GPO Box 2434 Brisbane Qld 4001 Australia*

Abstract

This paper investigates relationship between traffic conditions and the crash occurrence likelihood (COL) using the I-880 data. To remedy the data limitations and the methodological shortcomings suffered by previous studies, a multi-resolution data processing method is proposed and implemented, upon which binary logistic models were developed. The major findings of this paper are: 1) traffic conditions have significant impacts on COL at the study site; Specifically, COL in a congested (transitioning) traffic flow is about 6 (1.6) times of that in a free flow condition; 2) Speed variance alone is not sufficient to capture traffic dynamics' impact on COL; a traffic chaos indicator that integrates speed, speed variance, and flow is proposed and shows a promising performance; 3) Models based on aggregated data shall be interpreted with caution. Generally, conclusions obtained from such models shall not be generalized to individual vehicles (drivers) without further evidences using high-resolution data and it is dubious to either claim or disclaim speed kills based on aggregated data.

Keywords: Freeway Crash; Traffic Conditions; Traffic Characteristics

1. Introduction

Car crash is a big, if not the biggest, externality of the modern transportation systems. According to WHO, about 1.2 million people are killed in road crashes worldwide each year and as many as 50 million are injured. Even worse, if no effective methods of preventing car crash are developed and implemented, these numbers will increase by about 65% over the next 20 years (Peden, 2004). In 2003, about 117 persons were killed daily in vehicle crashes in the United States (NHTSA, 2006). In 2005, more than four Australian people were killed and about 84 Australian people were seriously injured each day in road crashes (ATSB, 2008). Road fatalities in developing countries are even worse. Road crashes also often induce traffic congestions (and then increase vehicle emissions). Researchers have shown that over 30 percent of daily delay is caused by car crashes (Kwon, 2006).

Therefore, the importance of identifying factors that have significant impacts on freeway crashes is self-evident. In the past 50 years, notable efforts were devoted to understanding the potential association

* Corresponding author. Tel.: +61-07-3138-9989; fax: +61-07-3138-1170.
E-mail address: *zuduo.zheng@qut.edu.au.*

between traffic flow characteristics and crash occurrence. Interest in this topic is recently renewed because of the availability of massive traffic flow data and crash data thanks to the technological advancement in data collection and communication. Although significant and continuous progresses are achieved, a widely-accepted conclusion is still lacking. Consequently, traffic condition's impact on road safety by and large remains elusive. A classic example is the lasting and ongoing debate on which matters more to road safety: speed or speed variance. Solomon probably is the first person who concluded that speed variance, instead of speed, had more impact on crash (Solomon, 1964). However, his finding got little attention until Lave (1985) claimed in his controversial paper that once one controls for the variation of speed on the road, increasing speed will have little impact upon highway safety. Triggered by this paper, a hot debate ensued (Levy & Asch 1989; Synder 1989; Fowles & Loeb 1989; Lave 1989). Inconsistent and sometimes contradictory findings are reported in the literature, which is mainly caused by data limitations and/or methodological deficiencies.

Meanwhile, no comprehensive studies investigated different traffic conditions' potential influence on the crash occurrence likelihood (COL). Most of the previous studies either analyzed traffic characteristics' impact on COL without distinguishing traffic conditions or only focused on a specific traffic condition, e.g., Zheng et al. (2010) concentrated on understanding traffic oscillations (a typical feature of the congested traffic)' impact on COL.

This paper fills the gap in the literature. Toward this end, the remaining of this paper is organized as follows. The following section briefly summarizes the major issues shared by most of the previous studies. Section 3 presents the methodology adopted in this study. Section 4 describes in detail the study site and the data processing procedure. Section 5 provides the detailed statistical analysis. Finally, Section 6 offers concluding remarks and suggestions for future research.

2. Literature Review

Note that instead of discussing the details of each previous study, their shared limitations are summarized in this section. For a comprehensive review on the major previous studies, see Zheng et al. (2010).

Since 50 years ago, researchers started investigating the relationship between crashes' characteristics (e.g. occurrence, type and severity) and the traffic characteristics. However, despite substantive progresses in understanding traffic condition's impact on road safety, most of the findings are undermined by using temporally and spatially highly-aggregated traffic data and/or crash data, e.g., data used in some previous studies were annually averaged for a large area (e.g., a county, or a state). Such low resolution can not sufficiently capture traffic dynamics' impact on road safety. Also, caution should be exercised in interpreting findings from using the aggregated data (Davis 2002). Basically, findings from the aggregated data shall not be generalized to individual vehicles without further evidences using high-resolution data. Detailed explanation is presented in Section 6.

Based on the premise that the traffic condition immediately before a crash's occurrence is a direct contributor, traffic characteristics in a certain time interval immediately before the crash's occurrence are often measured and linked to COL. First, the premise itself needs a justification. Furthermore, the time interval that is used to measure traffic characteristics is usually arbitrarily selected, e.g., 5 minutes, 10 minutes, 30 minutes, etc. (Kockelman & Ma 2004; Lee et al. 2002; Oh et al. 2001; Abdel-Aty et al. 2005; Golob & Recker 2003). Obviously, this time period should be different for each individual crash if the premise indeed holds.

Meanwhile, extracting accurate crash occurrence time is critical in linking traffic dynamics to crash occurrence. Unfortunately, for retrieving the crash occurrence time most previous research relied on police reports, which is generally acknowledged as inaccurate and sometimes erroneous. Using inaccurate crash occurrence time can potentially introduce ‘cause and effect’ ambiguities by mixing crash-free and crash-contaminated data.

To shed light on this important topic, this paper employs a novel data processing method to remedy the aforementioned issues suffered by the previous studies.

3. Methodology

To avoid issues previous studies suffered, this study proposes a novel method to process and analyze data, as elaborated below.

Our methodology is derived from the following assumption:

Assumption: By controlling other factors, COL is the same at each instant when the traffic condition is stable; however, COLs in different conditions are different.

This assumption has profound implications to modeling the relationship between traffic dynamics and crash occurrences, and essentially implies that traffic dynamics in the same traffic condition, which can be measured as speed variance or other indicators, do not significantly vary, thus its influence on COL remains constant. In contrast, with the changing of the traffic condition its influence on COL changes because different traffic conditions have different traffic dynamics. Of course, theoretically proving this assumption is extremely difficult. However, supportive evidences are found in this study, as elaborated below through an example.

Crash 1188 happened at detector station 18, southbound of I-880, California, at 08:44 a.m. of October 18th, 1993 (More information on the study site and the data is presented in the following section). Traffic was severely congested when this crash occurred. Speed and occupancy (a proxy of density) are plotted in Figure 1, which shows that the congestion started from 07:25 a.m. and ended around 08:55 a.m. During this congested period, either speed or occupancy was quite stable. The traffic condition did not show any significant fluctuation immediately before the crash occurred. Therefore, there is no reason to take the traffic condition in 5 minutes or 10 minutes etc. immediately before the crash occurrence as the direct contributor. The contribution of traffic dynamics from 07:25 a.m. to 07:30 a.m. may be the same as that from 07:25 a.m. to 08:44 a.m. The reason why Crash 1188 occurred at 08:44 a.m., not at other times, can be explained by other factors such as human error, lighting, etc. The same conclusion can be obtained by looking at other crashes in the I-880 database.

Therefore, to quantify traffic dynamics’ influence on crash occurrences, traffic characteristics should be measured from the time period where the traffic condition remains the same, instead of from an arbitrarily selected time period.

 This is a draft version; please cite it as: Zheng, Z., 2012. Empirical analysis on relationship between traffic conditions and crash occurrences. Procedia - Social and Behavioral Sciences 43, *302-312.*

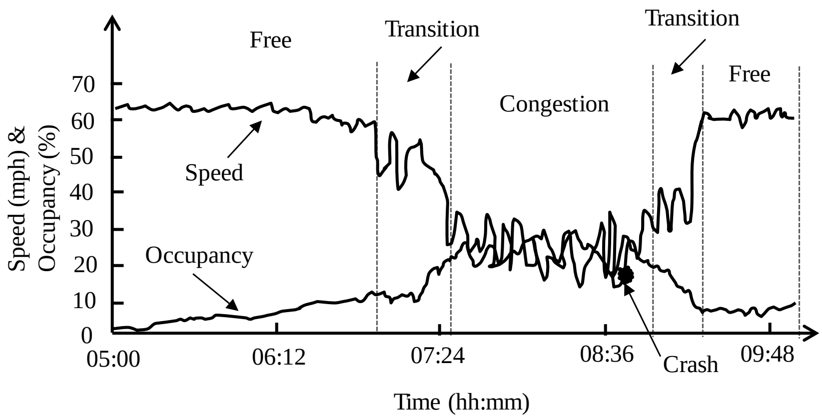


Fig.1. The speed & the occupancy time series for Crash 1188

Furthermore, a multi-resolution modeling structure is proposed to investigate the relationship between COL and traffic conditions, using the I-880 data. First, the crash data are linked to the corresponding traffic data; and using wavelet transform (WT), traffic conditions are automatically and systematically grouped into three different categories: free flow, transition (traffic is transitioning to a congested condition from a free flow condition and vice versa) and congestion. WT is a time-frequency decomposition tool that is particularly effective in extracting local information from non-stationary time series. Zheng et al. (2011a) demonstrate WT's applications in identifying bottleneck activations, phase transitions, and oscillation evolutions. They also utilized WT to microscopically investigate traffic oscillations' formation and propagation (Zheng et al. 2011b). For the theoretical background on WT's capability of detecting singular points and on selecting an optimal mother wavelet, see Zheng & Washington (2012). In this study, to avoid artifacts caused by using inappropriate data processing techniques, 1-minute traffic data are not further aggregated and WT is adopted to accurately and systematically detect the starting and the ending of a traffic condition.

After traffic conditions are identified and classified, characteristics of each traffic condition are measured using the averages and the variances of flow, occupancy and speed, respectively. A new variable (Y), crash occurrence, is created. For the traffic condition within which a crash occurs, Y is set to 1; otherwise, Y is set to 0 as illustrated in Figure 2. In Figure 2, Y=-1 means that data in that time period should be excluded to avoid any 'cause and effect' ambiguities.

Finally, binary logistic models are built with a multi-resolution framework to statistically investigate traffic condition's impact on freeway crashes by taking crash occurrence as the dependent variable. At the coarse resolution level, the independent variables are the traffic conditions (dummy variables); at the fine resolution level, the potential independent variables are the averages and the variances of speed, flow, and occupancy in each traffic condition.

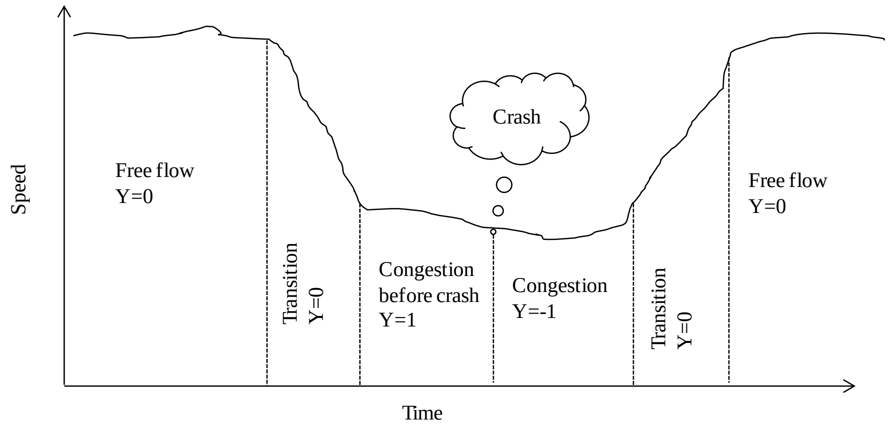


Fig. 2. Illustration of creating binary data

4. Study Site and Data Processing

The celebrated I-880 data in California (PeMS 2007) is used in this study. The I-880 data were collected from a section of the I-880 freeway in Hayward, California from February 16 to March 19, 1993 and from September 27 through October 29, 1993, respectively. The study section was 9.2 miles long and varied from 3 to 5 lanes with 17 detector stations on the southbound and 18 on the northbound. Detector stations were placed approximately 1/3 of a mile apart on the freeway mainline and on the on- and off-ramps. A schematic of the study site is provided in Figure 3.

The traffic measures collected by the detector stations are speed, occupancy, and vehicle count for every 1 minute. Their quality was checked using the special software (Petty 1995; Petty et al. 1996). The I-880 crash data was collected using 4 probe vehicles on the weekdays during the peak periods. 167 crashes in which more than one vehicle is involved were recorded, 14 of them were excluded because of lacking of the corresponding traffic data. In addition, some crashes, e.g. crashes 770 and 771, occurred at the same time of the same date at the same location, it is likely caused by misreporting. Therefore, they are taken as one crash. Then the total crash sample is reduced to 145.

Each crash is linked to the traffic data collected from the closest detector station by matching date, detector ID and direction. Note that the same method used in Zheng et al. (2010) to link crashes with traffic data is adopted here. Basically, due to the inherent limitation of loop detectors, which measure traffic conditions at specific points rather than over space, we define "influence areas" that are bounded by the midpoints between neighboring stations by assuming that traffic conditions are the same within each influence area and are represented by the measurements taken at the corresponding detector station. The soundness of this assumption was demonstrated in Zheng et al. (2010).

 This is a draft version; please cite it as: Zheng, Z., 2012. Empirical analysis on relationship between traffic conditions and crash occurrences. Procedia - Social and Behavioral Sciences 43, *302-312.*

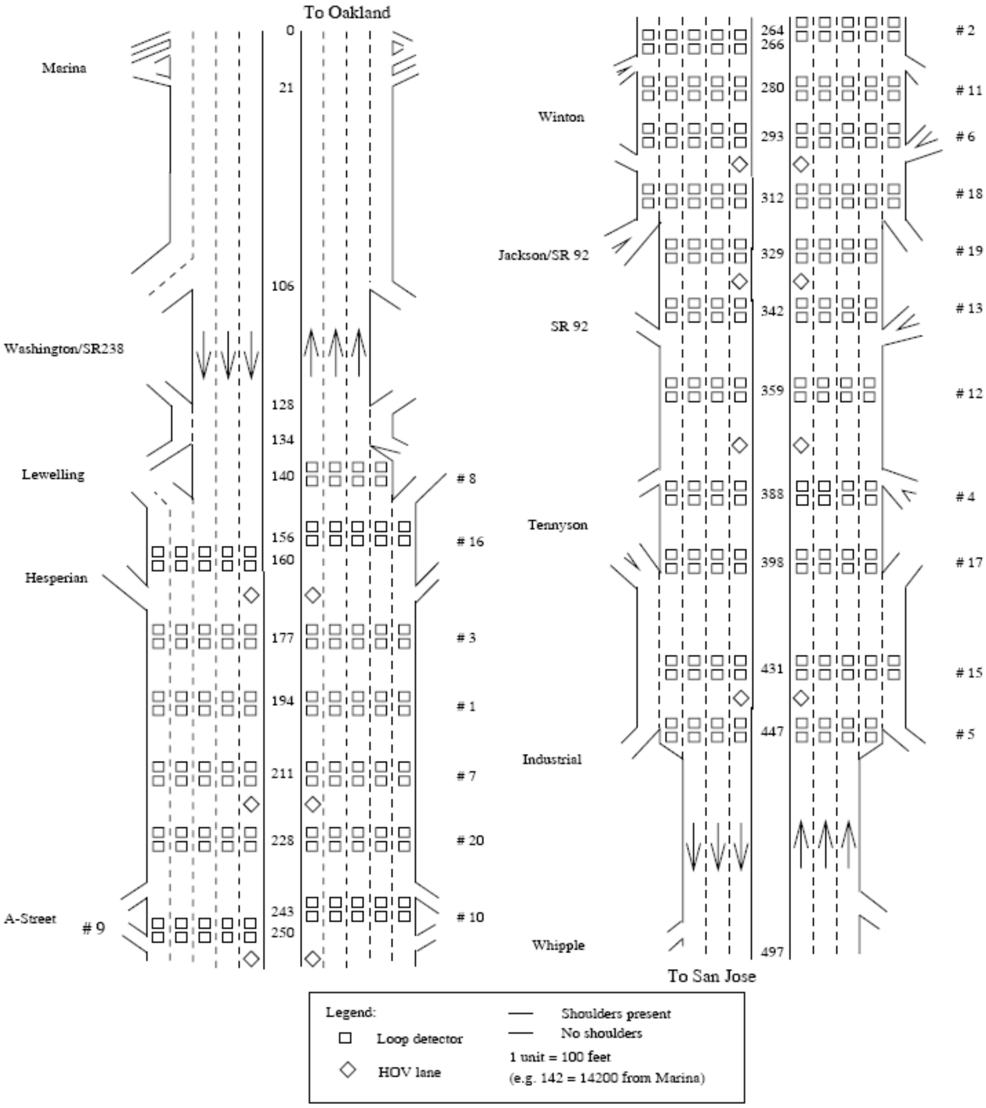
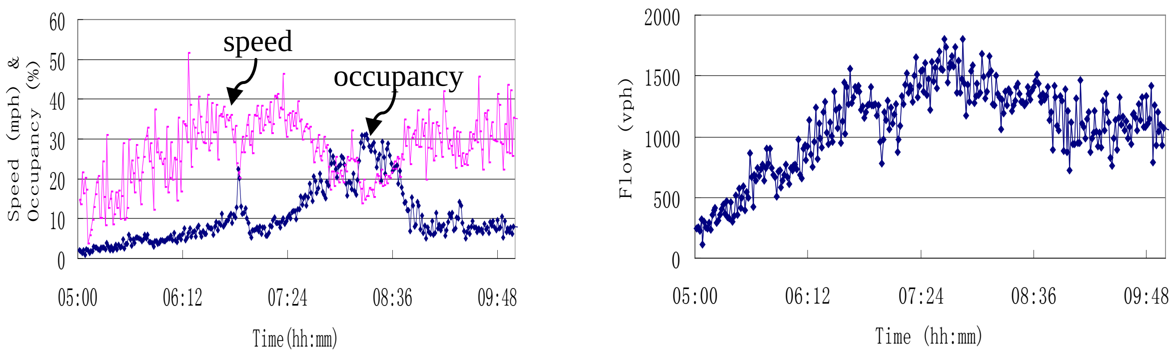


Fig.3. Schematic of the study site: A section of I-880, Hayward, California (Source: Petty (1995))

Then as previously discussed, traffic conditions are classified as free flow (coded as 0), transition (coded as 1), or congestion (coded as 2) using WT. Figure 1 is an example output from WT. Traffic between 05:00 a.m. and 06:57 a.m. is classified as the free flow because in this period the speeds were about 60 mph and the occupancies were mostly less than 0.1. Traffic between 7:25 a.m. and 08:55 a.m. is classified as the congested flow because in this period the speeds were mostly less than 40 mph and the occupancies were mostly above 0.20. Similarly, traffic in the period between 06:57 a.m. and 07:25 a.m. is classified as the transition because of the fluctuating speed and occupancy.

Extracting an accurate crash occurrence time is critical for obtaining unbiased results. The crash occurrence time in the I-880 dataset is verified or adjusted using the high-resolution traffic data based on the assumption that a crash in dense traffic will be reflected in the traffic characteristics (e.g., a sharp speed drop (increase) at an upstream (downstream) detector station). Note that this method was also used in Abdel-Aty et al. (2005) and Zheng et al. (2010).

Meanwhile, it is widely acknowledged that traffic data collected by loop detectors are error-prone due to various reasons. In this study, the erroneous traffic data are detected by using traffic diagrams and by plotting speed contours, as illustrated in figures 4-5.



(1) (2)

Fig. 4. A malfunctioning detector station: Station 1 on March 03 1993, Northbound, I-880; (1) The speed & occupancy diagram; (2) The flow diagram.

As shown in Figure 4, in the period between 05:00 a.m. to 06:00 a.m., flow and occupancy collected at station 1 on the northbound of I-880 on March 3rd 1993 were quite low. However, speed in the same period was also very low (below 30 mph), which is obviously abnormal. Therefore, this detector station was probably not working properly on that day. In order to verify this conclusion, the speed contour was produced (see Figure 5). Stations 3 (milepost 177), 1 (milepost 194), 7 (milepost 211), and 20 (milepost 228) locate at the same segment without any off/on- ramps, although an on-ramp exists before Station 20 and an off-ramp after Station 3. According to the flow conservation law, the traffic conditions at stations 1 and 7 should be similar except some time lag. However, as seen from Figure 5, the traffic condition at Station 1 was totally different from that at Station

1 on northbound was not working properly on that day. Therefore, the data collected by Station 1 on that day are excluded from any further analysis.

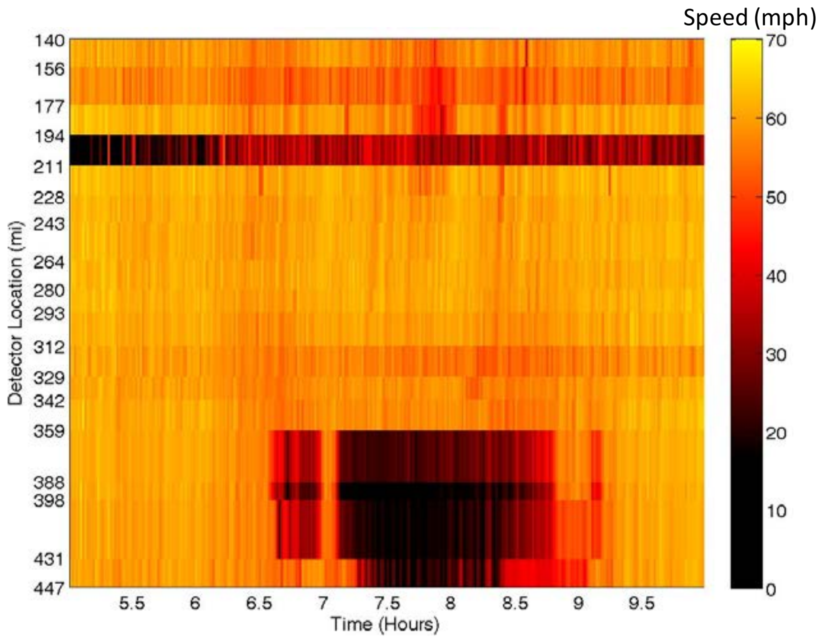


Fig. 5. Speed contour on March 03 1993, Northbound, I-880; colored according to speed: the dark region stands for low speeds and the bright region for high speeds.

Finally, the average and the variance (the standard deviation) of flow, speed and occupancy in each traffic condition are calculated. To avoid any "cause and effect" ambiguities, sequential traffic data belonging to the same traffic condition after the crash occurrence is excluded, as illustrated in Figure 2.

5. Modeling and Interpretation

The I-880 data show that traffic characteristics in the same traffic condition are relatively stable. The percentages that the speed variance is less than 10 mph in the free flow, in the transitioning condition, and in the congested traffic are 99%, 86%, and 94%, respectively. However, traffic characteristics across different traffic conditions vary notably, which has significant implications to road safety. As shown in Figure 6, among these 145 crashes, about 45% occurred in the congested condition, about 34% in the transitioning condition and only about 21% occurred in the free flow condition. It seems that crashes are more likely to occur in the congested or in the transitioning conditions than in the free flow because of stronger traffic dynamics in the former two conditions, which is confirmed by the statistical analysis as elaborated below.

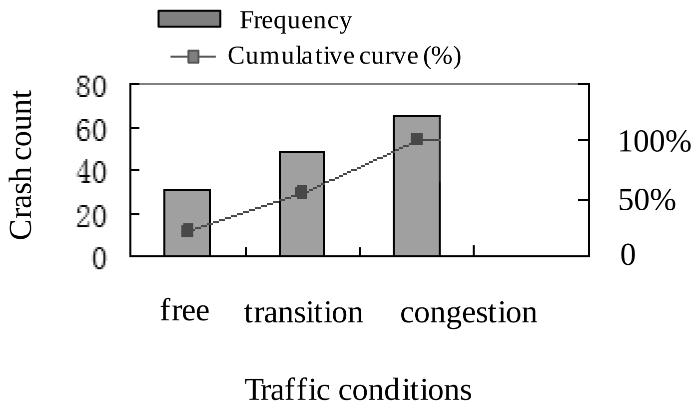


Fig. 6. The crash frequency distribution across traffic conditions

5.1. The Coarse Level Modeling: COL and Traffic Conditions

A logit model as defined below is built by taking crash occurrence as the dependent variable and traffic conditions as the independent variables.

$$log\frac{\pi(X)}{1-\pi(X)} = \beta_0 + \beta_1 x\ ,$$

where $\pi(X) = P(Y = 1|X)$, x is traffic conditions (a dummy variable)

The modeling results are summarized in Table 1. This model's overall performance is significant at the 99% confidence level.

Table 1 the modeling results for COL and traffic conditions

Crash occurrence	Coefficient	Standard error	p	95% confidence interval	
Transition	0.48	0.24	0.05	0.0034	0.96
Congestion	1.82	0.25	<0.001	1.33	2.3
Constant	-2.2	0.19	<0.001	-2.57	-1.83

Table 1 tells that the odds ratio of crash occurrence for traffic in the transitioning condition relative to traffic in the free condition is 1.62 (exp(0.483)) and that the odds ratio of crash occurrence for traffic in the congested condition relative to traffic in the free condition is 6.16, which strongly agrees with our assumption.

5.2. The Fine Level Modeling: COL and Traffic Characteristics

Again, the logit model is adopted. Occupancy, flow and their variances are excluded in the model because our preliminary analysis shows that they do not have significant influences on crash occurrences. In addition, the average speed is replaced by the median speed during the modeling procedure because the average speed and the speed variance are significantly correlated. Outliers' influence can also be avoided by using the median speed.

The modeling results are shown in Table 2. Note that although speed variance is not significant, this model's overall performance is significant at the 99% confidence level. Table 2 tells that multiplicative change in odds of crash occurrence with an additional 1 mph increase of average speed is 0.94 (exp(-0.058)). That means speed's increase is beneficial to the crash prevention, which seems anti-intuitive and inconsistent with the literature. However, this result again supports our assumption and is consistent with results from the coarse level modeling because here speed is actually reflecting traffic condition. Suppose speed is increased from 35 mph to 60 mph; consequentially, the traffic condition is changed from the congested condition to the free flow. According to results in Table 1, COL should probably decrease correspondingly.

Table 2 the modeling results for COL and traffic characteristics

Crash occurrence	Coefficient	Standard error	P	95% confidence interval	
Speed	-0.058	0.0067	<0.001	-0.071	-0.045
Speed variance	0.001	0.00083	0.22	-0.00062	0.0026
Constant	1.14	0.3	<0.001	0.55	1.73

Speed variance's impact on COL is not significant (p value is 0.22). However, this does not necessarily mean that traffic chaos has no any influence on COL. Perhaps, speed variance alone is not competent to sufficiently reflect traffic chaos and sequentially its adverse impact on traffic safety. Suppose there are three freeway segments A, B and C. They have the same physical features and the same speed variance 3 mph. However, speed on A is 25 mph and speed on B and C is 65 mph. Obviously, impact of the 3 mph speed variance on traffic at A is comparatively larger than that at B and C. if flow on A and B is 3000 vph, flow on C is 500 vph. Impact of the 3 mph speed variance on traffic at B is likely larger than that at C. briefly, traffic chaos is closely related to speed, speed variance and flow. To evaluate impact of traffic chaos on traffic safety, speed variance alone is not sufficient. Therefore, one new variable, the chaos index, is introduced, which is defined in the following equation.

$$chaos\ index = \frac{speed\ variance}{average\ speed} \times flow$$

Then, a logit model is built by taking crash occurrence as the dependent variable and the chaos index as the independent variable. The results are shown in Table 3.

Table 3 the modeling results for COL and chaos index

rash occurrence	Coefficient	Standard error	p	95% confidence interval	
Chaos	0.003	0.0005	<0.0001	0.002	0.004
Constant	-2.16	0.15	<0.0001	-2.46	-1.86

Table 3 indicates that multiplicative change in odds of crash occurrence with an additional one unit increase of the chaos index is 1.03 (exp (0.030)). The overall performance of the model and the coefficients are significant at the 99% confidence level. Therefore, there is strong evidence that the chaos index has the adverse impact on COL for the I-880 data.

6. Conclusions and Discussion

The modeling results based on the I-880 data show that: 1) traffic conditions have significant impact on COL. Briefly, COL in the congested condition is approximately six times of that in the free condition; COL in the transitional condition is approximately 1.6 times of that in the free condition; 2) speed has significant impact on COL. COL is likely to decrease when speed increases, which is inconsistent with the literature and anti-intuitive. However, the further analysis proves the soundness of this conclusion.

Researchers have been involved in the hot debate about speed kills or speed variance kills, which is triggered by Lave's controversial paper. Inconsistent and sometimes contradictory conclusions are reported. This phenomenon is partly caused by the complexity of this topic itself and by the data processing methodological limitations as mentioned before. Another reason is that interpretations of the modeling results are problematic. Traffic data used in most previous studies are not for each individual vehicle, but aggregated so they only measure collective characteristics of vehicles passing by a specific location during a certain time period. Individual vehicles' characteristics can hardly be revealed by such data. Therefore, findings from most, if not all, statistical models based on these data should not be applied to individual vehicles. For example, the coefficient for speed in Table 2 is -0.058, which should not be interpreted like this: for one vehicle, COL decreases when its speed increases. Of course, it is dangerously unreasonable to further conclude that speed limit should be cancelled or speeding should be encouraged. The correct interpretation is that when all vehicles' speeds increase, COL is likely to decrease thanks to the improved traffic condition. Briefly, it is dubious to either claim or disclaim speed kills based on the aggregated traffic data. One possible way to answer this question is to get the involved vehicle's speed immediately before the crash occurs as well as the aggregated traffic data for all vehicles on the site, and then build suitable models.

The modeling results also show that the speed variance has no any significant impact on COL for the I-880 data, which means that the speed variance alone is not competent to sufficiently reflect traffic's chaos and sequentially its adverse impact on traffic safety. Therefore, the chaos index has been introduced. The model incorporating the chaos index indicates that COL would increase with a larger chaos index.

Finally, models developed in this paper need to be validated and the aforementioned findings need to be verified by using independent dataset. Meanwhile, other factors (e.g., weather, geometric features, etc.) need to be considered during modeling.

References

ATSB, 2008. *Annual Road Crash Casualties and Rates from 1925 to 2005*. Australian Transport Safety Bureau.

Abdel-Aty, M., Pande, A., Uddin, N., Dilmore, J., Pemmanaboina, R., 2005. *Relating crash occurrence to freeway loop detectors data, Weather Conditions and Geometric Factors*. University of Central Florida.

Davis, G. A., 2002. Is the claim that 'variance kills' an ecological fallacy? *Accid. Anal. Prev.*, 34, 343-346.

Kockelman, K.M. and J. Ma, J., 2004. *Freeway speeds and speed variations preceding crashes, within and across lanes*. 14th Annual Meeting and Exposition, ITS America.

Kwon, J., Mauch, M., Varaiya, P., 2006. Components of Congestion: Delay from Incidents, Special Events, Lane Closures, Weather, Potential Ramp Metering Gain, and Excess Demand. *Transportation Research Record: Journal of the Transportation Research Board*, 1959, 84-91.

Lave, C. A., 1985. Speeding, Coordination, and the 55 MPH Limit. *Am. Econ. Rev.*, 75, 1159-1164.

Lave, C. A., 1989. Speeding, Coordination, and the 55-MPH Limit: Reply. *Am. Econ. Rev.*, 79, 926-931.

Levy, D. T. and Asch, P., 1989. Speeding, Coordination, and the 55-MPH Limit: Comment. *Am. Econ. Rev.*, 79, 913-915.

Fowles, R. and Loeb, P.D., 1989. Speeding, Coordination, and the 55-MPH Limit: Comment. *Am. Econ. Rev.*, 79, 916-921.

Garber, N.J., 2006. *Freeway crash predictions based on real-time pattern changes in traffic flow characteristics*, Tech. Rep. UVACTS-15-0-101, 01-20-06.

Golob, T.F., Recker, W. W., 2003. Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions, *J. Transp. Eng.*, 129 (4), 342-353.

Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of Crash Precursors on Instrumented Freeways. *Transp. Res. Rec.*, 1784, 1-8.

Martin, J. L., 2002. Relationship between crash rate and hourly traffic flow on interurban motorways, *Accid. Anal. Prev.*, 34, 619-629.

NHTSA. 2006. National Highway Traffic Safety Administration. http://www.nhtsa.gov/ (Dec. 15th).

Oh, C., Oh, J., Ritchie, S. G., Chang, M., 2001. *Real-Time Estimation of Freeway Accident Likelihood*. 80th Annual Meeting of Transportation Research Board, Washington, DC.

Peden, M., 2004. *World Report on Road Traffic Injury Prevention*. World Health Organization.

PeMS, 2007. Performance Measurement System. https://pems.eecs.berkeley.edu/?redirect=%2F%3Fdnode%3DState (December 2007)

Petty, K. F., Noeimi, H., Sanwal, K., Rydzewski, D., Skabardonis, A., Varaiya, P., Al-deek, H., 1996. The freeway service patrol evaluation project: Database support programs, and accessibility. *Transportation Research Part C*, 4, 71-85.

Petty, K., 1995. *The Analysis Software for the FSP Project*. Department of Electrical Engineering and Computer Science, University of California, Berkeley.

Solomon, D., 1964. *Accidents on Main Rural Highways Related to Speed, Driver and Vehicle*. US Department of Commerce & Bureau of Public Roads, Washington, DC.

Synder, D., 1989. Speeding, Coordination, and the 55-MPH Limit: Comment. *Am. Econ. Rev.*, 79, 922-925.

Zheng, Z., Ahn, S., Monsere, C. M., 2010. Impact of traffic oscillations on freeway crash occurrences. *Accid. Anal. Prev.*, 42, 626-636.

Zheng, Z., Ahn, S., Chen, D., Laval, J., 2011a. Applications of Wavelet Transform for Analysis of Freeway Traffic: Bottlenecks, Transient Traffic, and Traffic Oscillations. *Transportation Research Part B*, 45(2), 372-384.

Zheng, Z., Ahn, S., Chen, D., Laval, J. A., 2011b. Freeway traffic oscillations: microscopic analysis of formations and propagations using wavelet transform. *Transportation Research Part B*, 45(9), 1378-1388.

 This is a draft version; please cite it as: Zheng, Z., 2012. Empirical analysis on relationship between traffic conditions and crash occurrences. Procedia - Social and Behavioral Sciences 43*, 302-312.*

Zheng, Z. and Washington, S., 2012. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. Forthcoming in *Transportation Research Part C* .