Revisiting Classical Two and Three Phase Traffic Flow Theories for Pre-crash and Normal Traffic Conditions

Md. Mahmud HOSSAIN [a], Kazi Tahsin HUDA [b], Moinul HOSSAIN [c] and Yasunori MUROMACHI [d]

[a,b,c] *Department of Civil and Environmental Engineering, Islamic University of Technology (IUT), Board Bazar, Gazipur, 1704, Bangladesh.*
[a] *E-mail: mahmud31@iut-dhaka.edu*
[b] *E-mail: tahsinhuda@iut-dhaka.edu*
[c] *E-mail: moinul@iut-dhaka.edu*
[d] *Department of Built Environment, School of Engineering, Tokyo Institute of Technology, G3-5, 4259, Nagatuda-cho, Midori-ku, Yokohama, 226-8502, Japan.*
[d] *E-mail: ymuro@enveng.titech.ac.jp*

Abstract: Considerable research has been carried out to construct statistical and artificial intelligence based models to predict short-time crash probability taking basic traffic flow variables and/or their descriptive statistics and mathematical transformations. However, Little has been done to reveal whether and how the fundamental relationships of traffic flow theories differ for pre-crash and normal traffic conditions. This study revisits four classical two-phase traffic flow theories (Greenshields, Greenberg, Underwood and Bell-shaped models) and two methods of the three-phase traffic flow theory (detector based and correlation based) for pre-crash and normal traffic conditions and identifies their variations. The models developed with normal traffic data had better fit than those constructed with pre-crash data. The findings identify relatively high-velocity high-density traffic having association with pre-crash data both for two-phase and three-phase traffic flow theories. The hypothesis tests for the two-phase traffic flow theories affirm significant differences within the pre-crash and normal traffic condition data.

Keywords: two-phase traffic flow theory; three-phase traffic flow theory; pre-crash; normal traffic condition; real-time traffic data.

1. INTRODUCTION

Traffic flow theories are indispensable parts of the foundation of traffic science building the backbone of transportation engineering. They have various applications which represent a system in which many vehicles are interacting with each other (Velasco, 2012). Therefore, a basic distinction is made between microscopic and macroscopic models and their variables. Macroscopic model characterizes the traffic stream as a whole, and explains how the behavior of its parameters (speed, flow and density) changes with respect to one another. Microscopic model considers each of the individual vehicles in road network and how they interact with each other. These theories are inevitable for the fundamental characteristics of roadways. Around 80 years ago, the first macroscopic traffic flow theory was proposed by Greenshields *et al.* (1934) assuming a linear relationship between speed and density. Greenberg (1959) anticipated the same relationship to be simple logarithmic though not performing well in predicting speed at low density. Underwood (1961) assumed an exponential relationship but had reservations to be used under high density traffic conditions. Drake *et al.* (1967) relate

traffic flow model with bell-shaped normally distribution curve. Since then, many more complex models have followed. However, these four models are still widely used by researchers and practitioners. Rehborn *et al.* (2011) used four methods- detector based, correlation, flow-density and graphical method- for identifying the wide moving jam velocity of traffic flow in three-phase traffic flow theory. Two-phase classical theories fail to explain the phase transitions and the jam phases in actual traffic. Kerner's three phase traffic flow theory identifies these drawbacks and eliminates them (Kerner, 2004 and 2009b). Following empirical observations, it divides traffic flow into three distinct phases – free flow, synchronized flow and wide moving jam. The free flow phase occurs when a driver can drive at any speed he/she feels comfortable, considering speed limit and external factors like weather, number of passengers, type of vehicle, engine capacity, etc. Vehicle speed lower than the speed of the slowest vehicle in free flow state indicates the congested traffic. A wide moving jam evolves from bottle necks and maintains the mean velocity of the downstream jam front while moving upstream, even through any other traffic state or a freeway bottleneck (Rehborn *et al.*, 2011). Synchronized flow does not exhibit the property of wide moving jam and is often fixed at the bottleneck (Rehborn *et al.*, 2011). Practical online applications ASDA/FOTO are developed based upon these observations, in order to track and identify the different jam phases of a traffic flow theory (Kerner, 1999a, 2004 and 2009b; Kerner and Rehborn, 1998; Kerner *et al.*, 1998, 1999, 2004). Rehborn *et al.* (2011) evaluated universality of Kerner's theories and confirmed that the downstream jam velocity for a wide moving jam remains within the range of 10 km/h to 18 km/h.

Apart from being essential to the traffic flow theories, the basic traffic flow variables – speed, flow and density, are associated with many applications in traffic engineering, such as in road safety. Researchers are investigating the possibility to involve traffic flow variables for predicting road crashes in real time known as real-time crash prediction models. Some combinations of speed, flow and density can result in traffic conditions causing discomfort to drivers leading to mistakes that may result in as crashes (Hossain and Muramchi, 2011). In the developed world, a large number of access controlled roads such as freeways and expressways, equipped with substantial number of sensors and cameras, are used for yielding data of traffic flow variables in real-time that can serve as input to such models (Klein et al., 2006).

The research development in real-time crash prediction has mostly focused on distinguishing pre-crash data from normal traffic data through various kinds of prediction models (Lee *et al.*, 2002; Lee and Abdel-Aty, 2008; Abdel-Aty and Abdalla, 2004; Abdel-Aty *et al.*, 2006; Oh *et al.*, 2005; Lee *et al.*, 2003a), identifying their types (Golob *et al.*, 2004; Pande and Abdel-Aty, 2006a, 2006b), understanding crash mechanism (Lee *et al.*, 2002, 2003a, 2006; Luo and Garber, 2006) and exploring the possibility to introduce countermeasures to pacify hazardous traffic conditions (Abdel-Aty *et al.*, 2009; Lee and Abdel-Aty, 2003). Distinguishing hazardous traffic conditions from normal can be broadly classified into statistical approaches, such as, nonparametric Bayesian statistics (Oh *et al.*, 2005), aggregate log-linear model (Lee *et al.*, 2003b), generalized estimating equations (Abdel-Aty and Abdalla, 2004), matched case control logistic regression (Lee *et al.*, 2006; Abdel- Aty *et al.*, 2004), seemingly unrelated negative binomial regression (Abdel-Aty *et al.*, 2006) and artificial intelligence (AI) based approaches, such as, probabilistic neural network (Oh *et al.*, 2005; Abdel-Aty *et al.*, 2008), normalized radial basis function (Abdel-Aty *et al.*, 2008), naive Bayes (Luo and Garber, 2006), Bayesian Network (Hossain and Muramachi, 2013b), etc. Regarding studies concerning understanding crash mechanism, Hossain and Muramachi (2013a) identified clusters of hazardous traffic with Classification and Regression Trees. Abdel-Aty *et al.* (2005) found specific high-speed and low-speed regimes highly associated with crash. Pande and

Abdel-Aty (2006b) and Lee *et al*. (2006) ascertained that lane-changing related crashes are heavily influenced by average speeds at upstream and downstream along with difference in occupancy on adjacent lanes and standard deviation of volume and speed at a downstream location point of crash occurrence. All these studies provide ample evidence of a significant difference between pre-crash and normal traffic conditions. At present, attempts to introduce real-time interventions to smooth the traffic back to normal have commenced. Due to associated high risk, such studies cannot be conducted with real life data. Researchers have found it plausible to apply micro-simulation to develop pre-crash situation and apply various interventions, such as, variable speed limit (Abdel-Aty *et al.*, 2006a, 2008b; Lee and Abdel-Aty, 2008, Lee *et al.*, 2004), ramp metering (Abdul-Aty and Gayah, 2010; Lee *et al.*, 2006b), variable message signs (Al-Ghamdi, 2010; Lee and Abdel-Aty, 2008), etc. to bring the traffic condition back to normal. Therefore, a deeper understanding of the pre-crash traffic condition from the perspective of fundamental traffic flow theories is required. An attempt to understand the distinctions between pre-crash normal traffic conditions from the perspective of fundamental traffic flow theories is yet missing. The purpose of this research is to revisit traffic flow theories separately for pre-crash and normal traffic conditions. The specific objectives are to:

- Derive and compare classical two-phase traffic flow relationships for pre-crash and normal traffic condition data. This will include deriving four classical traffic flow theories – Greenshields, Greenberg, Underwood and Bell-shaped models, separately for pre-crash and normal traffic data.
- Evaluate longitudinal variation of traffic flow relationships both in time and space scale for pre-crash traffic data,
- Conduct hypothesis tests to detect differences in pre-crash and normal traffic conditions and discuss the variations, if any, and,
- Follow the same work flow for three-phase traffic flow theories (detector based and correlation based methods).

2 STUDY AREAS AND METHODOLOGY

Shibuya 3 and Shinjuku 4 routes of Tokyo Metropolitan Expressway, situated in the center of Tokyo Metropolitan area connecting and covering some important business and residential areas, are chosen as the study area. The length of Shibuya 3 and Shinjuku 4 routes are respectively 11.9 kilometers and 13.5 kilometers. This study area contains 210 detectors, mostly uniformly spaced, having average inter-detector spacing of around 250 meters. The detectors store data on speed, flow, occupancy and number of heavy vehicles. The crash data included information on date, time, location, vehicles involved, and lane number. Both crash and detector data were collected for 6 months (March 2014 to August 2014) of which 620 crash samples had complete data. Table 1 shows a brief description of the data (speed, flow and occupancy) used for both normal and pre-crash conditions. Figure 1 shows the box plot of the same data. Also, a map was provided which included information on location of ramps, detector positions, and section length (Figure 2). The routes being heavily access controlled ensure uninterrupted flow, which is essential for deriving classical two-phase traffic flow theories.

Table 1: Description of the data used

		Normal Conditions	Pre-crash Conditions
Occupancy (%)	Minimum	0.30	0.00
	First Quartile	4.50	6.30
	Median	7.50	13.15
	Third Quartile	14.05	28.85
	Maximum	69.80	100.00
Flow (veh/min)	Minimum	0.00	0.00
	First Quartile	20.00	20.00
	Median	32.00	28.00
	Third Quartile	44.00	38.00
	Maximum	69.00	69.00
Speed (km/hr)	Minimum	0.00	0.00
	First Quartile	53.70	16.10
	Median	72.17	38.90
	Third Quartile	81.36	68.50
	Maximum	117.85	117.85

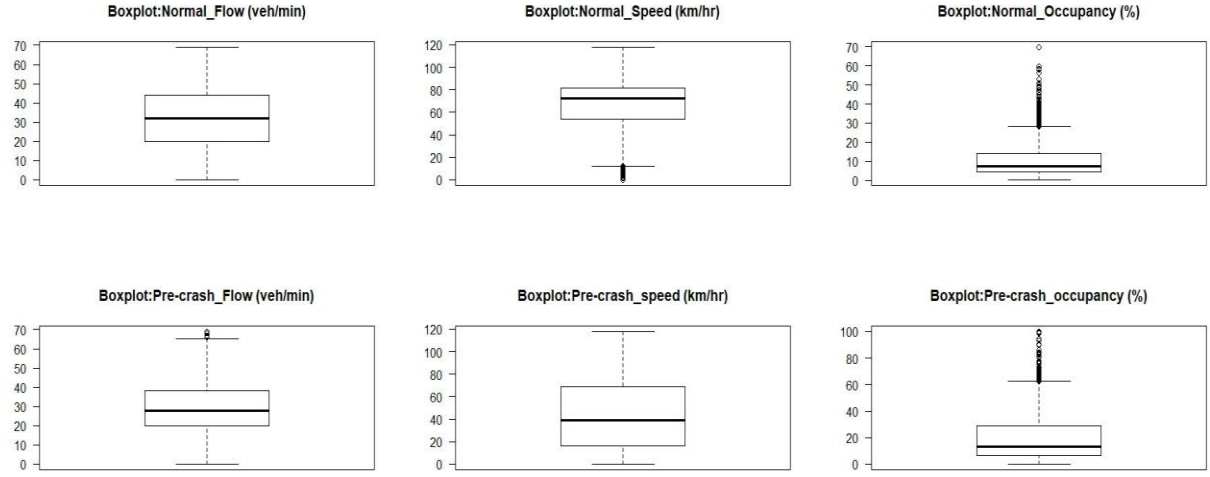

Figure 1: Boxplot of Normal and Pre-crash Data

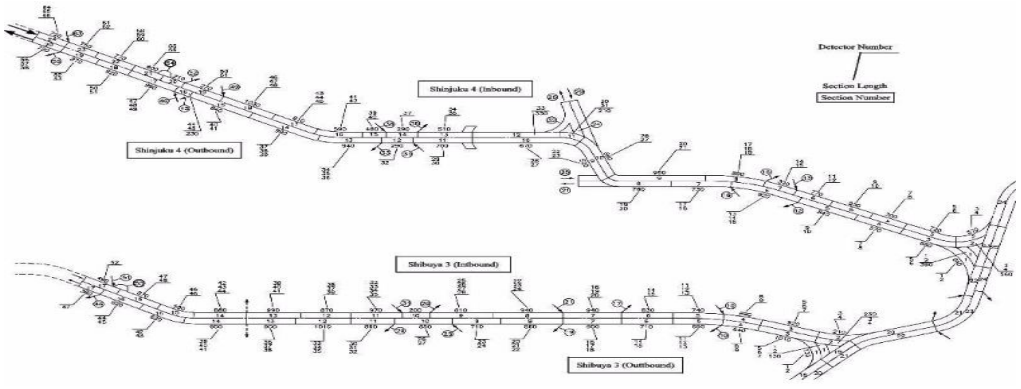

Figure 2: Shibuya 3 and Shinjuku 4 routes of Tokyo Metropolitan Expressway
Source: provided by Tokyo Metropolitan Expressway Company Limited

The overall methodology involves defining and extracting 'pre-crash' and 'normal' traffic condition data and then re-constructing the classical two-phase and three-phase traffic flow models using them for further comparison.

For two-phase theory, the 'speed-density' relationship was transformed into linear form and hypothesis tests was conducted to investigate if the differences in slopes are significant between 'pre-crash' and 'normal' traffic conditions. Finally, the differences in characteristics of the curves were explained based on traffic flow parameters. Traffic flow theories utilize speed, flow and density as the input variables. Here, although speed and flow values are obtained, the detectors provide occupancy values instead of density. Existing literature suggests that if the proportion of heavy vehicles is low (in the study area, it is around 8% only) and the detector length is roughly same (which is the case for the study area) then density and occupancy can be assumed to have a linear relationship (Kim and Hall, 2004). Hence, for model comparisons, the parameter values that were determined include, free flow speed, speed at maximum flow as well as occupancy factors during jam density and density at maximum flow conditions. For convenience and ease of understanding, the notations for the latter two parameters as well as occupancy have been kept as K_j , K_0 and *K*, however, the values should not be confused with actual jam density, density at maximum flow condition and density.

For three-phase theory the speeds of the wide moving jam at the downstream propagating upstream were calculated for pre-crash and normal traffic conditions for each crash and their values were compared.

2.1 Deriving linear forms for classical two-phase traffic flow theories

Fundamental relationship among three major variables of traffic flow theory: flow (q), speed (*V*) and density (*K*) is given below:

$$q = V \times K \tag{1}$$

Table 2 presents the fundamental non-linear relationships of Greehshields (1934), Greenberg (1959), Underwood, (1961) and Bell-Shaped (1967) classical traffic flow theories with respect to density and speed along with their linear transformations where K_j = jam density; V_f = free flow speed; K_0 = density at maximum flow; V_0= speed at maximum flow level; *K*= density and *V*=speed. As mentioned earlier, occupancy data were used instead of density and the notations K_j, K_0 and *K* used in this manuscript stand for occupancy factors during jam density, density at maximum flow conditions and occupancy. 'R-program' was used to develop the regression models.

Table 2: Transformation from speed-density function

Model	Speed-density Function	Transformation
Greenshields	$K = K_j(1 - \frac{V}{V_f})$	$K = K_j - K_j\frac{V}{V_f}$
Greenberg	$K = K_j e^{\frac{-V}{V_0}}$	$lnK = ln(K_j) + \left(-\frac{1}{V_0}\right)V$
Underwood	$K = K_0 ln(\frac{V_f}{V})$	$K = K_0\, ln(V_f) - K_0 ln(V)$
Bell Shape	$K = K_0(2ln\frac{V_f}{V})^{\frac{1}{2}}$	$K^2 = 2K_0^2\, ln(V_f) - 2K_0^2 ln(V)$

2.2 Methods for three-phase traffic flow theory

The flow-density method is ideally suited for long freeway sections without the presence of

on or off ramps. Our study area, being an urban expressway, is densely packed with ramps. The graphical method processes the image with a proprietary software called ASDA/FOTO. Hence, this study choses only detector based and the correlation methods.

2.2.1 Detector based Method

Two detectors *b* and *a* with sufficiently large distances are chosen which is then divided by the difference in the registration times of the downstream jam front of each detector. The velocity of the downstream jam front can be calculated as:

$$v_{gr} = \frac{Loc(b) - Loc(a)}{T_a(when\ v_{a(t)} > 30\ kmph) - T_b(when\ v_{b(t)} > 30\ kmph)} \quad (2)$$

Here *Loc(b)* and *Loc(a)* are the exact locations of the chosen detectors and T_b and T_a are the specific times in minutes when the real-time speeds are just greater than 30 km/h after a wide moving jam.

2.2.2 Correlation Method

The velocity of a downstream detector *b* can be correlated with its upstream counterpart *a* (Munoz and Daganzo, 2002) as follows:

$$\mathrm{correl}(q)_k^{(b,a)} = \frac{\frac{1}{n}\sum_{i=1}^{n}((q_i^a - \bar{q}^a)\times(q_{i+k}^b - \bar{q}^b))}{\sqrt{(\frac{1}{n}\sum_{i=1}^{n}(q_i^a - \bar{q}^a)^2)}\times\sqrt{(\frac{1}{n}\sum_{i=1}^{n}(q_{i+k}^b - \bar{q}_k^b)^2)}} \quad (3)$$

Here,
q_i^b = Flow of the downstream detector b
$\bar{q}^b$ = Average flow of the downstream detector b
q_{i+k}^a = Flow of the upstream detector a with a time lag of k minutes
$\bar{q}^a$ = Average flow of the upstream detector a
$\bar{q}_k^a$ = Average flow of the upstream detector a for the time lag of k minutes = $\frac{1}{n}\sum_{i=1}^{n} q_{i+k}^a$

(i=Time index, k= Time lag in minutes)
A cross-correlation function is then used for measuring the wave velocity for each amount of time lag used.

$$u = 60\left(\frac{Loc(b) - Loc(a)}{k}\right) \quad (4)$$

The value 60 converts *u* into km/h. Then a graph of flow correlation versus wave velocity is produced. The velocity, which has the highest correlation, is the downstream jam velocity for the traffic flow.

2.3 Comparing linear regression models of two independent samples

Comparing slopes of simple linear regression models is a common practice in statistics which

is conducted often to test whether the linear regression lines derived from different samples are coming from same population. Assume that two simple linear regression lines, having slopes b_1 and b_2 were constructed with two samples having form of Equation 5.

$$y = \beta x + c \tag{5}$$

To test if they come from the same population:

1. Test the null hypothesis: H_0: $= \beta_1 = \beta_2$, *i.e.*, $\beta_1 - \beta_2 = 0$.
2. The t-test statistics will be:
$$t = \frac{b_1 - b_2}{\sqrt{s_{b_1}^2 + s_{b_2}^2}} \sim T(n_1 + n_2 - 4) \tag{6}$$
3. If the null hypothesis is true then the difference in the slopes will form a normal distribution with mean to be zero, i.e.,
$\beta_1 - \beta_2 \sim N(0, s_{b_1 - b_2})$
where,
$$s_{b_1 - b_2} = \sqrt{s_{b_1}^2 + s_{b_2}^2} \tag{7}$$

Finally, a value for level of significance (α) will be set. If p-value is less than α then the difference in slope will be considered as significant. Otherwise, the null hypothesis cannot be rejected. 'R-program' was used to conduct the hypothesis test (Dalgard, 2008).

2.4 Detector Data Extraction and Preparation

2.4.1 For Two-Phase Traffic Flow Theory

The 'pre-crash' data were defined as crash data within the vicinity of the crash location just before its occurrence time where 'vicinity' refers to a zone bordered by the nearest upstream (UD) and the nearest downstream (DD) detector locations with respect to crash location as presented with Figure 3. The study area has two lanes in each direction and each lane has one detector installed, i.e., there are two detectors at each detector location. Hence, the values of speed and occupancy were averaged and flows were added to obtain the final values of speed, flow and occupancy at each detector location. For every crash case, data were extracted for speed, flow and occupancy for every minute up to the tenth minute prior to crash as XD_tY where X = U (Upstream) or D (Downstream), Y = S (Speed) or F (Flow) or O (Occupancy) and t = 1, 2, …., 9. Hence, UD_6S represents the speed at the nearest upstream detector 6 minutes before crash. For 'pre-crash' data, 18 datasets (2 for upstream and downstream X 9 for the 9 time slices) each containing 620 records were created and named as X_t where X = U (Upstream) or D (Downstream) and t = 1, 2, …., 9. For example, D_7 represents a dataset having speed, flow and occupancy data of all the 620 crashes collected from the nearest downstream detector seven minutes prior to crash.

The 'normal' traffic data were defined as speed, flow or occupancy data of any time at any location on the expressway where there was no crash taking place three hours before or after that time. To construct this dataset, data were extracted for each detector for each month in such way that it contained 24 hour data of at least one Sunday, Monday, Tuesday and so on for every month during the study period. To avoid bias, instead of selecting just one sample, four normal datasets were constructed by randomly selecting 1000 samples from the complete 'normal' traffic dataset.

The authority supplied raw detector data for each lane covering information on speed in kilometers per hour, flow in number of vehicles, occupancy in percentage and number of heavy vehicles aggregated for every minute and stored in separate files for each day. The data were further aggregated for two lanes for each detector location by taking the average speed and occupancy of the two detectors and adding the two flow values. Also, it was necessary to transpose and rearrange the dataset for further use as the original dataset kept data of all detectors for a specific one minute in just one record. A program was written in C++ to extract the data and arrange it in a single file under these headers – Route number, direction (inbound or outbound), detector ID, date, time, flow, speed and occupancy. Finally, the detector data as well as crash data were imported into PostgreSQL relational database system (Maymala, 2015) and the 'pre-crash' and 'normal' traffic condition datasets were generated after running a set of sub-queries fulfilling their definitions.

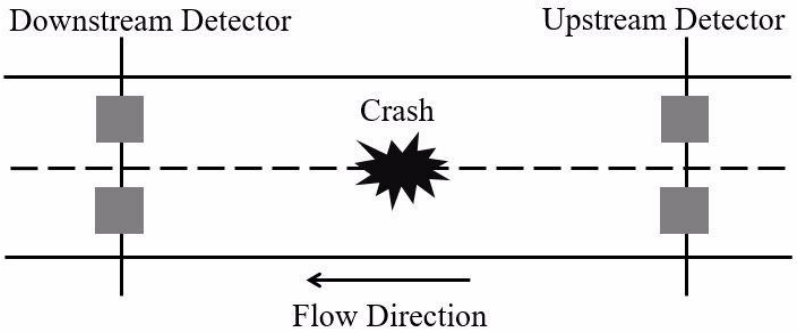

Figure 3: Detector location with respect to crash location

2.4.2 For Three-Phase Traffic Flow Theory

Literature suggests that for detector based methods, for A5 freeway of Germany, distance of 4.7 km was used for sections having no intersections or on or off ramps and a distance of 9 km was used for sections having one intersection. (Rehborn *et al.* 2011). In this study, at first, sections with no on and off ramp were selected randomly from both directions of Shibuya 3 route of Tokyo Metropolitan Expressway. A total of 148 crashes were found within those sections. Each of the chosen sections had a pair of detectors on each end. From the crash time, 1 hour traffic flow data from the corresponding detectors were compiled to represent the pre-crash data. Unlike the two phase analysis, 1 hour pre-crash data was used instead of 10 minutes. This is because, for our data it was observed that at least 1 hour is required for a wide moving jam to occur and then to dissolve to synchronized flow. To prepare the normal traffic dataset, from these chosen samples, randomly 1 hour samples were chosen when no crash took place in the corresponding direction of the route of the expressway at least 1 hour before and after that time.

3. ANALYSIS AND RESULTS

The results obtained following the outlined methodology are organized into five sub-sections of which 3.1-3.4 are for two-phase and 3.5 is for three-phase theory. For two-phase theory, firstly, a detailed analysis was conducted on the four sets of the normal traffic data. This included producing classical scattered plots of speed-flow-occupancy, their linear transformations representing the four classical relationships and finally, a hypothesis test to calculate the variations among the four datasets. Secondly, the classical traffic flow

relationships were derived for the 18 pre-crash datasets. Thirdly, the hypothesis tests were conducted to investigate if the 'pre-crash' relationships were significantly different from 'normal' relationships. Finally, the findings were combined and presented in a systematic manner for two-phase traffic flow theory

3.1 Deriving relationships for normal traffic condition

Figure 4 and 5 illustrate the matrix scatter plot and the regression lines developed using density and speed data for normal traffic condition (Dataset 1) for the four models (Greenshields, Greenberg, Underwood, Bell Shape). Table 3 and 4 present the regression equations, their classical forms as well as corresponding R^2, occupancy factor at jam density (K_j); free flow speed (V_f); occupancy factor for density at maximum flow (K_0) and speed at maximum flow (V_0) values for all four datasets. All the models were found to be significant as reflected by the p-values. For the three-phase traffic flow theory, the value of wide moving jam for every crash case were calculated and compared with values when the traffic condition was normal.

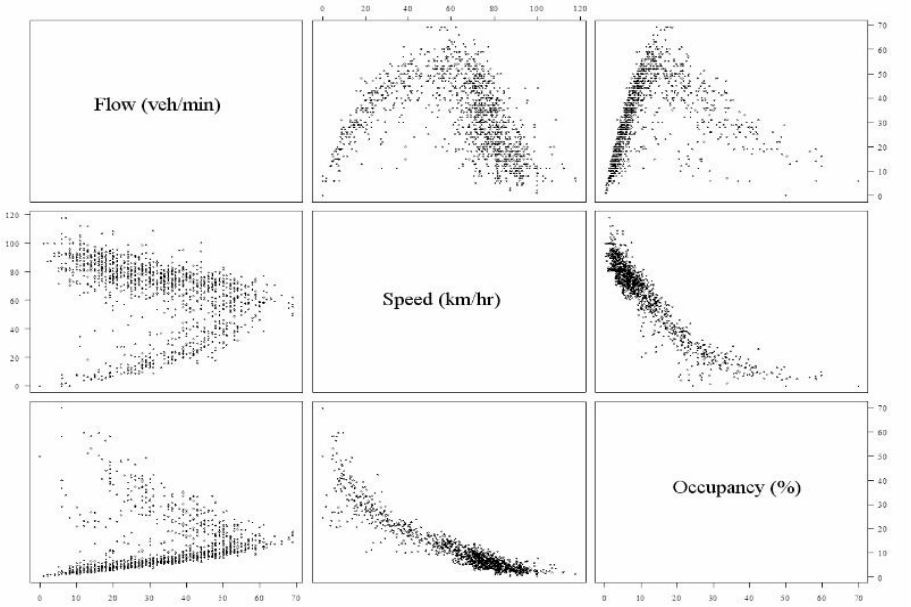

Figure 4: Matrix of scatter plots for normal traffic flow data (Dataset 1)

The speed-flow, speed-occupancy and flow-occupancy relationships exhibit trends following the findings of existing literature (Figure 4). Combining the outcomes of Figure 5 and Table 4, it can be observed that the values of coefficient of determination (R^2) are quite high for Greenshields, Underwood and Bell-shaped models, consistently explaining more than 75% of the variations of speed in estimating density. For these models, the values of R^2 ranged between 0.74-0.84, 0.79-0.86 and 0.71-0.83 respectively. However, the R^2 values were found to be low for Greenberg model ranging from 0.58 – 0.67, except for Dataset 1 where it was found to be 0.8. Expressways in general operate at high speed with low density which may have influenced the goodness of fit for Greenberg model as its performance is known to be unstable for such traffic conditions. Although the free flow speeds remain almost identical within the models for all four datasets, e.g., ranging from 94 – 96 kilometers per hour for Greenshields model, they exhibited variations among different models, ranging from as high as 120 kilometers per hour for the Dataset 1 for Underwood model to as low as 80 kilometres

per hour for Bell-shaped model developed with Dataset 2. In general, Underwood models estimated a high free flow speed, Bell shaped models estimated a low value and Greenshields model estimated a free flow value nearly the average of these two values. The speed at maximum flow conditions (V_0), which is an output of the Greenberg model, had slight variations among the four datasets, ranging between 30 to 34 kilometers per hour. The occupancy factors for maximum flow condition (K_0) remain for both Underwood (14.91 to 15.86) and Bell-shaped models remain close (16.45 to 17.67). Regarding the occupancy factors for jam density (K_j), the values vary substantially between Greenshields (31.41 to 36.57) and Greenberg (48.51 to 64.79) models, though, the variation within the four Greenshields models are slight.

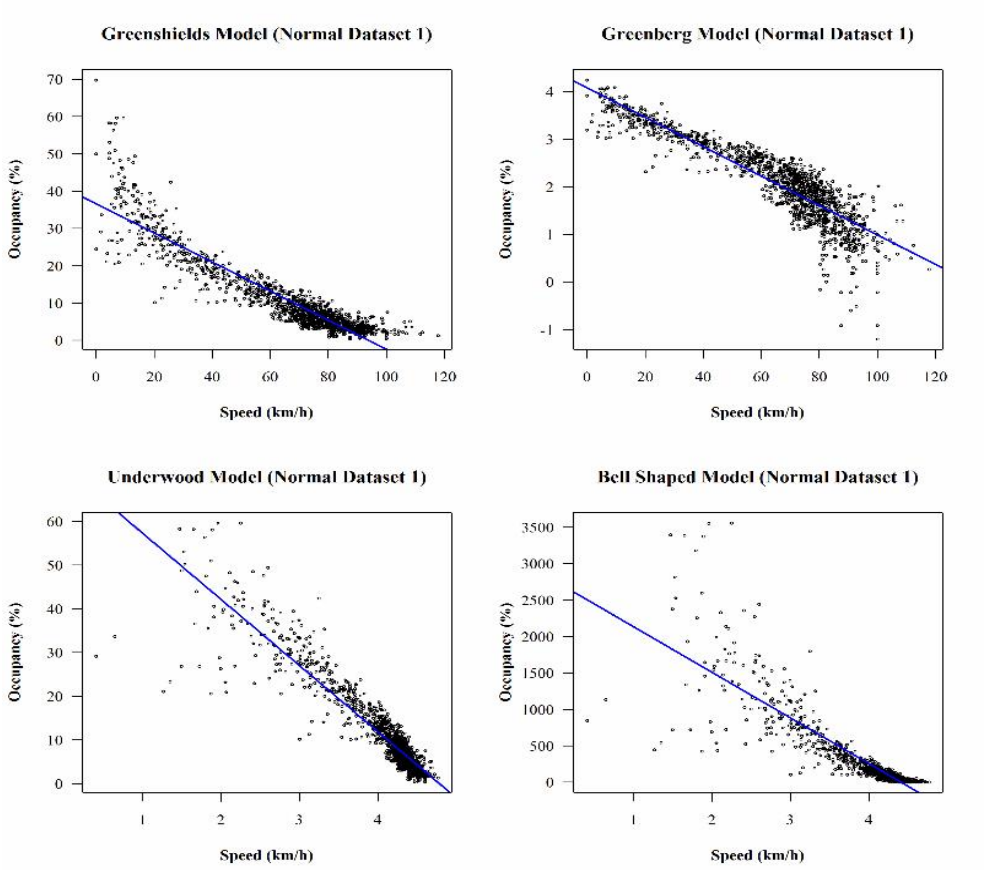

Figure 5: Regression lines in four models for Normal flow data (Dataset 1)

Table 3: Regression and classical equations for normal traffic condition

Dataset	Model	Regression line equation	Transformation
1	Greenshields	$K = 36.59 - 0.389V$	$K = 36.586(1 - \frac{V}{93.95})$
	Greenberg	$lnK = 4.082 - 0.0309V$	$K = 59.23e^{(-\frac{V}{32.34})}$
	Underwood	$K = 72.41 - 15.13lnV$	$K = 15.1288ln\frac{119.86}{V}$
	Bell Shape	$K^2 = 2754.9 - 624.42lnV$	$K = 17.67(2ln\frac{82.43}{V})^{\frac{1}{2}}$
2	Greenshields	$K = 32.36 - 0.344V$	$K = 32.3586(1 - \frac{V}{94.0486})$

	Greenberg	$lnK = 3.88 - 0.029V$	$K = 48.51e^{(-\frac{V}{34.4246})}$
	Underwood	$K = 70.79 - 14.905lnV$	$K = 14.9053ln\frac{111.716}{V}$
	Bell Shape	$K^2 = 2534 - 578.19lnV$	$K = 17.0028(2ln\frac{80.052}{V})^{\frac{1}{2}}$
3	Greenshields	$K = 31.41 - 0.329V$	$K = 31.4118(1 - \frac{V}{95.5757})$
	Greenberg	$lnK = 4.17 - 0.033V$	$K = 64.7907e^{(-\frac{V}{30.25})}$
	Underwood	$K = 73.2 - 15.46lnV$	$K = 15.459ln\frac{113.8755}{V}$
	Bell Shape	$K^2 = 2388.1 - 541.13lnV$	$K = 16.44886(2ln\frac{82.535}{V})^{\frac{1}{2}}$
4	Greenshields	$K = 32.56 - 0.346V$	$K = 32.56(1 - \frac{V}{94.227})$
	Greenberg	$lnK = 4.015 - 0.031V$	$K = 55.412e^{(-\frac{V}{32.188})}$
	Underwood	$K = 74.88 - 15.86lnV$	$K = 15.8594ln\frac{112.31}{V}$
	Bell Shape	$K^2 = 2545.2 - 579.53lnV$	$K = 17.0224(2ln\frac{80.7876}{V})^{\frac{1}{2}}$

Table 4: R^2 and estimated parameters value for normal condition

Dataset	Model	R^2	V_f	V_0	K_0	K_j
1	Greenshields	0.84	93.95			36.57
	Greenberg	0.80		32.34		59.24
	Underwood	0.86	119.86		15.13	
	Bell Shape	0.75	82.43		17.67	
2	Greenshields	0.76	94.05			32.36
	Greenberg	0.67		34.42		48.51
	Underwood	0.84	111.72		14.91	
	Bell Shape	0.75	80.05		17.00	
3	Greenshields	0.74	95.58			31.41
	Greenberg	0.58		30.25		64.79
	Underwood	0.79	113.88		15.46	
	Bell Shape	0.71	82.54		16.45	
4	Greenshields	0.76	94.23			32.56
	Greenberg	0.60		32.19		55.41
	Underwood	0.83	112.31		15.86	
	Bell Shape	0.83	80.79		17.02	

3.2 Comparing slopes for linear transformation of normal traffic condition data

The analysis results of the previous sub-section suggests the variation among four datasets to be low when compared within the four classical traffic flow theories separately, specially for Greenshields, Underwood and Bell-shaped models. However, to test if the variation between 'pre-crash' and 'normal' traffic conditions, it is important to verify that the four datasets of normal traffic conditions are representatives of the population of normal traffic. For that, hypothesis tests were conducted (α = 0.05) as outlined in Section 2.3. As initially none of the normal traffic condition datasets can be considered as the representative of the population, all possible combinations were tested by each time picking two datasets. Hence, a total of

$4 \times {}^{4}C_{2} = 24$ combinations were tested and the results of whether the null hypothesis, i.e., the slopes of the linear relationships are same or, the difference in slopes equal to zero, could be rejected are presented with Table 5 comparing the p-value and the level of significance (α = 0.05). Here, 'Y' represents that p-value is less than α , i.e., the difference is significant and 'X' represents otherwise. Also, ND1_ND2 suggests that the comparison was made between normal traffic dataset 1 and 2, and so on. The outcome suggests that there were no significant differences among the four datasets while linearly fitting the Greenshields and Underwood models. However, linear transformations of Greenberg models constructed with Dataset 2 and 3 as well as Bell-Shaped models with Dataset 3 and 4 were found to be significantly different suggesting that they may not have come from the same population. The poorer goodness-of-fit for Greenberg and Bell-Shaped models for Dataset 3 may further justify the findings (Table 4).

Table 5: Hypothesis test for slope comparison (normal traffic data)

Model	**ND1_ND2**	**ND1_ND3**	**ND1_ND4**	**ND2_ND3**	**ND2_ND4**	**ND3_ND4**
GS	X	X	X	X	X	X
GB	X	X	X	Y	X	X
UW	X	X	X	X	X	X
BS	X	Y	X	X	X	Y

3.3 Deriving relationships for pre-crash traffic condition

3.3.1 Pre-crash traffic condition at nearest upstream detector (UD)

From the time one minute before crash until ten minutes before the crash, data for the nearest upstream detector were extracted and nine separate models were constructed for each of the four classical traffic flow models. The resulting coefficient of determination (R^2) values and the models are presented with Table 6 and 7. All the models were found to be significant as reflected by the p-values.

Table 6: Coefficient of Determination (R^2) values for pre-crash conditions (UD)

Model Name	**T=10**	**T=9**	**T=8**	**T=7**	**T=6**
Greenshields	0.54	0.54	0.53	0.55	0.54
Greenberg	0.66	0.69	0.64	0.65	0.67
Underwood	0.73	0.70	0.73	0.71	0.73
Bell Shape	0.65	0.54	0.55	0.59	0.59
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	0.56	0.53	0.52	0.55	0.55
Greenberg	0.70	0.61	0.69	0.70	0.66
Underwood	0.74	0.75	0.71	0.78	0.76
Bell Shape	0.59	0.61	0.50	0.62	0.63

It is quite vivid from Table 6 that the overall goodness of fit for all the four models for all the time slices are substantially lower than that of the four sets of normal traffic data. Apart from the Underwood models, which performed the best in case of normal traffic conditions, the other three models had R^2 value less than 0.7 in almost all the cases. No specific trend can be

identified by observing the variations of R^2 values over the ten minutes for any of the models. From classical equations in Table 7 and parameter values in Table 4, it can be observed that the values of V_f for Greenshields and Bell-Shaped models are considerably lower than normal condition as it varies from 86.18 to 89.07 kilometers per hour for Greenshields and 67.36 to 73.03 kilometers per hour for Bell-Shaped for pre-crash conditions sensed at the nearest upstream detector from the crash location as compared to 93.95 to 95.58 kilometers per hour for Greenshields and 80.05 to 82.54 kilometers per hour for Bell-shaped models for normal traffic data. On the contrary, Underwood model estimates a higher value of V_f , ranging between 109.64 to 126.42 kilometers per hour whereas the normal condition suggested a value between 111.72 to 119.86 kilometers per hour. The speed at maximum flow, i.e., V_0 is an output of Greenberg model and it suggests a higher speed (35.32 to 37.12 kilometers per hour) during pre-crash situation as compared to normal traffic conditions (30.25 to 34.42 kilometers per hour). The K_0 values, obtained from Underwood and Bell-Shaped models, range between 14.1 to 15.11 and 18.62 to 20.37 respectively, suggesting a higher level of compactness for the pre-crash traffic conditions than the normal traffic conditions at maximum flow. Finally, the K_j values exhibit a sharp contrast between Greenshields and Greenberg model as the former (34.41 to 36.25) suggests a more compact condition at jam density but the later (41.97 to 44.95) indicates less compaction when compared with normal traffic conditions (31.41 to 36.57 for Greenshields and 48.51 to 64.79 for Greenshields). Combining the findings, it can be understood that as most of the crashes take place during congested periods for the study area under consideration, the free flow speed (V_f) was found to be lower for pre-crash conditions. Also, Greenshields, Underwood and Bell-Shaped models suggest a maximum flow condition which entertains a flow where the vehicles are moving more closely but at a higher speed than the normal traffic condition - which can be associated with higher crash risk.

Table 7: Classical Equations (speed-occupancy) for pre-crash data (UD)

Model Name	**T=9**	**T=8**	**T=7**
Greenshields	$K = 35.34(1 - \frac{V}{89.07})$	$36.25(1 - \frac{V}{87.42})$	$35.23(1 - \frac{V}{87.26})$
Greenberg	$K = 44.95e^{(-\frac{V}{36.57})}$	$43.6e^{(-\frac{V}{37.12})}$	$43.53e^{(-\frac{V}{36.57})}$
Underwood	$K = 14.1ln\frac{126.42}{V}$	$14.57ln\frac{121.15}{V}$	$14.13ln\frac{123.72}{V}$
Bell Shape	$K = 19.16(2ln\frac{71.92}{V})^{\frac{1}{2}}$	$20.37(2ln\frac{67.31}{V})^{\frac{1}{2}}$	$18.62(2ln\frac{73.03}{V})^{\frac{1}{2}}$
	T=6	**T=5**	**T=4**
Greenshields	$K = 36.04(1 - \frac{V}{86.89})$	$35.35(1 - \frac{V}{8.81})$	$35.24(1 - \frac{V}{86.23})$
Greenberg	$K = 44.04e^{(-\frac{V}{36.52})}$	$43.72e^{(-\frac{V}{36.16})}$	$42.06e^{(-\frac{V}{36.08})}$
Underwood	$K = 14.37ln\frac{120.67}{V}$	$14.51ln\frac{117.42}{V}$	$14.97ln\frac{112.29}{V}$
Bell Shape	$K = 19.15(2ln\frac{71.16}{V})^{\frac{1}{2}}$	$19.1(2ln\frac{70.61}{V})^{\frac{1}{2}}$	$19.63(2ln\frac{68.62}{V})^{\frac{1}{2}}$
	T=3	**T=2**	**T=1**
Greenshields	$K = 34.81(1 - \frac{V}{87.58})$	$35.72(1 - \frac{V}{86.18})$	$34.41(1 - \frac{V}{86.18})$

Greenberg	$K = 43.39e^{(-\frac{V}{35.57})}$	$43.5e^{(-\frac{V}{35.32})}$	$41.97e^{(-\frac{V}{35.67})}$
Underwood	$K = 14.55ln\frac{115.68}{V}$	$15.11ln\frac{109.64}{V}$	$14.63ln\frac{112.04}{V}$
Bell Shape	$K = 18.84(2ln\frac{72.72}{V})^{\frac{1}{2}}$	$19.84(2ln\frac{67.36}{V})^{\frac{1}{2}}$	$18.74(2ln\frac{69.97}{V})^{\frac{1}{2}}$

3.3.2 Pre-crash traffic condition at nearest downstream detector (DD)

Similar to Section 3.3.1, nine models were built with the pre-crash data collected from the nearest downstream detectors from the crash location and the results of their goodness-of-fit and classical formulations are presented with Table 8 and 9. The corresponding p-values suggest that all the models were significant.

Table 8: Coefficient of Determination (R^2) values for pre-crash conditions (DD)

Model Name	**T=10**	**T=9**	**T=8**	**T=7**	**T=6**
Greenshields	0.59	0.61	0.58	0.60	0.59
Greenberg	0.66	0.67	0.75	0.76	0.74
Underwood	0.78	0.82	0.79	0.79	0.77
Bell Shape	0.64	0.71	0.64	0.66	0.60
	T=5	**T=4**	**T=3**	**T=2**	**T=1**
Greenshields	0.60	0.59	0.63	0.60	0.60
Greenberg	0.73	0.70	0.78	0.74	0.74
Underwood	0.79	0.78	0.79	0.81	0.78
Bell Shape	0.65	0.66	0.64	0.68	0.59

Similar to the models built with upstream pre-crash data, these models also exhibit lower goodness-of-fit as compared to the datasets representing four normal traffic conditions. However, the R^2 values are slightly higher for all four classical models for the pre-crash data from downstream detectors. The Underwood models performed the best with R^2 values ranging from 0.78 to 0.82. Greenberg models also showed much improvement as their R^2 values varied from 0.66 to 0.78. However, no pattern could be seen in the performance variation among the models with time.

As compared to the models built with upstream data, these models (Table 9) exhibit a lower free-flow speed (V_f), i.e., substantially lower than that of normal traffic conditions, for all four classical models. The values for Greenshields, Underwood and Bell-Shaped models ranged between 82.26 to 88.38, 103.85 to 114.24 and 67.78 to 69.98 kilometers per hour. The speed at maximum flow (V_0), yielded by the Greenberg model, was also lower (32.57 to 36.46 kilometers per hour) than that with upstream detector data but still higher than the normal traffic condition data (30.25 to 34.42 kilometers per hour). Regarding occupancy factors for density at maximum flow (K_0), the downstream detectors suggest a higher level of compaction (15.53 to 16.21 for Underwood and 19.76 to 20.38 for Bell-Shaped) as compared to the upstream detectors for pre-crash traffic conditions. Following the trend for K_0, the Greenshields model present a higher level of compaction for the occupancy factors at jam density (K_j). However, the corresponding values for Greenberg models are higher than that of the upstream dataset but still considerably lower than that of normal traffic conditions.

Table 9: Classical Equations (speed-occupancy) for pre-crash data (DD)

Model Name	**T=9**	**T=8**	**T=7**

Greenshields	$K = 36.8(1 - \frac{V}{88.38})$	$36.82(1 - \frac{V}{87.57})$	$37.23(1 - \frac{V}{87.28})$
Greenberg	$K = 45.73e^{(-\frac{V}{36.46})}$	$47.65e^{(-\frac{V}{35.41})}$	$48.02e^{(-\frac{V}{34.81})}$
Underwood	$K = 15.67ln\frac{114.24}{V}$	$15.88ln\frac{112.11}{V}$	$15.88ln\frac{111.7}{V}$
Bell Shape	$K = 20.04(2ln\frac{69.62}{V})^{\frac{1}{2}}$	$20.38(2ln\frac{68.32}{V})^{\frac{1}{2}}$	$19.94(2ln\frac{69.96}{V})^{\frac{1}{2}}$
	T=6	**T=5**	**T=4**
Greenshields	$K = 38.45(1 - \frac{V}{85.78})$	$38.65(1 - \frac{V}{84.39})$	$37.45(1 - \frac{V}{85.26})$
Greenberg	$K = 48.86e^{(-\frac{V}{34.38})}$	$48.44e^{(-\frac{V}{33.69})}$	$45.91e^{(-\frac{V}{34.19})}$
Underwood	$K = 15.76ln\frac{113.38}{V}$	$15.78ln\frac{111.02}{V}$	$15.53ln\frac{109.13}{V}$
Bell Shape	$K = 20.25(2ln\frac{69.98}{V})^{\frac{1}{2}}$	$20.37(2ln\frac{68.24}{V})^{\frac{1}{2}}$	$19.76(2ln\frac{68.3}{V})^{\frac{1}{2}}$
	T=3	**T=2**	**T=1**
Greenshields	$K = 38.52(1 - \frac{V}{83.87})$	$38.24(1 - \frac{V}{83.03})$	$38.92(1 - \frac{V}{82.26})$
Greenberg	$K = 49.17e^{(-\frac{V}{32.57})}$	$46.86e^{(-\frac{V}{33.39})}$	$47.99e^{(-\frac{V}{32.64})}$
Underwood	$K = 15.97ln\frac{107.06}{V}$	$15.53ln\frac{107.26}{V}$	$16.21ln\frac{103.85}{V}$
Bell Shape	$K = 19.97(2ln\frac{68.9}{V})^{\frac{1}{2}}$	$20.48(2ln\frac{64.78}{V})^{\frac{1}{2}}$	$20.71(2ln\frac{65.64}{V})^{\frac{1}{2}}$

Although the trends for both the pre-crash datasets (upstream and downstream) have been very similar in nature, due to better goodness of fit and clearer distinction with normal traffic conditions, i.e., larger differences in values for free-flow speed (V_f), speed at maximum flow (V_0), density at maximum flow (K_0), and occupancy factors at jam density (K_j), the nearest downstream detectors with respect to crash locations can be considered to better explain traffic conditions relating to crashes. Among the four classical models, Underwood models for both upstream and downstream data exhibited better goodness-of-fit. At the same time, comparing results from both upstream and downstream detectors, it can be observed that several crashes on the study area were taking place where a densely packed but fast moving upstream was meeting a relatively denser and slower moving downstream traffic.

3.4 Comparing slopes of normal and pre-crash data of estimated parameters

Finally, hypothesis tests (α = 0.05) were conducted to investigate whether the differences in slopes of the models derived with pre-crash datasets and normal datasets are significantly different. As Section 3.2 suggests that all four normal traffic condition data except for Dataset 3 can be considered to come from the same population, 1000 records were randomly selected from Dataset 1, 2 and 4 to construct a new normal traffic condition dataset to produce four classical models. Lastly, each of these models were compared with the corresponding models for both upstream and downstream datasets for all nine pre-crash time segments. Hence, a total of 4 (classical models) X 2 (upstream and downstream detectors) X 9 (time segments) = 72 slope comparisons were made as presented with Table 10 containing the outcomes whether the null hypothesis was rejected (Y) or could not be rejected (X). The details of the process

are explained in Section 3.3 and 4.2. Upstream and Downstream detector positions along with the time-segment have been represented as XD_t where X = U (upstream) or D (downstream), and t = 1,2,….,9. The findings from Table 10 illustrates that other than the linear transformation of Underwood model with data taken from the immediate upstream detectors of the crash locations just two minutes before the crashes, all other models for all other time segments were found to be significantly different from traffic conditions that existed during times when there were no crash hazards in the study area.

Table 10: Hypothesis test for slope comparison (normal vs. pre-crash)

Model	**UD_1**	**UD_2**	**UD_3**	**UD_4**	**UD_5**	**UD_6**	**UD_7**	**UD_8**	**UD_9**
GS	Y	Y	Y	Y	Y	Y	Y	Y	Y
GB	Y	Y	Y	Y	Y	Y	Y	Y	Y
UW	Y	X	Y	Y	Y	Y	Y	Y	Y
BS	Y	Y	Y	Y	Y	Y	Y	Y	Y
	DD_1	**DD_2**	**DD_3**	**DD_4**	**DD_5**	**DD_6**	**DD_7**	**DD_8**	**DD_9**
GS	Y	Y	Y	Y	Y	Y	Y	Y	Y
GB	Y	Y	Y	Y	Y	Y	Y	Y	Y
UW	Y	Y	Y	Y	Y	Y	Y	Y	Y
BS	Y	Y	Y	Y	Y	Y	Y	Y	Y

* GS = Greenshields, GB = Greenberg, UW = Underwood, BS = Bell-Shaped

3.5 Comparison of Pre-Crash and Normal traffic data for three-phase traffic flow theory

For each of the pre-crash data, flow versus time graphs and their corresponding speed versus time graphs are produced. Time stamps are numbers assigned consecutively to each of the time intervals within the chosen duration. These graphs are used to identify the wide moving jam phase within the traffic for a particular duration. A wide moving jam is associated with a decrease in flow as well as a decrease in speed. One set of example is demonstrated with Figure 6. Here, duration from 15:40 to 16:40 was chosen with 1 minute time intervals. Hence, the first interval "15:40 -15:41" is assigned a time stamp of 1 and the last interval "16:39-16:40" is assigned a time stamp of 60. Also, we can observe from the figure that after the 40th minute, a major dip in both the flow and speed is observed simultaneously, indicating a wide moving jam.

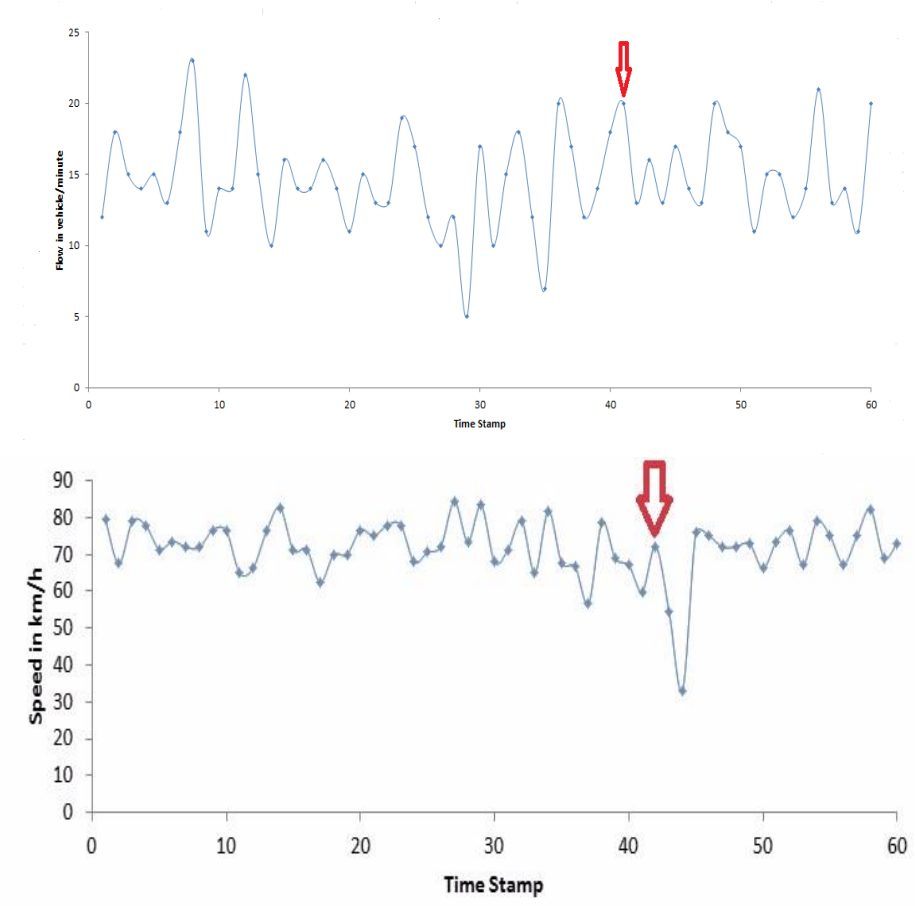

Figure 6: (a) flow versus time graph for one hour interval, (b) its corresponding speed versus time graph.

After identifying the wide moving jam in the same way as detector based methods, time duration of 25 minutes is chosen from within the jam duration. Then the flow correlation between the upstream and the downstream detector and the corresponding wave velocity for a time lag of 1 minute to 15 minutes are calculated by using equation (3) and equation (4). Then, using these results, a figure of correlation values versus velocity is plotted as illustrated by Figure 7. From the figure, the velocity with highest correlation value would be the downstream jam velocity. As demonstrated with Figure 7, the maximum correlation of 0.55 is at velocity of 13.89 km/h.

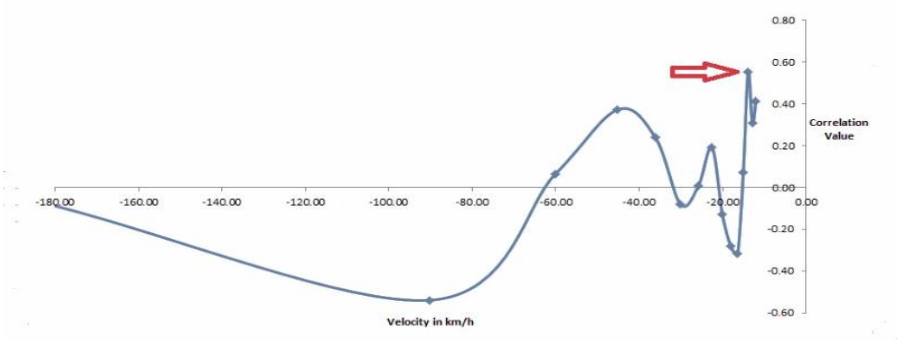

Figure 7: Correlation versus velocity diagram with the highest correlation value of 0.55 at around 13.89 km/h

The same calculation was carried out for all 148 crash samples and their corresponding pre-crash and normal traffic data. Table 11 presents that the minimum, maximum, average downstream jam velocity and their standard deviations for different methods as well as different conditions. As all the values lay between -10 to -18 kph, just like two-phase traffic flow theory, both the pre-crash and normal traffic conditions are valid for three-phase traffic flow theory as well. It can be observed that the results yielded by both the methods are quite consistent where the pre-crash traffic exhibits a higher wide moving jam velocity propagating from downstream to upstream through traffic than the normal traffic condition.

Table 11: Maximum, minimum, average and the standard deviation of the wide moving jam velocities for different conditions for detector based and correlation based method

	Detector Based Method		**Correlation Based Method**	
	Pre-crash	**Normal**	**Pre-crash**	**Normal**
Maximum	-18.60	-13.89	-18.06	-12.04
Minimum	-11.10	-10.62	-11.93	-11.29
Average	-14.38	-12.37	-14.32	-11.67
Standard Deviation	2.27	1.34	2.33	0.31

4 CONCLUSIONS AND DISCUSSION

The findings suggest that all four classical models of two-phase traffic flow theory exhibit high goodness of fit for normal traffic data. Apart from Greenberg models, the estimated R^2 value for Greenshields, Underwood and Bell-Shaped models were over 0.75 for all four normal traffic data samples. The nearest downstream detectors from the crash locations fit the pre-crash data better than the upstream detectors. For pre-crash conditions, data were collected from both nearest upstream and downstream detectors from the crash locations for every minute, up to nine minutes before crash. A total of 72 models were built considering nine time segments, two locations (upstream and downstream) and four classical traffic flow theory models. A sharp drop in goodness-of-fit as compared to the normal traffic condition were seen for all these 72 models, though, just like models with normal traffic data, all the models with pre-crash data were found to be statistically significant as well. Underwood and Greenberg models had higher goodness-of-fit for both upstream and downstream pre-crash data and Greenshields performed the worst for the same datasets. The values for the free-flow speed, speed at maximum flow, and parameters corresponding to jam density and density at maximum flow suggest that vehicle stream is more compacted for pre-crash conditions as compared to normal conditions. Also, at maximum flow conditions, the pre-crash traffic stream moves at a higher speed than the normal traffic speed. The pattern was visible for all four classical models and suggests a traffic condition which may require higher level of cognitive ability to negotiate from the perspective of drivers. At the same time, the tendency of a faster but compact upstream meeting a compact but relatively slower downstream just before crash was also identified. Finally, a hypothesis test was carried out to test whether the traffic flow models developed for pre-crash conditions had any significant difference with normal traffic conditions. The results suggested for both upstream and downstream pre-crash

conditions that the null hypothesis that there was no difference in the pre-crash and normal conditions could be rejected with 95% confidence interval.
Similar findings were also obtained from the three-phase traffic flow theory as both pre-crash and normal traffic data satisfied the range for wide moving jam in both the methods of calculation. Supporting the findings from two-phase theory, it was also observed that the wide moving jam moves faster from downstream to upstream for pre-crash conditions as compared to normal traffic conditions. Also, pre-crash data exhibited greater standard deviation for the calculated wide moving jam velocities than the normal traffic condition.

Acknowledgements: The authors would like to thank Tokyo Metropolitan Expressway Company Limited for providing necessary data to conduct the study.

Funding: The research project was supported by JSPS KAKENHI Grant [KIBAN(C)-#25420535]

REFERENCES

Abdel-aty, M., Abdalla, F. (2004) Linking Roadway Geometrics and Real-Time Traffic Characteristics to Model Daytime Freeway Crashes Generalized Estimating Equations for Correlated Data. *Transp. Res. Rec. J. Transp. Res. Board.* 1897, 106–115. doi:10.3141/1897-14.

Abdel-Aty, M., Cunningham, R.J., Gayah, V. V., Hsia, L. (2009) Dynamic Variable Speed Limit Strategies for Real-Time Crash Risk Reduction on Freeways. *Transp. Res. Rec. J. Transp. Res. Board.* 2078, 108–116. doi:10.3141/2078-15.

Abdel-Aty, M., Pande, A., Das, A., Knibbe, W.J. (2008) Assessing safety on Dutch freeways with data from infrastructure-based intelligent transportation systems, *Transp. Res. Rec.* 2083, 153–161. doi:10.3141/2083-18.

Abdel-Aty, M., Pemmanaboina, R., Hsia, L. (2006) Assessing Crash Occurrence on Urban Freeways by Applying a System of Interrelated Equations. *Transp. Res. Rec.* J. Transp. Res. Board. 1953, pp 1–9. doi:10.3141/1953-01.

Abdel-Aty, M., Uddin, N., Pande, A. (2005) Split Models for Predicting Multivehicle Crashes During High-Speed and Low-Speed Operating Conditions on Freeways. *Transp. Res. Rec.* 1908, 51–58. doi:10.3141/1908-07.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, M.F., Hsia, L. (2004) Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transp. Res. Rec.*, 88–95.

Dalgaard, P. (2008) Introductory Statistics with R, New York.

Drake, J.S., Schofer, J.L., May Jr., A.D. (1967) A Statistical Analysis Of Speed-Density Hypotheses. In Vehicular Traffic Science, *Highw. Res. Rec.*, 112–117. http://trid.trb.org/view.aspx?id=693312.

Golob, T.F, Recker, W.W., Alvarez, V.M. (2004) Freeway safety as a function of traffic flow. *Accid. Anal. Prev.* 36, 933–946. doi:10.1016/j.aap.2003.09.006.

Greenberg, H. (1959) An Analysis of Traffic Flow. *Oper. Res.* 7, 79-85 doi:10.1287/opre.7.1.79.

Greenshields, B.D., Thompson, J.T., Dickinson, H.C., Swinton, R.S. (1934) The Photographic Method Of Studying Traffic Behavior. *Highw. Res. Board Proc.* 13, 382–399. http://trid.trb.org/view.aspx?id=120821.

Hossain, M., Muromachi, Y. (2013a) Understanding crash mechanism on urban expressways using high-resolution traffic data. *Accid. Anal. Prev.* 57, 17–29.

Hossain, M., Muromachi, Y. (2013b) A real-time crash prediction model for the ramp vicinities of urban expressways, *IATSS Res.* 37, 68–79.

Hossain, M., Muromachi, Y. (2011) Understanding Crash Mechanisms and Selecting Interventions to Mitigate Real-Time Hazards on Urban Expressways. *Transp. Res. Rec. J. Transp. Res. Board.* 2213, 53–62. http://trb.metapress.com/openurl.asp?genre=article&id=doi:10.3141/2213-08.

Kerner, B.S. (1999a) Method for monitoring the condition of traffic for a traffic network comprising effective narrow points. *German Patent* DE 199 44 075 C2, *US Patent* US6813555B1; *Japan Patent*: JP 2002117481.

Kerner, B.S. (1999b) Congested traffic flow: Observations and theory. *Transportation Research Record* 1678, 160-167.

Kerner, B.S. (2004) The Physics of Traffic. *Springer*, Berlin, New York.

Kerner, B.S. (2009b) Introduction to Modern Traffic Flow Theory and Control. *Springer*, Berlin, New York.

Kerner, B.S., and Rehborn, H. (1998) Traffic surveillance method and vehicle flow control in a road network. *German Patent Publication* DE 198 35 979 A1, US Patent: US 6587779B1.

Kerner, B.S., Rehborn, H., Aleksic, M., and Haug, A. (2004) Recognition and Tracing of Spatial-Temporal Congested Traffic Patterns on Freeways. *Transportation Research C*, Vol. 12, 369-400.

Kim, Y., Hall, F. (2004) Relationships Between Occupancy and Density Reflecting Average Vehicle Lengths. *Transp. Res. Rec. J. Transp. Res. Board.* 1883, 85–93. doi:10.3141/1883-10.

Klein, L. a, Mills, M.K., Gibson, D.R.P. (2006) Traffic Detector Handbook. *Oper. Res.* II,462 doi:FHWA-HRT-06-108 October.

Lee, C., Abdel-Aty, M. (2008) Testing Effects of Warning Messages and Variable Speed Limits on Driver Behavior Using Driving Simulator. *Transp. Res. Rec. J. Transp. Res. Board.* 2069, 55–64. doi:10.3141/2069-08.

Lee, C., Abdel-Aty, M., Hsia, L. (2006) Potential Real-Time Indicators of Sideswipe Crashes on Freeways. *Transp. Res. Rec.* 1953, 41–49. doi:10.3141/1953-05.

Lee, C., Hellinga, B., Saccomanno F. (2003a) Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. *Transp. Res. Rec.* 1840, 67–77. doi:10.3141/1840-08.

Lee, C., Saccomanno, F., Hellinga, B. (2002) Analysis of Crash Precursors on Instrumented Freeways. *Transp. Res. Rec.* 1784, 1–8. doi:10.3141/1784-01.

Lee, C.,Hellinga, B., Saccomanno, F. (2003) Proactive freeway crash prevention using real-time traffic control. *Can. J. Civ. Eng.* 30, 1034–1041. doi:10.1139/l03-040.

Luo, L., Garber, N. (2006) Freeway Crash Predictions Based on Real-Time Freeway Crash Predictions Based on Real-Time Pattern Changes in Traffic Flow Characteristics By : Lili Luo. A Res. Proj. Rep. Intell. Transp. Syst. Implement. Center, UVA Cent. *Transp. Stud. Research R*.

Maymala, J. (2015) PostgreSQL for Data Architects, Packt Publishing.

Oh, J.-S., Oh, C., Ritchie, S.G., Chang, M. (2005) Real-Time Estimation of Accident Likelihood for Safety Enhancement. *J. Transp. Eng.* 131, 358–363.doi:10.1061/(ASCE)0733-947X(2005)131:5(358).

Oh, C., Oh, J.S., Ritchie, S.G. (2005) Real-time hazardous traffic condition warning system: Framework and evaluation, *IEEE Trans. Intell. Transp. Syst.* 6, 265–272. doi:10.1109/TITS.2005.853693.

Pande, A., Abdel-Aty, M. (2006b) Assessment of freeway traffic parameters leading to

lane-change related collisions. *Accid. Anal. Prev.* 38, 936–948. doi:10.1016/j.aap.2006.03.004.

Pande, A., Abdel-Aty, M. (2206a) Comprehensive Analysis of the Relationship Between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways. *Transp. Res. Rec.* 1953, 31–40. doi:10.3141/1953-04.

Rehborn, H., Klenov, S.L., Palmer, J. (2011) Common traffic congestion features studied in USA, UK, and Germany based on Kerner's three-phase traffic theory. *IEEE Intelligent Vehicles Symposium*, Proceedings, 19–24.

Rehborn, H., Palmer, J. (2008) ASDA/FOTO based on Kerner's Three-Phase Traffic Theory in North Rhine-Westphalia and its Integration into Vehicles. *Proceedings of the IEEE intelligent vehicles symposium*, Eindhoven, Netherland.

Underwood, R.T. (1961) Speed, volume, and density relationships, in: Qualty Theory Traffic Flow. *A Symp.*, Yale University. Bureau of Highway Traffic, pp: 141–187. http://trid.trb.org/view.aspx?id=115231.

Velasco, R.M. (2012) Clusters in Macroscopic Traffic Flow Models. *World J. Mech.* 02, 51–60. doi:10.4236/wjm.2012.21007.