Available online at www.sciencedirect.com

ScienceDirect

Safety Science 46 (2008) 1306–1333

SAFETY SCIENCE

www.elsevier.com/locate/ssci

Probabilistic models of freeway safety performance using traffic flow data as predictors

Thomas F. Golob [a], Will Recker [b,*], Yannis Pavlis [b]

[a] Institute of Transportation Studies, University of California, 522 Social Science Tower, Irvine, CA 92697-3600, USA
[b] Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, 522 Social Science Tower, Irvine, CA 92697-3600, USA

Received 6 July 2006; received in revised form 17 August 2007; accepted 24 August 2007

Abstract

In this paper we lay the groundwork for gauging the level of safety of any type of traffic flow on a freeway, based on data from single loop detectors; the procedure can be implemented wherever such data are monitored or simulated. Our analyses are based on loop detector data for each of the freeway lanes for a short period of time preceding teach of over 1700 accidents in our case study. This case study covers the six major freeways in Orange County, California, for a six-month period in 2001.

Recognizing that loop detector data at a specific time and place cannot be converted to speed, because it is not possible to know effective vehicle length at such a detailed level (that is, the mix of long and short vehicles is unknown at a specific place for a short period of time), we avoid using any direct speed or density measures among the parameters. Rather, we employ explanatory parameters that include not only central tendencies (means and medians), but variations, and measures of systematic and synchronized traits that capture patterns in short period of loop detector data. Such patterns include breakdown from free flow to congested operations or recovery back to free flow, and differences in traffic conditions across lanes. In the analysis, we uncover an extensive set of statistical parameters that capture those aspects of traffic flow that are strongly related to accident potential. We demonstrate that the parameters can account for speed and density, even though these are not used directly. Moreover, the parameters account for important differences among the types of accidents that occur under different types of traffic flow.

Keywords: Freeway safety; Traffic flow; Loop detector data; Accidents; Statistical analysis; Multinomial logit; Factor analysis

1. Introduction

Our goal is to calibrate and verify a model that translates traffic flow, as measured by ubiquitous single loop detectors, into safety performance. By quantifying the safety benefits accrued from smooth and efficient traffic

* Corresponding author.
E-mail addresses: tgolob@uci.edu (T.F. Golob), wwrecker@uci.edu (W. Recker), ipavlis@translab.its.uci.edu (Y. Pavlis).

0925-7535/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.ssci.2007.08.007

operations, transportation management agencies will be able to incorporate safety measures in assessment of performance gains resulting from deployment of Intelligent Transportation System (ITS) measures, such as system-wide ramp metering (SWARM), freeway service patrol (FSP) and other incident response measures, and advanced driver information. In the research reported here, our objective was to capture the relationships between traffic flow, as measured by an extensive set of statistical parameters, and the types of accidents that occur under different types of traffic flow conditions. The work builds upon previous work by the authors (Golob and Recker, 2003, 2004; Golob et al., 2002, 2004a, 2004b).

Two datasets are used in this research: (1) accident data from the Traffic Accident Surveillance and Analysis System (TASAS) database (Caltrans, 1993), which covers all police-reported, on the California State Highway System and (2) traffic flow data from the front end processors of the Vehicle Detection System (VDS), received directly from a real-time data intertie with a California Department of Transportation (Caltrans) District office in Orange County, California.

2. Data

2.1. Accident data

The TASAS database (Caltrans, 1993; FHWA, 2000) covers police-reported crashes that occur on the California State Highway System. Most of the crashes included in the TASAS database were investigated in the field, but some were reported after the fact. The database does not cover crashes for which there are no police reports. We are concerned only with highway (mainline) crashes on well-defined urban freeways.

The TASAS data available to us contain the following types of crash characteristics: (1) the type of collision (rear-end, sideswipe, broadside, head-on, overturn), (2) the collision factor (e.g., speeding, following too close, illegal turn, alcohol), (3) number of vehicles and other parties involved, (4) the movements of each vehicle prior to collision (e.g., proceeding straight ahead, slowing, stopping, turning), (5) the location of the collision involving each vehicle (e.g., left lane, interior lanes, right lane, right shoulder area, off-road beyond right shoulder area), (6) the object struck by each vehicle (e.g., another vehicle, guardrail, bridge abutment), (7) number of injured and fatally injured persons per vehicle, and (8) environmental conditions, such as lighting, weather, and pavement conditions. No information was available concerning drivers and extent of injuries.

The time of each crash is not known with precision. An inspection of the crash times, presumably obtained from eyewitness accounts documented in police reports, reveals that almost 88% of the accidents in our study have reported times in minutes that fall precisely on the 12 five-minute intervals that comprise an hour. Thus, reported crash times must be treated as likely being rounded to the nearest five-minute interval, with a lesser secondary rounding to the nearest quarter hour. Since it is important that the traffic data in this study represent pre-crash conditions (rather than conditions arising from the crash itself), the period of the traffic-flow data used in the analysis was cut off 2.5 min before the nominal crash time—under the presumption that any "rounding" is to the nearest five-minute mark, this ensures that in cases in which "rounding up" occurs data used are from the pre-accident conditions (for the "rounding down" case, there are no data conflict).

2.2. Traffic flow data from single loop detectors

Our traffic flow data are drawn from the Vehicle Detection System (VDS) 30-s single loop detector data. There are approximately 8000 VDS locations on California freeways, typically spaced one-third to one-half mile apart (Varaiya, 2005). These loop detectors record volume (flow) and occupancy (the percent of time a vehicle is within the detection field of a loop) for each freeway lane at 30 s (30-s) intervals. From volume and occupancy, traffic density and point speed can be derived, but only under very restrictive assumptions of uniform speed and average vehicle length, and taking into account the physical installation of each loop. Such assumptions are relevant for aggregated data over extended periods of time. For our disaggregated purposes, there is no accurate information on average vehicle lengths for each 30-s interval, or for any aggregation of data over the periods of time (e.g., 20 min) needed to relate accident hazards to traffic flow conditions. Consequently, we assume that absolute density and speed are *not* determinable. These are declared to be prohibited variables due to the absence of accurate effective average vehicle length for each 30-s observation.

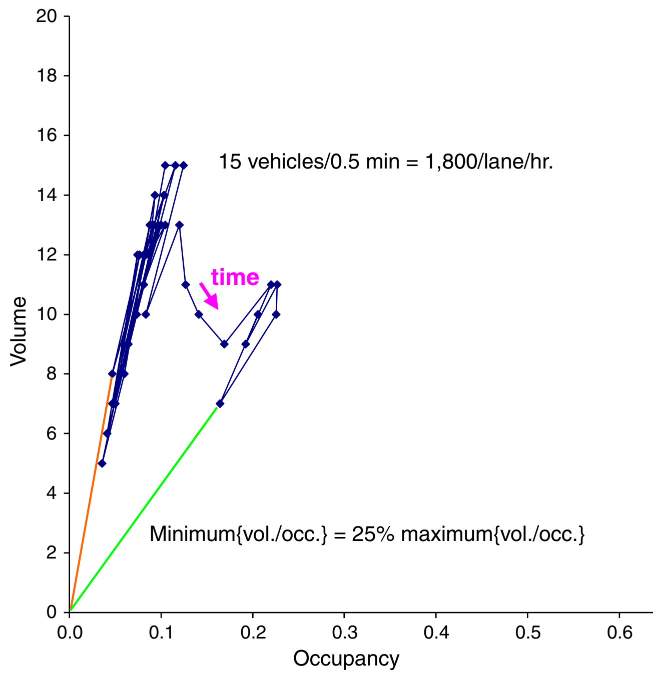


Fig. 1. Trace of 20 min of 30-s loop detector data showing transition from free flow to congested flow operation.

A 20-min time trace of 30-s data is shown in Fig. 1. For the first 15 min or so, the lane is operating in free flow. Volumes vary from 5 to 15 vehicles per 30-s (equivalent to 600–1800 vehicles per hour), and occupancies are roughly in a constant proportion to volume. However, after 15 min there is a transition to congested flow. Volumes initially fall, while occupancies increase substantially. The last six 30-s observations exhibit a similar ratio of volume to occupancy, but this ratio is approximately 25% of the ratio of volume to occupancy observed during free flow operation. It is impossible to know if speeds vary in this same factor of 4:1, since it is not known whether average vehicle lengths were the same at all times.

These distinctions between free, transitional, and congested flows do not require assuming effective average vehicle lengths in order to calculate speeds. Rather, it is based on the pattern of the raw data itself. We intend to demonstrate that, with a sufficient number of 30-s observations (approximately 30–40) across multiple lanes of a freeway, we can capture important traffic flow characteristics without making tenuous assumptions about clearly unattainable data about average vehicle lengths. While some types of traffic flow detectors are able to identify the proportions of large trucks (long vehicles) for a given duration of traffic flow on different lanes, such data are not available at the time and location of accidents.

2.3. Scope of the case study

Data used in the study are drawn from six major Orange County freeways: Interstate Route (I) 5, State Route (SR) 22, SR-55, SR-57, SR-91, I-405. The period of study is March to August, 2001 (six full months). Of the 4412 accidents on the mainlines of these freeways recorded in TASAS, we have sufficient loop detector data for analysis for 1712 (about 39%) of these accidents. Chi-square tests performed on contingency tables revealed that the subset of accidents with sufficient traffic flow data is a random selection with respect to: (1) type of collision (e.g., rear end, sideswipe, hit object), (2) the number of vehicles involved, (3) the types of vehicles involved (e.g., automobile, panel or pickup truck, large truck), (4) location (which lane or side of road where the primary collision was located), (5) timing (time of day, day of week), and accident severity (injury or fatality versus property damage only (PDO)). The fortunate conclusion is that the subset of accidents with sufficient traffic flow data for analysis are a random sample of all reported accidents in the case study area over six months of 2001.

2.4. Creation of 36 traffic flow parameters

From the loop detectors, repeated measurements over time at 30-s intervals of volume (flow) and occupancy for each of the freeway lanes at the site of an accident were drawn to capture temporal variations. Sensitivity analyses revealed that we need 20 consecutive minutes of data (40 consecutive 30-s observations) in order to estimate the traffic flow parameters used in our models; missing data for up to ten 30-s observations were tolerated in computing the various statistics. To allow consistent definitions for all locations within our case study network, we use data for three lanes: (1) the leftmost, number one, or median lane (excluding any HOV lanes, designated "1"): (2) one interior lane[1] (whichever has least missing data, designated "M"), and the rightmost or curb lane (designated "R").

Four types of parameters were found to be useful: (1) C*oefficients of variation*, the ratio of standard deviation and mean; (2) *Correlations* of traffic conditions across the lanes; (3) *Autocorrelations*, the correlation of a variable at one 30-s interval with the value of the same variable in the previous 30-s interval, for all adjacent time intervals in the 20-min period; and (4) *Means and standard deviations* of volumes and relative speed, where relative speed is defined as the ratio of volume to occupancy divided by the maximum ratio of volume to occupancy over the entire 20-min time period. The variables constructed from these statistics are summarized in Table 1. In this and succeeding tables, the labels 1, M, and R are used to designate the leftmost (adjacent the freeway median), interior, and rightmost (curb/shoulder) lanes, respectively.

3. Data reduction

Owing to the redundancy in these 36 traffic flow parameters, principal components analysis is used to eliminate such redundancy by finding a smaller number of orthogonal (i.e., statistically independent) linear combinations (called principal components, or "Factors") of the original variables such that the least amount of information is lost. The number of Factors needed to describe the unique information in the original variables is classically determined based purely on sufficiency of explanation of relationships among the original variables (i.e., the percentage of variance in the original variables accounted for by the set of Factors selected). In this application, we added two more criteria for selecting the appropriate number of Factors: (1) interpretability and (2) effectiveness in describing derived parameters (speed and density).

Eight Factors were found to account for approximately 79% of the variance in the original variables. The percent of variance of each of the original 36 variables that is accounted for by the eight Factors together (the R^2 of the regression of each of the 36 parameters on all eight Factors) is known as the communality of that variable. For the eight-factor solution, all communalities were in excess of 0.72, with the single exception of "standard deviation of volume in the right lane", which was 0.67. In order to evaluate the effectiveness in describing derived parameters (speed and density), each of prohibited scaled measures was regressed on a set of 44 variables, made up of the eight Factors, plus eight factor quadratic terms (the products of two like Factors), plus 28 factor interactions (the products of any two different Factors). The regression results (not shown) for the means and standard deviations of occupancy and the ratio of volume to occupancy for each of the three lanes indicate that the Factors do very well explaining all of the prohibited variables that are proportional to traffic flow density and speed. The conclusion is that all of the original variables are sufficiently described by the eight Factors.

These Factors are expressed in terms of their correlations with the original variables, called factor loadings. To these factors we applied a varimax orthogonal rotation (i.e., a rotation that maximizes the sum of the variances of the factor loadings) in the 36-dimensional space of the original variables in order to improve interpretability by driving loadings as far as possible towards the extreme values of unity (or minus unity) and zero. Scores on each of the eight rotated Factors were calculated for each accident based on the factor loadings, which express the factor in terms of a linear combination (weighted average) of the original 36 traffic flow variables. The scores for each factor are standardized (they are centered at zero with unity standard deviation).

[1] There are strong correlations of data across all interior lanes.

Table 1
The 36 Traffic Flow parameters

Variable type	Measurement	Lanes	Variable
Coefficients of variation	Volume	Left (1)	Coef. of var. volume 1
		Middle (M)	Coef. of var. volume M
		Right (R)	Coef. of var. volume R
	Occupancy	1	Coef. of var.occupancy 1
		M	Coef. of var.occupancy M
		R	Coef. of var.occupancy R
	Volume/occupancy	1	Coef. of var. vol./occ. 1
		M	Coef. of var. vol./occ. M
		R	Coef. of var. vol./occ. R
Correlations across lanes	Volume	Left (1) vs. middle (M)	Correlation volume 1 vs. M
		Left (1) vs. right(R)	Correlation volume 1 vs. R
		Middle (M) vs. right (R)	Correlation volume M vs. R
	Occupancy	1 vs. M	Correlation occupancy 1 vs. M
		1 vs. R	Correlation occupancy 1 vs. R
		M vs. R	Correlation occupancy M vs. R
	Volume/occupancy	1 vs. M	Correlation vol./occ. 1 vs. M
		1 vs. R	Correlation vol./occ. 1 vs. R
		M vs. R	Correlation vol./occ. M vs. R
Autocorrelation	Volume	1	Autocorrelation volume 1
		M	Autocorrelation volume M
		R	Autocorrelation volume R
	Occupancy	1	Autocorrelation occupancy 1
		M	Autocorrelation occupancy M
		R	Autocorrelation occupancy R
Central tendencies and variations	Volume	1	Mean volume 1
		M	Mean volume M
		R	Mean volume R
		1	Standard deviation volume 1
		M	Standard deviation volume M
		R	Standard deviation volume R
	Relative speed: (vol./occ.)/{MAX(vol./occ.)}	1	Mean relative speed 1
		M	Mean relative speed M
		R	Mean relative speed R
		1	Std. dev. relative speed 1
		M	Std. dev. relative speed M
		R	Std. dev. relative speed R

The factor loadings of the original variables on the rotated Factors (i.e., the correlations between the variables and Factors) form the basis for the interpretation of the Factors. This is pursued in the following section.

The factor loadings are listed in Table 2. To aid in interpretation, only factor loads with absolute value greater than 0.20 are shown. The most indicative parameter for each factor is identified by a box around its loading. These eight indicative parameters come from each of the four types of parameters, indicating that each type provides some unique information: two of the indicative parameters are coefficients of variation (volume in lane 1 and volume/occupancy in lane M); three indicative parameters are correlations across lanes (2 volumes and 1 occupancy, one for each pair of lanes (1 vs. M, 1 vs. R, M vs. R)); one indicative parameter is an autocorrelation (volume for lane M); and two are standard deviations (volume for lane M and relative speed for lane R).

The strongest loadings for each factor are listed in Table 3. As a further aid in interpretation, the means for each factor for all accidents occurring within five mutually exclusive time periods are graphed in Fig. 2. Interpretations of each factor follow.

Table 2
Rotated factor loadings for the eight-factor solution for the principal components for the factor analysis of the traffic flow parameters

	Factor							
	1	2	3	4	5	6	7	8
CV_vol1		−0.89						
CV_volM		−0.85						
CV_volR		−0.75	−0.22	0.39				
CV_occ1		−0.85						
CV_occM	0.26	−0.80	0.26					
CV_occR		−0.70		0.47		−0.21		
cv_volocc1	0.79							
cv_voloccM	0.88		0.21					
cv_voloccR	0.53			0.73				
corr_vol1M					0.22	0.28		0.81
corr_vol1R						0.80		
corr_volMR						0.82		
corr_occ1M			0.57					0.64
corr_occ1R			0.76			0.40		
corr_occMR			0.74			0.38		
corr_volocc1M	0.43	−0.27	0.64					0.24
corr_volocc1R	0.43		0.74					
corr_voloccMR	0.45		0.69					
autocorr_vol1	0.48		0.24				0.61	0.22
autocorr_volM	0.47		0.24				0.62	
autocorr_volR	0.31			0.33		0.30	0.55	
autocorr_occ1	0.43		0.64				0.37	
autocorr_occM	0.44		0.64				0.41	
autocorr_occR	0.40		0.61	0.21			0.37	
mu_vol1		0.76	0.28		0.34			
mu_volM	−0.21	0.78	0.28		0.29			
mu_volR		0.66	0.33	−0.36	0.25			−0.25
sd_vol1					0.73			0.41
sd_volM					0.82			0.21
sd_volR		0.29			0.68			−0.24
mu_rel_speed1	−0.82				0.21			
mu_rel_speedM	−0.87							
mu_rel_speedR	−0.52			−0.71				
sd_rel_speed1	0.74		0.43				0.23	
sd_rel_speedM	0.79		0.43					
sd_rel_speedR	0.42		0.32	0.70				

4. Interpretations of the traffic flow factors

4.1. Factor 1: outer (left and interior) lanes congestion

High scores on Factor 1 indicate that traffic conditions are varying widely within the region of loop detector data that generally denotes congested operation (see example depicted in Fig. 3), while low scores indicate that conditions are exclusively free flow, as in the example of Fig. 4. High scores on Factor 1 are most likely to occur during the evening peak period, as shown in Fig. 2. Low scores are most likely to occur on weekends and at night.

Table 3
Highest factor loadings for the eight traffic flow factors

Highly loaded variables	Factor interpretation
Coefficient of variation: volume/occupancy lane 1 Coefficient of variation: volume/occupancy lane M Mean relative speed lane 1 (negative) Mean relative speed lane M (negative) Std. dev. relative speed lane 1 Std. dev. relative speed lane M	1. Outer lanes congestion
Coefficient of variation: volume lane 1 (negative) Coefficient of variation: volume lane M (negative) Coefficient of variation: volume lane R (negative) Coefficient of variation: occupancy lane 1 (negative) Coefficient of variation: occupancy lane M (negative) Coefficient of variation: occupancy lane R (negative) Mean: volume lane 1 Mean: volume lane M Mean: volume lane R	2. Volume level
Correlation occupancy: 1 vs. R Correlation: occupancy M vs. R Correlation: volume/occupancy lane 1 vs. lane M Correlation: volume/occupancy lane 1 vs. lane R Correlation: volume/occupancy lane M vs. lane R Autocorrelation: occupancy lane 1 Autocorrelation: occupancy lane M Autocorrelation: occupancy lane R	3. Synchronized lane conditions
Coefficient of variation: volume/occupancy lane R Mean: relative speed lane R (negative) Standard deviation: relative speed lane R	4. Curb lane perturbation
Standard deviation: volume lane 1 Standard deviation: volume lane M Standard deviation: volume lane R	5. Volume variation
Correlation: volume lane 1 vs. lane R Correlation: volume lane M vs. lane R	6. Conforming curb volumes
Autocorrelation: volume lane 1 Autocorrelation: volume lane M Autocorrelation: volume lane R	7. Systematic volume changes
Correlation: volume lane 1 vs. lane M Correlation: occupancy lane 1 vs. lane M	8. Synchronized outer flow

4.2. Factor 2: volume level

High scores on Factor 2 indicate high-volume conditions (Fig. 5), more likely at night and on weekends (Fig. 2), while low scores on Factor 2 indicate high-volume free-flow conditions (Fig. 6) and are more likely during the afternoon weekday peak period.

4.3. Factor 3: synchronized lane conditions

High scores on Factor 3 indicate traffic conditions that are changing in the same manner on all lanes. As shown in Fig. 2, this can happen during any period, but is more likely during the evening and morning peak periods. Detector data are plotted in Fig. 7 for both the left and interior lanes for a situation with a high score on this Factor. For most of this period (the first 15 min), conditions were free-flow. However, during the last 5 min, congestion set in, and speeds diminished in all lanes (including the right lane, not shown here). To

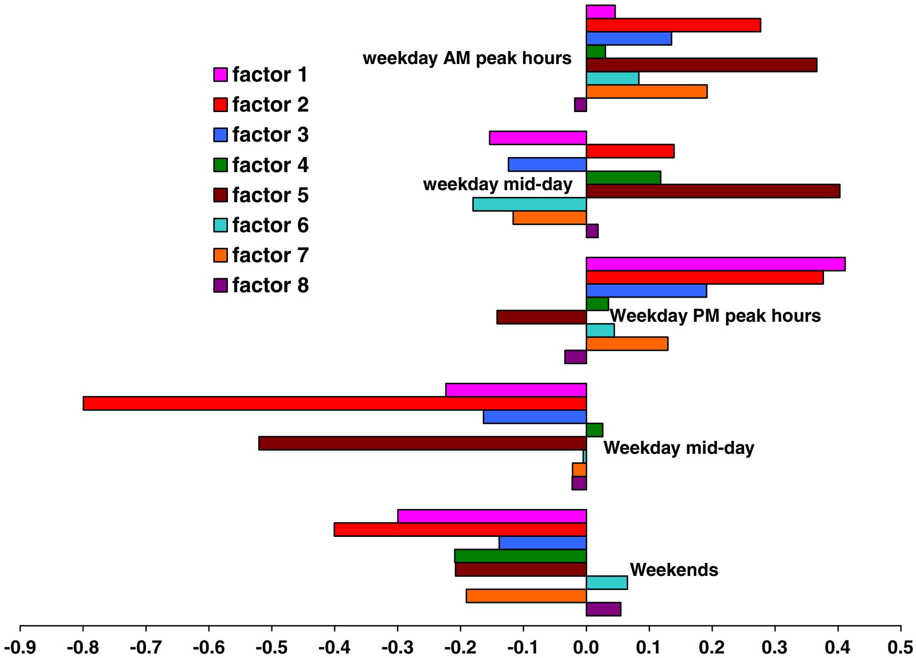


Fig. 2. Mean values of the eight traffic flow factors for all accidents occurring within five time periods.

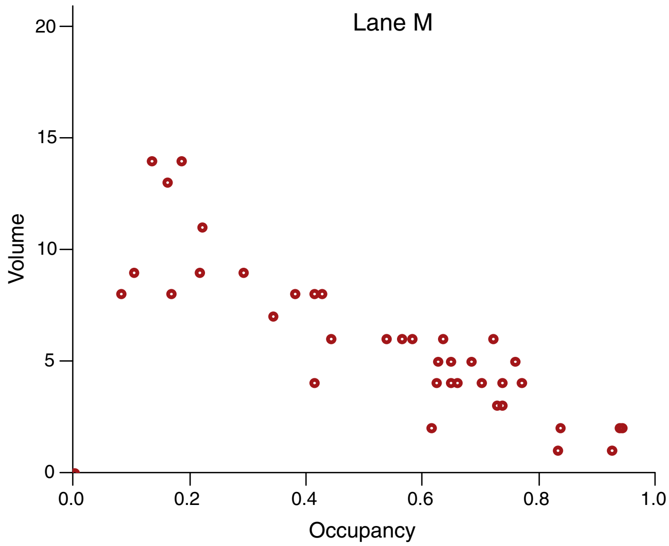


Fig. 3. Twenty minutes of loop detector data for one lane for an observation (PM 15.26 on WB SR-91 on 04/09/01 prior to 15:10) with a ***high*** score on Factor 1: outer lanes congestion.

visualize the breakdown period for this same observation, the traces of detector data for the crucial 3 min of transition are graphed in Fig. 10 for all three lanes. The traces for all three lanes follow a similar pattern of declining volume to occupancy ratios, which are roughly proportional to speeds (see Fig. 8).

Conversely, low scores on Factor 3 indicate traffic conditions that are changing differently across the lanes. Two lanes of detector data are plotted in Fig. 9 for an observation with a low score on Factor 3. Traces of detector data for 3 min during this heavily congested period are graphed in Fig. 10 for each of the three lanes.

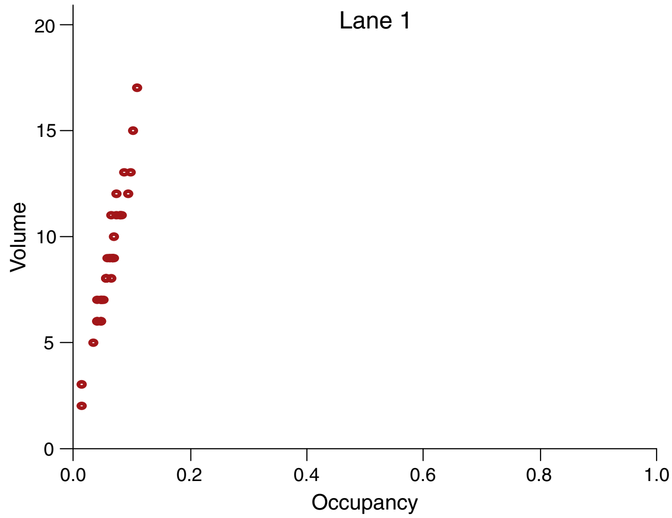


Fig. 4. Twenty minutes of loop detector data for one lane for an observation (PM 27.68 on SB SR-55 on 07/09/01 prior to 17:45) with a ***low*** score on Factor 1: outer lanes congestion.

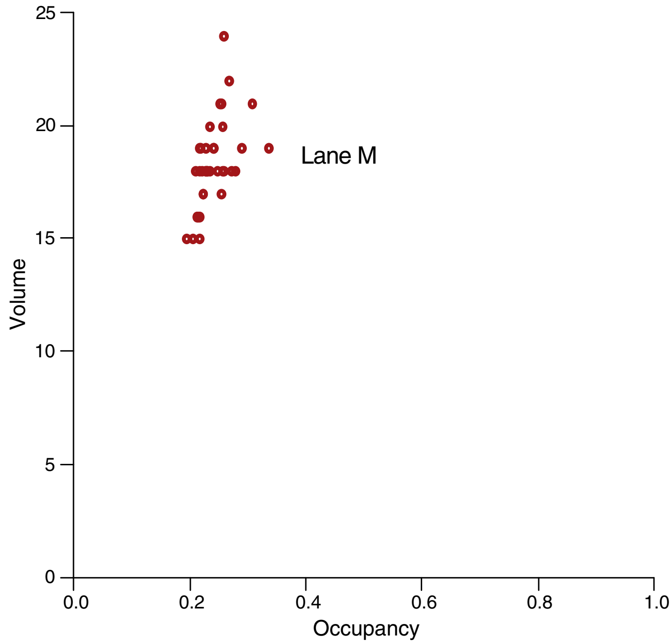


Fig. 5. Twenty minutes of loop detector data for one lane for an observation (PM 30.06 on NB I-5 on 03/14/01 prior to 07:25) with a ***high*** score on Factor 2: volume level.

These traces are show that, in any 30-s interval, speeds and densities in a given lanes is not linked to speeds and densities in either of the other two lanes.

4.4. Factor 4: curb lane perturbation

High scores on Factor 4 – indicating curb (right) lane loop detector data in the congested region (Fig. 1) – can occur during any time period, but are most likely during the mid-day period (Fig. 2). An extreme example

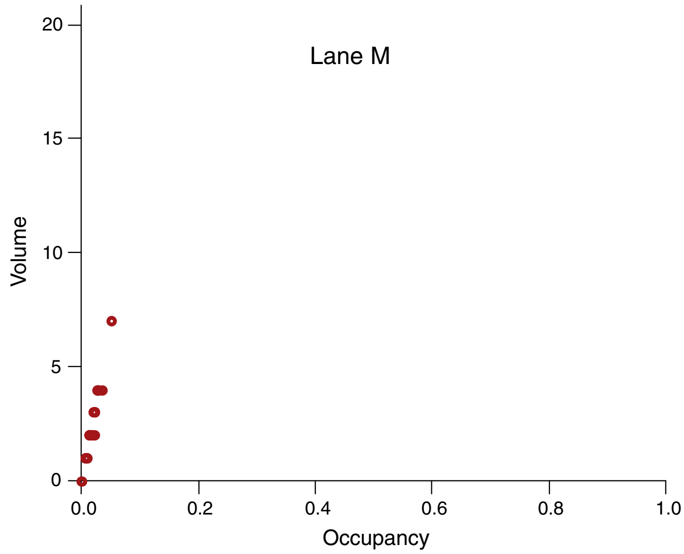


Fig. 6. Twenty minutes of loop detector data for one lane for an observation (PM 0.90 on NB I-5 on 06/17/01 prior to 02:40) with a ***low*** score on Factor 2: volume level.

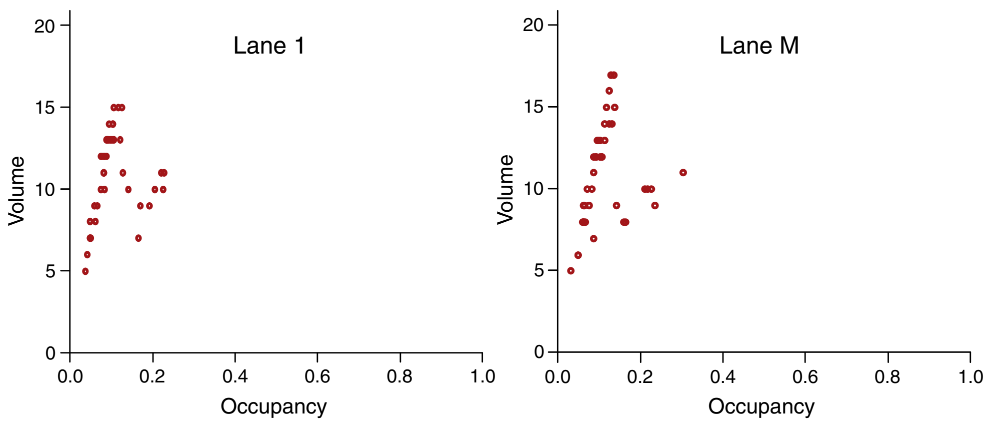


Fig. 7. Twenty minutes of loop detector data for two lanes for an observation (PM 4.91 on WB SR-22 on 07/07/01 prior to 19:50) with a *high* value on Factor 3: synchronized lane conditions.

is that of a right lane that exhibits the entire range of speed from free flow to virtually stopped traffic (Fig. 11). Low values of Factor 4 will be found for free flowing traffic. For example, Fig. 12 displays a case in which right lane traffic volume is high but speeds are consistently fast. The spread in the domain of occupancy is likely due in part to differences in average vehicle lengths over the 30-s observations, as there is likely to be a mix of trucks in this curb-lane flow.

4.5. Factor 5: volume variation

Factor 5 measures the extent to which volume is varying across the entire road, particularly in the non-curb lanes. Shown in Fig. 13 is a situation with a high score on Factor 5. This interior lane exhibits 30-s volumes ranging from 1 to 20 (120–2400 vehicles per lane per hour). This example also has a high score on Factor 1: variations in non-curb speeds, but any two Factors, by definition, are uncorrelated, so the score for one factor does not predict the score on any other factor. To demonstrate this independence, Fig. 14 shows a situation with a high score on Factor 1, but a relatively low score on Factor 5.

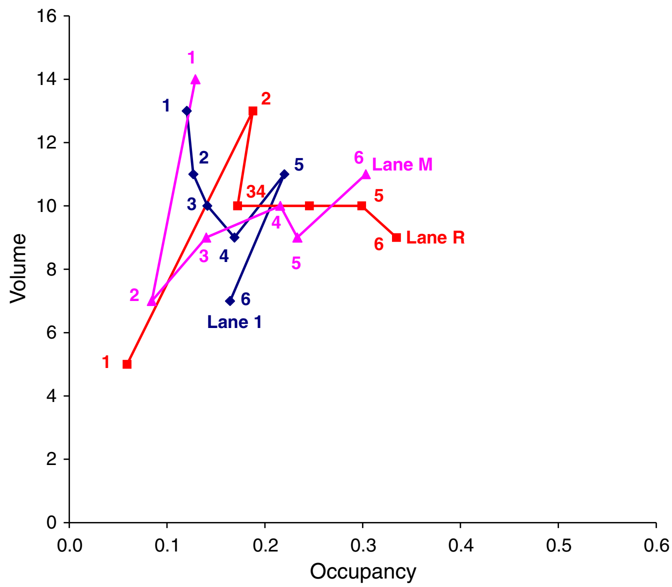


Fig. 8. Three minutes of loop detector data for all three lanes for an observation (PM 4.91 on WB SR-22 on 07/07/01 prior to 19:50) with a ***high*** score on Factor 3: synchronized lane conditions.

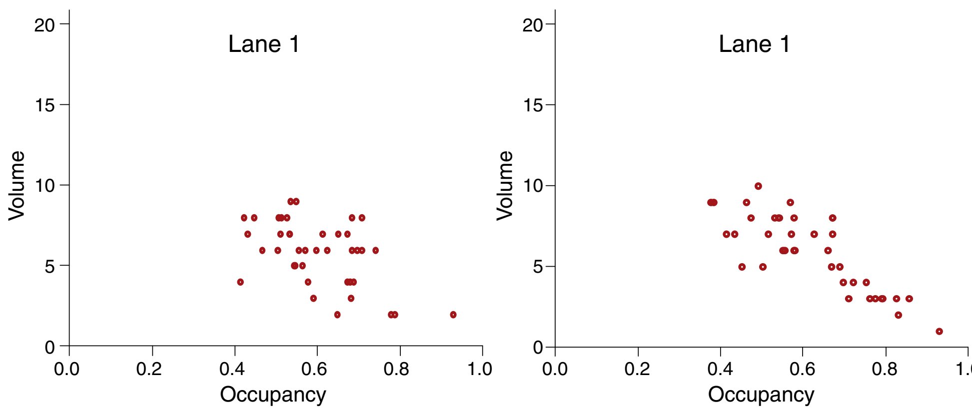


Fig. 9. Twenty minutes of loop detector data for two lanes for an observation (PM 9.99 on EB SR-91 on 04/24/01 prior to 16:15) with a ***low*** score on Factor 3: synchronized lane conditions.

Fig. 15 shows a situation with a very low score on factor 5 and a relatively low score on Factor 1. Traffic conditions measuring high on Factor 5 tend to occur during the morning peak hours and during the mid-day period. These are indicative of variable levels of congestion. Conditions measuring high on Factor 1, on the other hand, is more likely during the afternoon peak hours, and these are manifestations of heavy congestion.

4.6. Factor 6: conforming curb volumes

Factor 6 measures the degree to which volumes in the curb lane are related to volumes in the interior and left lanes. Fig. 16 is an example of a situation with a high score on Factor 6, and Fig. 17 is an example of a situation with a low score. Low scores, which indicate that either a curb or non-curb lane is experiencing congestion while the other lane is not, are more likely to occur during the mid-day period.

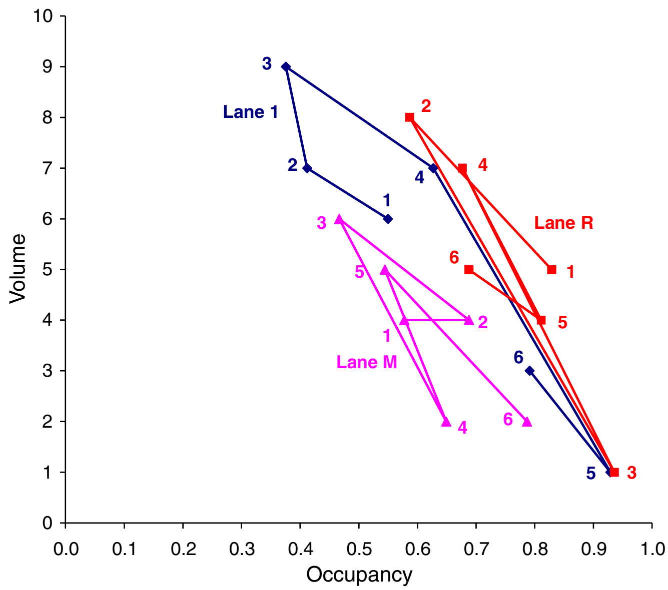


Fig. 10. Three minutes of loop detector data for all three lanes for an observation (PM 4.91 on WB SR-22 on 07/07/01 prior to 19:50) with a ***high*** score on Factor 3: synchronization of all lanes.

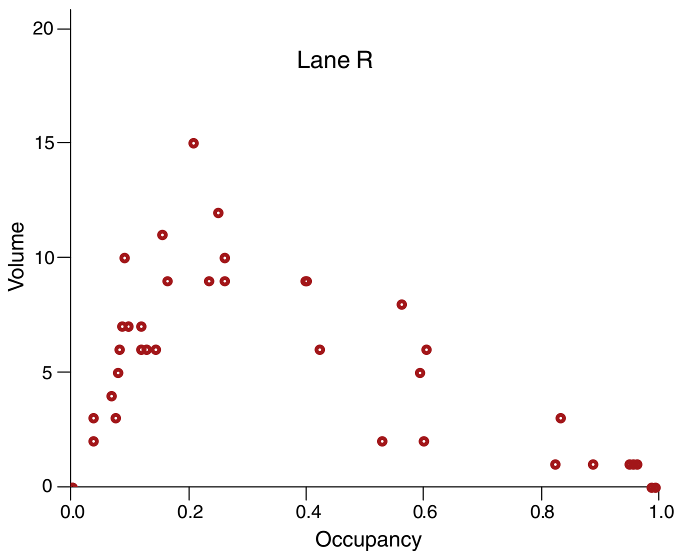


Fig. 11. Twenty minutes of loop detector data for one lane for an observation (PM 6.63 on SB I-5 on 05/07/01 prior to 09:35) with a ***high*** score on Factor 4: curb lane perturbation.

4.7. Factor 7: systematic volume change

Factor 6 measures the degree to which volumes change systematically, as opposed to random fluctuation. Highly systematic situations occur when a road shifts from free flow to some congestion, and such a case is shown in Fig. 18. These situations tend to occur during the morning and evening peak periods, more so in the morning peak (Fig. 2). Conversely, randomly fluctuating volumes, such as the situation depicted in Fig. 19, tend to occur on weekends, and during the weekday mid-day period.

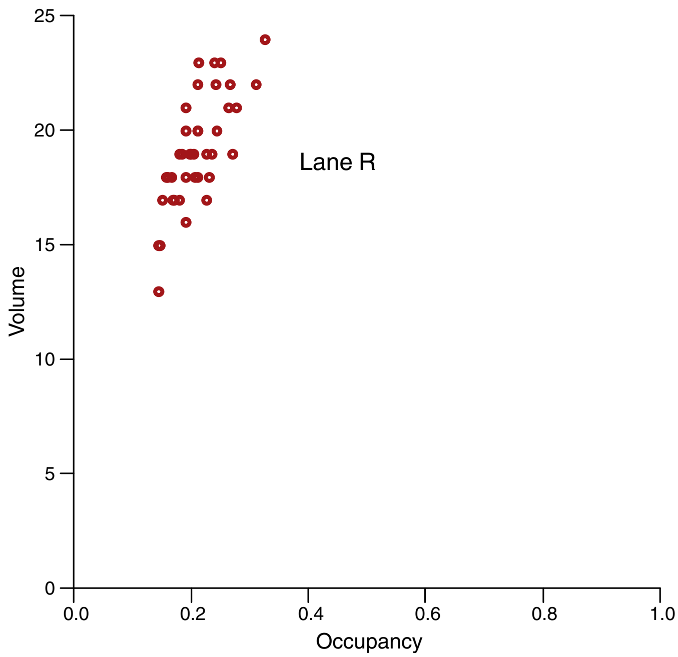


Fig. 12. Twenty minutes of loop detector data for one lane for an observation (PM 30.82 on SB I-5 on 08/06/01 prior to 08:30) with a ***low*** score on Factor 4: curb lane perturbation.

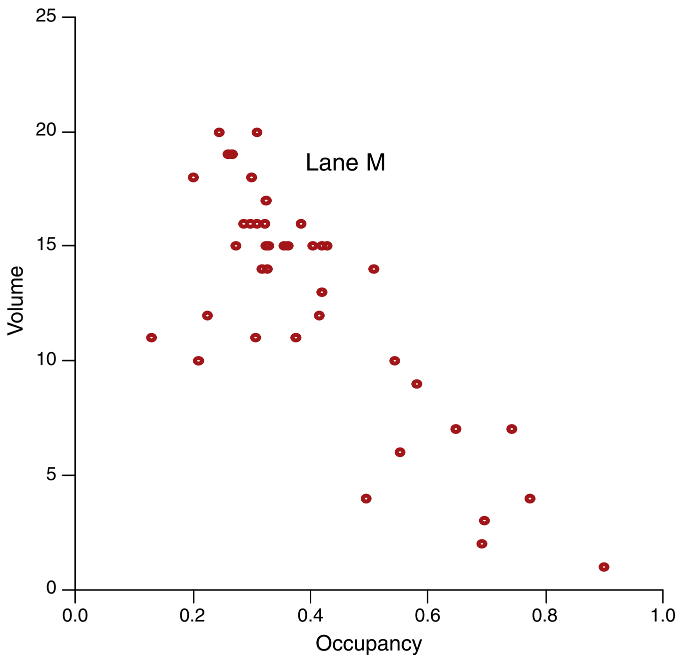


Fig. 13. Twenty minutes of loop detector data for one lane for an observation (PM 13.33 on SB SR-57 on 05/13/01 prior to 16:00) with a ***high*** score on Factor 5: volume variation.

4.8. Factor 8: synchronized outer flow

Factor 8 measures the degree to which volumes and densities in the interior and left lanes are synchronized. High or low levels of synchronization are likely to occur anytime (Fig. 2). A high level of non-curb

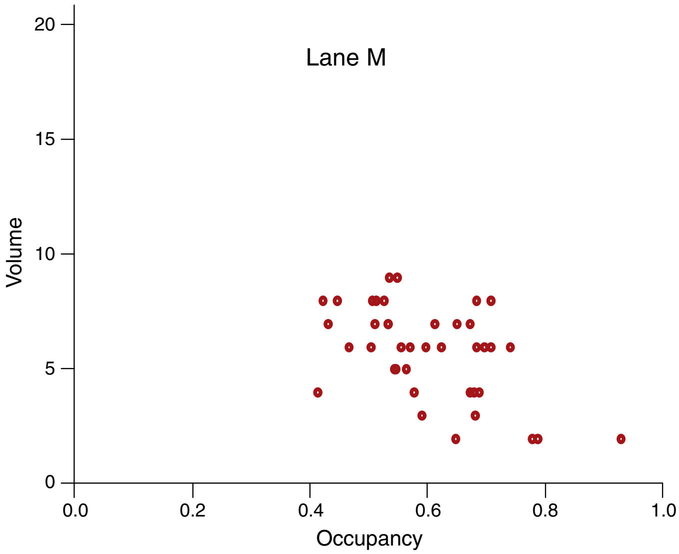


Fig. 14. Twenty minutes of loop detector data for one lane for an observation (PM 9.99 on EB SR-91 on 04/24/01 at 16:15) with a ***high*** score on Factor 1: variation in non-curb conditions, but a relatively **low** score on Factor 5: volume variation.

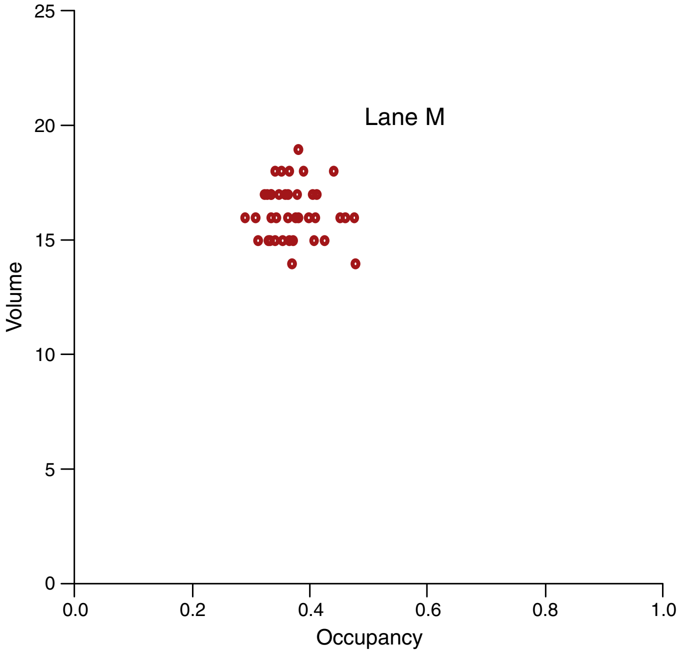


Fig. 15. Twenty Minutes of loop detector data for one lane for an observation (PM 33.31 on SB I-5 on 03/22/01 prior to 07:25) with a ***low*** score on Factor 5: volume variation.

synchronization is shown in Fig. 20. In this situation, volumes and densities in the left and interior lanes move in unison.

An example of a situation with a low score on Factor 8 is shown in Fig. 21. Here, the left lane is operating consistently in free flow mode with only minor perturbations. However, the interior lane is operating in congested mode for a major part of the 20-min period.

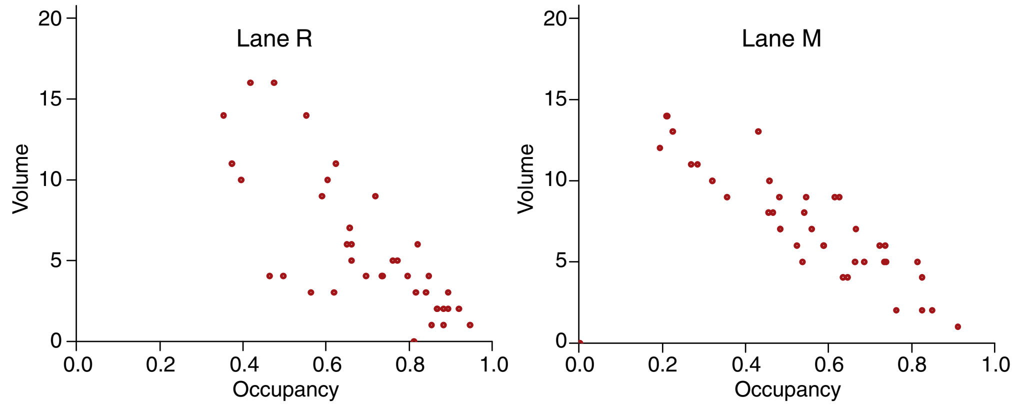


Fig. 16. Twenty minutes of loop detector data for two lanes for an observation (PM 15.96 on SB I-405 on 04/23/01 prior to 07:55) with a **high** score on Factor 6: conforming curb volumes.

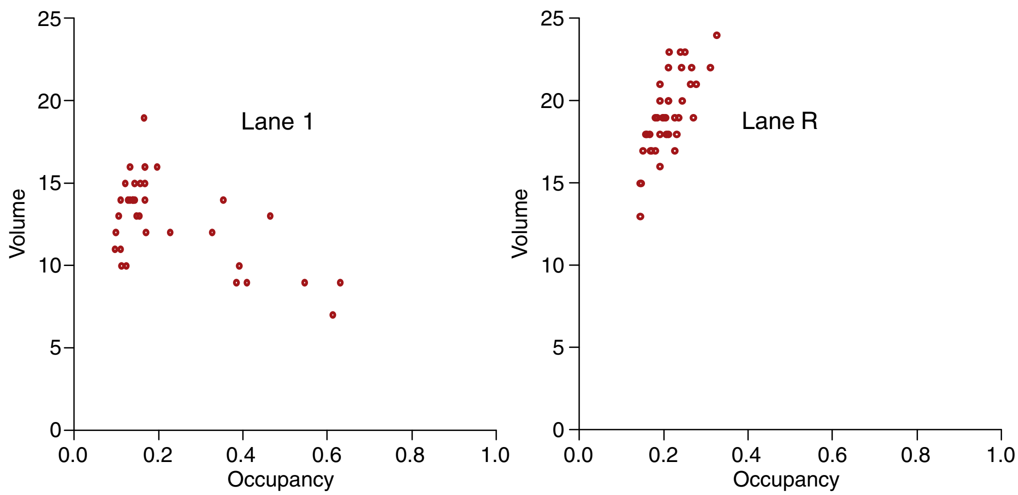


Fig. 17. Twenty minutes of loop detector data for two lanes for an observation (PM 30.82 on SB I-5 on 08/06/01 prior to 08:30) with a low value on Factor 6: conforming curb volumes.

5. Traffic Flow Factors related to accident propensity

The Traffic Flow Factors identified above were used to describe accident potential. Four accident variables were analyzed: (1) accident severity, (2) collision type, (3) collision location, and (4) number of involved vehicles. Logit (logistic regression) models are used to capture the relationships between the Traffic Flow Factors and their second-degree interactions, and the probabilities of occurrence of an event. Binomial (or binary) logistic regression is used in the case of a dichotomy, e.g., accident severity. Multinomial logit is used in all other cases of dependent variables with more than two categories of outcome. Logit models apply maximum likelihood estimation after transforming the dependent variables into natural logarithms of the odds of whether or not an outcome occurs. The exponential function of each coefficient for each dependent category in a logit model gives the multiplicative effect of that variable on the odds of occurrence of the event in question.

5.1. Accident severity

About one-quarter of all accidents (25.3%) in our case study led to an injury, the rest are property damage only (PDO). Results of a binomial logit regression model for severity are listed in Table 4. The dependent

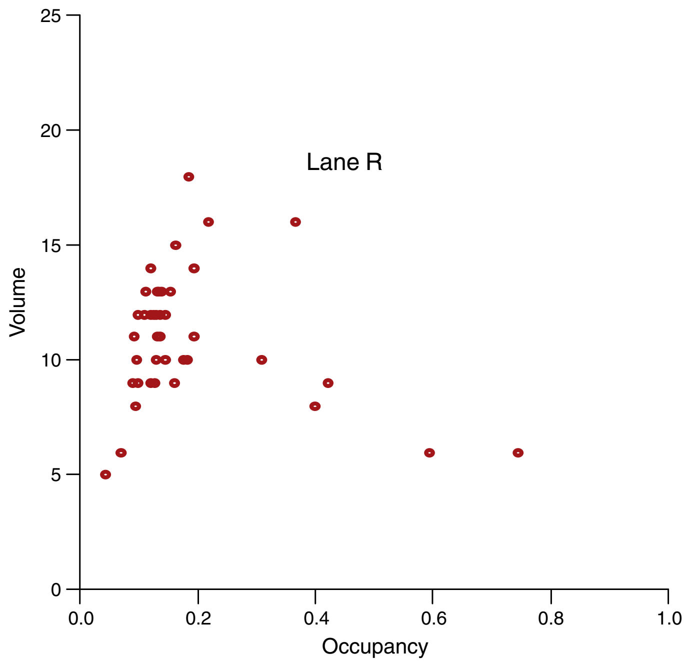


Fig. 18. Twenty minutes of loop detector data for one lane for an observation (PM 18.03 on SB I-405 on 03/08/01 prior to 06:45) with a ***high*** value on Factor 7: systematic volume change.

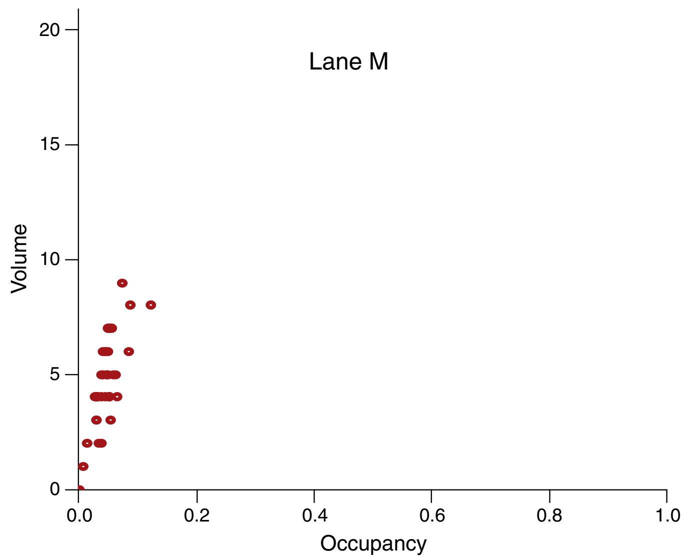


Fig. 19. Twenty minutes of loop detector data for one lane for an observation (PM 24.89 on NB I-5 on 05/05/01 prior to 09:10) with a ***low*** value on Factor 6: systematic volume change.

variable is encoded 1 for injury and 0 for PDO, so that a positive coefficient indicates that injury accidents are more likely for higher levels of the independent variable. Although eight independent variables were significant at the 95% confidence level (three Factors and five factor interactions), the overall fit of the model, as measured by the Nagelkerke Pseudo-R^2, an analogy to R^2 in linear regression, is 0.044, indicating that severity is only marginally associated with traffic conditions.

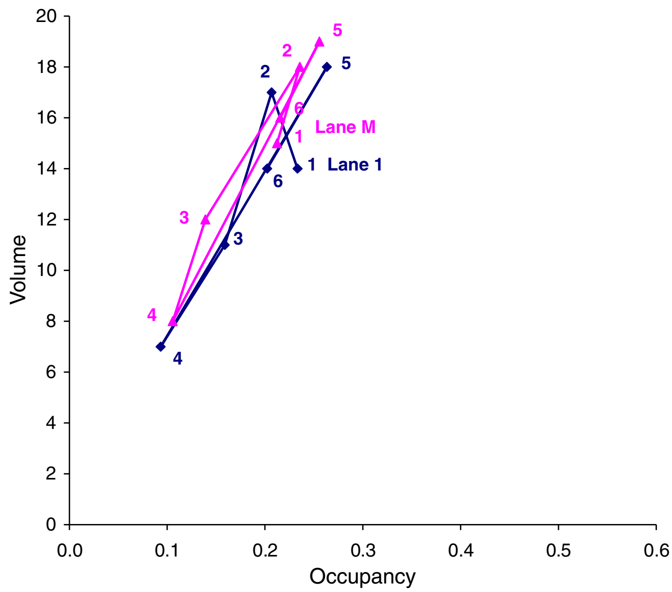


Fig. 20. Three minutes of loop detector data for two lanes for an observation (PM 27.45 on NB I-5 on 03/20/01 prior to 10:05) with a ***high*** value on Factor 8: synchronized outer flow.

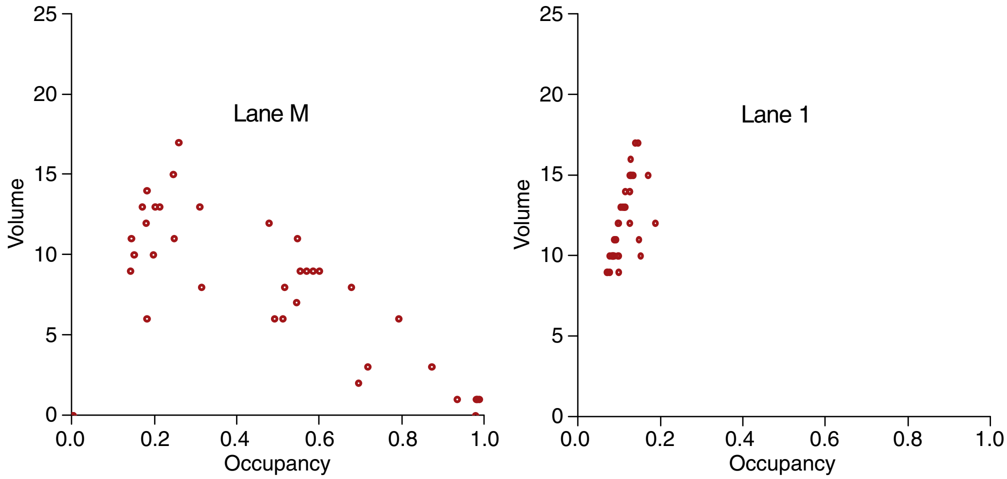


Fig. 21. Twenty minutes of loop detector data for two lanes for an observation (PM 11.70 on SB SR-55 on 07/09/01 prior to 17:45) with a ***low*** value on Factor 8: synchronized outer flow.

As expected, congestion in the outer (left and interior) lanes, on its own, leads to a lower likelihood of injury accidents, since lower speeds in these lanes are generally associated with the presence of congested conditions throughout the particular section of freeway. However, to the extent that the pattern of congestion is specific to the outer lanes, there are also two interaction terms involving this Traffic Flow Factor. If right lane volumes track those of the left and interior lanes, the effect of congestion on reducing severity is more than doubled. Such conditions represent a spatial and temporal uniformity in traffic that both discourages traffic maneuvers designed to enhance one's position in the traffic stream as well as making any such maneuvers (e.g., lane changing) less complex owing to the relatively small differentials in relative speed between lanes and the stable gaps in the traffic stream.

Table 4
Logit model of accident severity as a function of statistically significant traffic flow factors

Explanatory variable	Coefficient	*t*-Statistic	Probability
1. Outer lanes congestion	−0.162	−2.532	0.011
2. Volume level	−0.201	−3.676	0.000
3. Synchronized lane conditions	−0.141	−2.401	0.016
(1. Outer lanes congestion) × (6. Conforming curb volumes)	−0.181	−3.337	0.001
(1. Outer lanes congestion) × (7. Systematic volume changes)	0.137	2.262	0.024
(2. Volume level) × (4. Curb lane perturbation)	−0.124	−2.624	0.009
(4. Curb lane perturbation) × (6. Conforming curb volumes)	0.130	2.370	0.018
(5. Volume variation) × (6. Conforming curb volumes)	0.113	2.063	0.039
Constant	−1.124	−19.815	0.000
Reference category: Property damage only			

Alternatively, if there are systematic changes in volumes, for example if the road is transitioning from free flow to congested conditions, or conversely, this compensates for the negative effect of congestion in the left and interior lanes on severity, reducing the effect essentially to zero. The interpretation here is that, while the lower speeds in the left and interior lanes are generally indicative of less severe collisions, this effect is negated under conditions in which the traffic flow is unstable.

Congestion in the curb lane leads to a lower level of injury accidents if curb lane volume conforms to volumes in the left and interior lanes. That is, if the entire road is congested, accidents are more likely to be PDO, rather than injury; again, under such conditions, it can be expected that most collisions that occur would be at low relative speeds among vehicles. Congestion in the curb lane only has little effect on accident severity. The probable explanation is that, while congestion in the left and interior lanes may induce significant lane-changing among vehicles traveling at significantly different speeds, congestion in the curb lane more likely affects either drivers who are among the less aggressive in the traffic stream and/or those vehicles that are consigned to the curb lane because of exiting maneuvers; in both cases, the population of vehicles in this lane would be less likely to alter behavior in a way that would cause significant risk of collision with vehicles in other, uncongested, lanes.

Controlling for whether or not the road is operating under free flow or congested conditions, higher levels of traffic flow are related to a lower likelihood of injury accidents. This is consistent with the generally-accepted inverse relationship between speed and density. The interaction term involving Factor 5 and Factor 6 indicates that higher levels of variation in volume lead to more severe accidents if volume in the right lane is similar to volume in the other lanes. Such conditions are more likely to occur on sections of freeway not in the vicinity of entrance or exit ramps—conditions under which the propensity for lane-changing would not be lessened by ramp-related curb lane restriction; attendant lane changing would, owing to the relatively high variation in traffic volumes in the non-curb lanes, be into (and out of) traffic that would both (a) be expected to be traveling at different speeds and (b) traffic gaps that, according to accepted traffic flow principles, would be accompanied by subsequent accelerations/decelerations required to accommodate an altered density (or, alternatively, spacing).

Controlling for both volume and whether or not the road is operating under free flow or congested conditions, if conditions are relatively the same in all lanes, accidents are more likely to be PDO—again, an obvious consequence of vehicles having low relative speeds with respect to each other. Loss of synchronization leads to a higher likelihood of injury accidents.

5.2. Type of collision

There are four major types of primary collision. Rear-end collisions are most common (59.6%), followed by sideswipes (20.9%) and hit-object collisions (15.1%); the remaining (4.4%) are categorized as "other." Results of a multinomial logit model for collision type, using "other" as the base category, are listed in Table 5. Variables are included in this model if their inclusion leads to a significant overall improvement in the explanatory

Table 5
Multinomial logit model of collision type as a function of traffic flow factors (*t*-statistics in parenthesis)

Explanatory variable	Collision type		
	Sideswipe	Rear end	Hit object
1. Outer lanes congestion	**0.601 (2.75)**	**1.067 (5.00)**	0.011 (0.05)
2. Volume level	0.326 (1.46)	**0.942 (4.36)**	0.026 (0.11)
3. Synchronized lane conditions	0.141 (0.86)	**0.782 (4.93)**	0.102 (0.59)
4. Curb lane perturbation	**0.272 (2.01)**	**0.314 (2.37)**	0.063 (0.45)
7. Systematic volume changes	0.169 (1.22)	**0.385 (2.87)**	−0.044 (−0.31)
(2. Volume level)^2 (squared)	−0.026 (−0.28)	**0.169 (2.00)**	0.073 (0.84)
(3. Synchronized lane conditions)^2	**−0.281 (−2.73)**	**−0.342 (−3.59)**	−0.136 (−1.32)
(5. Volume variation)^2	**0.249 (1.87)**	**0.251 (1.91)**	0.132 (0.95)
(4. Curb lane perturbation) × (6. Conforming curb volumes)	**0.256 (2.27)**	**0.332 (3.00)**	**0.369 (3.06)**
Constant	**2.098 (7.54)**	**1.067 (5.00)**	**1.327 (4.54)**
Reference category: Other type of collision (e.g., overturn, broadside)			

power of the model. The overall fit of the model is very good for models of this type, the Nagelkerke Pseudo-R^2 being 0.283. Statistically significant coefficients are in bold in Table 5 and subsequent tables.

Congestion in the left and interior lanes, on its own, is strongly associated with a greater likelihood of rear-end collisions, all else held constant. To a lesser degree, the probability that an accident is the result of a sideswipe collision also increases with congestion in the left and interior lanes. The explanation here is that the left lanes generally can be expected to comprise the greatest concentration of more "aggressive" drivers, and also feature higher speeds and propensity for maneuvering designed to advance one's relative position in the traffic stream. As such, we would expect that, under conditions in which these lanes are relatively congested, vehicle maneuvers both into and out of these lanes would result in the need/desire for corresponding acceleration/deceleration of vehicles in the traffic stream—conditions conducive to the potential for rear-end collisions. At the same time, vehicles attempting to move out of congested lanes to adjacent, presumably less-congested lanes increase the risk of sideswipe collisions. Right lane perturbation is associated with both rear-end and sideswipe accidents, especially if curb lane volume conforms to the left and interior lane volumes. Right lane perturbation in and of itself, particularly in the vicinity of ramps, would be expected to produce conditions in which non-ramp-bound curb lane traffic would tend to cross lanes to escape such perturbation. The latter case in which curb lane traffic conforms to other lane volumes is more likely associated with sections of freeway not in the vicinity of ramps (which tend to distort the natural tendency toward uniformity across lanes); instability (perturbations) in the curb lane would, in such cases, impact the behavior of a greater number of drivers than in the case in which a significant portion of the curb lane volume was confined to remain in that lane because of the ramp. Conforming curb lane volume combined with curb lane congestion also leads to a higher likelihood of hit-object collisions. This effect is most probably attributable to the tendency of vehicles in congested curb lanes to turn into the shoulder in an evasive maneuver to avoid either rear-end or sideswipe collisions.

Controlling for whether the road is operating under free flow or is congested, the likelihood that an accident is a rear-end collision increases with increasing levels of volume—an obvious consequence of there simply being more candidates available. Variation in volume leads to a marginally greater likelihood of both rear-end and sideswipe collisions. The effect of volume level on the odds of a rear-end collision is given by the exponential function of the sum of the linear and quadratic terms for volume level. This is graphed in Fig. 22. The odds of a rear-end collision increase at an increasing rate for the volume level Factor.

Systematic changes in volume, such as a transition between free flow and congested conditions positively affects the odds of a rear-end collision, confirming the accepted notion that it is the so-called "shoulders of congestion" (both temporal and spatial) that produce the conditions most likely precipitate such accidents.

Controlling for both volume and whether the road is operating under free flow or congested conditions, synchronization of traffic flow conditions across all freeway lanes is related to the odds of both rear-end and sideswipe collisions. Both relationships are nonlinear. The odds of a sideswipe collision increase at a decreasing rate for a little more than half the range of Factor 3, reach a maximum, then decrease, as graphed in Fig. 23. Sideswipe collisions are most likely at about an average level of synchronized conditions, all else

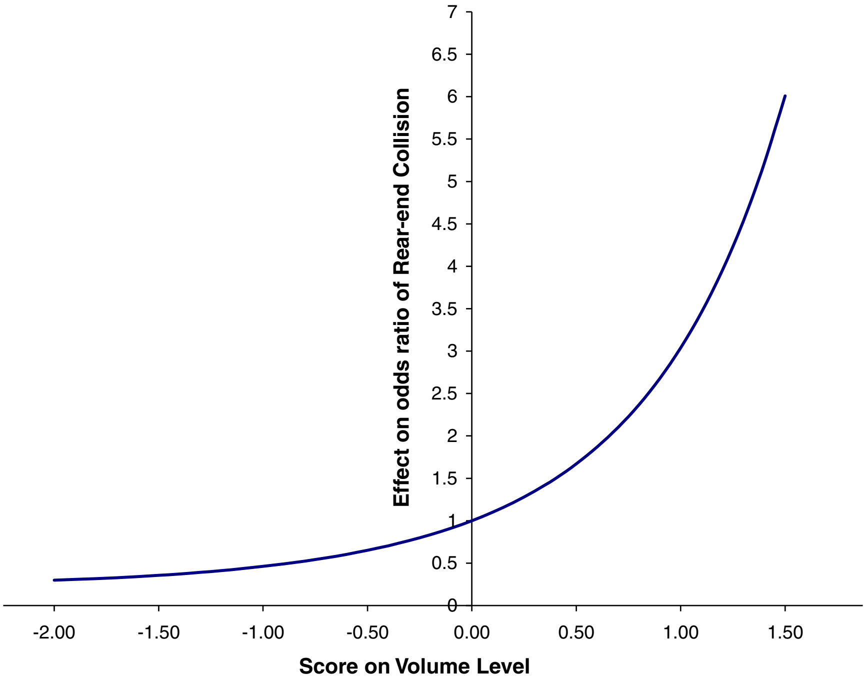


Fig. 22. Effects on the odds of a rear-end collision of Factor 2: volume level.

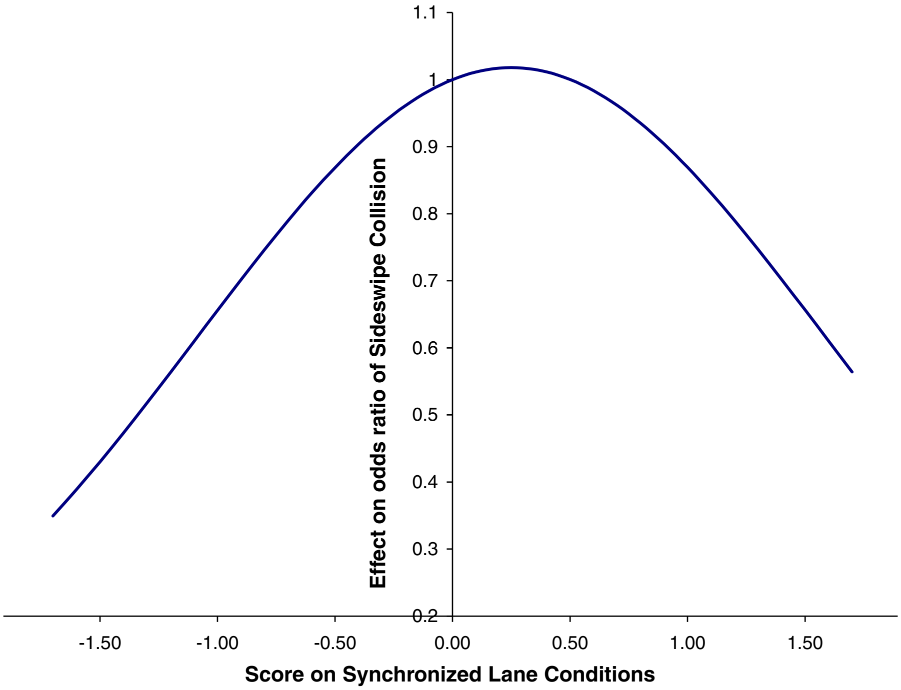


Fig. 23. Effects on the odds of a sideswipe collision of Factor 3: synchronized lane conditions.

held constant. Such collisions are less likely if traffic conditions are either highly synchronized across lanes, or if conditions are highly chaotic across lanes. In the former case, there is relatively little advantage to be expected by changing lanes, and those lane changes that do occur are between lanes in which traffic conditions are very similar; in the latter case, the chaotic nature of the traffic likely discourages a high degree of lane changing, and with it, the potential for collisions. The nonlinear effects of synchronized lane conditions on the odds of a rear-end collision are graphed in Fig. 24. The odds increase with increasing score on Factor 3 over the lower 85% of the range of this Factor. For the highest 15% of scores on Factor 3 (above the value of positive 1.2 standard deviation from the mean), the odds of a rear-end collision fall with increasing synchronization. The lower range of synchronization is generally associated with relatively light to moderate, uncongested, traffic regimes in which there can be expected to be significant mixing of traffic between lanes moving at different relative speeds. At the low end of the synchronization factor, sparse traffic conditions prevail in which there are relatively few collision opportunities; as traffic intensifies, collision opportunities increase as does synchronization (due to well-known car-following principles coming into play). At some point (at about score value 1.25 on Factor 3) the “calming” synchronization effects overcome the speed differentials between lanes, and the risk of rear-end collisions falls.

5.3. Collision location

Location of the primary collision of an accident is broken down into five categories. Collisions in the Interior lane(s) are most common (36.3%), followed in order by left-lane collisions (28.3%), right-lane collisions (17.1%), run-offs to the drivers’ left (11.6%) and run-offs to the drivers’ right (6.7%). A multinomial logit model was estimated to determine the relationships between the Traffic Flow Factors and collision location, and the results are listed in Table 6. The base category is “Interior lane(s)”. Although not as good as that obtained in the previous collision type model, the overall fit of the model indicated by the Nagelkerke Pseudo-R^2 of 0.140, is considered good by accepted standards.

The results indicate that congestion has less effect on collision location than it does on accident severity and collision type. Only congestion in the left and interior lanes is a statistically significant predictor of collision location. A higher degree of congestion in the left and interior lanes is related to increased odds of an accident being in one of the lanes, as opposed to off-road—this a result of the majority of off-road accidents being asso-

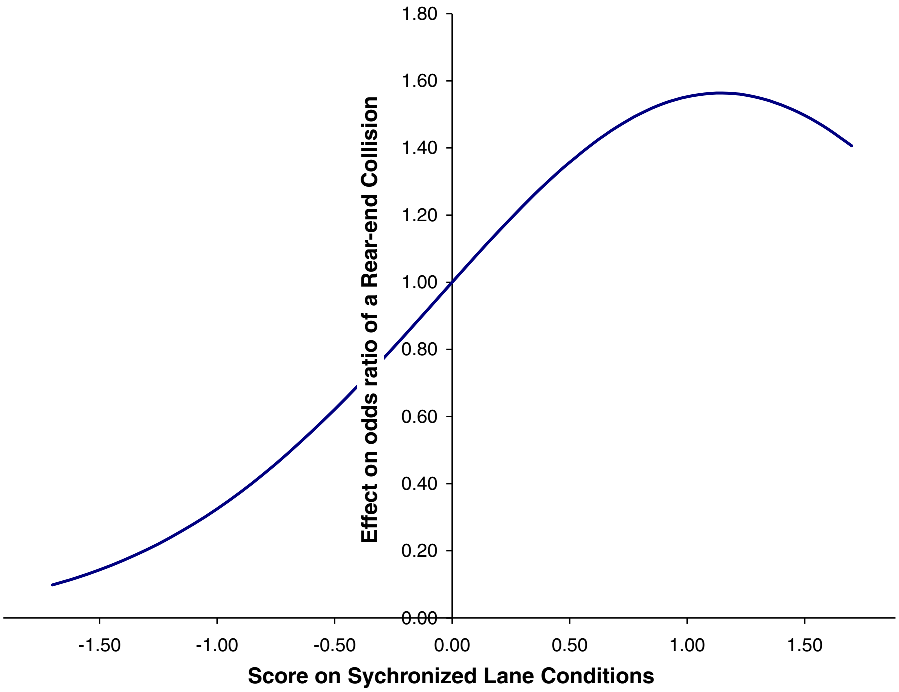


Fig. 24. Effects on the odds of a rear-end collision of Factor 3: synchronized lane conditions.

Table 6
Multinomial logit model of collision location as a function of traffic flow factors (*t*-statistics in parenthesis)

Explanatory variable	Collision location			
	Off-road left	Left lane	Right lane	Off-road right
1. Outer lanes congestion	**−0.349 (−3.85)**	**0.182 (2.96)**	−0.045 (−0.61)	**−0.638 (−4.70)**
2. Volume level	**−0.188 (−2.51)**	**0.226 (3.10)**	0.053 (0.73)	**−0.347 (−4.05)**
3. Synchronized lane conditions	−0.027 (−0.30)	**0.368 (5.99)**	0.136 (1.90)	−0.164 (−1.30)
7. Systematic volume changes	0.027 (0.34)	**0.249 (3.92)**	0.044 (0.60)	**−0.267 (−2.41)**
8. Synchronized outer flow	**0.236 (2.74)**	**0.279 (4.07)**	−0.016 (−0.18)	0.223 (1.92)
(8. Synchronized outer flow)^2	**0.107 (2.36)**	**0.097 (2.41)**	−0.063 (−1.15)	−0.007 (−0.09)
(6. Conforming curb volumes) × (7. Systematic volume changes)	−0.066 (−0.97)	**−0.193 (−3.34)**	−0.045 (−0.69)	0.045 (0.47)
(6. Conforming curb volumes) × (8. Synchronized outer flow)	0.045 (0.58)	**−0.145 (−2.31)**	0.131 (1.70)	0.146 (1.33)
Constant	**−1.321 (−13.34)**	**−0.435 (−5.78)**	**−0.702 (−8.46)**	**−2.008 (−13.91)**
Reference category: Interior lane(s)				

ciated with very sparse, late-night, traffic. Expectedly, congestion in the left and interior lanes has no significant effect on whether or not an accident is located in the right lane.

Controlling for whether the road is operating under free flow or congested conditions, volume level has a positive effect on the odds of an accident being located in the left lane, versus off-road, to either side. This characteristic is consistent with the observation noted above that off-road accidents occur predominantly under extremely sparse traffic. If the road is undergoing systematic changes in volume, controlling for volume level, there is a greater likelihood that an accident will be located in the left lane, not in the right lane. Most likely, this can be explained by the propensity for traffic to seek out the presumably faster-moving lanes as traffic volumes steadily increase toward peak period values. This effect is enhanced if right lane volumes are nonconforming—perhaps a result of increased ramp activity as the peak period approaches—but diminished if right lane volumes are conforming.

Controlling for both volume and whether the road is operating under free flow or congested conditions, the degree to which traffic conditions are synchronized over the lanes has a substantial and complex effect on where an accident is likely to occur. In this case, the statistically significant predictor of collision location is Factor 8: Synchronized Outer Flow, whereas in the case of collision type it was Factor 3: Synchronized Lane Conditions. The location categories affected are off-road left and left lane.

The effects of synchronized outer flow on the probability of an off-road-left location are graphed in Fig. 25. The odds of an accident being located off-road left are reduced for below-average levels of synchronization. For above-average levels, the odds multiplier increases with the level of synchronization in the left and interior lanes at an increasing rate.

The effects of synchronized outer flow on the likelihood of a left-lane location are shown in Fig. 26. Although similar in shape to the effect noted in Fig. 25 for off-road-left accidents, the effect for left-lane accidents is coupled to Factor 8. The interaction involving Factor 8 and Factor 6 is captured by parameterizing the effect by the level of conforming curb volume. Three curves are graphed: a below-average conforming curb volume score of minus one standard deviation, an average score, and above-average score of plus one standard deviation. Non-conforming curb volumes accentuate the effect of synchronous outer flow on the likelihood of a left-lane accident location.

5.4. Number of involved vehicles

Although the majority of case study accidents (59%) involved two vehicles, there were a sufficient number of accidents involving three vehicles (20.1%), single vehicles (13.2%), and four or more vehicles (7.8%) to allow four categories in a multinomial logit model of vehicle involvement as a function of Traffic Flow Factors. The base category for the model presented in Table 7 is "three vehicles". The overall fit of the model (Pseudo-R^2 of 0.159) is considered good.

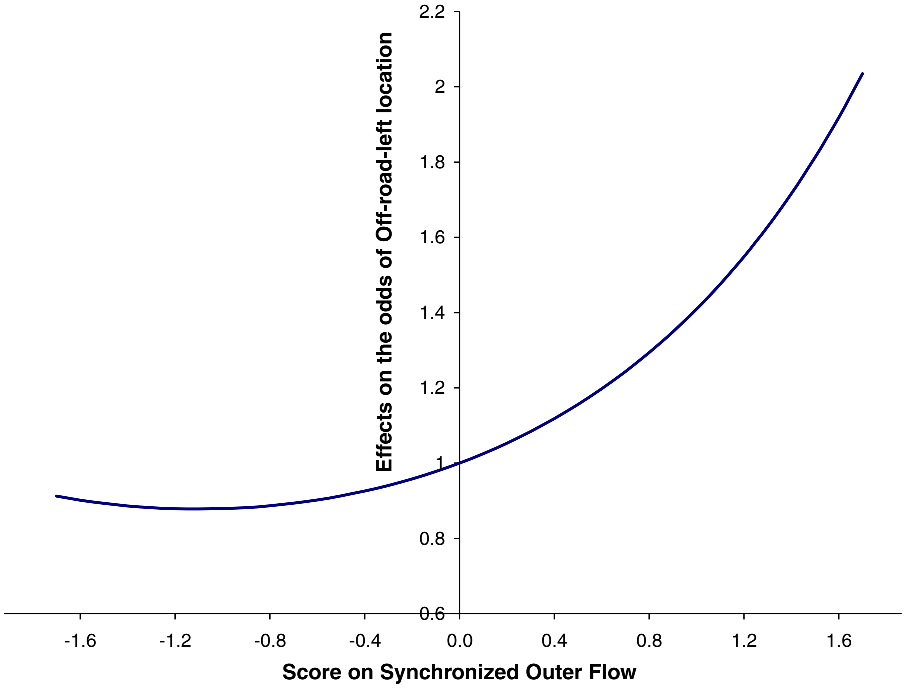


Fig. 25. Effects on the odds of a off-road-left location of Factor 3: synchronized outer flow.

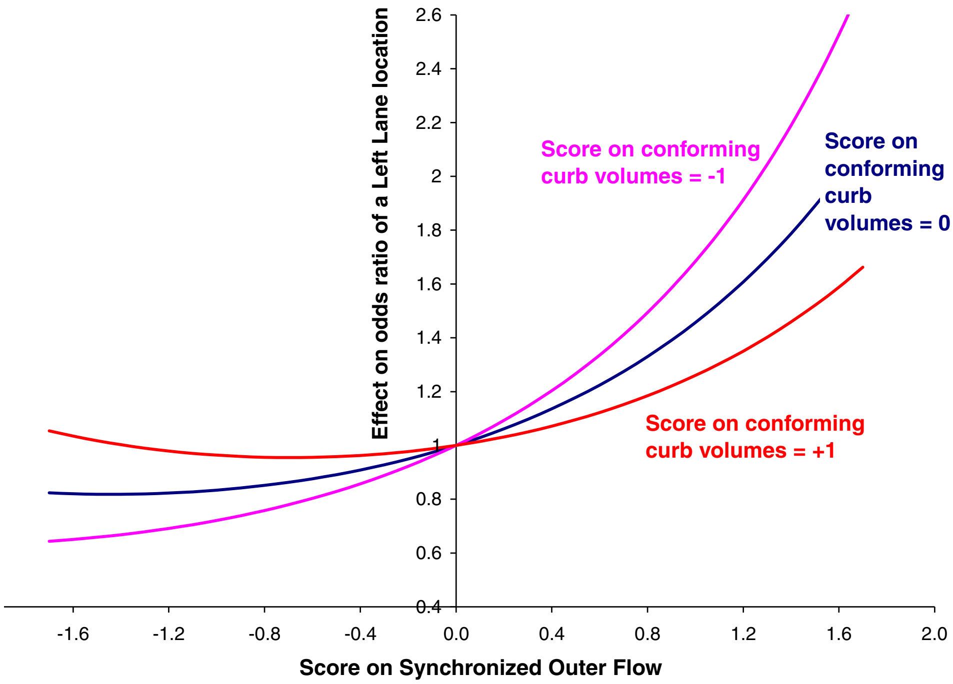


Fig. 26. Effects on the odds of a left-lane location of Factor 8: synchronized outer flow, and Factor 6: conforming curb volumes.

As expected, congestion has a considerable influence on vehicle involvement. Congestion in the left and interior lanes distinguishes single vehicle crashes from multi-vehicle crashes. The logarithm of the odds of a

Table 7
Multinomial logit model of number of involved vehicles as a function of traffic flow factors (*t*-statistics in parenthesis)

Explanatory variable	Number of involved vehicles		
	Single vehicles	Two vehicles	⩾Four vehicles
1. Outer lanes congestion	**−0.754 (−6.31)**	0.062 (0.97)	−0.030 (−0.27)
2. Volume level	**−0.728 (−8.12)**	−0.131 (−1.82)	0.228 (1.61)
3. Synchronized lane conditions	**−0.414 (−3.97)**	−0.080 (−1.32)	0.122 (1.23)
5. Volume variation	0.156 (1.66)	0.007 (0.11)	**0.331 (2.97)**
6. Conforming curb volumes	0.082 (0.83)	**0.182 (2.91)**	−0.138 (−1.29)
7. Systematic volume changes	**−0.414 (−4.56)**	−0.112 (−1.76)	0.189 (1.83)
(4. Curb lane perturbation)^2	**−0.159 (−2.88)**	**−0.105 (−2.89)**	−0.031 (−0.55)
(5. Volume variation) × (6. Conforming curb volumes)	−0.065 (−0.71)	0.094 (1.47)	**0.249 (2.45)**
Constant	**−0.685 (−5.53)**	**0.062 (0.97)**	**−1.066 (−.03)**
Reference category: Three vehicles			

single-vehicle accident is a simple (negative) linear function of Factor 1: Outer Lanes Congestion. As noted earlier, accidents involving a single vehicle are predominantly associated with late-night hit-object and run-off-road accidents that occur during periods of extremely light traffic conditions. While this Factor is the dominant determinant of single vehicle accidents, it does not significantly distinguish between the likelihood of different numbers of vehicles in multi-vehicle accidents, which, by their nature, rarely fall into the categories of hit-object and run-off-road accidents.

Curb lane perturbation is related to vehicle involvement in a more complex manner. There are statistically significant nonlinear effects for both single-vehicle and two-vehicle involvement. In both cases, increased curb lane perturbation decreases the odds of such involvement; i.e., increases the odds of multiple-vehicle involvement. This is expected, since high levels of such perturbations are typically associated with congested traffic regimes. The multiplicative effect on the odds of a single vehicle being involved in any accident is graphed in Fig. 27 as a function of the score on Factor 4: Curb Lane Perturbation. For extreme low values of curb

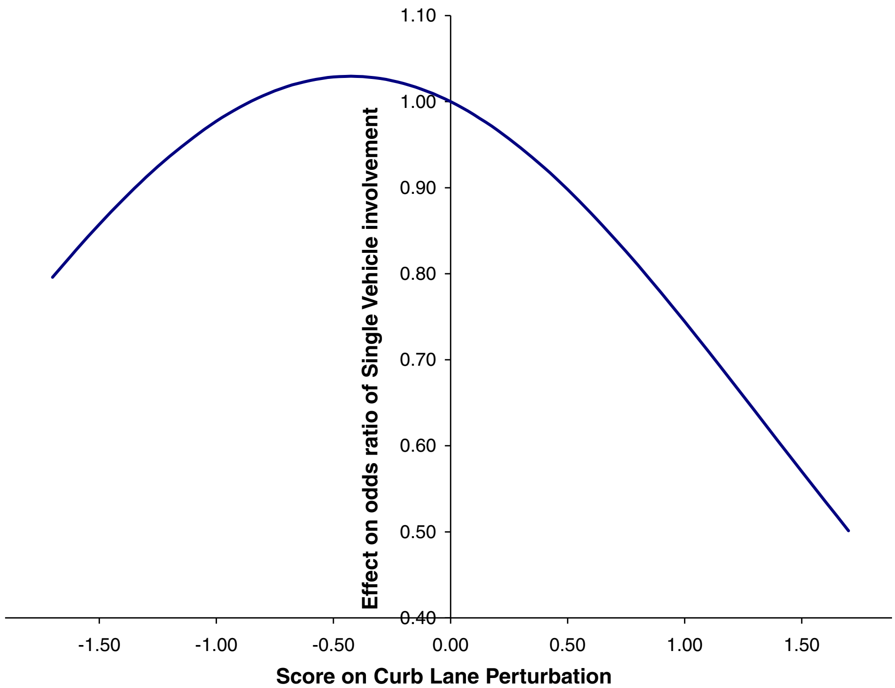


Fig. 27. Effects on the odds of single-vehicle involvement of Factor 4: curb lane perturbation.

lane perturbation—scores less than about −0.85 standard deviations—the odds of a single-vehicle accident are reduced in approximate proportion to the value of the Factor score. For Factor scores between −0.85, and −0.40 the odds of a single-vehicle accident decrease at a declining rate, peaking at the score of −0.40. For positive scores, these odds decrease in approximate proportion to the score. The interpretation here is that, while increasing perturbation in the curb lane most certainly produces corresponding increased likelihood of accidents involving vehicles traveling in that lane, perturbations as being more closely associated with heavier traffic concentrations make more likely multiple vehicle collisions than single vehicle (run-off-road and/or hit-object) accidents. This effect toward multiple-vehicle accidents (and, away from single-vehicle accidents) is heightened in the midrange of the Factor score, and lessens at the extremes—ostensibly at the lower extreme owing to such values being associated with the light traffic conditions present during periods of relatively high numbers of late-night run-off road accidents, and at the higher extreme to high perturbations in dense, congested, traffic in which there is a heightened tendency for vehicles in the curb lane to execute escape maneuvers that involve moving onto the shoulder.

The effects of curb lane perturbation on the odds of two vehicles being involved are plotted in Fig. 28. For negative factor scores, these odds decrease in rough proportion to the absolute value of the score. Below-average curb lane perturbation leads to a lower probability of two-vehicle accidents. For positive scores, the effect is to increase the likelihood of two-vehicle accidents slightly over the effective domain, with a maximum effect at a score of 0.90 standard deviations. Above-average curb lane perturbation leads to a slightly higher probability of two-vehicle accidents.

Controlling for whether or not the road is operating under free flow or congested conditions, volume level distinguishes among all of the levels of vehicle involvement. As expected, higher volume levels lead to a diminished probability of single-vehicle accidents, but volume level itself does not substantially differentiate among the levels of multi-vehicle collisions. Two-vehicle accidents are more likely under higher levels of conforming curb volumes; two-vehicle accidents are more likely when volumes are similar in all lanes. Finally, large scale accidents (those involving four or more vehicles) are positively related to volume variation and the interaction of volume variation and conforming curb volumes. Large scale accidents are more likely to occur when volumes are similar in all lanes *and* there are high levels of variation in these volumes.

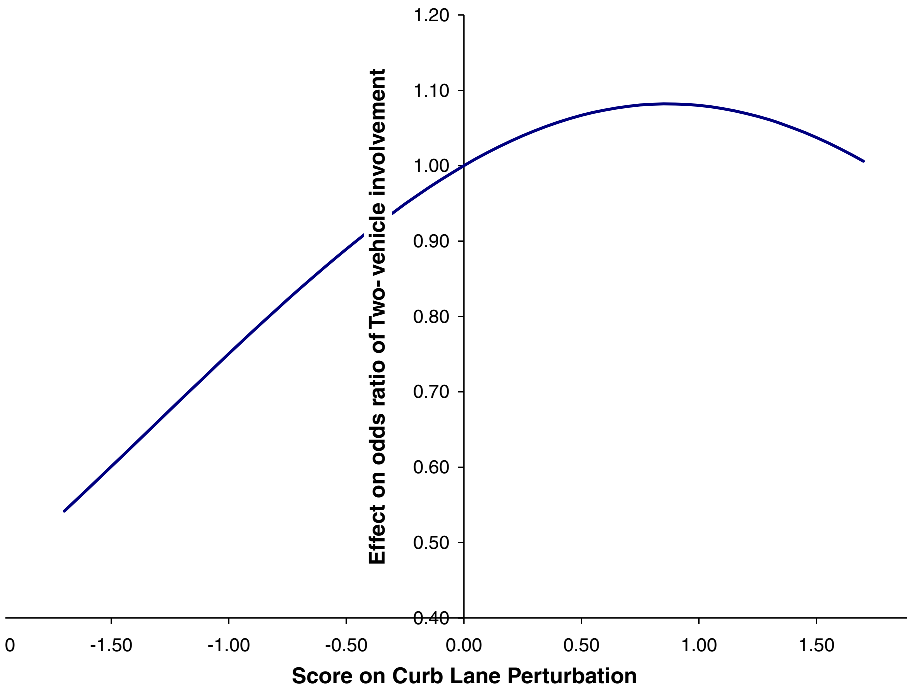
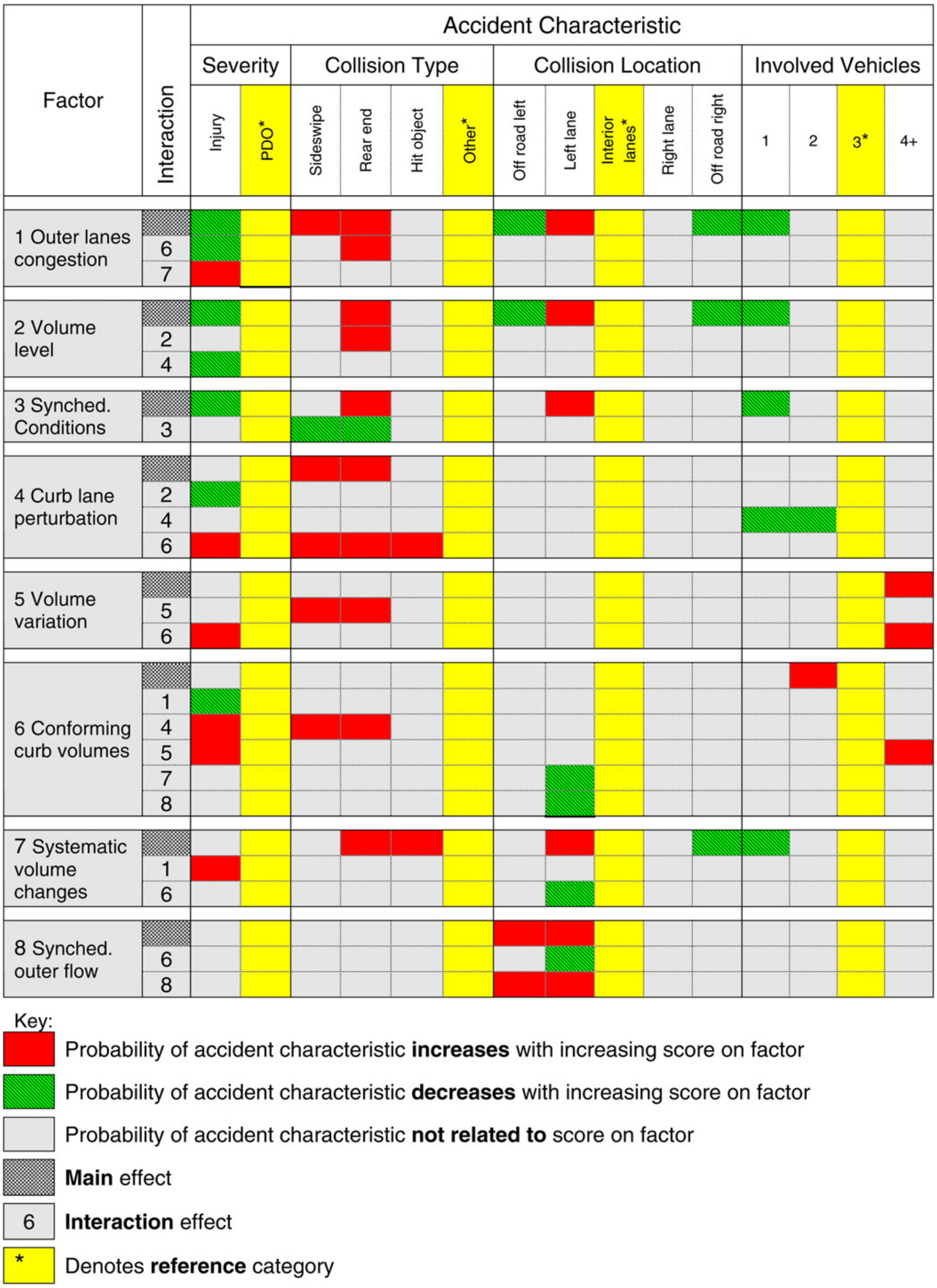


Fig. 28. Effects on the odds of two-vehicle involvement of Factor 4: curb lane perturbation.

Synchronized lane conditions, controlling for volume and whether or not the road is operating under free flow or congested conditions, affects only the likelihood of single-vehicle versus multi-vehicle accidents. The

Table 8
Summary of Key Results from Logit Models of Accident Characteristics as a Function of Traffic Flow Factors



higher the level of synchronization, the higher the probability that an accident will involve more than one vehicle.

5.5. Summary of Traffic Flow Factors and accident characteristics

A summary of the main results of the analysis of accident propensity as a function of traffic flow is presented in Table 8. Each of the eight Traffic Flow factors is effectively related to at least two of the four sets of accident characteristics. This sensitivity bodes well for continued research into the development of hazard functions in the eight-dimensional space of the Factors and their second-level interactions.

Because the determinants of safety are explicitly related to traffic conditions, their values—and with them the corresponding conditions conducive to propensity for accidents—on any section of freeway under prevailing conditions can ostensibly be manipulated by applying traffic engineering principles known to produce certain results. For example, an investigation of high accident rates during peak hour traffic in the vicinity of an entry ramp may reveal significant perturbations in traffic in the curb lane (Factor 4), raising the probabilities of both rear-end and sideswipe collisions to an unacceptable level. Using the models contained herein, traffic engineers could evaluate the expected effectiveness of mitigating the problem either through capital construction (e.g., adding an extended auxiliary lane) or through traffic management (e.g., ramp metering).

6. Conclusions

Understanding the benefits of improved traffic flow (reduced congestion) is critical to the assessment of investments in infrastructure or traffic management and control. The manner in which safety is improved by smoothing traffic flow is not well understood at the present time. However, the models described in this paper indicate that we now have the capability to implement a tool that can be used either in real-time monitoring of the safety level of any freeway traffic flow, or for forecasting the safety aspects of changes in traffic flows.

It has been demonstrated that an extensive set of statistical parameters—36 in total—can be extracted from 20 min of loop detector data for three lanes at a specific location and time, without recourse to untenable assumptions that convert loop detector data to densities and speeds. These statistical parameters can be reduced to a set of eight weighted averages (called Factors) with minimal loss of information. These Traffic Flow Factors perform well in explaining different modes of traffic flow, as uncovered in a series of visualizations of loop detector data. The Factors also perform well in terms of explaining differences among accident characteristics.

The objective of this research has been to develop a real-time tool for safety analysis of freeways that could be implemented with minimal effort, relying only on data commonly available from standard inductance loop detectors. When deployed as an on-line, real-time, application, the models can be used to alert transportation management center (TMC) personnel to the onset of conditions that may be conducive to the occurrence of freeway accidents, much in the same manner that conventional loop-detector-based incident detection models are used to alert such personnel to possible blockages that affect freeway performance. Of course, since the actual probabilities of any single accident event occurring at a specific location during any 30-s time interval remain extremely small, the model outputs would only identify general, relative, tendencies toward potentially dangerous conditions. However, this information could prove extremely valuable to TMC operators, for example, in terms of which among a multitude of CCTV surveillance cameras to activate and monitor at any given time, or where to concentrate freeway service patrol vehicles.

A primary goal of this effort is to provide an easily accessible tool for use in assessing the safety performance of freeway operations and to evaluate and document improvements to safety arising from such Intelligent Transportation System (ITS) deployments as system-wide ramp metering (SWARM), freeway service patrol (FSP) and other incident response measures, and driver information. By quantifying the safety benefits accrued from smooth and efficient traffic operations, freeway operating agencies will be able to incorporate safety measures in assessment of performance gains resulting from ITS deployment. Another application will be to forecast the safety implications of proposed projects by evaluating the levels of safety implied by traffic simulation model outputs. The tool can also be used to forecast the safety consequences of doing nothing. The models provide evidence of how traffic flow can be affected in order to reduce freeway crashes—translating traffic flow, as measured by ubiquitous single loop detectors, into safety performance in terms of expected numbers of crashes by type of crash per exposed vehicle mile of travel.

Acknowledgements

This research was funded in part by the California Partners for Advanced Transit and Highways (PATH) and the California Department of Transportation (Caltrans). The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the University of California, California PATH, or the California Department of Transportation.

References

Caltrans, 1993. Manual of Traffic Accident Surveillance and Analysis System. California Department of Transportation, Sacramento.

FHWA, 2000. Guidebook for the California state Data Files. Report FHWA-RD-00-137, Federal Highway Administration, US Department of Transportation. <http://www.hsisinfo.org/pdf/00-137.pdf> (accessed May 23, 2005).

Golob, T.F., Recker, W.W., 2003. Relationships among urban freeway accidents, traffic flow, weather and lighting conditions. ASCE Journal of Transportation Engineering 129, 342–353.

Golob, T.F., Recker, W.W., 2004. A method for relating type of crash to traffic flow characteristics on urban freeways. Transportation Research, Part A 38, 53–80.

Golob, T.F., Recker, W.W., Alvarez, V.M., 2002. Freeway Safety as a Function of Traffic Flow: The FITS Tool for Evaluating ATMS Operations. Final Report Prepared for California Partners for Advanced transit and Highways (PATH). Institute of Transportation Studies, University of California, Irvine, CA.

Golob, T.F., Recker, W.W., Alvarez, V.M., 2004a. A tool to evaluate the safety effects of changes in freeway traffic flow. ASCE Journal of Transportation Engineering 130, 222–230.

Golob, T.F., Recker, W.W., Alvarez, V.M., 2004b. Freeway safety as a function of traffic flow. Accident Analysis and Prevention. 36, 933–946.

Varaiya, P., 2005. What We Have Learned about Highway Congestion. April 10, 2005. <http://paleale.eecs.berkeley.edu/~varaiya/papers_ps.dir/accessF05v2.pdf> (accessed May 12, 2005).