Implementation of a Tool for Measuring ITS Impacts on Freeway Safety Performance

Thomas F. Golob

Will Recker

Ioannis Pavlis

Institute of Transportation Studies
University of California Irvine

ATMS Testbed Technical Report TTR3-04

This work was performed as part of the ATMS Testbed Research and Development Program of the University of California, Irvine.* The contents of this report 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 of polices of the State of California. This report does not constitute a standard, specification, or regulation.

June 2005

* In cooperation with the California Partners for Advanced Transit and Highways

Implementation of a Tool for Measuring ITS Impacts on Freeway Safety Performance

Thomas F. Golob

Wilfred W. Recker

Ioannis Pavlis

Institute of Transportation Studies
University of California Irvine

Executive Summary

In this first year of a two-year project, we lay the groundwork for the development of a performance tool that gauges the level of safety of any type of traffic flow on a California freeway. The inputs to this tool are data from single loop detectors, so the tool 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 1,700 accidents in our case study. This case study covers the six major freeways in Orange County for a six-month period in 2001. In this first year of the project we have uncovered an extensive set of statistical parameters that capture those aspects of traffic flow that are strongly related to accident potential.

In this work we recognize 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). Consequently, we avoid using any direct speed or density measures among the parameters. These parameters 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. 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 type of traffic flow.

The second year of the project will be devoted to estimating exposure to traffic conditions to establish accident rates, followed by the coding of the tool that will allow engineers and planners to evaluate the safety consequences of traffic flow, either as a real-time heads-up of potentially unsafe conditions, or for use in evaluating projects.

1 PROJECT OBJECTIVES

The objective of this project is to implement a real-time tool for safety analysis. The overall project goal is to calibrate and verify a tool that translates 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. This tool can be used in monitoring the safety performance of freeway operations and to evaluate and document improvements to safety arising from such ITS deployment 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, Caltrans should 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 safety aspects of costs and benefits can be assessed by comparing the levels of safety estimated by the tool for traffic flows before and after implementation of a treatment, such as a component of an intelligent transportation system (ITS) or infrastructure project. It can also be used to forecast the safety consequences of doing nothing. It is meant to complement performance measurement systems that focus on travel times and delay (PeMS, 2005).

In the first year of the project, reported here, our objective has been 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. This represents a search for evidence of how traffic flow can be affected in order to reduce freeway crashes.

In the second year of the project, the goal will be to test the model's ability to distinguish locations and conditions with high accident rates from those with low accident rates. Because the methodology does not depend directly on specific geometric characteristics, but rather is based on the traffic conditions arising from both roadway layout and demand, the goal is to demonstrate that the tool can be readily transferred to any urban freeway that is fully instrumented with loop detectors without the need for extensive calibration. Once validated, code will be developed to deploy the model, first as a stand-alone on the Testbed website using data from the Caltrans District 12 FEP as input. Eventually the tool could provide the safety element of a performance measurement system such as PeMS.

2 BACKGROUND

The present project builds upon a previous PATH project, which involved the development of FITS (Flow Impacts on Traffic Safety) (Golob, Recker, and Alvarez, 2002). In Golob, Recker and Alvarez (2004), we stated that a common aim of transportation management and control projects on urban freeways is to increase productivity by reducing congestion. Reducing congestion ostensibly leads to reductions in travel time, vehicle emissions and fuel usage, and improved travel time reliability. PeMS has been implemented to measure the real-time performance of any instrumented segment of freeway in terms of throughput: travel time per vehicle, average speed or total delay (Chen, *et al.*, 2001; Choe, Skabardonis, Varaiya, 2002; Varaiya, 2001). The inputs to these tools are typically total flows and mean speeds computed from volume and occupancy data from single inductive loop detectors, typically for intervals of 30-seconds or more. Increasingly, such single loop detectors are distributed throughout the freeway system. Data from more accurate but less ubiquitous sensors, such as double loops and video cameras, is often used to adjust or calibrate single loop measurements, but the primary source of real-time surveillance data for traffic management is likely to remain the single loop detector for the foreseeable future.

Reduced congestion and smoothed traffic flow are also likely to improve safety, as well as reduce psychological stress on drivers. Concentrating on the safety issue, our objective in this paper is to demonstrate that researchers are beginning to understand the relationship between safety and improved traffic flow. Recent developments indicate that the time is right to refine and implement analytical tools that can be used in real-time monitoring of the safety level of the traffic flow on any instrumented segment of freeway. As opposed to tools that measure freeway performance in terms of throughput or travel time, we found that the key elements of traffic flow affecting safety are not only mean volume and speed, but also variations in volume and speed. We further determined that it is important to capture variations in speed and flows separately across freeway lanes, and that such information is useful in differentiating types of crashes.

3 DATA

Two datasets are used in this project: (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 Vehicle Detection system (VDS). The VDS traffic flow data we need have been received directly from front-end processors (FEP) using the UCI ATMIS Testbed Intertie with Caltrans District 12. These data have been used in two other PATH projects: “Development of a Path Flow Estimator for Deriving Steady-State and Time-Dependent Origin-Destination Table Trips,” and “A Tool for the Incorporation of Non-Recurrent Congestion Costs of Freeway Accidents in Performance Management.” The data are also being used in the project “Modeling Matched Traffic and Accident Datasets to Significantly Improve Safety and Efficiency of Urban Freeway Operations,” funded by the National Science Foundation. In this first phase of the analysis, accident data from TASAS are linked with these traffic flow data. In analyses that will be conducted during the second year of the project, the traffic flow data associated with accidents will be compared to traffic flow data corresponding to times and places at which no accident was reported. This will allow us to develop models to assess the hazards of different types of traffic flow.

3.1 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 percent of the accidents in our study have reported times in minutes that fall precisely on the twelve 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 needs to be cut off 2.5 minutes before the nominal crash time.

3.2 Traffic Flow Data from Single Loop Detectors

Our traffic flow data are drawn from the Vehicle Detection System (VDS) thirty-second single loop detector data. There are approximately 8,000 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 thirty second (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. Thus, combined with calibration studies of average vehicle length, volume and occupancy can be converted to average speed and corresponding travel time, which then forms the basis of performance monitoring tools, such as PeMS (2005). However, for our disaggregated purposes, there is no accurate information on average vehicle lengths for each 30-second interval, or for any aggregation of data over the periods of time (e.g., twenty minutes) 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. Consequently, we make different use of VDS data than is done in PeMS and similar aggregate studies.

A scatter plot of raw 30-s loop detector data for one lane for an extended period of time is shown in Figure 1. Over this period the lane was operating in each of the two basic traffic flow regimes: (1) free flow, and (2) congested flow. Free flow operation corresponds to the cone in which the slope of the rays from the origin are steep, and this slope varies very little from observation to observation. In statistical terms, volume can vary substantially, but occupancy will vary within a small range of the domain from zero to one. Two 30-s points might have a slightly different ratio of volume to occupancy (different slopes), but it is unknown as to whether this difference is due to different mean speeds or different average vehicle lengths. In the congested flow regime the vector slope is small compared to the free flow regime, and the varies substantially from observation to observation; occupancy typically varies at least as much as volume. Of course, there is a gray area between free and congested flow.

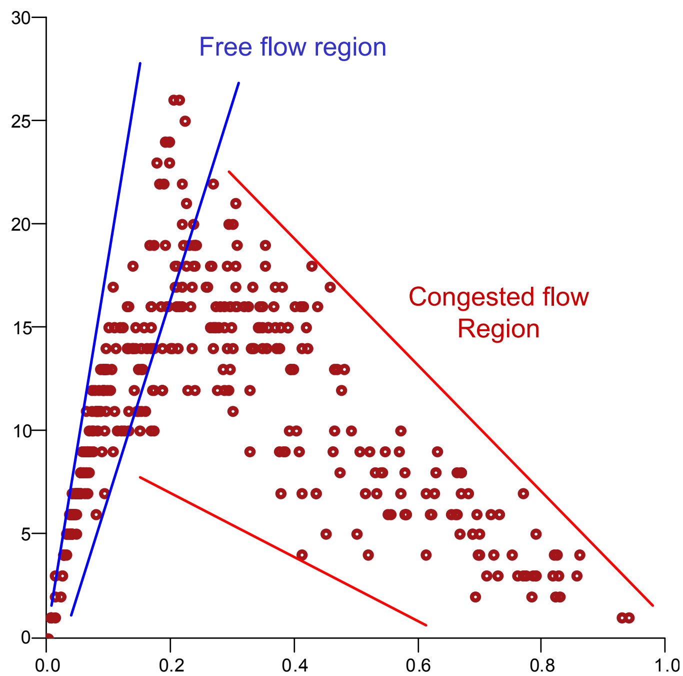

Figure 1 A Typical Pattern of Raw Loop Detector Data Showing Both Observations of Free Flow and Congested Flow Operation

A twenty-minute time trace of 30-s data is shown In Figure 2. For the first fifteen minutes or so, the lane is operating in free flow. Volumes vary from 5 to 15 vehicles per 30-s (equivalent to 600 to 1,800 vehicles per hour), and occupancies are roughly in a constant proportion to volume. However, after fifteen minutes 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.

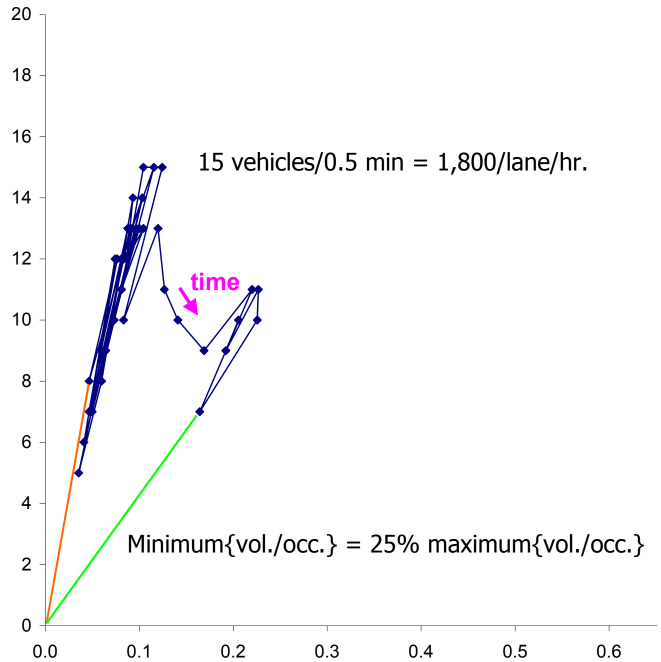

Figure 2 Trace of Twenty Minutes of 30-Second Loop Detector Data Showing Transition from Free Flow to Congested Flow Operation

Note that 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 to 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. The use of hourly …

3.3 Data Limitations and Scope of the Case Study

The network used in the project is comprised of the six major Orange County (non-toll) freeways (I-5, SR-22, SR-55, SR-57, SR-91, I-405). The period of study is March through August, 2001 (six full months). This is the period for which we have both traffic flow and accident data for the case study network. The six routes are divided into segments, each with an adequate number of accidents to allow identification of route segments in subsequent analyses.

The TASAS database records 4,412 accidents on the mainlines of these freeways recorded in TASAS. Of these, we have some loop detector data for 2,924, or 66% of the accidents. We have sufficient data for analysis for 769, or 40% of the TASAS accidents. For reasons unknown to the authors, 2001 loop detector data are lacking for four segments of the six Orange County Freeways, comprising 31.6 centerline miles. These segments, which must be excluded from the analyses due to poor data coverage, are: (1) I-5 from the SR-22 and SR-57 junction to the Los Angeles County line, (2) SR-55 from its junction with SR-22 to its end at SR-91, (3) SR-91 from the Los Angeles County line to SR-55, (4) I-405 from SR-22 eastbound to the L.A. County line. As these problematic segments are all at the ends of the routes, the resulting case study network, comprising 97.0 centerline miles, is contiguous. For the final case study network, we have sufficient loop detector data on 51.4% of all accidents recorded during the six month study period.

Table 1 The Case Study Orange County Freeway Network, also Showing Segments That Must be Excluded from Future Analyses (Listed in Red Bold Font).

			Accidents March-August 2001					
		can segment be included?	centerline miles	insufficient loop data		sufficient loop data		total
Route	Route segment			number	percent	number	percent	
5	San Diego Co. to SR-74 eastbound	yes	9.6	71	39.7%	108	60.3%	179
5	SR-74 to I-405	yes	11.7	147	45.8%	174	54.2%	321
5	I-405 to SR-55	yes	9.0	92	45.1%	112	54.9%	204
5	SR-55 to SR-22/SR-57	yes	3.8	66	31.1%	146	68.9%	212
5	**SR-22/SR-57 to Los Angeles Co.**	**no**	**10.4**	**265**	**97.4%**	**7**	**2.6%**	**272**
22	GoldenWest to about Tustin Ave	yes	10.2	193	50.4%	190	49.6%	383
55	Victoria/22 St. to SR-22	yes	11.0	202	49.6%	205	50.4%	407
55	**SR-22 to SR-91**	**no**	**4.9**	**122**	**89.1%**	**15**	**10.9%**	**137**
57	Chapman to SR-91	yes	4.4	100	51.0%	96	49.0%	196
57	SR-91 to Los Angeles Co.	yes	7.0	138	50.2%	137	49.8%	275
91	**L.A. County to SR-55**	**no**	**12.9**	**487**	**94.0%**	**31**	**6.0%**	**518**
91	SR-55 southbound to Riverside Co.	yes	9.7	112	30.8%	252	69.2%	364
405	I-5 to SR-73	yes	10.1	203	63.8%	115	36.2%	318
405	SR-73 to SR-22 eastbound	yes	10.5	297	62.7%	177	37.3%	474
405	**SR-22 to L.A. County**	**no**	**3.4**	**148**	**97.4%**	**4**	**2.6%**	**152**
	included in analysis		97.0	1,621	48.6%	1,712	51.4%	3,333
	exclusing some segments		**31.6**	**1,022**	**94.7%**	**57**	**5.3%**	**1,079**
	Total		128.6	2,643	59.9%	1,769	40.1%	4,412

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 our accidents with traffic flow data are a random sample of all reported accidents. Our analyses are representative of all accidents that were reported in case study area over six months of 2001.

3.4 Creation of Thirty-Six Traffic Flow Parameters

While there are only two raw data measurements – volume (flow) and occupancy – for each of the freeway lanes at the site of an accident, repeated measurements over time at 30-second intervals allows the opportunity of capturing temporal variations. The available of the same data for multiple lanes allows comparisons across lanes. Sensitivity analyses revealed that we need twenty minutes of data (40 30-s observations) in order to estimate our traffic flow parameters. Furthermore, in calculating these parameters, we an accept five minutes of missing data over the twenty minute period. That is, we need a minimum of 30 observations over 20 minutes. Also, due to strong correlations of data across all interior lanes, we only need data for three lanes: (1) the leftmost, number one, or median lane (excluding any HOV lanes, designated "1"): (2) one interior lane (whichever has least missing data, designated "M"), and the rightmost or curb lane (designated "R"). The use of only three lanes allows consistent definitions for all locations within our case study network, as all of our freeway sections have at least three directional lanes.

Four types of parameters were found to be useful; these are summarized in Table 2:

1. C*oefficients of variation* are dimensionless indices of variation in traffic flow conditions over time. They are defined as the standard deviation of any variable measured in terms of its mean (i.e., the ratio of standard deviation and mean). The minimum sample size for these calculations is 30, the maximum is 40 (30-s observations). Coefficients of variation were calculated for three variables: volume, occupancy, and the ratio of volume to occupancy. Each of these measures was computed for each of the three lanes (Lane 1, Lane M, and Lane R), leading to nine coefficients of variation.
2. *Correlations of traffic conditions across the lanes* are dimensionless indices of lane-to-lane coordination in traffic flow. Three pair-wise comparisons are available: lane 1 vs. lane M, lane 1 vs. lane R, and lane M vs. lane R. These were calculated for volume, occupancy, and volume/occupancy, leading again to nine parameters.
3. *Autocorrelations* are dimensionless measures of the temporal consistency of traffic flow conditions. They are defined as 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 minute period. Autocorrelations were computed for volume and occupancy for each of the three lanes, leading to six parameters. Autocorrelations were not computed for the ratio of volume to occupancy, because we detected instability and encountered a large number of missing observations due to zero values of occupancy.
4. Means and standard deviations are the only parameters that are not dimensionless. Effective average vehicle length is not known, so we cannot calculate these parameters for occupancy and the ratio of volume to occupancy, because we are unable to convert to, respectively, density and speed. However we can calculate means and standard deviations for volume and for 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-minute time period. Relative speed only assumes similar effective average vehicle lengths all intervals in the 20-minute period. Means and standard deviations were computed for volume and relative speed for each of the three lanes, leading to six parameters.

The total number of traffic flow parameters listed in Table 2 thus comes to thirty-six. All of these parameters are measured without recourse to highly inaccurate computations of the derived variables: density and speed.

Table 2 The Thirty-six 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
	Realtive 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

4 DATA REDUCTION

There will be a great deal of redundancy in these thirty-six traffic flow parameters. In statistical terms, this redundancy is expressed in terms of correlations with high absolute values. Principal components analysis (a form of what is called factor analysis) is designed to eliminate such redundancy by finding a smaller number of linear combinations (weighted averages) of the original variables such that the least amount of information is lost. This is based on a singular decomposition (eigenvalue and eigenvector solution) of the correlation matrix of the original variables. The linear combinations, called principal components, or Factors, are statistically independent (orthogonal). 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 of describing derived parameters (speed and density).

4.1 Extraction of Traffic Flow Factors

Principal components analysis of the thirty-six traffic flow parameters (Table 2) resulted in the selection of eight Factors. These eight Factors account for approximately 79% of the variance in the original variables, as shown in Table 3. Figure 3 shows a plot of the cumulative explained percentage of variance as a function of the number of extracted Factors. The eight-factor solution is justified on the basis of interpretability and description of derived parameters.

Table 3 Variance Explained as a Function of Number of Principal Components for the Factor Analysis of the Thirty-six Traffic Flow Parameters

Component	Initial Eigenvalues			Rotated		
	Total variance	% of Variance	Cumulative %	Total variance	% of Variance	Cumulative %
1	11.55	32.08	32.08	6.67	18.53	18.53
2	6.98	19.40	51.47	5.91	16.43	34.95
3	2.98	8.27	59.74	5.43	15.09	50.04
4	1.92	5.34	65.08	2.43	6.74	56.78
5	1.51	4.20	69.28	2.28	6.34	63.12
6	1.38	3.84	73.12	2.13	5.91	69.04
7	1.28	3.55	76.67	1.86	5.18	74.21
8	**0.90**	**2.49**	**79.16**	**1.78**	**4.94**	**79.16**
9	0.72	1.99	81.15			
10	0.69	1.93	83.08			
11	0.63	1.74	84.82			
12	0.54	1.50	86.32			
13	0.48	1.34	87.66			
14	0.45	1.25	88.90			
15	0.39	1.07	89.98			

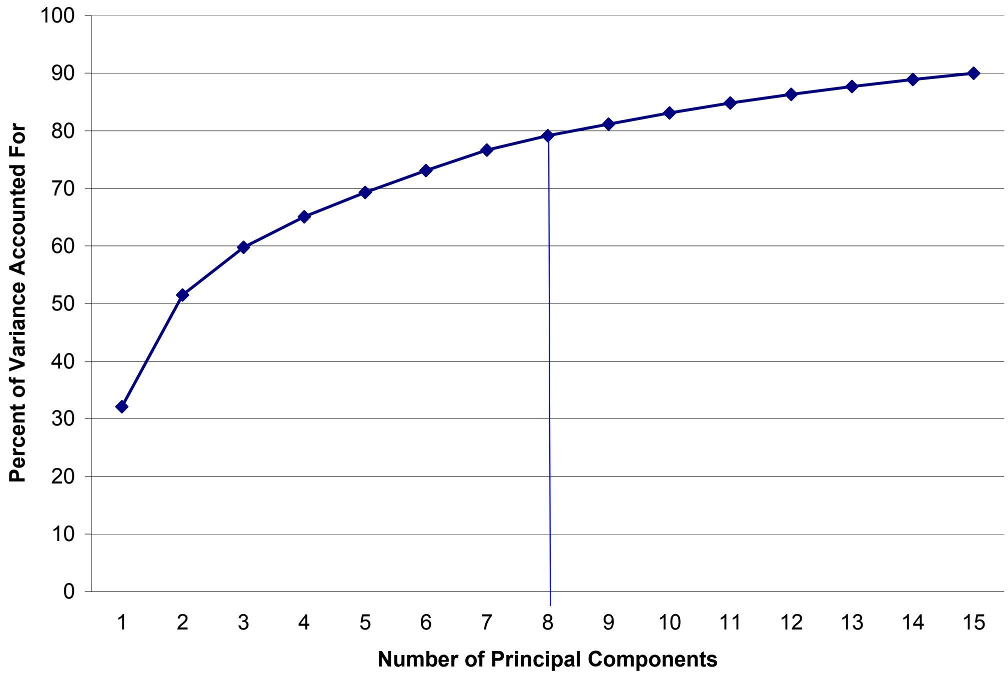

Figure 3 Variance Explained as a Function of Number of Principal Components for the Factor Analysis of the Thirty-six Traffic Flow Parameters

The percent of variance of each of the original thirty-six variables that is accounted for by the eight Factors together (the R^2 of the regression of each of the thirty-six 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. 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; these correlations are called factor loadings. Any orthogonal rotation can be applied to the Factors in the thirty-six dimensional space of the original variables, and such a rotation redistributes the percent of variance explained but does not change the overall level of explanation. The most commonly used rotation, called varimax, maximizes the sum of the variances of the factor loadings. This improves interpretability by driving ladings as far as possible towards the extreme values of unity (or minus unity) and zero. The redistributed percent of variance accounted for by the eight rotated Factors is listed in Table 3. 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 next section.

Scores on each of the eight 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 thirty-six traffic flow variables. The scores for each factor are standardized (they are centered at zero with unity standard deviation).

4.2 Interpretations of the Traffic Flow Factors

The factor loadings are listed in Table 4. 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).

Table 4 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**				

The strongest loadings for each factor are listed in Table 5. As a further aid in interpretation, the means for each factor for all accidents occurring within five mutually exclusive time periods are graphed in Figure 3. Interpretations of each factor follow.

Table 5 Highest factor Loadings for the Eight Traffic Flow Factors

Highly Loaded Variables	Factor Interpreation
coefficiant of variation: volume/occupancy lane 1 coefficiant of variation: volume/occupancy lane M mean relative speed lane 1 (negative) mean relative speed lane M (negative) std. dev. realtive speed lane 1 std. dev. realtive speed lane M	1. Outer lanes congestion
coefficiant of variation: volume lane 1 (negative) coefficiant of variation: volume lane M (negative) coefficiant of variation: volume lane R (negative) coefficiant of variation: occupancy lane 1 (negative) coefficiant of variation: occupancy lane M (negative) coefficiant 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
coeficiant. of variation: volume/occupancy lane R mean: relative speed lane R (negative) standard deviation: realtive 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

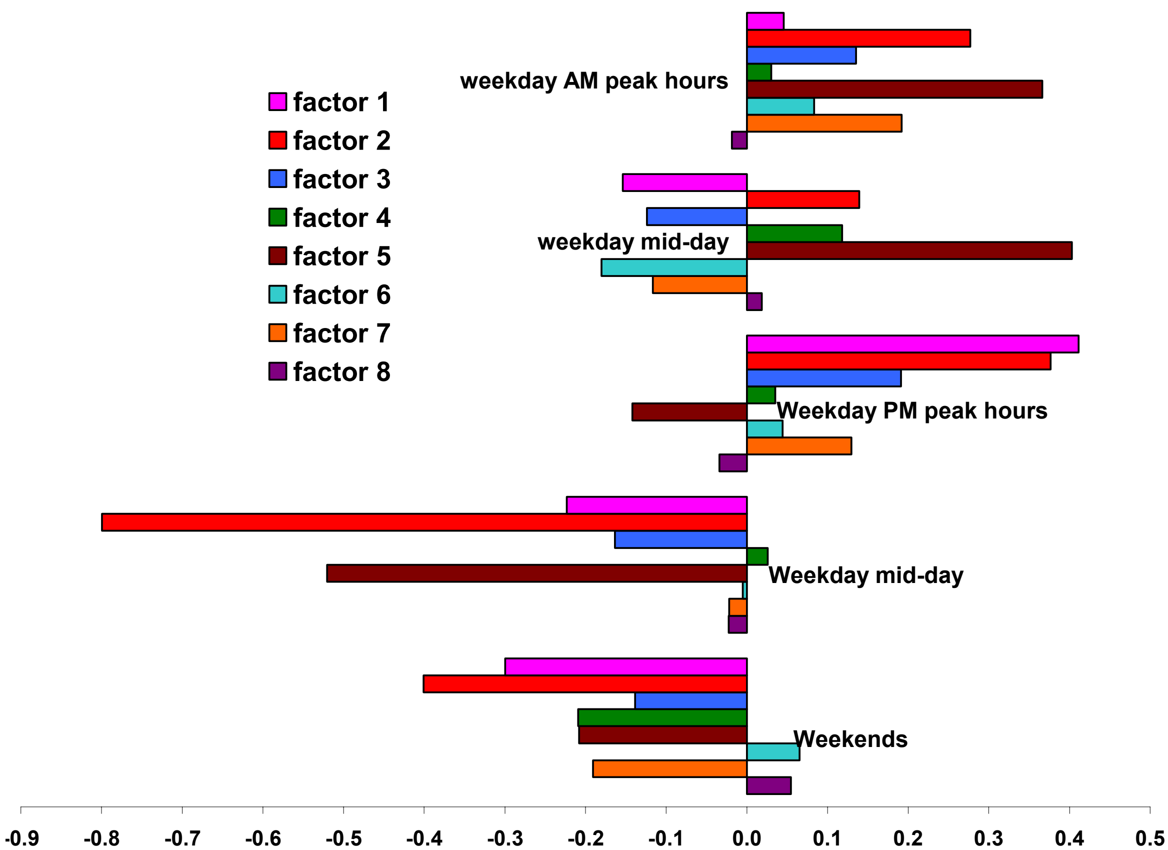

Figure 4 Mean Values of the Eight Traffic Flow Factors for All Accidents Occurring Within Five Time Periods

4.2.1 Factor 1: Outer Lanes Congestion

High scores on Factor 1 indicate that traffic conditions are varying widely within the region of loop detector data generally denoting congested operation (depicted in Figure 1). An example of a twenty-minute period of 20-s loop detector observations for with a high value of Factor 1 is plotted in Figure 5. Low scores on Factor 1 indicate that conditions are exclusively free flow, as in the example of Figure 6. High scores on Factor 1 are most likely to occur during the evening peak period, as shown in Figure 4. Low scores are most likely to occur on weekends and at night.

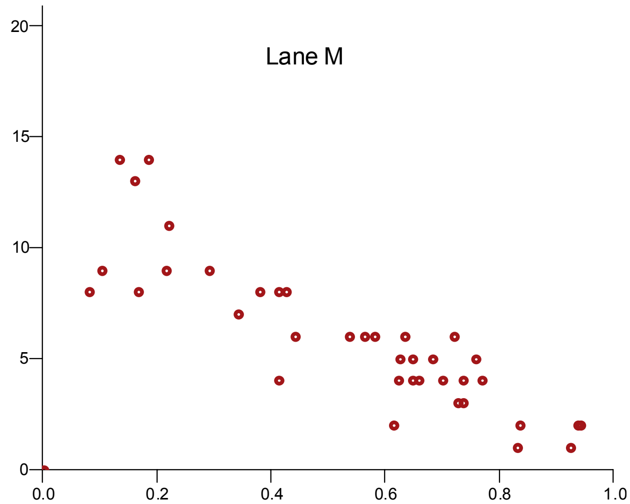

Figure 5 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 15.26 on WB SR-91 on 04/09/01 prior to 15:10) with a **High** Score on Factor 1: Outer Lanes Congestion.

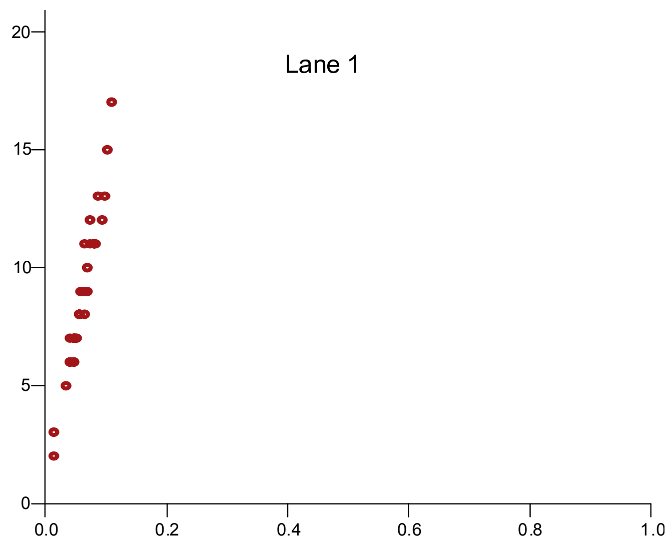

Figure 6 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 27.68 on SB SR-55 on 07/09/01 prior to 17:45) with a **Low** Score on Factor 1: Outer Lanes Congestion.

4.2.2 Factor 2: Volume Level

High scores on Factor 2 indicate high-volume conditions. Such conditions are more likely at night and on weekends (Figure 4). The period plotted in Figure 7 is a good example. Low scores on Factor 2 indicate high-volume free-flow conditions (Figure 8). Such conditions are more likely during the afternoon weekday peak period.

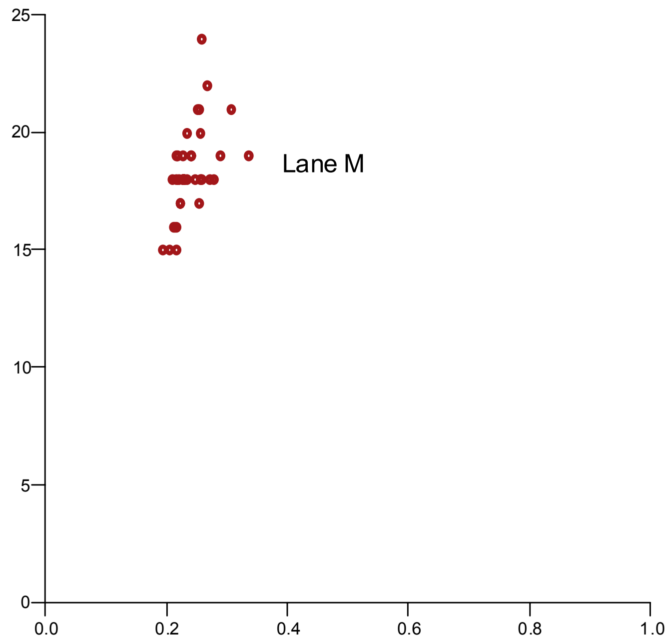

Figure 7 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 30.06 on NB I-5 on 03/14/01 prior to 07:25) with a **High** Score on Factor 2: Volume Level.

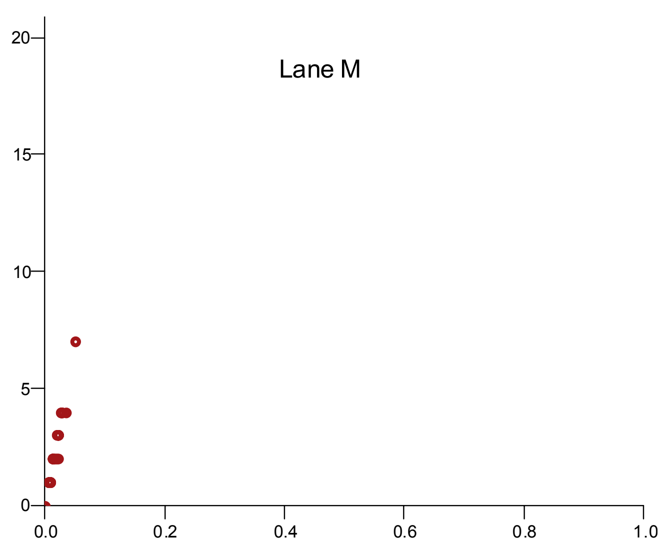

Figure 8 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 0.90 on NB I-5 on 06/17/01 prior to 02:40) with a **Low** Score on Factor 2: Volume Level.

4.2.3 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 Figure 4, this can happen during any period, but is more likely during the evening and morning peak periods. Plots of loop detector data for at least two lanes is needed to demonstrate a high score on this Factor. Detector data is plotted in Figure 9 for both the left and middle lanes for a situation with a high score on this Factor. For most of this period (the first fifteen minutes), conditions were free-flow. However, during the last five minutes, congestion set in, and speeds diminished in all lanes (including the right lane, not shown here). To visualize the breakdown period for this same observation, the traces of detector data for the crucial three minutes of transition are graphed in Figure 10 for all three lanes. The traces for all three lanes follow a similar pattern of declining speeds.

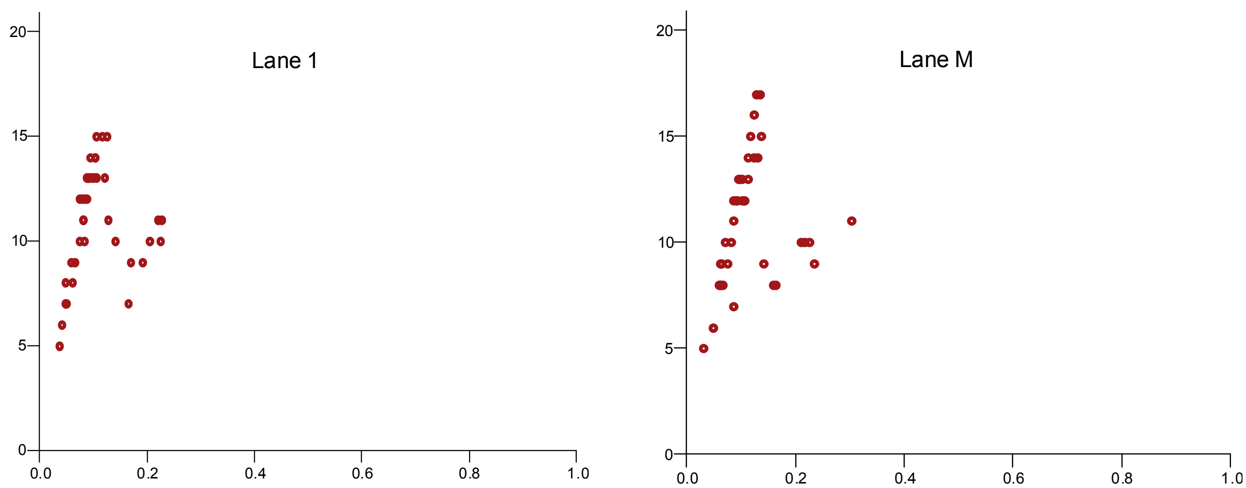

Figure 9 Twenty Minutes of Loop Detector Data for Two lanes for an Observation (PM 4.91 on WB SR-22 on 07/07/01 prior to 19:50) with a High Value on Factor 3: Synchronized Lane Conditions

Conversely, low scores on Factor 3 indicate traffic conditions that are changing differently across the lanes. Two lanes of detector data is plotted in Figure 11 for an observation with a low score on Factor 3. Traces of detector data for three minutes during this heavily congested period are graphed in Figure 12 for each of the three lanes. 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.

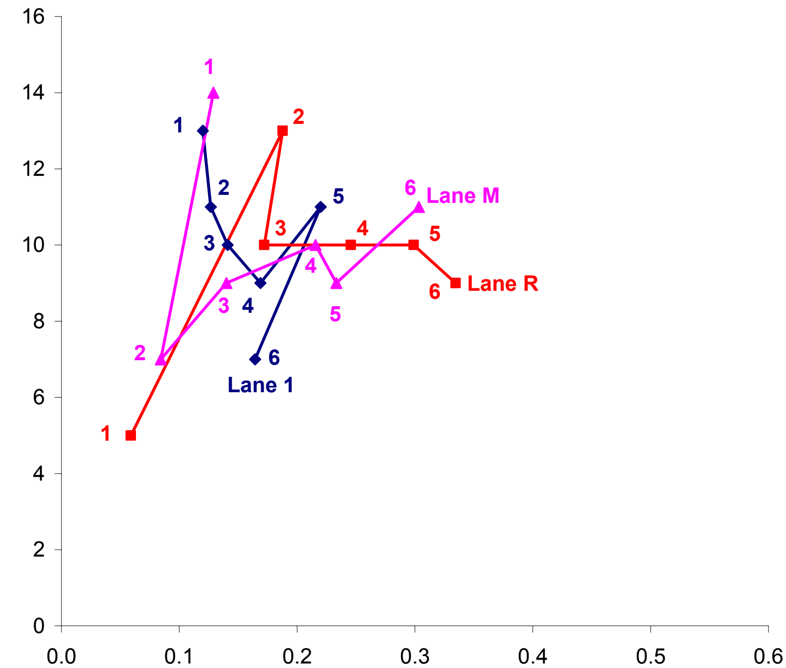

Figure 10 Three Minutes of Loop Detector Data for All Three Lanes for an Observation (PM 4.91 on WB SR-22 on 07/07/01 prior to 19:50) with a **High** Score on Factor 3: Synchronized Lane Conditions

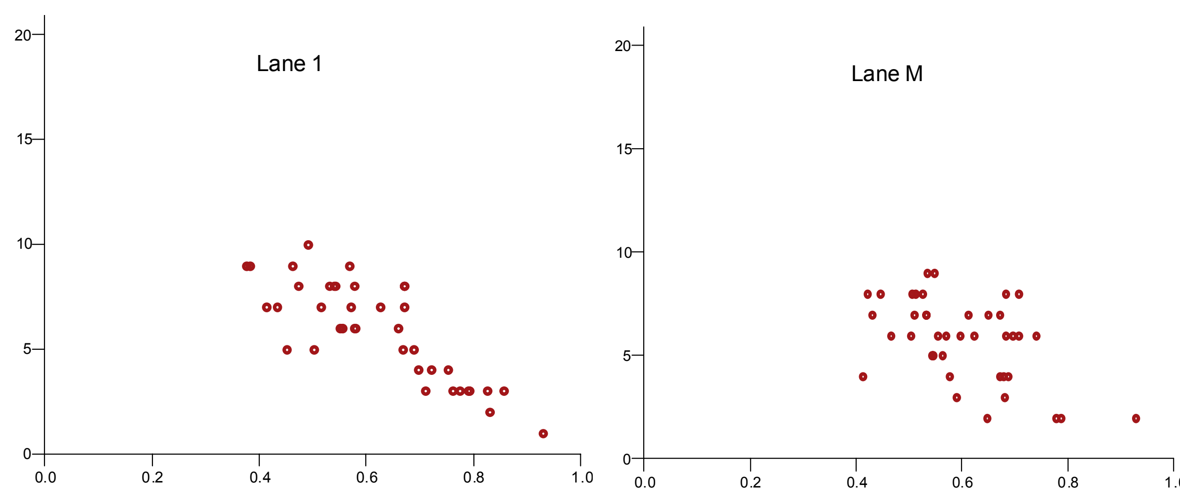

Figure 11 Twenty Minutes of Loop Detector Data for Two lanes for an Observation (PM 9.99 on EB SR-91 on 04/24/01 prior to 16:15) with a **Low** Score on Factor 3: Synchronized Lane Conditions

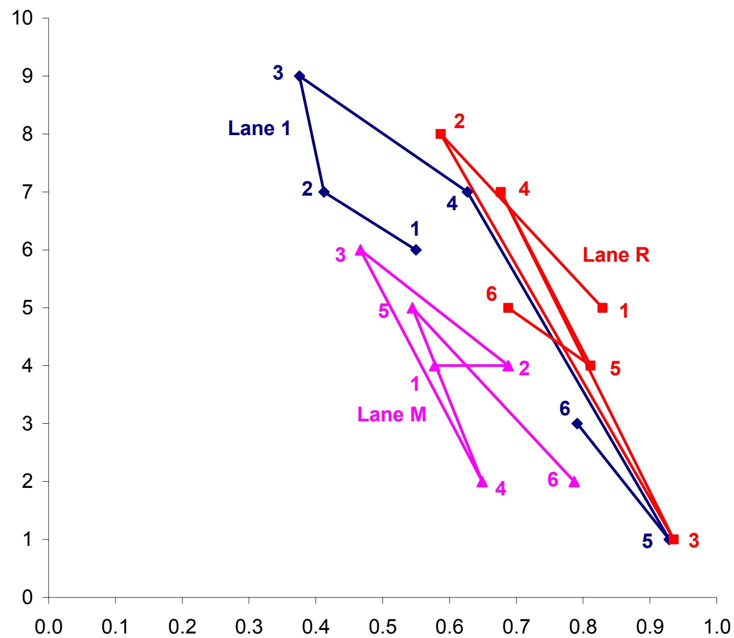

Figure 12 Three Minutes of Loop Detector Data for All Three Lanes for an Observation (PM 4.91 on WB SR-22 on 07/07/01 prior to 19:50) with a **High** Score on Factor 3: Synchronization of All Lanes

4.2.4 Factor 4: Curb Lane Perturbation

High scores on Factor 4 – indicating curb (right) lane loop detector data in the congested region (Figure 1) – can occur during any time period, but are most likely during the mid-day period (Figure 4). An extreme example is that of a right lane that exhibits the entire range of speed from free flow to virtually stopped traffic (Figure 13). Low values of Factor 4 will be found for free flowing traffic

An example of a twenty-minute observation with a low score on Factor 4 is shown in Figure 14. Here, at a time late in the morning peak period, 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.

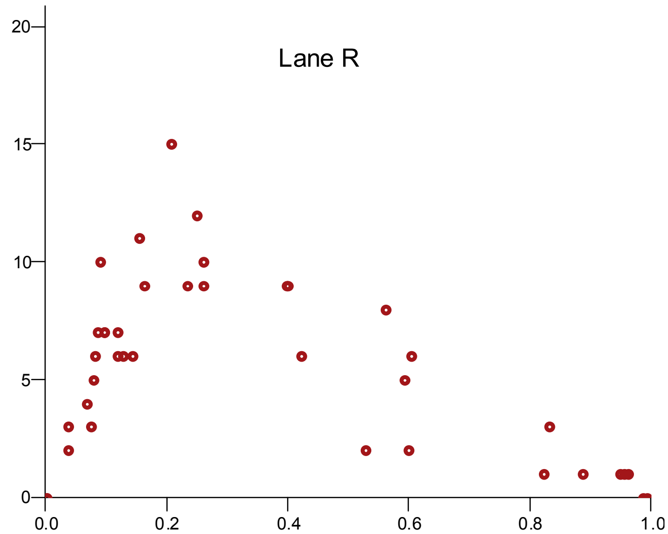

Figure 13 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 6.63 on SB I-5 on 05/07/01 prior to 09:35) with a **High** Score on Factor 4: Curb Lane Perturbation.

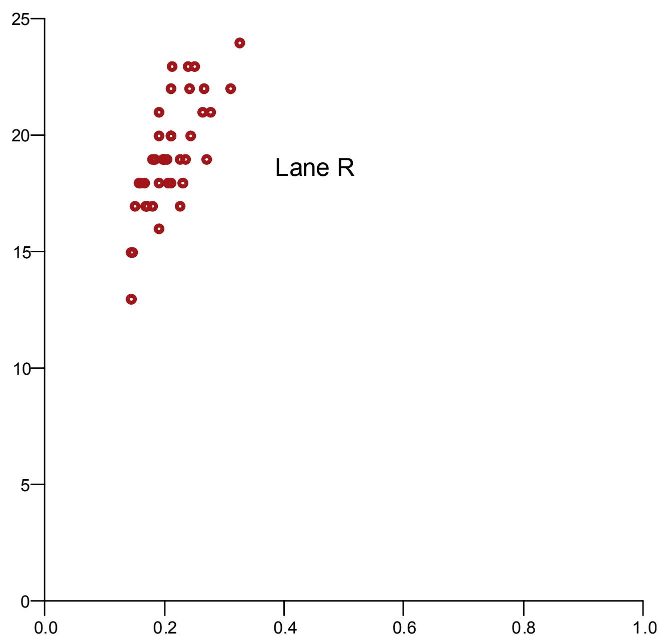

Figure 14 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 30.82 on SB I-5 on 08/06/01 prior to 08:30) with a Low Score on Factor 4: Curb Lane Perturbation.

4.2.5 *Factor 5: Volume Variation*

Factor 5 measures the extend to which volume is varying across the entire road, particularly in the non-curb lanes. Shown in Figure 15 is a situation with a high score on Factor 5. This interior lane exhibits 30-s volumes ranging from one to twenty (120 to 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, Figure 16 shows a situation with a high score on Factor 1, but a relatively low score on Factor 5. Figure 17 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.

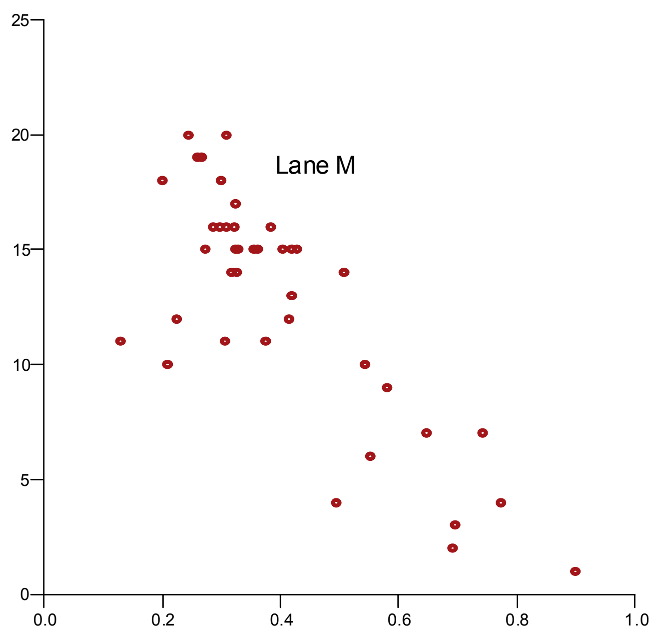

Figure 15 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 13.33 on SB SR-57 on 05/13/01 prior to 16:00) with a **High** Score on Factor 5: Volume Variation.

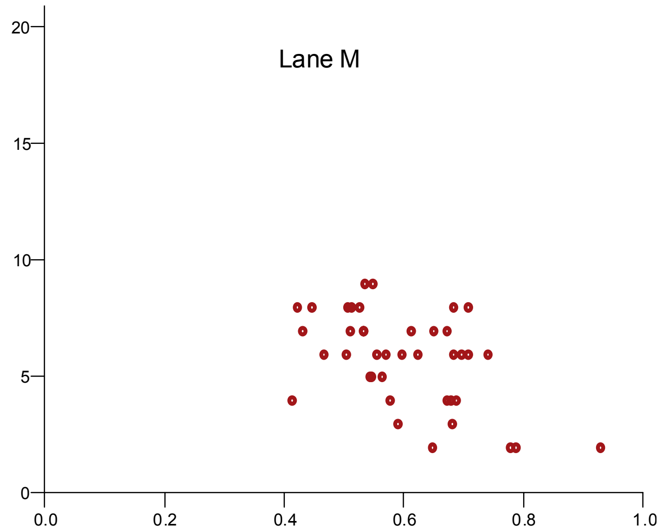

Figure 16 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 9.99 on EB SR-91 on 04/24/01 at 16:15) with a **High** Score on Factor 1: Variation in Non-curb Conditions, but a Relatively Low Score on Factor 5: Volume Variation.

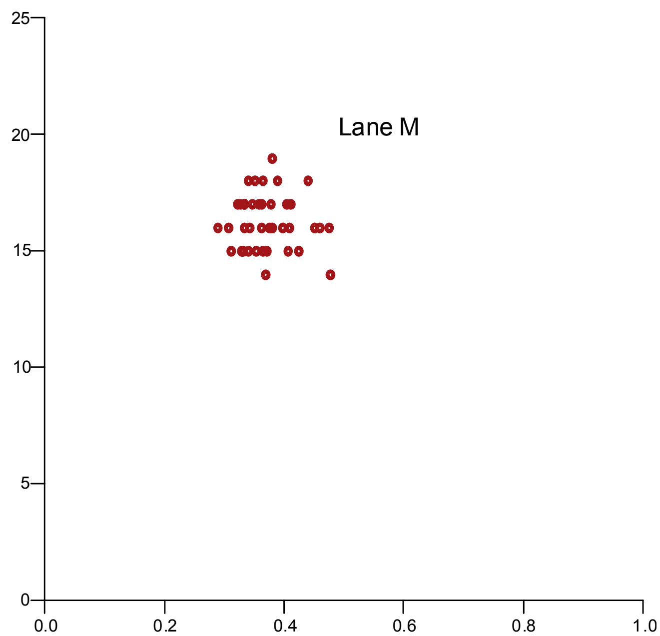

Figure 17 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 33.31 on SB I-5 on 03/22/01 prior to 07:25) with a **Low** Score on Factor 5: Volume Variation.

4.2.6 *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. Figure 18 is an example of a situation with a high score on Factor 6, and Figure 19 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, is more likely to occur during the mid-day period.

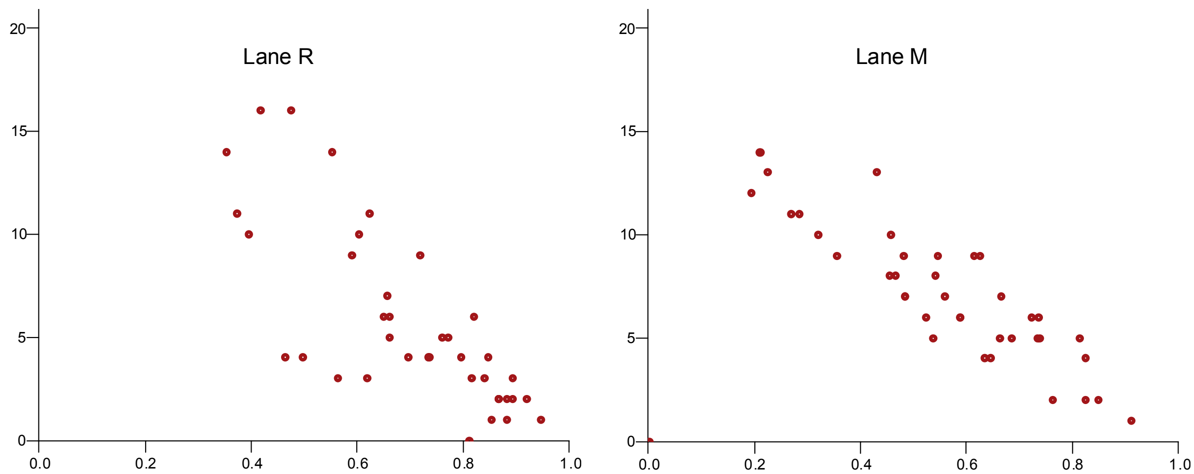

Figure 18 Twenty Minutes of Loop Detector Data for Two lanes for an Observation (PM 15.96 on SB I-405 on 04/23/01 prior to 07:55) with a High Score on Factor 6: Conforming Curb Volumes.

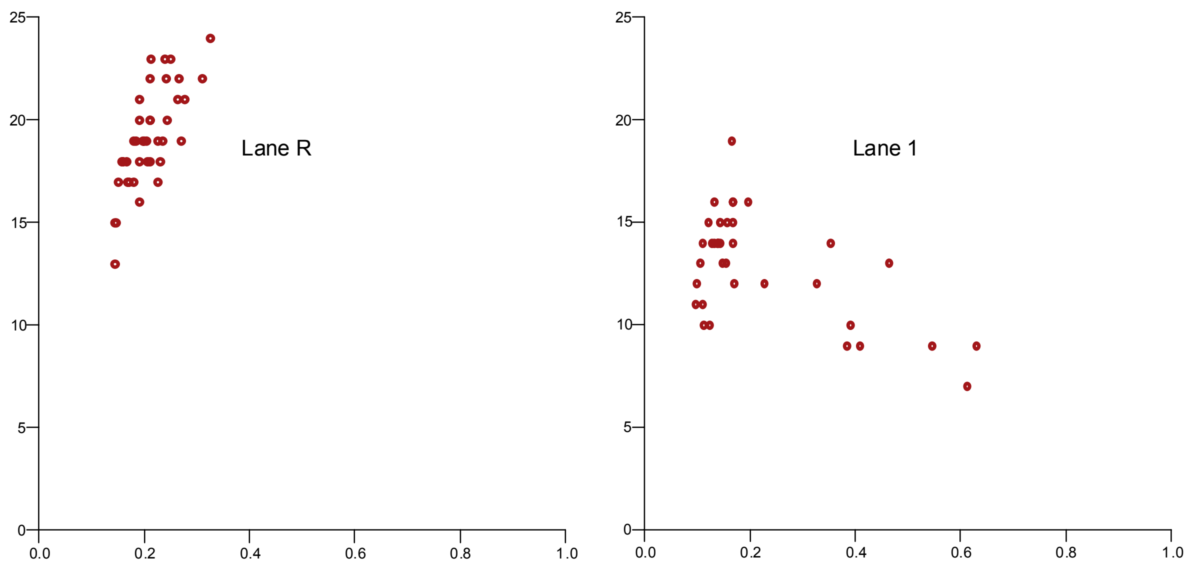

Figure 19 Twenty Minutes of Loop Detector Data for Two lanes for an Observation (PM 30.82 on SB I-5 on 08/06/01 prior to 08:30) with a Low Value on Factor 6: Conforming Curb Volumes.

4.2.7 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 the a road shifts from free flow to some congestion, and such a case is shown in Figure 20. These situations tend to occur during the morning and evening peak periods, more so in the morning peak (Figure 4). Conversely, randomly fluctuating volumes, such as the situation depicted in Figure 21, tend to occur on weekends, and during the weekday mid-day period.

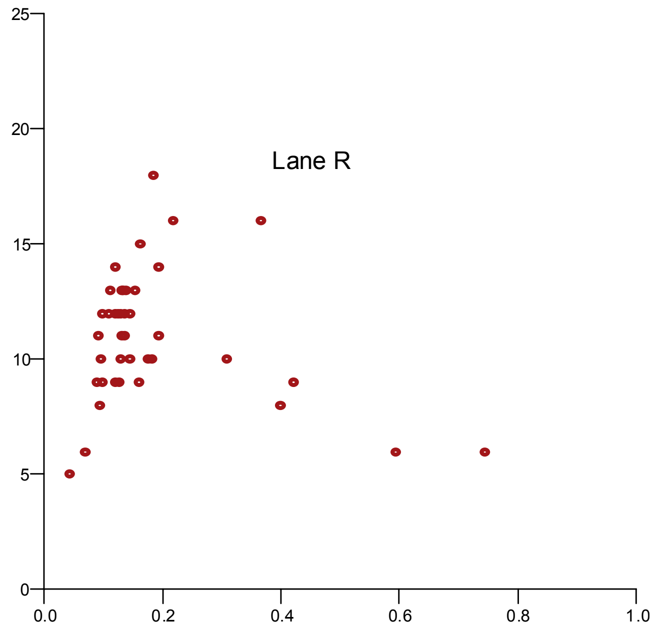

Figure 20 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 18.03 on SB I-405 on 03/08/01 prior to 06:45) with a **High** Value on Factor 7: Systematic Volume Change.

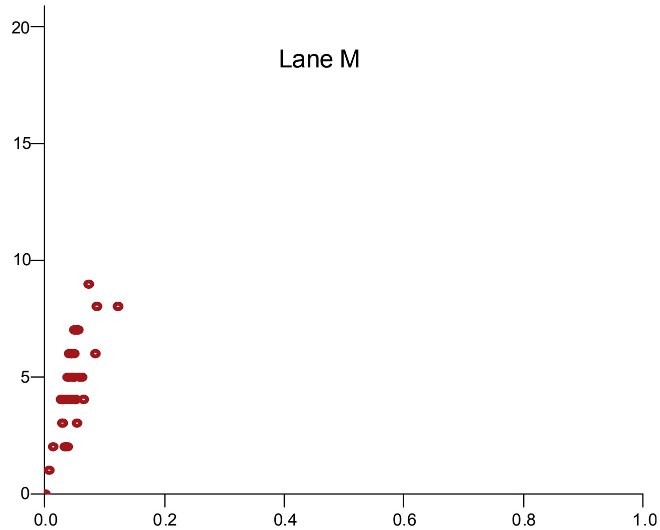

Figure 21 Twenty Minutes of Loop Detector Data for One lane for an Observation (PM 24.89 on NB I-5 on 05/05/01 prior to 09:10) with a **Low** Value on Factor 6: Systematic Volume Change.

4.2.8 Factor 8: Synchronized Outer Flow

Factor 8 measures the degree to which volumes and densities in the interior and left lanes are synchronized. There are very little differences among the mean values of this factor for accidents occurring during various time periods (Figure 4), which means high or low levels of synchronization are likely to occur anytime. A high level of non-curb synchronization is shown in Figure 22. 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 Figure 23. 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 twenty-minute period.

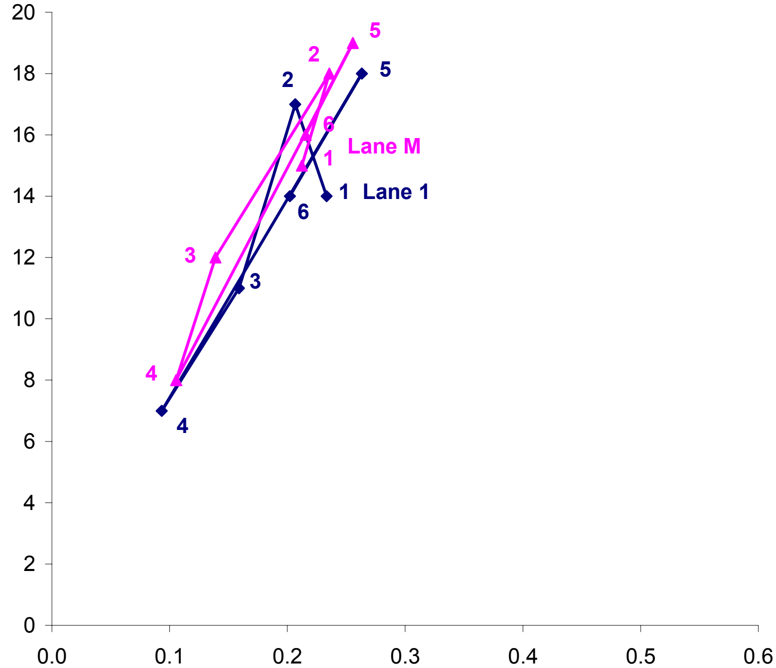

Figure 22 Three Minutes of Loop Detector Data for Two Lanes for an Observation (PM 27.45 on NB I-5 on 03/20/01 prior to 10:05) with a **High** Value on Factor 8: Synchronized Outer Flow.

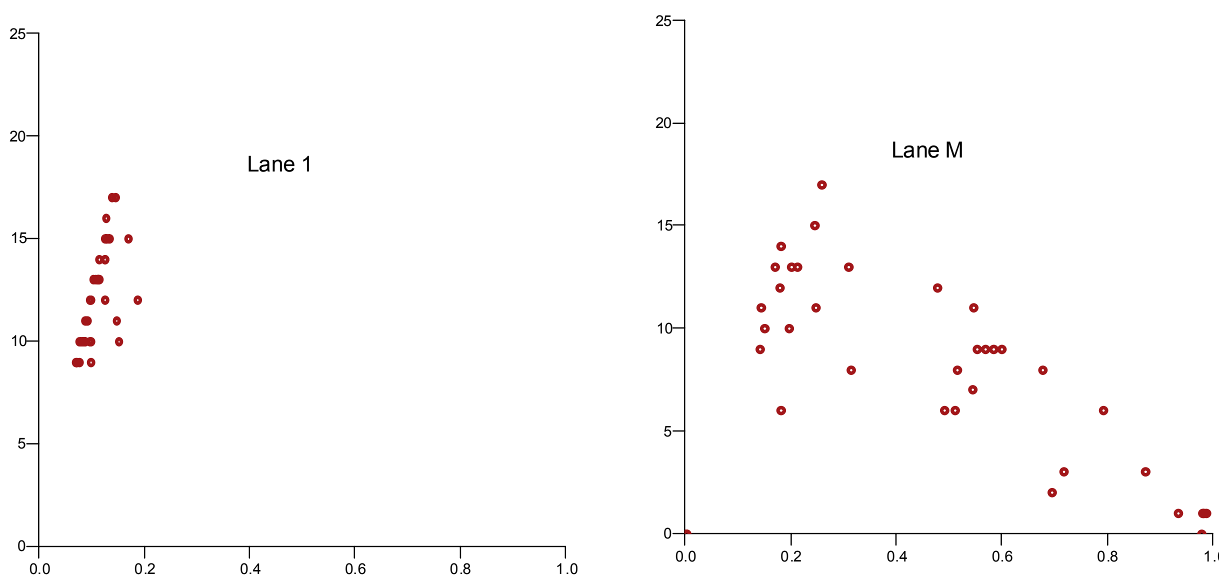

Figure 23 Twenty Minutes of Loop Detector Data for Two lanes for an Observation (PM 11.70 on SB SR-55 on 07/09/01 prior to 17:45) with a **Low** Value on Factor 8: Synchronized Outer Flow.

4.3 Traffic Flow Factors as Descriptors of Variables Proportional to Speed and Density

A criterion in the determination of the necessary number of Factors was the ability of these Factors to account for the variables measured directly in terms of vehicular density and speed. We are prohibited from using these variables, because effective vehicle lengths are not available for the time and place of each accident. Without effective vehicle length for each observation, we are unable to convert occupancy to density, and consequently we are unable to compute speed as the ratio of volume and density. The question is, how well are these prohibited variables explained by the Factors, including second degree factor interactions. Second degree interactions account for potential nonlinear effects, allowing projection of a dependent variable onto a quadratic surface, rather than simply a hyperplane, in the space of the eight Factors.

Each of prohibited scaled measures was regressed on a set of forty-four variables, made up of the eight Factors, plus eight factor quadratic terms (the products of two like Factors), plus twenty-eight factor interactions (the products of any two different Factors). The results for the means and standard deviations of occupancy for each of the three lanes are shown in Table 6. The Factors do very well explaining all six of the prohibited variables that are potentially proportional to traffic flow density.

Both the Factors and the second-degree interaction terms are important in explaining each of these potential density variables. For example, in the regression for standard deviation of occupancy in the left lane, the least well described dependent variable (82% of variance), all eight of the factor linear terms were significant at the 95% confidence level, as were five of the eight quadratic terms, and thirteen of the twenty-eight factor interactions. The most important variables were: (1) Factor 1: Outer Lanes Congestion (positive), (2) Factor 3: Synchronized Lane Conditions (positive), (3) Factor 7: Systematic Volume Change (positive), (4) The square of Factor 2 (Volume Level) (positive), and (5) the interaction of Factor 2 (Volume Level) and Factor 3 (Synchronization of All lanes) (negative).

Table 6 Variance Explained in Regressions of Six Prohibited Density Variables on the Eight Factors

Prohibited variable	Adjusted R^2
Mean occupancy left lane	0.846
Mean occupancy interior lane	0.863
Mean occupancy right lane	0.827
Standard deviation occupancy left lane	0.820
Standard deviation occupancy interior lane	0.881
Standard deviation occupancy right lane	0.870

Similar regression results are displayed for the ratio of volume to occupancy in Table 7. All ratios are well explained, with the exception of left-lane standard deviation of the ratio volume to occupancy. As effective vehicle lengths are unlikely to change across the 30-s observations for the left lane, this variable is probably proportional to the standard deviation of left-lane speed. While 58% is a decent percent explained variance, this reveals that variation in left-lane speed is the most difficult of traffic flow parameters to capture with parameters that avoid untenable assumptions regarding effective vehicle length. The most important variable in explaining left-lane standard deviation of volume/occupancy is the interaction of Factor 1: Outer Lanes Congestion and Factor 8: Synchronized Outer Flow (with a positive coefficient).

Table 7 Regressions of Six Prohibited Speed Variables on the Eight Factors

Prohibited Variable	Adjusted R^2
Mean volume/occupancy left lane	0.778
Mean volume/occupancy interior lane	0.797
Mean volume/occupancy right lane	0.714
Standard deviation volume/occupancy left lane	0.581
Standard deviation volume/occupancy interior lane	0.668
Standard deviation volume/occupancy right lane	0.699

5 TRAFFIC FLOW FACTORS RELATED TO ACCIDENT PROPENSITY

A major question is: how well do the Traffic Flow Factors describe accident potential? As an initial step in answering this question, in the first year of the project we investigated the extent to which the Factors describe differences among the types of accidents that occur under different types of traffic flow conditions. Four accident variables were analyzed: (1) accident severity, (2) collision type, (3) collision location, and (4) number of involved vehicles. Each of these is the subject of a subsection that follows.

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, in this case 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 lead to an injury, the rest are property damage only (PDO). Results of a binomial logit regression model for severity are listed in Table 8. The dependent 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. Eight independent variables were significant at the 95% confidence level, being 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 modestly associated with traffic conditions. These results are interpreted as follows.

5.1.1 Accident Severity and Congestion

Congestion in the outer lanes, on its own, leads to a lower likelihood of injury accidents, since speeds will be lower during congested conditions. However, there are also two interaction terms involving this traffic flow Factor. If right lane volumes are track those of the outer lanes, the effect of congestion on reducing severity is more than doubled. On the other hand, 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 outer lanes congestion on severity, reducing the effect essentially to zero.

Congestion in the curb lane leads to a lower level of injury accidents if curb lane volume conforms to outer lane volumes. That is, if the entire road is congested, accidents are more likely to be PDO, rather than injury. Congestion in the curb lane only has little effect on accident severity.

Table 8 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 perturbation x 6. Conforming curb volumes	-0.181	3.337	0.001
1. Outer lanes perturbation x 7. Systematic volume changes	0.137	-2.262	0.024
2. Volume level x 4. Curb lane perturbation	-0.124	2.624	0.009
4. Curb lane perturbation x 6. Conforming curb volumes	0.130	2.370	0.018
5. Volume variation x 6. Conforming curb volumes	0.113	2.063	0.039
Constant	-1.124	-19.815	0.000
Dependent variable: 1 = injury of fatality, 0 = property damage only			

5.1.2 Accident Severity and Aspects of Traffic Volume

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. 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.

5.1.3 Accident Severity and Lane Synchronization

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. Loss of synchronization leads to a higher likelihood of injury accidents.

5.2 Type of Collision

There are four major types of primary collision, as listed in Table 9. Rear end collision are most common, followed by sideswipes and hit-object collisions. Results of a multinomial logit model for collision type are listed in Table 10. The base category is "other." Variables are included in this model if their inclusion leads to a significant overall improvement in the explanatory 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 10 and subsequent tables.

Table 9 Breakdown of Collision Type

Collision type	Frequency	Percent
Sideswipe	369	20.9%
Rear end	1054	59.6%
Hit object	268	15.1%
Other	78	4.4%

5.2.1 Collision Type and Congestion

Congestion in the outer 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 outer lane congestion. Right lane perturbation is associated with both rear-end and sideswipe accidents, especially if curb lane volume conforms to outer lane volumes. Conforming curb lane volume combined with curb lane congestion also leads to a higher likelihood of hit-object collisions.

5.2.2 Collision Type and Aspects of Traffic Volume

Controlling for whether or not the road is operating under free flow or congested, the likelihood that a accident is a rear-end collision increases with increasing levels of volume. Variation in volume leads marginally to a 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 Figure 24. 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.

Table 10 Logit Model of Collision Type as a Function of Traffic Flow Factors

Explanatory variable	Coefficient	t-statistic	Probability
Sideswipe collision			
1. Outer lanes congestion	**0.601**	**2.75**	**0.006**
2. Volume level	0.326	1.46	0.144
3. Synchronized lane conditions	0.141	0.86	0.391
4. Curb lane perturbation	**0.272**	**2.01**	**0.044**
7. Systematic volume changes	0.169	1.22	0.221
2^2. Volume level (squared)	-0.026	-0.28	0.776
3^2. Synchronized lane conditions (squared)	**-0.281**	**-2.73**	**0.006**
5^2. Volume variation (squared)	0.249	1.87	0.061
1. Outer lane congestion x 2. Volume level	0.053	0.32	0.746
4. Curb lane perturbation x 6. Conforming curb volumes	**0.256**	**2.27**	**0.023**
Constant	2.098	7.54	0.000
Rear end collision			
1. Outer lanes congestion	**1.067**	**5.00**	**0.000**
2. Volume level	**0.942**	**4.36**	**0.000**
3. Synchronized lane conditions	**0.782**	**4.93**	**0.000**
4. Curb lane perturbation	**0.314**	**2.37**	**0.018**
7. Systematic volume changes	**0.385**	**2.87**	**0.004**
2^2. Volume level (squared)	**0.169**	**2.00**	**0.045**
3^2. Synchronized lane conditions (squared)	**-0.342**	**-3.59**	**0.000**
5^2. Volume variation (squared)	0.251	1.91	0.056
1. Outer lane congestion x 2. Volume level	0.243	1.55	0.120
4. Curb lane perturbation x 6. Conforming curb volumes	**0.332**	**3.00**	**0.003**
Constant	1.067	5.00	0.000
Hit object			
1. Outer lanes congestion	0.011	0.05	0.962
2. Volume level	0.026	0.11	0.910
3. Synchronized lane conditions	0.102	0.59	0.555
4. Curb lane perturbation	0.063	0.45	0.654
7. Systematic volume changes	-0.044	-0.31	0.757
2^2. Volume level (squared)	0.073	0.84	0.399
3^2. Synchronized lane conditions (squared)	-0.136	-1.32	0.188
5^2. Volume variation (squared)	0.132	0.95	0.340
1. Outer lane congestion x 2. Volume level	0.175	1.08	0.282
4. Curb lane perturbation x 6. Conforming curb volumes	**0.369**	**3.06**	**0.002**
Constant	1.327	4.54	0.000
Reference category: other type of collision (e.g., overturn, broadside)			

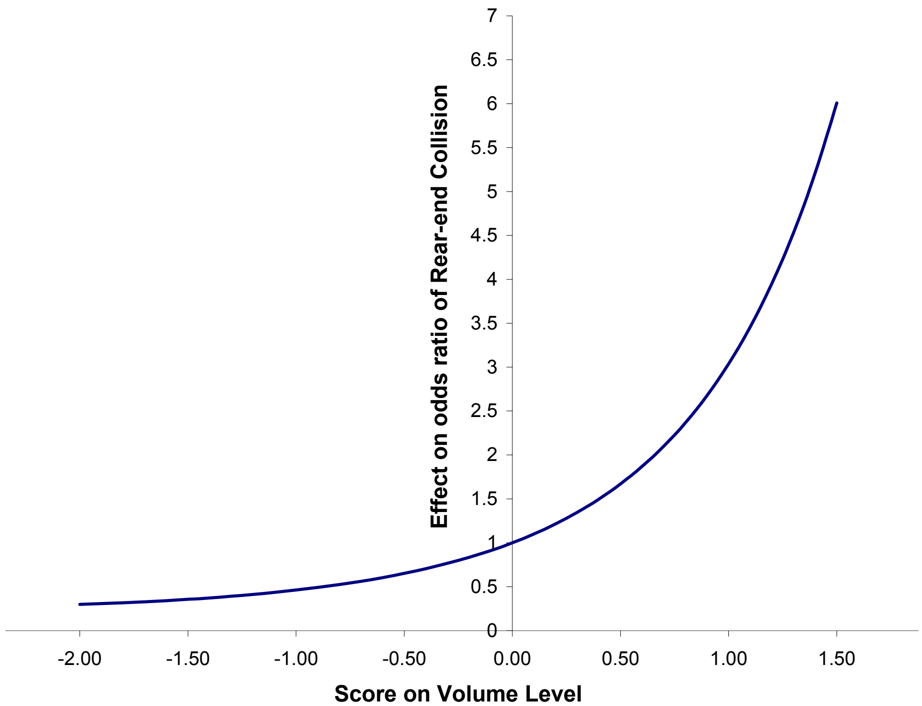

Figure 24 Effects on the Odds of a Rear-end Collision of Factor 2: Volume Level

5.2.3 Collision Type and Lane Synchronization

Controlling for both volume and whether or not 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 an decreasing rate for a little more than half the range of Factor 3, reach a maximum, then decrease, as graphed in Figure 25. Sideswipe collisions are most likely at about an average level of synchronized conditions, all else held constant. Such collisions are less likely if traffic conditions are either highly synchronized across lanes, or if conditions are highly chaotic across lanes.

The nonlinear effects of synchronized lane conditions on the odds of a rear-end collision are graphed in Figure 26. The odds increase with increasing score on Factor 3 for 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.

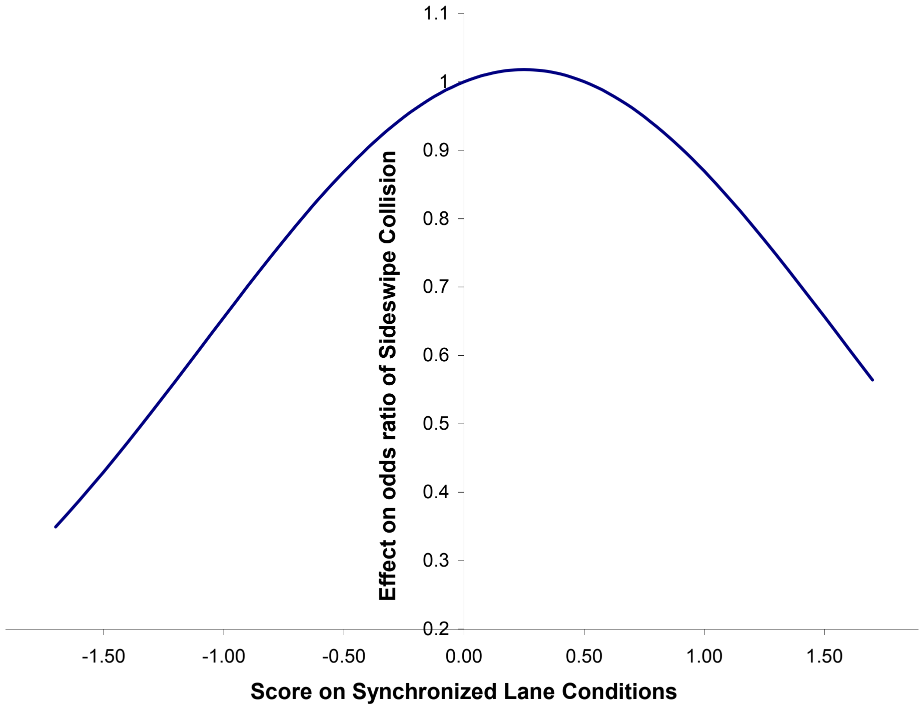

Figure 25 Effects on the Odds of a Sideswipe Collision of Factor 3: Synchronized lane Conditions

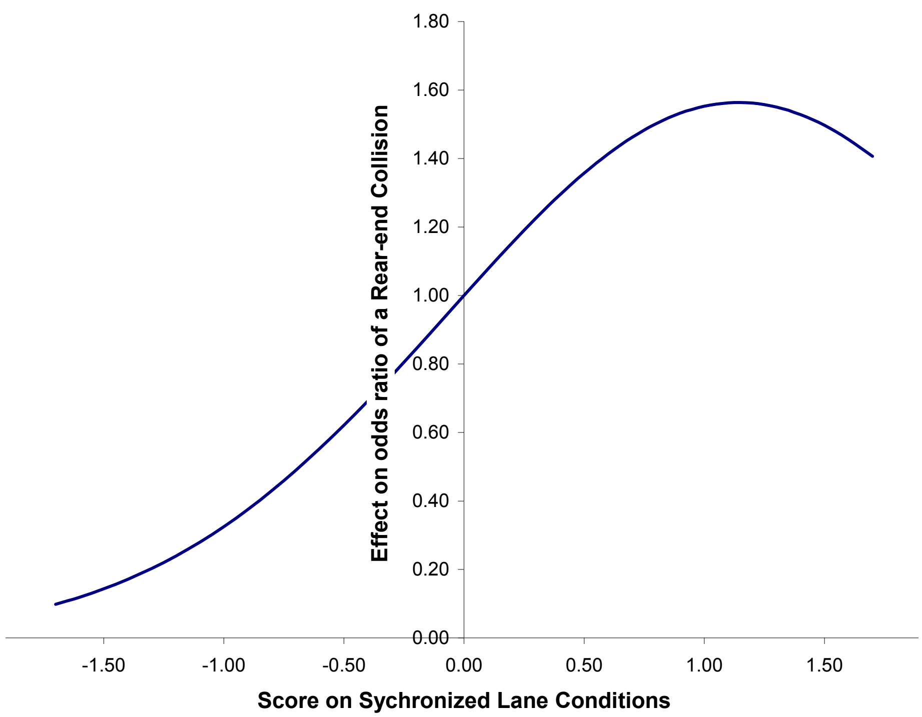

Figure 26 Effects on the Odds of a Rear-end Collision of Factor 3: Synchronized lane Conditions

5.3 Collision Location

Location of the primary collision of an accident is broken down into the five categories listed in Table 11. Collisions in the Interior lane(s) are most common, followed by left-lane collisions. 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 12. The base category is “Interior lane(s)”. The overall fit of the model indicated by the Nagelkerke Pseudo-R^2 of 0.140, is good, but not as good as that obtained in the previous collision type model.

Table 11 Breakdown of Collision Location

Collision type	Frequency	Percent
Off road to drivers’ left	205	11.6%
Left lane	500	28.3%
Interior lane(s)	643	36.3%
Right lane	303	17.1%
Off road to drivers’ right	118	6.7%

5.3.1 Collision Location and Congestion

Congestion has less effect on collision location than it does on accident severity and collision type. Only congestion in the outer lanes is a statistically significant predictor of collision location. A higher degree of congestion on the outer lanes is related to increased odds of an accident being in one of the lanes, as opposed to off-road. Outer lanes congestion has no significant effect on whether or not an accident is located in the right lane.

5.3.2 Collision Location and Aspects of Traffic Volume

Controlling for whether or not 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. 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. This effect is enhanced if right lane volumes are nonconforming, but diminished if right lane volume are conforming.

Table 12 Logit Model of Collision Location as a Function of Traffic Flow Factors

Explanatory variable	Coefficient	t-statistic	Probability
Off road to drivers' left			
1. Outer lanes congestion	**-0.349**	**-3.85**	**0.000**
2. Volume level	**-0.188**	**-2.51**	**0.012**
3. Synchronized lane conditions	-0.027	-0.30	0.767
7. Systematic volume changes	0.027	0.34	0.731
8. Synchronized outer flow	**0.236**	**2.74**	**0.006**
8^2. Synchronized outer flow (squared)	**0.107**	**2.36**	**0.018**
6. Conforming curb volumes x 7. Systematic volume changes	-0.066	-0.97	0.334
6. Conforming curb volumes x 8. Synchronized outer flow	0.045	0.58	0.564
Constant	-1.321	-13.34	0.000
Left lane			
1. Outer lanes congestion	**0.182**	**2.96**	**0.003**
2. Volume level	**0.226**	**3.10**	**0.002**
3. Synchronized lane conditions	**0.368**	**5.99**	**0.000**
7. Systematic volume changes	**0.249**	**3.92**	**0.000**
8. Synchronized outer flow	**0.279**	**4.07**	**0.000**
8^2. Synchronized outer flow (squared)	**0.097**	**2.41**	**0.016**
6. Conforming curb volumes x 7. Systematic volume changes	**-0.193**	**-3.34**	**0.001**
6. Conforming curb volumes x 8. Synchronized outer flow	**-0.145**	**-2.31**	**0.021**
Constant	-0.435	-5.78	0.000
Right lane			
1. Outer lanes congestion	-0.045	-0.61	0.540
2. Volume level	0.053	0.73	0.466
3. Synchronized lane conditions	0.136	1.90	0.057
7. Systematic volume changes	0.044	0.60	0.552
8. Synchronized outer flow	-0.016	-0.18	0.854
8^2. Synchronized outer flow (squared)	-0.063	-1.15	0.248
6. Conforming curb volumes x 7. Systematic volume changes	-0.045	-0.69	0.492
6. Conforming curb volumes x 8. Synchronized outer flow	0.131	1.70	0.089
Constant	-0.702	-8.46	0.000
Off road to drivers' right			
1. Outer lanes congestion	**-0.638**	**-4.70**	**0.000**
2. Volume level	**-0.347**	**-4.05**	**0.000**
3. Synchronized lane conditions	-0.164	-1.30	0.193
7. Systematic volume changes	**-0.267**	**-2.41**	**0.016**
8. Synchronized outer flow	0.223	1.92	0.055
8^2. Synchronized outer flow (squared)	-0.007	-0.09	0.928
6. Conforming curb volumes x 7. Systematic volume changes	0.045	0.47	0.636
6. Conforming curb volumes x 8. Synchronized outer flow	0.146	1.33	0.182
Constant	-2.008	-13.91	0.000
Reference category: Interior lane(s)			

5.3.3 Collision Location and Lane Synchronization

Controlling for both volume and whether or not 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. In the case of 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 Figure 27. 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 level of outer lanes synchronization at an increasing rate.

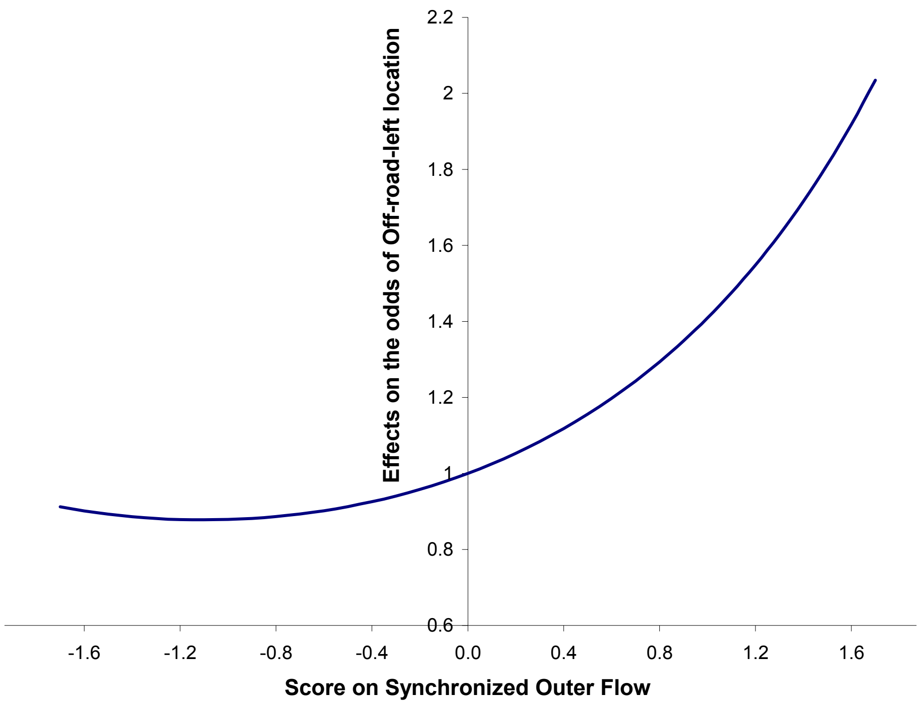

Figure 27 Effects on the Odds of a Off-road-left Location of Factor 3: Synchronized Outer Flow

The effects of synchronized outer flow on the likelihood of a left-lane location are shown in Figure 27. These effects are parameterized by the level of conforming curb volume because there is a significant interaction term involving Factor 8 and Factor 6. 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.

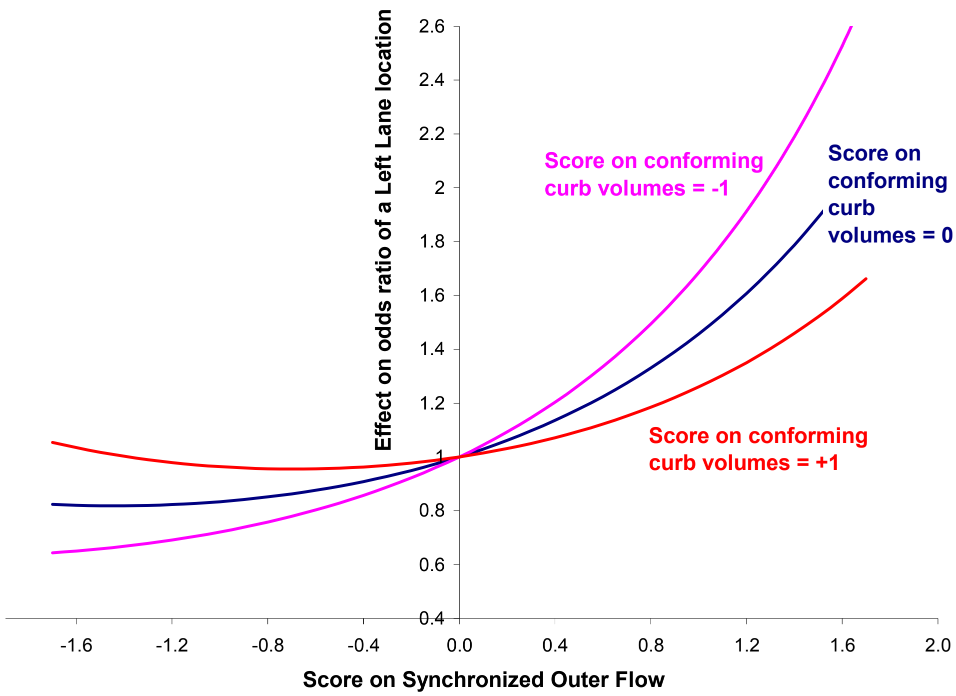

Figure 28 Effects on the Odds of a Left-lane Location of Factor 8: Synchronized Outer Flow, and Factor 6: Conforming Curb Volumes

5.4 Number of Involved Vehicles

Most case study accidents (59%) involved two vehicles, as shown in Table 13. However, there were sufficient number of accidents involved four or more vehicles 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 14 is “three vehicles”. The overall fit of the model (Pseudo-R^2 of 0.159) is good.

Table 13 Breakdown of Number of Involved Vehicles

Collision type	Frequency	Percent
Single Vehicle	233	13.2%
Two vehicles	1043	59.0%
Three vehicles	355	20.1%
Four or more vehicles	138	7.8%

5.4.1 Involved Vehicles and Congestion

As expected, congestion has a considerable influence on vehicle involvement. Congestion on the through lanes distinguishes single vehicle crashes from multi-vehicle crashes. The logarithm of the odds of a single-vehicle accident is a simple linear function of Factor 1: Outer Lanes Congestion. However, this Factor does not significantly distinguish between the likelihood of different numbers of vehicles in multi-vehicle 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. The multiplicative effect on the odds of a single vehicle being involved in any accident is graphed in Figure 29 as a function of the score on Factor 8: Curb Lane Perturbation. For extreme low values of curb lane perturbation, scores less than -0.85 standard deviations, the odds of a single-vehicle accident are reduced in approximate proportion to the value of the difference between the Factor score and –the critical value of 0.85. For Factor scores between -0.85 and the mean of zero, the odds of a single-vehicle accident are slightly increased, peaking at the score of -0.40. Finally, for positive scores, these odds decrease in approximate proportion to the score.

The effects of curb lane perturbation on the odds of two vehicles being involved are plotted in Figure 30. 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.

Table 14 Logit Model of Number of Involved Vehicles as a Function of Traffic Flow Factors

Explanatory variable	Coefficient	t-statistic	Probability
Single vehicle			
1. Outer lanes congestion	**-0.754**	**-6.31**	**0.000**
2. Volume level	**-0.728**	**-8.12**	**0.000**
3. Synchronized lane conditions	**-0.414**	**-3.97**	**0.000**
4. Curb lane perturbation	-0.136	-1.39	0.165
5. Volume variation	0.156	1.66	0.096
6. Conforming curb volumes	0.082	0.83	0.406
7. Systematic volume changes	**-0.414**	**-4.56**	**0.000**
4^2. Curb lane perturbation (squared)	**-0.159**	**-2.88**	**0.004**
5. Volume variation x 6. Conforming curb volumes	-0.065	-0.71	0.479
Constant	-0.685	-5.53	0.000
Two vehicles			
1. Outer lanes congestion	0.062	0.97	0.333
2. Volume level	-0.131	-1.82	0.069
3. Synchronized lane conditions	-0.080	-1.32	0.188
4. Curb lane perturbation	0.033	0.52	0.605
5. Volume variation	0.007	0.11	0.912
6. Conforming curb volumes	**0.182**	**2.91**	**0.004**
7. Systematic volume changes	-0.112	-1.76	0.079
4^2. Curb lane perturbation (squared)	**-0.105**	**-2.89**	**0.004**
5. Volume variation x 6. Conforming curb volumes	0.094	1.47	0.141
Constant	0.062	0.97	0.333
Four or more vehicles			
1. Outer lanes congestion	-0.030	-0.27	0.787
2. Volume level	0.228	1.61	0.107
3. Synchronized lane conditions	0.122	1.23	0.218
4. Curb lane perturbation	0.015	0.14	0.885
5. Volume variation	**0.331**	**2.97**	**0.003**
6. Conforming curb volumes	-0.138	-1.29	0.196
7. Systematic volume changes	0.189	1.83	0.067
4^2. Curb lane perturbation (squared)	-0.031	-0.55	0.585
5. Volume variation x 6. Conforming curb volumes	**0.249**	**2.45**	**0.014**
Constant	-1.066	-8.03	0.000
Reference category: Three vehicles			

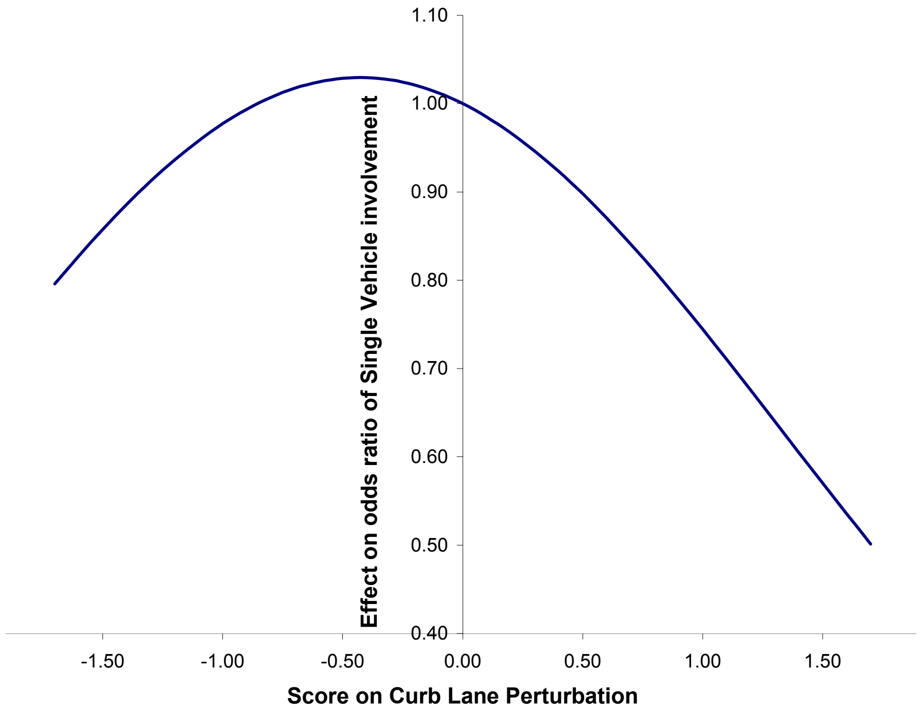

Figure 29 Effects on the Odds of Single-vehicle Involvement of Factor 4: Curb Lane Perturbation

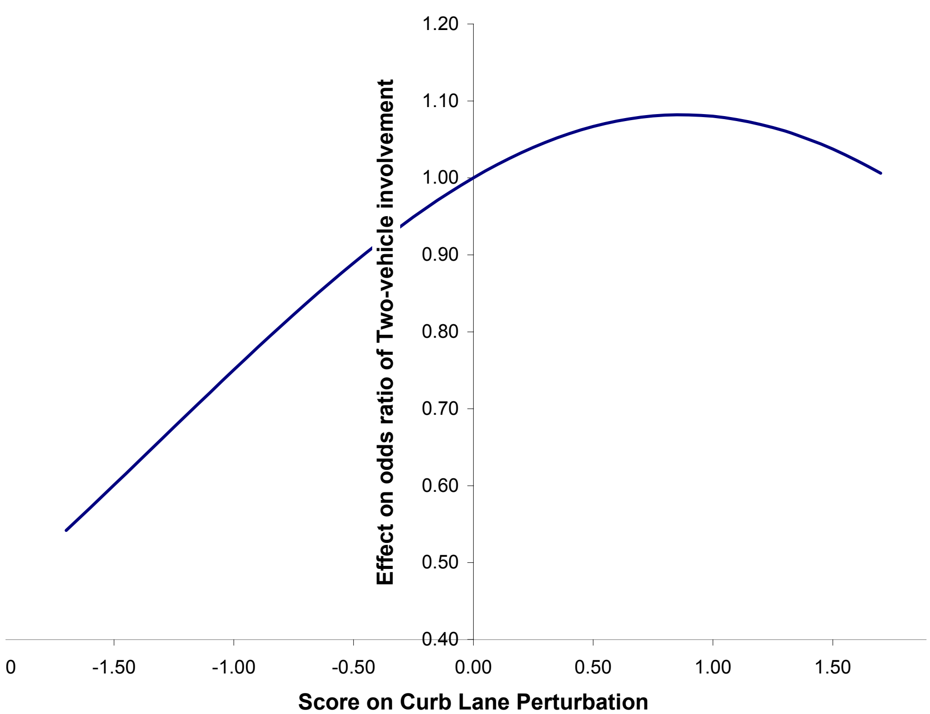

Figure 30 Effects on the Odds of Two-vehicle Involvement of Factor 4: Curb Lane Perturbation

5.4.2 Involved Vehicles and Aspects of Traffic Volume

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.

5.4.3 Involved Vehicles and Lane Synchronization

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 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 15. 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.

Table 15 Summary of Key Results from Logit Models of Accident Characteristics as a Function of Traffic Flow Factors

Factor	Severity	Collision type	Collision location	Involved vehicles
1. Outer lanes congestion	PDO PDO x F6 PDO x F7	Rear ends Sideswipes	Left lane Not off-road	Multi-vehicle
2. Volume level	PDO PDO x F4	Rear ends (increasing rate)	Left lane Not off-road	Multi-vehicle
3. Synchronized lane conditions	PDO	Sideswipes nonlinear; Mostly more rear ends	Left lane	Multi-vehicle
4. Curb lane perturbation	Injury x F6	Rear ends Sideswipes Rear ends x F6		Mostly less single-vehicle and more 2-vehicle
5. Volume variation	Injury x F6			4+ vehicles 4+ vehicles x F6
6. Conforming curb volumes	PDO x F1 PDO x F4 Injury x F5	Hit object x F4 Rear end x F4	Less left lane x F7 Less left lane x F8	Two-vehicle
7. Systematic volume changes	Injury x F1	Rear ends	Less left lane x F6	Multi-vehicle
8. Synchronized outer flow			Off-road left; Left lane (increasing rate) Less Left lane x F6	

6 Conclusions and Forthcoming Research

It has been demonstrated that an extensive set of statistical parameters – 36 in total – can be extracted from twenty minutes 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 objectives in the second year of the project are to test the model's ability to distinguish locations and conditions with high accident rates from those with low accident rates. In accomplishing this, we have extracted a random sample of traffic flow conditions at times and places that *do not* correspond to accidents, for the same case study network and time period. We intend to establish accident rates in terms of vehicle exposure to different traffic conditions. Once exposure rates are established, code will be developed to deploy the model. The initial application will be as a stand-alone tool on the Testbed website using data from the Caltrans District 12 FEP as input.

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.

Chen, C; K. F. Petty, A. Skabardonis, P.P. Varaiya, and Jia, Z. (2001). Freeway performance measurement system: mining loop detector data. *Transportation Research Record 1748*: 96-102.

Choe, T., Skabardonis, A., and Varaiya, P.P. (2002). Freeway performance measurement system (PeMS): an operational analysis tool. Presented at Annual Meeting of Transportation Research Board, January 13-17, Washington, DC.

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

Golob, T.F. and 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. and Recker, W.W. (2004). A Method for relating type of cash to traffic flow characteristics on urban freeways. *Transportation Research,* 38: 53-80.

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

PeMS (2005). Freeway Performance Measurement System. Department of Electrical Engineering and Computer Science, University of California Berkeley http://pems.eecs.berkeley.edu/Public/ (accessed May 11, 2005).

Varaiya, P. (2001). Freeway Performance Measurement System, PeMS V3, Phase 1: Final Report. Report UCB-ITS-PWP-2001-17, California PATH Program, Institute of Transportation Studies, University of California Berkeley, CA.

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).