Title

Freeway Safety as a Function of Traffic Flow: The FITS Tool for Evaluating ATMS Operations

Authors

Golob, Thomas F.
Recker, Wilfred W.
Alvarez, Veronica

Publication Date

2002-09-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Freeway Safety as a Function of Traffic Flow: The FITS Tool for Evaluating ATMS Operations

Thomas F. Golob, Wilfred W. Recker, Veronica Alvarez
University of California, Irvine

California PATH Research Report
UCB-ITS-PRR-2002-28

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

Final Report for TO 4117

September 2002

ISSN 1055-1425

Freeway Safety as a Function of Traffic Flow:
The FITS Tool for Evaluating ATMS Operations

Thomas F. Golob

Institute of Transportation Studies
University of California, Irvine
tgolob@uci.edu

Wilfred W. Recker

Department of Civil and Environmental Engineering *and*
Institute of Transportation Studies
University of California, Irvine
wwrecker@uci.edu

Veronica M. Alvarez

Institute of Transportation Studies
University of California, Irvine
valvarez@uci.edu

September 2002

Acknowledgement

This research was funded 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.

TABLE OF CONTENTS

1 Objective and Scope

Understanding the benefits of improved traffic flow (reduced congestion) is critical to the assessment of investments in infrastructure or traffic management and control. Improved flow should lead to reductions in travel time, vehicle emissions, fuel usage, psychological stress on drivers, and improved safety. However, the manner in which safety is improved by smoothing traffic flow is not well understood. The documented research is aimed at shedding light on the complex relationships between traffic flow and traffic accidents (crashes).

The overall objective of the project is to develop an evaluation tool that uses relationships between traffic flow and crash characteristics to assess the safety benefits that are likely to be realized under specific ATMS implementations. A program, called FITS (Flow Impacts on Traffic Safety) has been written that uses a data stream of 30-second observations from single inductive loop detectors to forecast the types of crashes that are most likely to occur for the flow conditions being monitored. The program is based on analyses of crash characteristics as a function of traffic flow, using data from the Caltrans Traffic Accident Surveillance and Analysis System (TASAS), in conjunction with loop-detector traffic data from the VDS System. The crash and traffic flow data are for Orange County California in 1998. The traffic flow conditions for a thirty-minute time period immediately preceding more than 1,000 crashes on six major freeways are analyzed against the characteristics of the crashes themselves.

Following a brief background on previous research in Section 2, the data are discussed in Section 3. In Section 4 we present results of the analyses in which we determine traffic flow regimes that best describe differences in safety conditions, controlling for weather and ambient lighting conditions. In Section 5 we describe how the traffic flow regimes defined in Section 4 are distributed over time and space for Orange County freeways for calendar year 1998. In Chapter 6, we apply the tool in a case study of a section of one freeway for one week. We close with conclusions and a discussion of future research in Section 7.

2 Background

Benefit/cost comparisons have long been a standard in assessing the effectiveness of investment of limited resources, and have served as an essential element in determining the most effective allocation of such resources. Developing these comparisons has presented a very perplexing problem to Caltrans Operations in presenting arguments for resources, primarily because hard numbers for benefit/cost ratios associated with traffic management operations can not be obtained practically. For example, the costs of such management strategies as ramp metering or freeway service patrols (FSP) are easily determined. However, a true measurement of the benefits of these strategies can be determined only by shutting down all the ramp metering or curtailing FSP, say for a day so, and measuring any adverse consequences. This direct approach, of course, is not feasible due to liability reasons. This measurement problem is heightened dramatically when issues of safety are involved, yet one of the most compelling arguments for implementation of ITS elements is their presumed enhancement of traffic safety.

Assessment of benefits of ATMS and ITS improvements largely translates into a problem of quantifying the benefits of improved traffic flow. Improved flow ostensibly leads to reductions in travel time, vehicle emissions, fuel usage, psychological stress on drivers, and improved safety. However, the manner in which safety is improved by smoothing traffic flow is not well understood at this time. The present research is aimed at shedding light on the complex relationships between traffic flow and traffic crashes.

There is strong empirical evidence of functional relationships between crash rates and traffic flow, conditional upon roadway characteristics (e.g., Aljanahi, *et al.*, 1999, Cedar and Livneh, 1982, Frantzeskakis and Iordanis, 1987, Garber and Gadiraju, 1990, Gwynn, 1967, Hall and Pendleton, 1989, Maher and Summersgill, 1996, Sandhu and Al-Kazily, 1996, Stokes and Mutabazi, 1996, Sullivan, 1990, Sullivan and Hsu, 1988, and Zhou and Sisiopiku, 1997). A series of studies have dealt with quantification of the safety component of the marginal costs of roadway use, as a function of traffic speed and density or flow (Dickerson, Peirson and Vickerman, 2000, Jansson, 1994, Johansson, 1996, Jones-Lee, 1990, Newberry, 1988, O'Reilly, *et al.*, 1994, Shefer and Rietveld, 1997, Vickery, 1969, and Vitaliano and Held, 1991).

These previous studies all used such aggregate traffic flow data as hourly traffic counts and volume to capacity measures. Types of crashes are generally not distinguished, except that crashes with fatalities traditionally are studied independently, and in rare instances injury and property-damage crashes are separated. Largely due to limited data availability, quantitative specification of a relationship between crash characteristics and traffic flow, as measured by commonly available monitoring devices, has remained elusive. By using traffic flow data prevailing just prior to the time of each crash and by including the conditions of the crash in the analysis, we are able to avoid the two problems of averaging cited by Mensah and Hauer (1998) – "argument" averaging and "function" averaging. The "argument" averaging problem is caused by using aggregate traffic flow data, rather than data measuring traffic conditions at the time of the crash; this problem is overcome in the current study by incorporating actual traffic flow measures, together with their spatial and temporal gradients, prevailing immediately prior to the associated crash. The second of the two problems is function averaging, which is caused by using the same functional relationship for all types of crashes under all conditions. This is addressed by explicitly incorporating type of crash in the analyses and by segmenting according to different environmental conditions. The resulting analyses show clear patterns emerging from the relationships between crash characteristics and prevailing flow conditions. However, as discussed in the concluding Section 8, complementary work remains to be completed before these analyses can be translated into crash rates for different traffic conditions.

3 Data Description

3.1 Fusion of Crash and Traffic Flow Data

The TASAS database (Caltrans, 1993) covers crashes that occur on the California State Highway System for which there are police reports. 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. Due to the large number of jurisdictions involved, typically there is at least a half year lag in entering all crashes into the TASAS database. Our case study is of Orange County California, and 1998 is the most recent year with a mature TASAS database and loop-detector traffic data available for the entire year (loop detector data were unavailable for much of 1999 due to relocation of the Caltrans Orange County Transportation Management Center and system modifications).

The TASAS database structure distinguishes three types of crashes: highway crashes, ramp crashes, and intersection crashes. In calendar year 1998, the database contains 11,958 highway crashes, 2,357 ramp crashes, and 894 intersection crashes for sixteen routes in Orange County. Here we are concerned only with highway (mainline) crashes on well-defined (i.e., those of substantial length) urban freeways. A total of 9,341 crashes, or 78% of all highway crashes, occurred on six freeway routes: Interstate 5 (the Santa Ana Freeway and the southern section of the San Diego Freeway), State Route 22 (Garden Grove Freeway), State Route 55 (Costa Mesa Freeway), State Route 57 (Orange Freeway), State Route 91 (Riverside and Artesia Freeways), and Interstate 405 (the northern section of the San Diego Freeway in Orange County). These crashes are the subject of this study. The remainder of the highway crashes occurred on arterial routes or on (short or newly opened) freeway segments with fewer than 200 crashes.

traffic flow data for the time period leading up to each crash were matched to each crash. These data come from an archived database of 30-second observations from single inductance loop detectors maintained throughout the State Highway System. Each observation provides count and occupancy averages for a 30-second time slice. It was arbitrarily decided to collect data for 30 minutes prior to the reported time of each crash at four loop locations – those being the closest two loops upstream and downstream of the crash. The analysis reported here uses data only for that loop detector station closest to location of the crash. For other analyses being conducted as part of the overall study, we use data from all four loop detector stations, which allows incorporation of spatially longitudinal differences in traffic flow conditions.

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 85.6% of the 9,341 crashes have reported times in minutes that fall precisely on the twelve five-minute intervals that comprise an hour. Because of this obvious reporting bias, 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 observations used in the analysis is cut off 3 minutes before the “nominal” crash time to remove “cause and effect” ambiguities associated with round-off. Consequently, for each crash, pre-crash traffic conditions are measured by up to 55 thirty-second loop-detector observations, beginning 30 minutes before the nominal crash time.

The focus is on crashes that involved vehicles traveling on the mainline freeway lanes, so all traffic flow data comes from mainline single inductive loop detectors. At each mainline loop detector station, data typically are collected for each freeway lane; the minimum number of lanes at any mainline freeway section in Orange County in 1998 was three. In order to standardize traffic flow data for all crashes independent of the number of freeway lanes involved, data were compiled for three lane designations: (a) the left lane, always being the lane designated as being the number one lane according to standard nomenclature; (b) an interior lane, being lane two on three- and four-lane freeway sections and lane three on five- and six-lane sections; and (c) the right lane, always being the highest numbered (right-most) lane. The corresponding total number of loop detector observations sought for the analysis reported here is given by the product of 9,341 crashes, 55 time slices, and 3 lanes per location, or 1,541,265 distinct 30-second counts and occupancies.

Use of loop detector data is reliant upon the performance of the data retrieval and processing system. Missing data proved a major problem in dealing with the loop detector data. Complete data for all 55 time slices (a 22.5-minute period) was available for 24.5% of the stations; another 11.4% of the stations had missing data for one or more of the 55 time slices. The remaining 64.1% of the loop detector stations had no data at all for the entire 27.5-minute period. Presumably these latter stations were inoperative at that time, or there was some other problem in retrieving the data.

Filtering of observations was still necessary for the loop detector stations with full or partial data. We reviewed all data sequences based on time series deviations, deviations across lanes, and logical rules derived from feasible volume and occupancy relationships (i.e., from properties of plausible fundamental diagrams). Based on these tests, approximately 16% of the available 30-second loop-detector observations were identified as being potentially invalid. In situations where one 30-second observation was missing or out-of-bounds but the data for the adjacent time slices were valid, the data for the missing time slice were interpolated from the adjacent observations.

Implementation of the filtering and interpolation operations resulted in a sample of 1,192 crashes with a full 27.5 minutes of ostensibly valid loop detector data for the designated three lanes at the closet detector station. This represents 12.8% of the 9,341 highway crashes on the six major Orange County freeways that are recorded in the TASAS database for 1998. For this final sample, the average distance from the crash location to the closest detector station is 0.17 miles and the median distance is 0.12 miles. Fully 78% of the 1,015 crashes were located within 0.25 miles of the detector station.

3.2 Crash Characteristics

The following crash characteristics are available in the TASAS dataset: (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) injuries and fatalities per vehicle, and (8) environmental conditions, such as lighting, weather, and pavement conditions. No information was available concerning drivers. Based on exploratory analyses, only the three characteristics listed in Table 1 were found to be useful in the analysis.

Table 1 Crash Characteristics Used in the Analyses

	Percent of sample (N = 1192)
Collision type	
Single vehicle hit object or overturn	14.2
Multiple vehicle hit object or overturn	5.9
Two-vehicle weaving crash [a]	19.3
Three-or-more-vehicle weaving crash [a]	5.5
Two-vehicle straight-on rear end	33.8
Three-or-more-vehicle straight-on rear end	21.3
Collision Location	
Off-road, driver's left	13.8
Left lane	25.8
Interior lane(s)	32.7
Right lane	19.3
Off road, driver's right	8.3
Severity	
Property damage only	71.9
Injury or fatality [b]	28.1

[a] Sideswipe or rear end crash involving lane change or other turning maneuver

[b] There were only five fatal accidents

The first crash characteristic in Table 1 is our new six-category coding of collision type. This variable incorporates the most important information from three TASAS variables: collision type, movement prior to collision, and number of vehicles. The new coding avoids problems of structural relationships if the three TASAS variables were used separately (e.g., rear-end and sideswipe crashes by definition involve more than one vehicle, and rear-end crashes almost always involve a vehicle slowing or stopped. The second crash characteristic in Table 1 recodes the TASAS variable defining location of the primary collision into five categories by combining all drivers-left and driver-right off-road categories. Finally, the third crash characteristic distinguishes crashes with injuries or fatalities from crashes that entail property damage only. There were only five fatal crashes out of a total of 1192, too few to include as a separate category in the analyses.

3.3 Traffic Flow Variables

This research uses raw detector data that provide information on two variables: count and occupancy for each thirty-second interval. Although these two variables can be used (under very restrictive assumptions of uniform speed and average vehicle length, and taking into account the physical installation of each loop) to infer estimates of point speeds, we avoid making any such assumptions, and use only these direct measurements in the analysis that follows.

Based on preliminary analyses, four blocks of three variables (one measure for each of the three lane type designations, left, interior, and right) were identified as being potentially related to typology of crash. The first of these blocks is an indicator of prevailing traffic speed, the second the temporal variation of the prevailing speed, the third the traffic volume, and the fourth the temporal variation in the traffic volume. The four blocks of three variables are listed in Table 2.

Table 2 Traffic Flow Variables

Block 1 Central tendency of speed	Median volume/occupancy - left lane
	Median volume/occupancy - interior lane
	Median volume/occupancy - right lane
Block 2 Variation in speed	Difference between 90th and 50th percentiles of volume/occupancy - left lane
	Difference between 90th and 50th percentiles of volume/occupancy - interior lane
	Difference between 90th and 50th percentiles of volume/occupancy - right lane
Block 3 Central tendency of volume	Mean volume left lane
	Mean volume interior lane
	Mean volume right lane
Block 4 Variation in volume	Standard deviation of volume left lane
	Standard deviation of volume left lane
	Standard deviation of volume left lane

The interpretation of these traffic flow variables is as follows.

- The first block measures the central tendency of the ratio of volume to occupancy, an approximate proportional indicator of space mean speed. Median, rather than mean, is used in order to avoid the influence of outlying observations that can be due to failure of the loop detectors
- The second block represents the variation in the ratio of volume to occupancy over the entire period. Because the variable is defined as a ratio and we wish to minimize the influence of potentially invalid observations and the effects of outliers, we use the difference of the 90th percentile and 50th percentile to capture variation.
- The third block measures the central tendency of volume over the entire 22.5 minute period. Volume alone is not as sensitive to outliers as the ratio of volume to occupancy is, so mean is used rather than median. Mean and median values are quite similar for these data, so either can be used without affecting results.
- The fourth block measures variation in volume over the entire period. Here we use standard deviation, but the difference between the 90th percentile and 50th percentiles can be used without affecting the results.

4 Traffic Flow and Safety for Different Weather and Lighting Conditions

The objective is to find the best explanation of patterns in the three crash characteristics listed in Table 1 as a function of the flow characteristics listed in Table 2. To avoid the problem of argument averaging for the resulting functional relationships, we assume that the relationships between crash typology and traffic flow conditions will depend upon driving conditions defined, at a minimum, by weather and ambient lighting (Fridstrøm, *et al.*, 1995).

4.1 Segmentation Based on Weather and Lighting Conditions

Our data can be used to distinguish six sets of environmental conditions, defined by the combination of weather and ambient lighting, as shown in Table 3. Several of the cells in Table 2 contain too few observations to support the nonparametric analysis. In the nonparametric analyses used in this study, the total number of variables for which results are sought and interpreted is not simply the total number of categorical variables in the study, but rather the total number of categories by which these variables are described, minus the total number of variables. Also, many types of nonparametric analyses are sensitive to spurious effects caused by variable categories with relatively small frequencies. (This is analogous to the outliers problem in conventional linear methods.) Thus, there is a tradeoff between the total number of categories in all of the nominal variables (Table 1) and the size of the analysis sample. After several tests, we determined that our method could support a minimum sample size of approximately 120. Only two of the cells in Table 3 satisfy this criterion; it became necessary to combine or eliminate the remaining cells in Table 3 in order to meet this minimum sample size requirement. To accomplish this, we determined an optimal segmentation strategy based on similarities and differences among the categories defined by the cells in Table 3.

Table 3 Segmentation Based on Weather and Ambient Lighting Conditions

Lighting	Weather [a]		Total by lighting
	Dry	Wet	
Daylight	789	101	890
Dusk or dawn	30	3	33
Darkness	217	52	269
Total by weather condition	1036	156	1192

[a] Based on condition of the roadway surface (wet or dry)

A canonical correlation analysis (CCA) was conducted in which the single exogenous variable represented weather and lighting conditions with five categories defined by each of the cells in Table 3, with the exception of the cell representing dusk-dawn crashes on wet roads (the three wet dusk-dawn crashes were excluded from this analysis). The minimum category size was thus 30, which is sufficient to support the CCA analysis. The endogenous side of the problem was composed of the three crash characteristics listed in Table 1 (i.e., Collision Type, Collision Location, and Severity). The objective in this analysis is to determine similarities among the five segments of weather and ambient lighting conditions identified in Table 3 in terms of their explanation of the crash typology identified in Table 1.

If all of the variables were numerical (measured on a scale with equal intervals), and all functional forms expected to be linear, this could be accomplished using conventional linear CCA. CCA is an expansion of regression analysis to more than one dependent variable, and the objective is to find a linear combination of the variables in each of two or more sets, so that the correlations among the linear combinations in each set are as high as possible. Depending on the number of sets and the number of variables in each set, multiple linear combinations (called canonical variates) can be found that have maximum correlations subject to the conditions that all canonical variates are mutually orthogonal (uncorrelated).

The present CCA problem involves nonparametric (nonlinear), rather than numerical variables. The variables defining the five segments of weather and lighting conditions (eliminating the “dusk or dawn – wet” category) and the two crash characteristics with more than two categories (i.e., “collision type” and “collision location”) are nominal (categorical) by definition.[2] Also, because we expect to find nonlinear relationships involving the traffic flow variables, these variables are also assumed to be nonlinear (either nominal or ordinal) in determining the optimal functional forms. The nonparametric CCA problem is more complex than its linear counterpart, because the optimal linear combination of the variables is undefined until the categories of each crash characteristic are quantified and the most effective nonlinear transformations of the traffic flow variables are determined. The variable categories must be optimally quantified (scaled) while simultaneously solving the traditional linear CCA problem of finding variable weights (van de Geer, 1986, van Buren and Heiser, 1989, ver Boon, 1996).

An elegant solution to the nonparametric (nonlinear) CCA problem was first proposed by researchers at the Department of Data Theory of Leiden University in the Netherlands. The Leiden University team developed a suite of nonparametric methods for conducting canonical correlation analysis (CCA), principal components analysis, and homogeneity analysis with variables of mixed scale types: nominal, ordinal, and interval. Their nonlinear CCA (NLCCA) method was operationalized in a program called CANALS (Canonical Analysis by Alternating Least Squares), later extended to generalized canonical analysis with more than two sets of variables in a program called OVERALS. The Leiden method for nonlinear CCA is described in de Leeuw (1985), van der Burg and de Leeuw (1983), Israëls (1987), Michailidis and de Leeuw (1998), and (most extensively) in Gifi (1990). The method simultaneously determines both (1) optimal re-scaling of the nominal and ordinal variables and (2) variable weights (coefficients), such that the linear combinations of the weighted re-scaled variables in all sets are maximally correlated. The variable weights and optimal category scores are determined as an eigenvalue problem related to minimizing a loss function derived from the concept of "meet" in lattice theory.

A two-dimensional NLCCA solution was chosen. The original weather – lighting segmentation variable category quantifications (five, in all) were allowed to have optimal scaling independently on each of the two dimensions (called canonical variates). The first dimension arising from the solution mostly explains collision type (correlation = ρ = 0.68) and, to some extent, severity (ρ = -.40). The second dimension mostly explains collision location (ρ = 0.72). The optimally scaled segmentation variable has correlations of 0.79 and 0.76 with the two dimensions, respectively. The optimal category quantifications for the segmentation variable on the two dimensions are plotted in Figure 1. The similarity between any two categories (segments) in terms of their explanation of crash typology is indicated by the Euclidean distance between the two categories in the orthogonal space of Figure 1 (Ter Braak, 1990).

[2] A nominal variable with only two categories is a special case of a numerical variable, a dummy variable, since only one interval is involved.

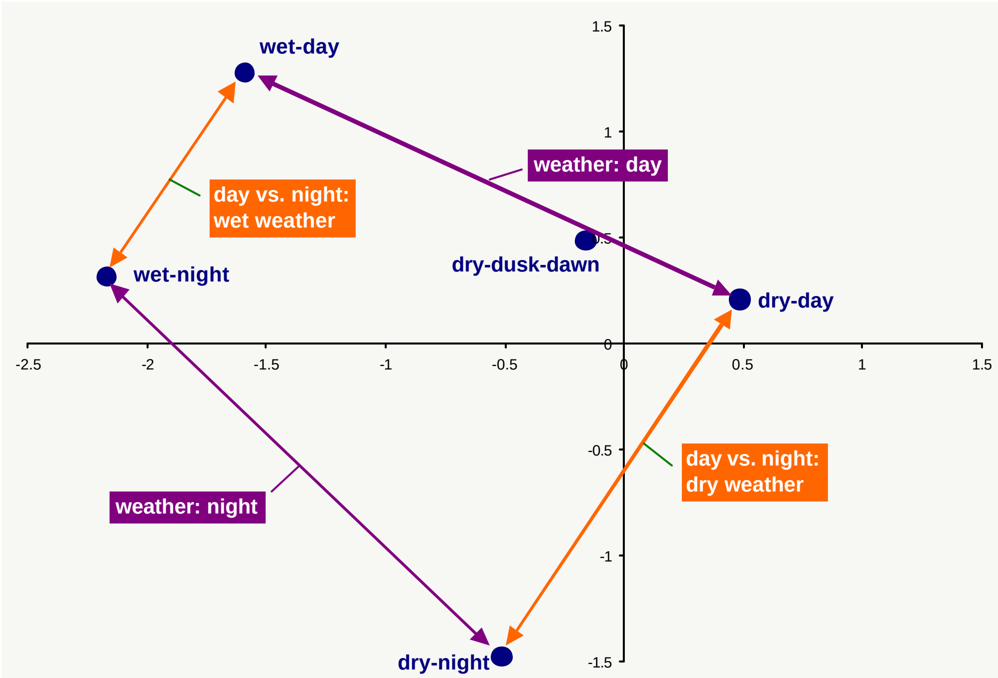

Figure 1 Nonlinear Canonical Correlation Analysis of Weather and Lighting Conditions Versus Crash Characteristics: Category Quantifications for Weather and Lighting Segments in

The category quantifications reveal that day and night crashes for the same weather conditions are more similar than wet and dry crashes for the same lighting conditions. That is, weather has a greater influence than lighting on crash typology. This result is intuitive, in that weather conditions directly impact the braking and handling performance characteristics of the vehicle, which under prevailing driving conditions tend to have a relatively narrow margin for error, while lighting conditions are mediated through a broader spectrum of capable human factor response. Additionally, adverse weather conditions are likely to be compounded by the same sort of visibility issues as are associated with diminished lighting conditions. The two pairs of categories – (1) dry-day versus dry-night and (2) wet-day versus wet-night – are separated by two nearly-parallel axes shown as double-arrows in Figure 1. The two axes separating wet and dry crashes for the same lighting condition are approximately orthogonal to the lighting axes; these are also shown as solid double-arrows in Figure 1. Because the two lighting conditions are more similar in wet weather than in dry weather, the geometric pattern of relationships in Figure is that of a trapezoid.

The relationship between the weather and lighting segments and crash type is depicted in Figure 2. Collision type is almost entirely explained by the first canonical variate, as shown by the alignment vector of the projections of the category quantifications for the optimally scaled crash type variable. Thus, crash type is a function of both lighting and weather. Single-vehicle hit-object crashes are more likely to occur on wet roads at night, while rear end crashes are associated with dry-daylight conditions.

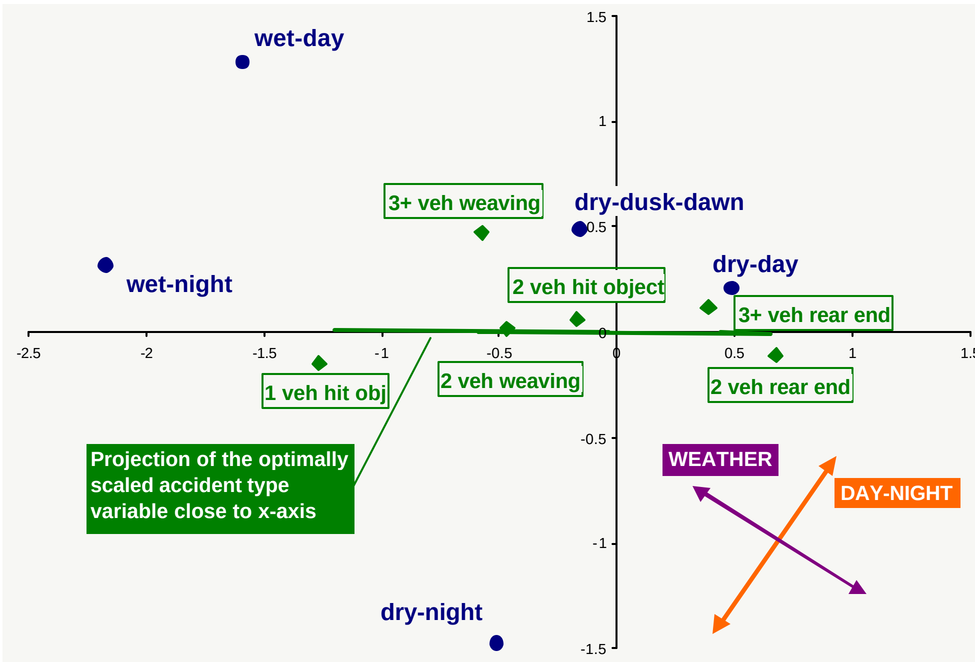

Figure 2 Nonlinear Canonical Correlation Analysis of Weather and Lighting Conditions Versus Crash Characteristics: Category Quantifications for Crash Type and Weather and Lighting Segments

The relationship between the weather and lighting segments and crash location is depicted in Figure 3. Location is explained primarily by the second canonical variate, and the vector of projections of category quantifications is oriented in the same direction as the lighting axis. Crash location is more related to lighting conditions than to weather conditions. In particular, crash locations on the left side of the freeway (left lane or off-road to drivers' left) are more likely during daylight, while crash locations on the right side are more likely during nighttime.

The relationship between the weather and lighting segments and crash severity is depicted in Figure 4. Severity is explained primarily by the first canonical variate and is aligned with the weather axis. Injury crashes are more likely during wet weather, both during the day and at night. Property-damage-only crashes are more likely during dry weather.

Dry dusk-dawn crashes are most similar to dry day crashes, but they vary from dry-day crashes in a direction towards wet-day crashes. The interpretation is that crashes occurring during dusk and dawn periods are more similar to daytime crashes than they are to nighttime crashes, but they have some of the characteristics of daytime wet-weather crashes, perhaps a reflection of diminished daytime visibility common to the dusk/dawn hours and during rainy conditions.

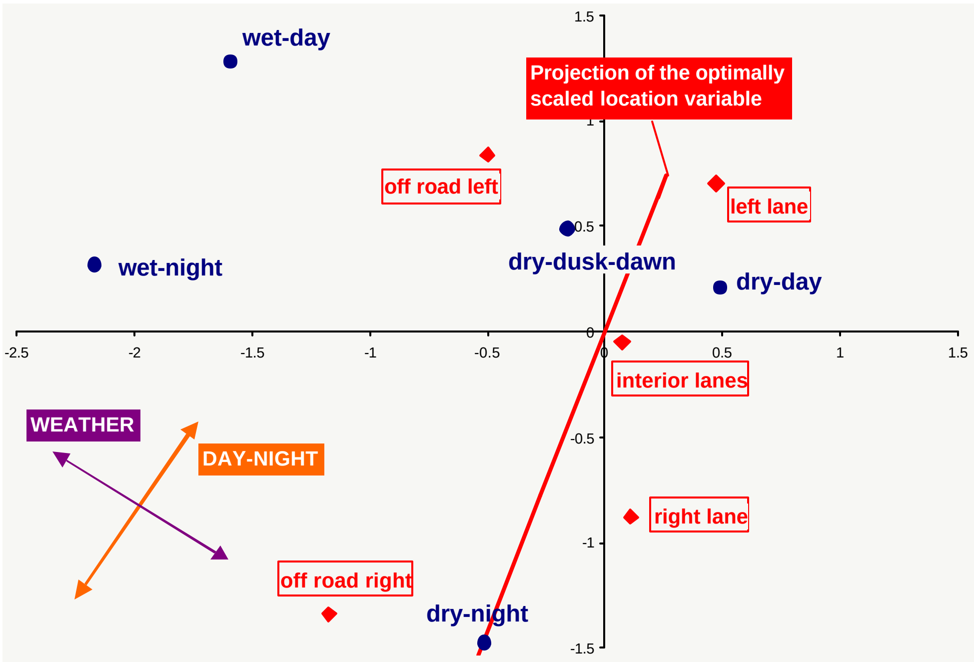

Figure 3 Nonlinear Canonical Correlation Analysis of Weather and Lighting Conditions Versus Crash Characteristics: Category Quantifications for Crash Location and Weather and Lighting Segments

Based on these NLCCA results, it is best to combine the Wet-Night and Wet-Day segments into a single Wet segment. (This implies that the very few Wet-Dusk-Dawn observations should be included in this Wet segment as well.) This also solves the problem of insufficient sample size for the original Wet-Night segment. Secondly, the relatively sparse Dry-Dusk-Dawn segment can be combined with the adjacent Dry-Day segment. The resulting segmentation is (1) Dry-Day (including Dusk-Dawn): 819 crashes, (2) Dry-Night, 217 crashes, and (3) Wet (any lighting condition): 156 crashes. This segmentation scheme satisfies the minimum cell size criterion of 120, and is used in the remaining analyses.

The breakdowns of the three crash characteristics for each weather and lighting segment are listed in Table 4. Descriptive statistics for the traffic flow variables described in Table 2 are listed in Table 5 for the three weather and lighting segments.

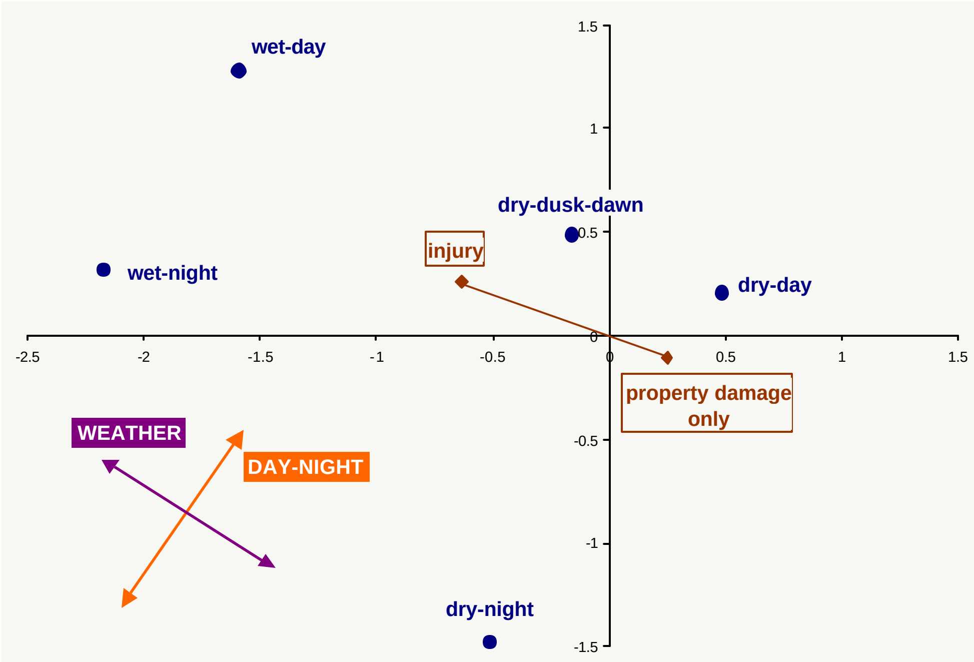

Figure 4 Nonlinear Canonical Correlation Analysis of Weather and Lighting Conditions Versus Crash Characteristics: Category Quantifications for Crash Severity and Weather and Lighting Segments

Table 4 Characteristics of the Crashes that Occurred Under Three Conditions (percentage breakdowns)

	Daylight-Dry N = 819	Dark-Dry N = 217	Wet road N = 156
Collision type			
Single vehicle hit object or overturn	10.5	18.9	26.9
Multiple vehicle hit object or overturn	5.6	6.5	6.4
Two-vehicle weaving crash [a]	17.8	20.3	25.6
Three-or-more-vehicle weaving crash [a]	5.1	4.1	9.6
Two-vehicle straight-on rear end	38.2	30.0	16.0
Three-or-more-vehicle straight-on rear end	22.7	20.3	15.4
Collision Location			
Off-road, driver's left	12.3	11.5	25.0
Left lane	30.4	15.7	16.0
Interior lane(s)	32.5	32.3	34.6
Right lane	18.7	26.3	12.8
Off road, driver's right	6.1	14.3	11.5
Severity			
Property damage only	75.0	70.5	57.7
Injury or fatality	25.0	29.5	42.3

[a] Sideswipe or rear end crash involving lane change or other turning maneuver

Table 5 Descriptive Statistics for the Traffic Flow Variables for Three Segments

		Daylight crashes on dry roads		Crashes during darkness on dry roads		Crashes on wet roads	
		Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
Block 1	Median volume/occupancy (V/O) left lane	96.0	47.3	100.6	54.0	120.8	58.1
	Median volume/occupancy (V/O) interior lane	93.6	46.6	107.6	49.7	110.4	43.1
	Median volume/occupancy (V/O) right lane	91.1	99.5	103.0	38.9	94.7	43.9
Block 2	Variation: difference between 90th and 50th percentiles of V/O left lane	23.6	16.1	33.8	39.7	18.9	10.3
	Variation: difference between 90th and 50th percentiles of V/O interior lane	24.7	14.5	26.9	18.1	20.5	10.5
	Variation: difference between 90th and 50th percentiles of V/O right lane	30.4	16.1	40.8	51.1	28.3	19.3
Block 3	Mean volume left lane	13.7	3.2	9.7	5.1	11.0	4.0
	Mean volume interior lane	13.4	3.0	10.0	4.6	11.3	3.5
	Mean volume right lane	11.2	3.2	7.9	4.5	9.4	3.8
Block 4	Variation: standard deviation of volume left lane	3.2	0.8	2.5	0.9	2.8	0.7
	Variation: standard deviation of volume interior lane	2.9	0.7	2.5	0.7	2.5	0.6
	Variation: standard deviation of volume right lane	2.8	0.6	2.2	0.8	2.5	0.6

4.2 Daylight Crashes on Dry Roads

4.2.1 Principal Components Analysis of the Traffic Flow Variables

We expect that some pairs of our traffic flow variables will be highly correlated, potentially clouding interpretation of results from any analysis in which the full complement of traffic flow variables are related to safety analysis. Specifically, the three variables in each of the four blocks (Tables 2 and 5) might be highly correlated if the flow characteristic being measured is consistent across all three freeway lanes. However, it is not known how strongly these particular measures of traffic flow are linked across lanes, and this is especially true of speed and volume variances. To minimize this potential problem, principal components analysis (PCA) was performed to identify linear combinations of strongly related variables. Our objective was to extract a sufficient number of factors to identify independent traffic flow variables while simultaneously discarding as little of the information in the original variables as possible.

A PCA was performed on the twelve traffic flow variables for the group of crashes that occurred during daylight or dusk-dawn on dry roads; six factors accounted for 86.8% of the variance in the original twelve variables. The six-factor PCA solution is invariant under orthogonal rotations; varimax rotation, a standard technique in factor analyses, was performed to aid in interpreting these factors. The rotation results in a redistribution of the explanatory power of each factor while preserving the cumulative variance explained by all retained factors. The factor loadings, which are the correlations between the original variables and the rotated factors, are listed in Table 6. Also listed in Table 6 are the variances accounted for by each factor. For ease of interpretation, one variable was then selected to represent each factor in the subsequent stages of the analysis.

The factor loadings show that the central tendency of speed (Variable Block 1) is consistent across all three lanes. Based on consistent PCA results for the other two lighting and weather segments (reported in Sections 4.3.1 and 4.4.1), the variable chosen to represent this central tendency of speed factor is median volume/occupancy in the interior lane. The correlation between this variable and its factor is shown underlined in bold in Table 6.

A single factor also encompasses the central tendency of volume (Variable Block 3) in all three lanes, but the factor is more representative of volumes in the left and interior lanes than in the right lane, as witnessed by the lower correlation between this factor and right lane mean volume (0.635). Mean volume in the left lane was chosen to represent this factor in all further analyses. (Although the factor loading for mean volume in the interior lane is greater, its higher correlations with other factors, not shown, resulted in the choice of volume in the left lane to represent this factor.)

Table 6 Rotated PCA Loadings for Traffic Flow Variables for the Daylight, Dry Roads (showing only absolute values > 0.3)

		Principal component					
	Traffic flow variable	1	2	3	4	5	6
	Percentage of original variance accounted for	22.5%	17.8%	14.6%	13.6%	9.4%	9.0%
Block 1	Median volume/occupancy left lane	0.904					
	Median volume/occupancy interior lane	**0.892**					
	Median volume/occupancy right lane	0.921					
Block 2	Variation in volume/occupancy left lane	-.308		0.832			
	Variation in volume/occupancy interior lane			**0.853**			
	Variation in volume/occupancy right lane						**0.911**
Block 3	Mean volume left lane		**0.920**				
	Mean volume interior lane		0.929				
	Mean volume right lane		0.635			0.392	-.418
Block 4	Variation in volume left lane				**0.902**		
	Variation in volume interior lane				0.821	0.323	
	Variation in volume right lane					**0.914**	

Factor three represents the temporal variation in speed on the left and interior lanes only. Variation in speed in the right lane is captured by a separate, sixth, factor. Variation in speed is represented by variation in the volume to occupancy ratio on the left and interior lanes in further analyses, and variation in volume to occupancy ratio for the right lane represents the sixth factor. The implication here is that the variation in speed in the rightmost lane, which may be influenced significantly by merging behavior in the vicinity of freeway on- and off-ramps, relates to crash characteristics in a fundamentally different way than does the variation in speed that is attributable primarily to mainline freeway flow.

Finally, the PCA results also show that temporal variations in volumes on the three lanes is partitioned into two factors: variations in volume on the left and interior lanes (factor 4), and variation in volume on the right lane (factor 5). The left lane is again chosen to represent the former factor. Here again, The implication is that the volume in the rightmost lane, which has a direct influence on the level of service in the vicinity of freeway on- and off-ramps, relates to crash characteristics in a fundamentally different way than does the mainline freeway flow.

The six variables chosen as input to the FITS program are listed in Table 7. These same variables are used for all lighting and weather conditions.

Table 7 Loop Detector Variables Used to as Input to the FITS Toll for all Lighting and Weather Conditions

Specific Variable	Representing
Median volume/occupancy interior lane	Mean speed (on all lanes)
90th%tile - 50th%tile of volume/occupancy interior lane	Variation in speed – all but right lane
90th%tile - 50th%tile of volume/occupancy right lane	Variation in speed – right lane
Mean volume left lane	Mean volume (on all lanes)
Standard deviation of volume interior lane	Variation in volume – all but right lane
Standard deviation of volume right lane	Variation in volume – right lane

4.2.2 Cluster Analysis in the Space of Six Traffic Flow Variables

Cluster analyses were performed in the space of these six principal traffic flow variables In order to establish relatively homogenous traffic flow regimes. A k-means clustering algorithm was used. The objective was to determine the best grouping of observations into a specified number of clusters, such that the pooled within groups variance is as small as possible compared to the between group variance given by the distances between the cluster centers. We repeated runs of the clustering algorithm with different initial cluster centers to avoid local optima.

The optimal number of clusters is usually determined by inspecting various clustering criteria, most of which are developed from eigenvalues of the characteristic equation involving the ratio of the pooled within-groups and between-groups variance matrices. Two of the commonly used criteria, (1) Wilk's Lambda, given by the ratio of the determinants of the within-groups and total variance matrices (equivalent to the product of the eigenvalues of the characteristic equation), and (2) Hotelling's Trace, given by the sum of these eigenvalues, are graphed in Figure 5. It can be seen that selection of the optimal number of groups using such criteria is relatively arbitrary, although there is some indication that between eight to ten groups (a point at which the Wilk's Lambda begins to exhibit asymptotic behavior, and where there is a break in the linear behavior of the Hotelling's Trace) would be a good choice. As in many applications with well-distributed continuous data on many variables, there is no natural number of clusters based on clustering criteria alone.

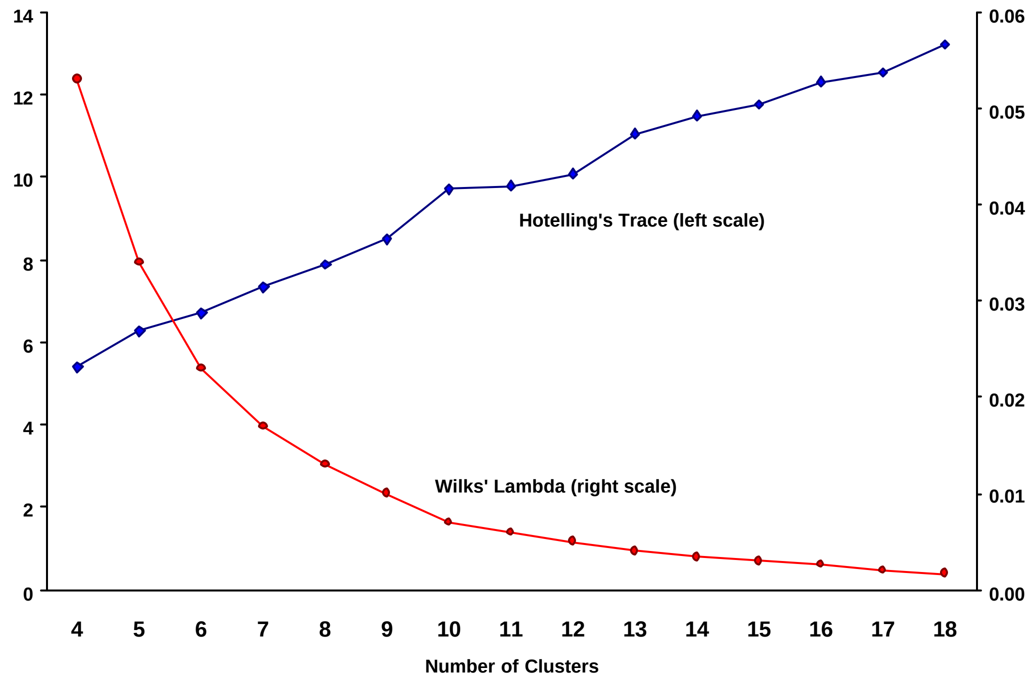

Figure 5 Clustering Criteria by Number of Clusters – Daylight, Dry Roads

Fortunately, in the present application, we can apply an external criterion to the clustering problem to identify the optimal number of clusters. We conducted NLCCA for each clustering solution (from 4 to 18 clusters). The NLCCA problem was configured with the multiple nominal cluster variable on one side and the three single nominal crash variables described in Table 4 on the other side (NLCCA methodology is described in Section 3). The criteria that describe how well each of the cluster variables explained the crash characteristics are the canonical correlations between the two sets of variables, one for each of the variates of the two-dimensional solution. The results are shown in Figure 6. It can be seen that the canonical correlations for the first dimension reaches a maximum at eight clusters. The fit for the second dimension does not improve until the 13-cluster level is reached. Based on these results, and the results shown in Figure 5, eight clusters (representing eight distinct traffic flow regimes, hereafter simply referred to as “Regimes”) were selected. It is not possible to use 14 clusters, because cluster sizes become too small.

The distribution of the 819 dry-day crashes over the eight Regimes is as shown in Table 8. The flow characteristics of the Regimes are described in the next Section.

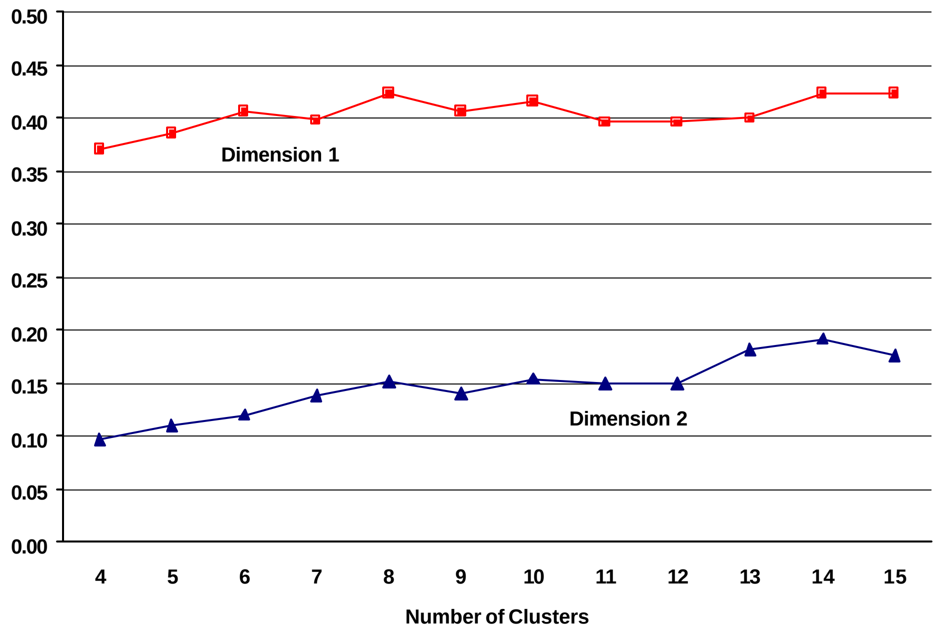

Figure 6 Canonical Correlations for Two-dimensional Nonlinear Discriminant Analysis Solutions for Different Number of Clusters – Daylight, Dry Roads

Table 8 Sample Distribution Among the Eight Traffic Flow Regimes – Daylight, Dry Roads

Regime	Membership	% of all dry daylight crashes
D1	71	8.7
D2	68	8.3
D3	99	12.1
D4	85	10.4
D5	159	19.4
D6	148	18.1
D7	81	9.9
D8	108	13.2

4.2.3 Description of the Eight Dry-Day Traffic Flow Regimes

We begin by investigating the mean values for each traffic flow Regime on each of the six salient traffic flow variables. In Figures 7 through 14 below, the Regime means are scaled in terms of standard deviations from each variable's grand mean for the entire sample of dry daylight crashes (N = 819). Each of the figures highlights the means for a specific Regime while showing the other seven Regimes in the background for comparison purposes. Distributions of numbers of crashes across hours of the day are investigated in conjunction with the group means to aid in interpreting the Regimes.

Regime D1 (Figure 7) is characterized by very low mean volume, high speed, and average variances, except for lower than average variance in right-lane speeds. We can describe Regime D1 as "light free-flow."

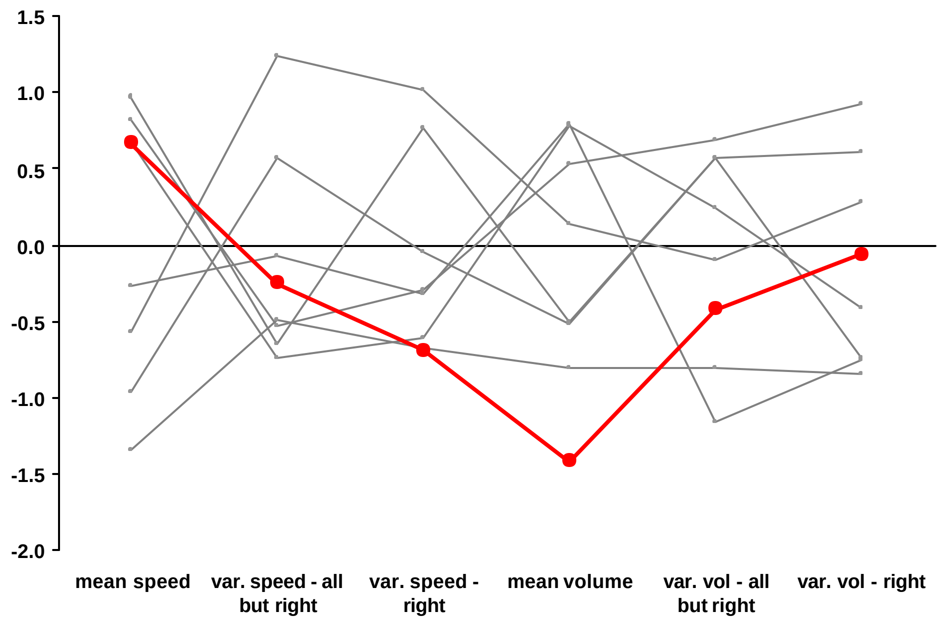

Figure 7 Means for Regime D1 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

The cluster means for Regime D2 are shown in Figure 8. Regime D2 exhibits the lowest mean speed for any Regime, low speed variances, especially for the right lane, low volumes, and low volume variances. We can describe this Regime as "heavily congested flow."

Regime D3 (Figure 9) is characterized by low mean speed and mean volume, and high variances in volume and (mid- and left-lane) speed. These conditions are indicative of "congested flow."

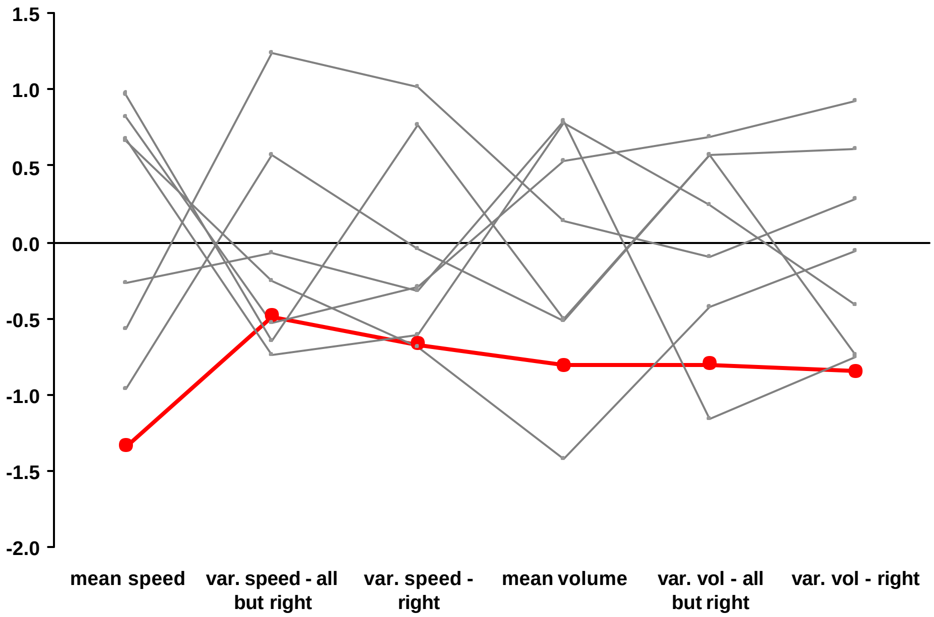

Figure 8 Means for Regime D2 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

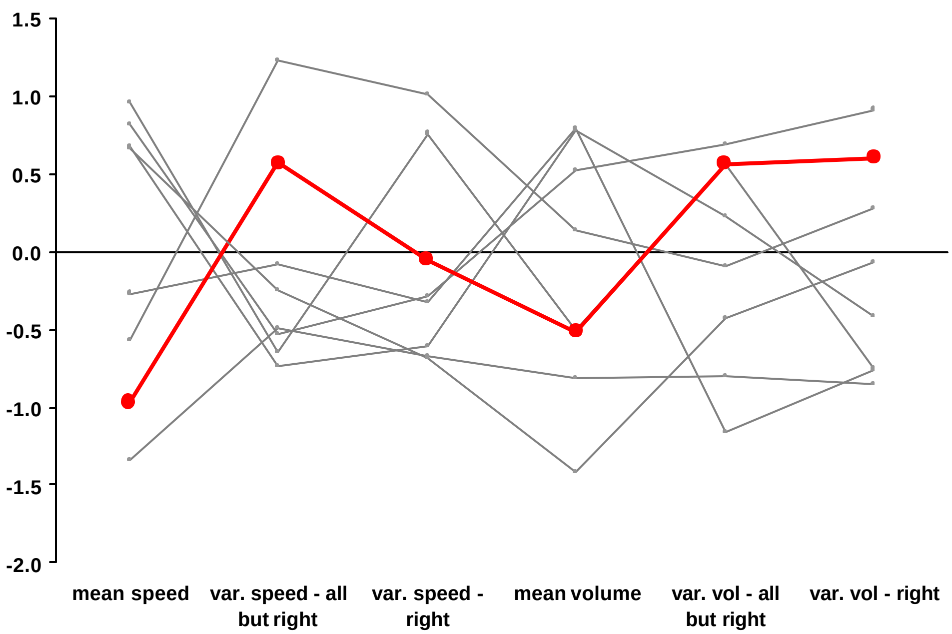

Figure 9 Means for Regime D3 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

The cluster means for Regime D4 are shown in Figure 10. This Regime is characterized by high mean speeds and low mean volume. However, temporal variances of volume and speed are quite different in the right versus the interior and left lanes. In the right lane, the variance of speed is high and, correspondingly, the variance of volume is low. In the left and interior lanes, the opposite is true, with the variances of speed low and the variances of volume high. This may be explained by relatively low weekend-type traffic characterized by a higher relative percentage of activity in the vicinity of on- and off-ramps (a by-product of shorter, non-commuter trips). Under such a scenario, the additional traffic either entering or exiting the freeway would be expected to cause a relatively high variance in freeway volume, but within the range of the fundamental diagram where freeway speed is relatively insensitive to traffic volume. Conversely, relatively high merging of ramp traffic with fast-moving freeway traffic in the rightmost lane would be expected to result in correspondingly high variation in speed in the rightmost lane. The Regime can be labeled as “light, right-variable flow.”

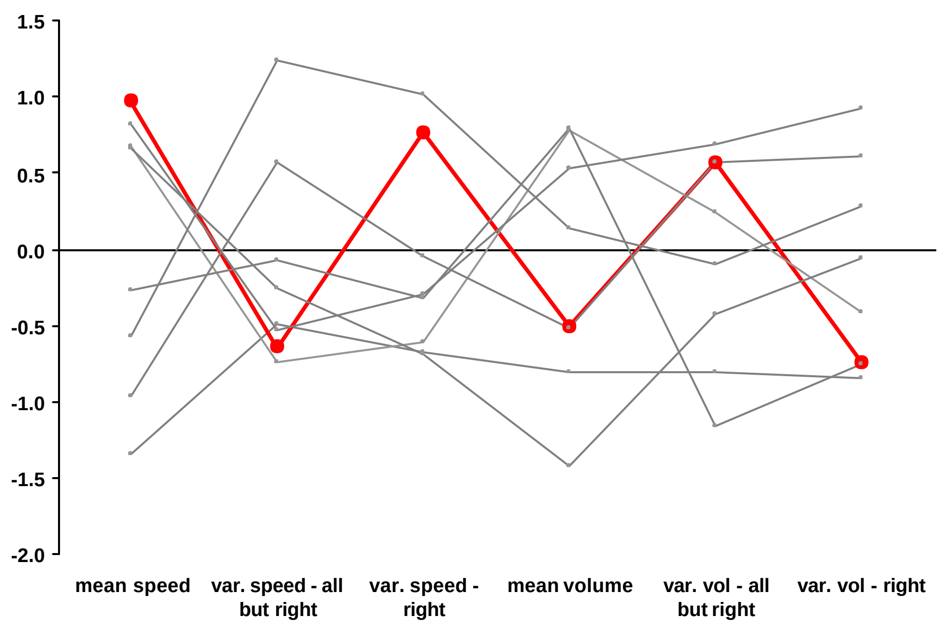

Figure 10 Means for Regime D4 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

Regime D5 (Figure 11) is characterized by very high variances in speed, average volumes, and moderately low average speed. These are characteristics of “flow at capacity,” where speeds vary substantially over time and space, but volumes remain fairly constant.

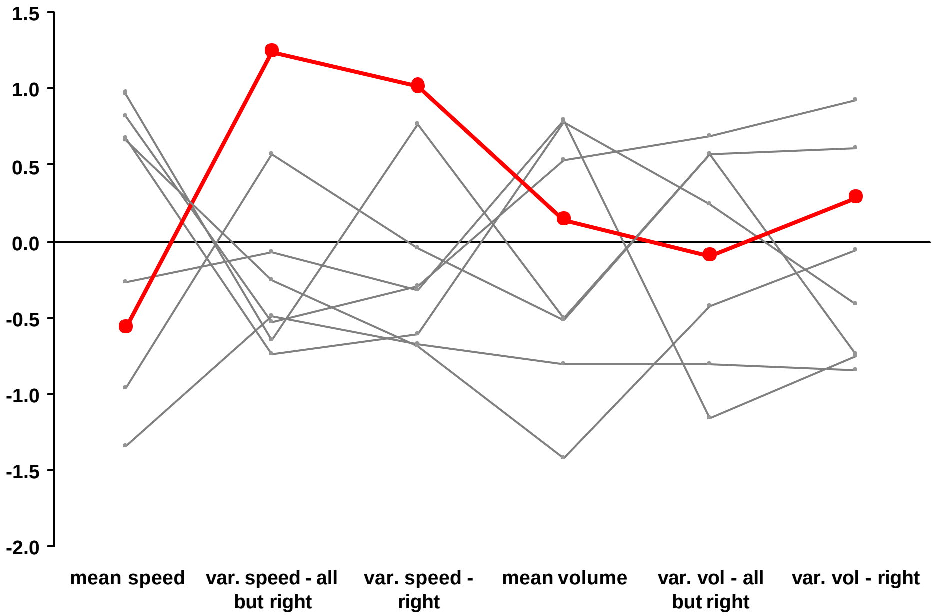

Figure 11 Means for Regime D5 (in bold) in standard deviations from grand mean) – Daylight, Dry Roads

Regime D6 (Figure 12) comprises moderately high volumes and the highest volume variances, particularly in the right-lane. Regime D6 has high mean speeds and relatively low temporal variances of speed, particularly in the right and interior lanes. This Regime identifies “heavy variable flow.”

The cluster means for the Regime D7 (Figure 13) reveal that this Regime represents traffic flow that is both high-volume and high-speed, with low variances of speed in all lanes and near-average variances of volume. We can label Regime D7 as “heavy, steady flow.”

Finally, Regime D8 (Figure 14) is also indicative of very high flow with relatively low variances in flow in all lanes. Speed is about the overall average for all clusters, with average variation in speed. These conditions reflect a freeway operating at near maximum capacity.

Table 9 summarizes the relative levels of the standardized cluster means for the eight traffic flow Regimes. There is a clear distinction between the Regimes in terms of the six traffic flow characteristics. The traffic flow conditions defining the Regimes are interpreted in Table 10.

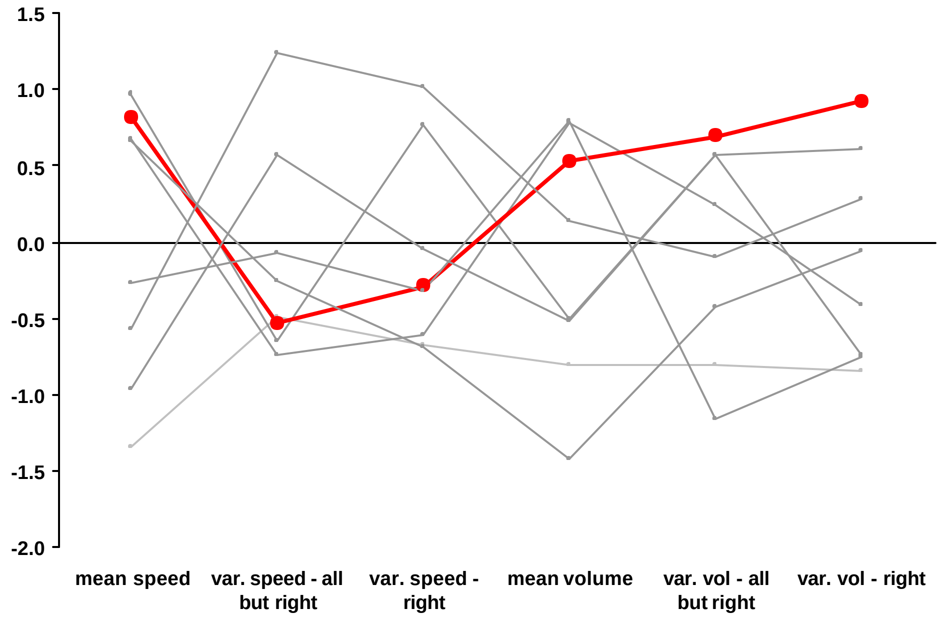

Figure 12 Means for Regime D6 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

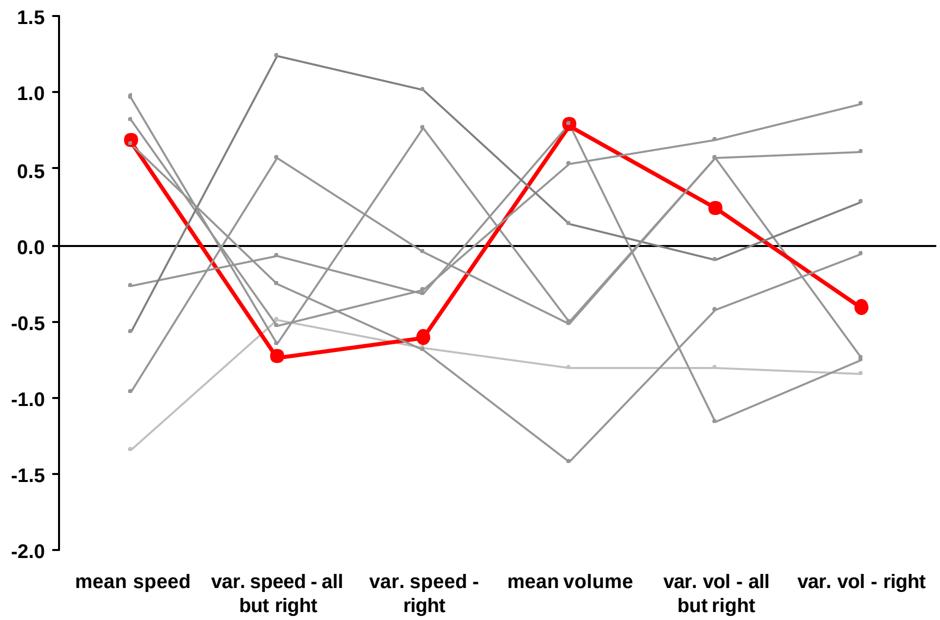

Figure 13 Means for Regime D7 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

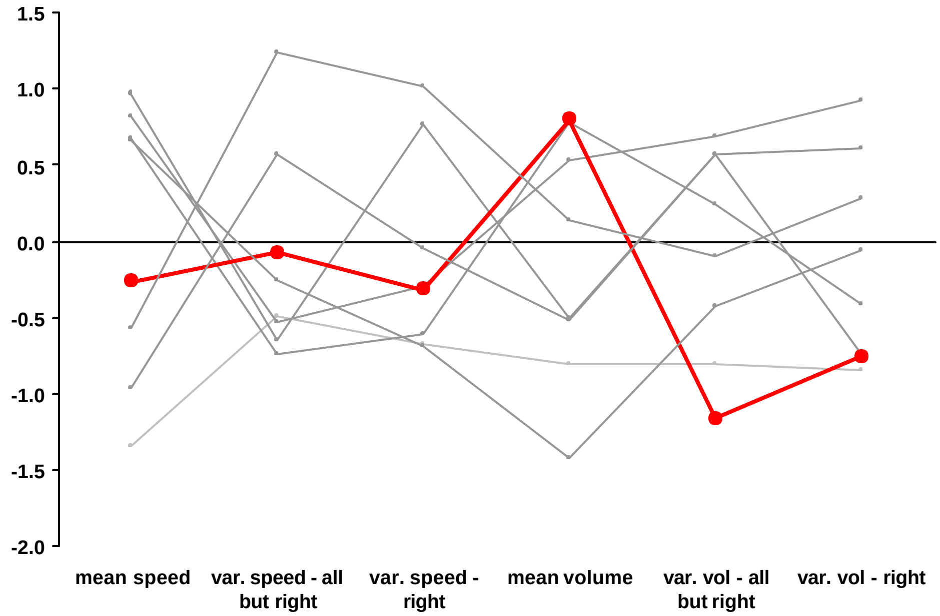

Figure 14 Means for Regime D8 (in bold) in standard deviations from grand mean – Daylight, Dry Roads

Table 9 Summary of Traffic Flow Variables for Each of the Eight Traffic Flow Regimes – Daylight Crashes on Dry Roads

Regime	Mean speed	Var. speed all but right	Var. speed right	Mean volume	Var. volume all but right	Var. volume right
D1	**High**	**Slightly low**	**Very low**	**Very low**	**Low**	**Average**
D2	**Lowest**	**Low**	**Very low**	**Low**	**Very low**	**Very low**
D3	**Very Low**	**High**	**Average**	**Slightly low**	**Very high**	**Very high**
D4	**Highest**	**Low**	**High**	**Slightly low**	**Very high**	**Very low**
D5	**Low**	**Very high**	**Very high**	**Slightly high**	**Average**	**Slightly high**
D6	**High**	**Low**	**Low**	**High**	**Highest**	**Highest**
D7	**High**	**Lowest**	**Very Low**	**Very High**	**High**	**Low**
D8	**Slightly low**	**Average**	**Low**	**Very High**	**Lowest**	**Very low**

Table 10 Summary of the Eight Traffic Flow Regimes, in Order of Average Volumes (from lowest to highest) – Daylight, Dry Roads

Regime	Traffic flow conditions
D1	**Light free-flow:** Very low volume, high average speed, low variance of speed in the right lane and about average variances of speed in the other lanes.
D2	**Heavily congested flow:** Low volume and very low speeds. Low variances of volume in all lanes. Low variance of speed, particularly in right lane.
D3	**Congested flow:** Moderately low average volume and low average speed. High variances in volumes and high variance in speeds except for the right lane.
D4	**Light, right-variable flow:** High mean speeds and moderately low mean volumes. Left and interior lanes free-flowing, but right lane speed variance high and volume variance low.
D5	**Flow at capacity:** Very high variances in speed, average volumes and variances in volume, and moderately low average speeds.
D6	**Heavy, variable flow:** Very high volume variances, particularly in the right-lane, and moderately high volumes. High mean speeds and relatively low speed variances.
D7	**Heavy, steady flow:** High volume and high mean speed, with low temporal variances of speed on all lanes and near-average volume variances.
D8	**Flow near capacity:** High volume, and low volume variances. Speed and speed variations about average to moderately below average.

4.2.4 Crash Typology Explained by Traffic Flow Regime

Nonlinear canonical correlation analysis (NLCCA) of the 8-category traffic Regime variable versus the three crash characteristics shows how the traffic flow Regimes are related to patterns of crash characteristics. Another way to view the problem is to ask how the crash characteristics distinguish among traffic flow Regimes. Indeed, NLCCA with a single categorical (segmentation) variable in one set is equivalent to nonlinear (nonparametric) discriminant analysis. The relationships between the traffic flow Regimes and the categories of each of the three crash characteristics is captured graphically by a joint plot of the locations of the category centroids of each variable in the space of the canonical variates. These are graphed in Figures 15 through 17.

Focusing first on the locations of the traffic Regimes in the two-dimensional space of the canonical variates, which is constant in Figures 15 through 17, we see that the first canonical variate, the x-dimension in these Figures, captures primarily (negative) mean speed, and secondarily flow. In the negative domain of the first variate, the Regimes are ordered from low to high in terms of decreasing mean speed in the middle lane (D4, then D1, then D7 and D6). The four Regimes that score in the positive domain of the first variate are more similar to one another; they all represent heavy traffic, and their ordering from low to high is according to mean volume, rather than mean speed. The first dimension captures aspects of the density (concentration) dimension of the fundamental diagram of traffic flow versus traffic density (Prigogine and Herman, 1971).

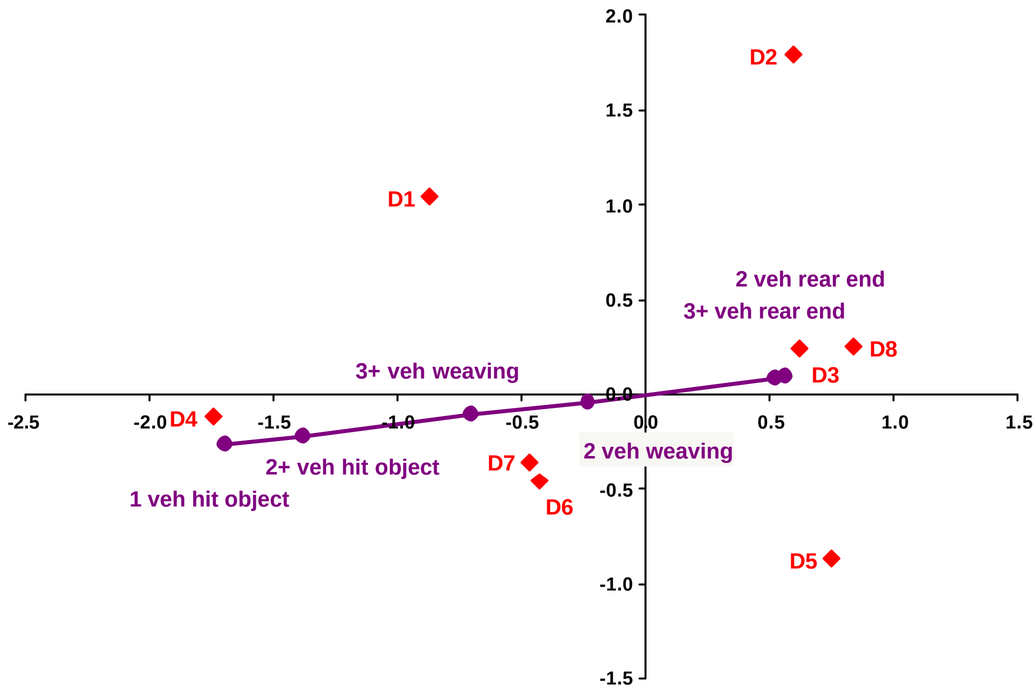

Figure 15 Category Centroids for the Traffic Flow Regime and Collision Type Variables – Daylight, Dry Roads

The second canonical variate, which is independent of the first in terms of its functional relationships with the two sets of variables, primarily distinguishes high-flow Regimes D5, D6 and D7, from low-flow Regimes D2, and D1. This dimension captures, to a considerable degree, the flow dimension of the fundamental diagram.

The relationship between traffic flow Regime and crash type is depicted in Figure 15. Collision type is almost entirely explained by the first canonical variate, which resembles the density dimension of the fundamental diagram. The optimal scaling of the crash type categories contrasts hit-object versus rear-end crashes, with weaving crashes in between. Thus, as expected, rear-end crashes are associated with high density traffic, and hit-object crashes are associated with low density traffic. Weaving crashes (sideswipes and rear-ends caused by lane-change maneuvers) are associated with intermediate density traffic.

High-density Regimes D8, D3 and D5 are most associated with rear-end crashes, while low-density Regimes D4 and D1 are associated with hit-object crashes. Intermediate-density Regimes D6 and D7 are most associated with crashes involving weaving maneuvers.

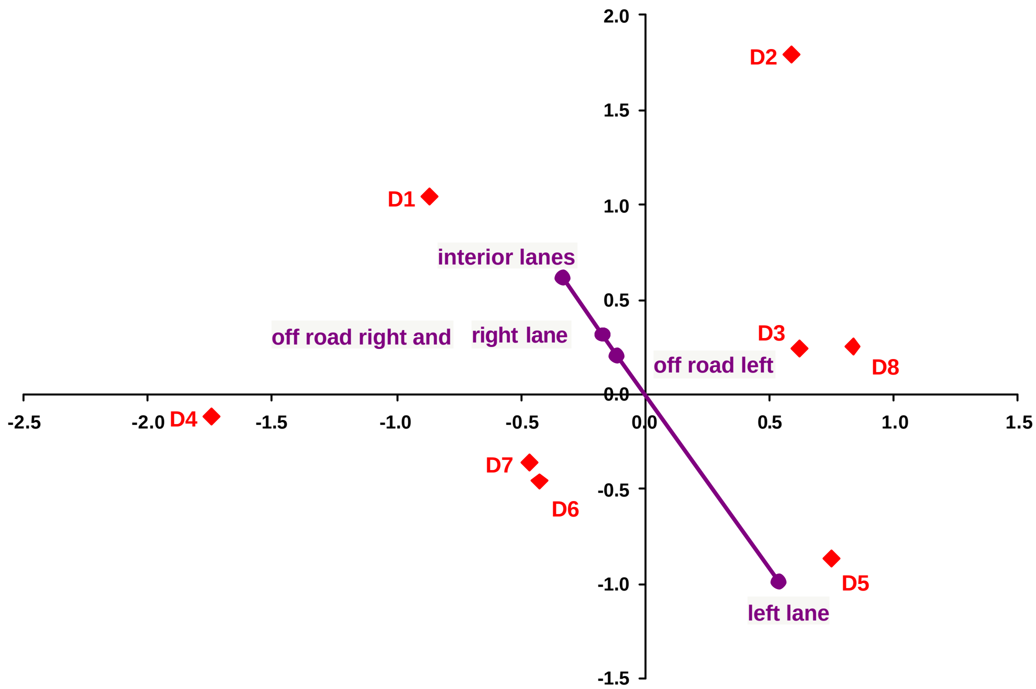

Figure 16 Category Centroids for the Traffic Flow Regime and Collision Location Variables – Daylight, Dry Roads

Collision location (Figure 16) is explained by both dimensions, with the second canonical variate being stronger. We can interpret this to mean that collision location is primarily a flow phenomenon, and secondarily a density phenomenon. The optimal scaling of the categories of the location variable shows that left-lane crashes are associated with high density and high flow conditions, while other locations, especially interior lane crashes, are associated with low density and low flow conditions. Regime D5 is associated with left lane crashes, while Regimes D1 and D4 are associated with off-road crashes.

Finally, crash severity (Figure 17) also is explained by both dimensions, on an approximately equal basis. Thus, the difference between property-damage and injury crashes is a function both of flow and density. Injury crashes are more likely to occur in lower density conditions, and in higher flow conditions. Regimes D2 and D4 have the most extreme projections onto the vector defined by the category quantifications of the severity variable. Thus, the NLCCA model predicts that Regime D4 will have a higher proportion of injury crashes, and Regime D2 will have a higher proportion of property-damage-only crashes.

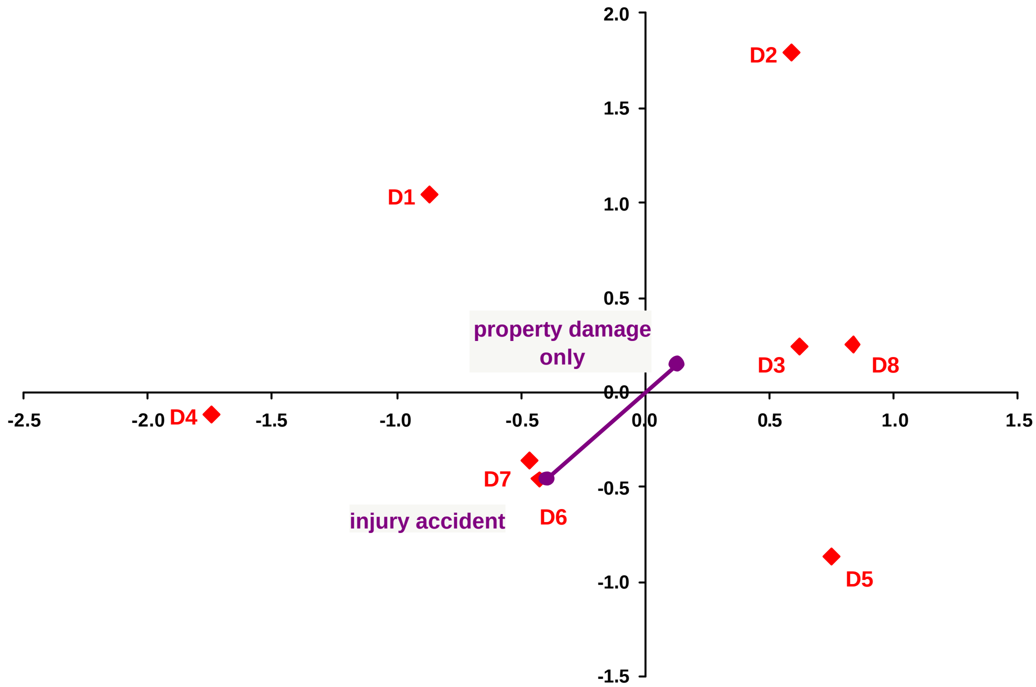

Figure 17 Category Centroids for the Traffic Flow Regime and Crash Severity Variables – Daylight, Dry Roads

The results of the NLCCA model were verified and refined by cross-tabulating each crash characteristic against the eight-category Regime segmentation variable. The results were consistent. Table 11 summarizes the main results of the combined NLCCA and cross-tabulation analyses.

Table 11 Typology of Crashes Occurring During Each of the Eight Traffic Flow Regimes – Daylight, Dry Roads

Regime	Collision types		Collision locations		Crash severity	
	More likely	Less likely	More likely	Less likely	More likely	Less likely
D1	**1 veh hit obj**	**3+ veh rear end**	**off road right**	**Left lane**		
D2		**1 veh hit obj**			**Property damage only**	**injury**
D3	**2 veh rear end**	**1 veh hit obj**				
D4	**Any hit obj**	**Any rear end**	**off road**	**left lane**	**Injury**	**Property damage only**
D5	**3+ veh rear end**	**Any weaving** **1 veh hit obj**	**left lane**	**off road right**		
D6						
D7	**1 veh hit obj** **3+ veh weaving**					
D8	**2 veh rear end**	**any hit obj**		**off road right**		

4.2.5 Summary of Results for Daylight, Dry Road Conditions

The traffic flow conditions that define the eight traffic flow Regimes for daylight, dry road conditions are summarized in Table 12, along with the crash typology.

Table 12 Summary of the Eight Traffic Flow Regimes – Daylight, Dry Roads

	Traffic flow characteristics	Most likely types of crashes
D1	**Light free-flow:** Very low volume, high average speed, low variance of speed in the right lane and about average variances of speed in the other lanes.	**Right-side run-offs**: Single-vehicle hit-object crashes. More off-road right (fewer left-lane crashes).
D2	**Heavily congested flow:** Low volume and very low speeds. Low variances of volume in all lanes. Low variance of speed, particularly in right lane.	**Multi-vehicle:** All types of property damage crashes, except single-vehicle hit-object crashes. Fewer injury crashes.
D3	**Congested flow:** Moderately low average volume and low average speed. High variances in volumes and high variance in speeds except for the right lane.	**Two-vehicle rear-ends:** Rear-end crashes, especially those with two vehicles.
D4	**Light, right-variable flow:** High mean speeds and moderately low mean volumes. Left and interior lanes high speed; right lane speed variance high and volume variance low.	**Serious Run-offs:** Hit-object, off-road, and injury crashes more likely (fewer left-lane crashes and less rear-ends).
D5	**Flow at capacity:** Very high variances in speed, average volumes and variances in volume, and moderately low average speeds.	**Left lane rear-ends:** Rear-end crashes, especially those with 3 or more vehicles, more left lane crashes.
D6	**Heavy, variable flow:** Very high volume variances, particularly in the right-lane, and moderately high volumes. High mean speeds and relatively low speed variances.	**Mixed:** No prevailing types.
D7	**Heavy, steady flow:** High volume and high mean speed, with low temporal variances of speed on all lanes and near-average volume variances.	**Lateral navigation:** Single-vehicle hit-object crashes and 3+ vehicle crashes caused by weaving.
D8	**Flow near capacity:** High volume, and low volume variances. Speed and speed variations about average to moderately below average.	**Two-vehicle rear-ends:** Rear-end crashes, especially those with two vehicles.

4.3 Crashes During Darkness on Dry Roads

4.3.1 Principal Components Analysis of the Traffic Flow Variables

A similar series of analyses was performed for all crashes that occurred during darkness on dry roads. Here six factors account for 87.7% of the variance in the twelve original traffic flow variables, which is a slightly more effective solution than for daylight crashes. The factor loadings and breakdown of the explained variance are listed in Table 13. The factor structure is similar to that found for daylight conditions on dry roads (Table 6). The variables chosen to represent the factors are the same as before (Table 7), and the factor loadings for these variables are underlined and in bold in Table 13.

Table 13 Rotated PCA Loadings for Traffic Flow Variables for Nighttime Crashes on Dry Roads (showing only absolute values > 3.0)

	Traffic flow variable	Principal component 1	2	3	4	5	6
	Percentage of original variance accounted for	21.4%	19.9%	15.6%	15.1%	7.9%	7.8%
Block 1	Median volume/occupancy left lane		0.872				
	Median volume/occupancy interior lane		**0.874**				
	Median volume/occupancy right lane		0.854				
Block 2	Variation in volume/occupancy left lane			0.828			
	Variation in volume/occupancy interior lane			**0.909**			
	Variation in volume/occupancy right lane			0.493			**0.816**
Block 3	Mean volume left lane	**0.912**					
	Mean volume interior lane	0.945					
	Mean volume right lane	0.786				0.303	-.329
Block 4	Variation in volume left lane				**0.925**		
	Variation in volume interior lane				0.844	0.327	
	Variation in volume right lane				0.397	**0.832**	

4.3.2 Cluster Analysis in the Space of Six Traffic Flow Variables

Once again we clustered the crashes in the space of the six variables. The optimal number of clusters is six clusters, based on the internal clustering criteria (Figure 18) and the explanation of crash characteristics (Figure 19). Sample size limitations prevented using more than nine clusters.

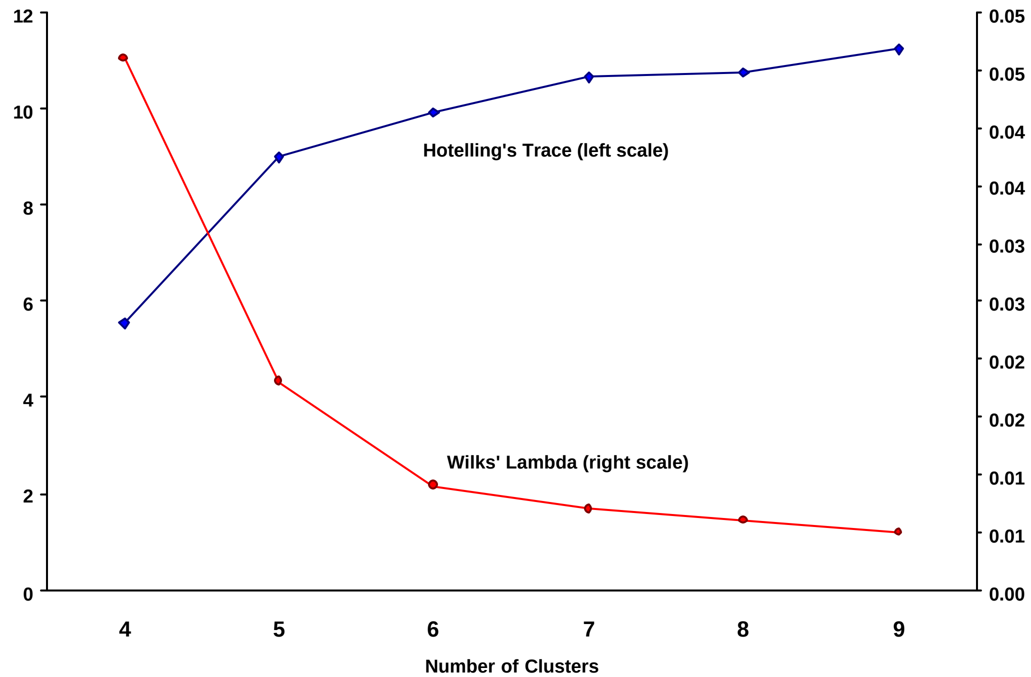

Figure 18 Clustering Criteria by Number of Clusters – Nighttime, Dry Roads

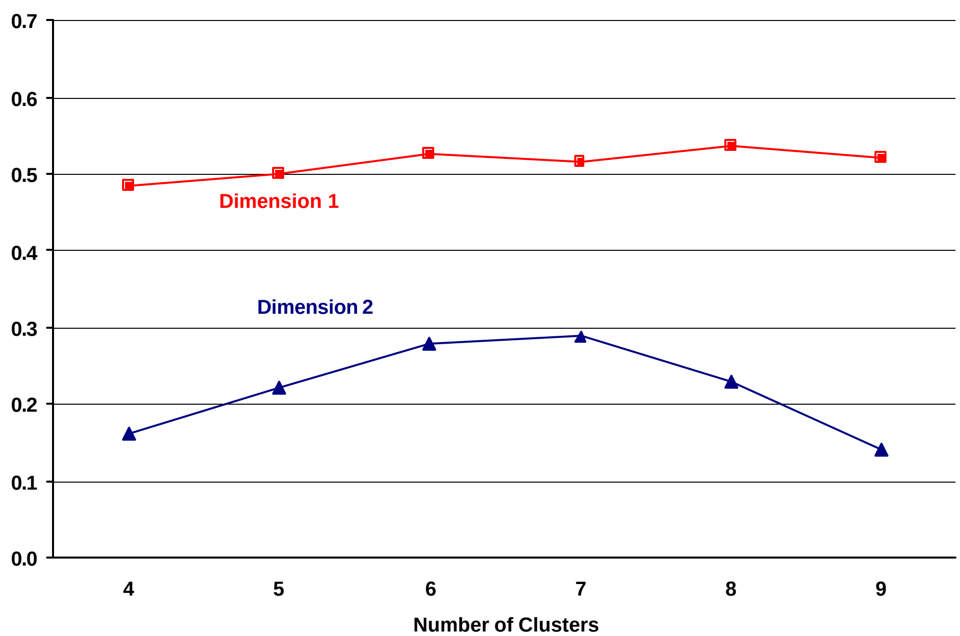

Figure 19 Canonical Correlations for Two-dimensional Nonlinear Discriminant Analysis Solutions for Different Number of Clusters – Nighttime, Dry Roads

The distribution of the 217 dry-night crashes over the six Regimes is as shown in Table 14. The flow characteristics of these Regimes are described in the next Section. The Regimes are numbered in order of increasing average volume.

Table 14 Sample Distribution Among the Six Traffic Flow Regimes – Nighttime, Dry Roads

Regime	Membership	% of all crashes during darkness on dry roads
N1	49	22.6
N2	47	21.7
N3	23	10.6
N4	30	13.8
N5	32	14.7
N6	36	16.6

4.3.3 Description of the Six Dry-Night Traffic Flow Regimes

Table 15 summarizes the relative levels of the standardized cluster means for the six traffic flow Regimes for dry-night conditions. Once again, there is a clear distinction among the Regimes in terms of the six traffic flow characteristics. Regime N1 has very low mean volume and low variances of volumes, with high speed and high variances of speeds. Regime N2 has moderately high speed and moderately low volumes and below average or average variances of speeds and volumes. Regime N3 is characterized exclusively by very low mean speeds, a slightly greater than average volume, low variances of speed and average variances of volume. Regime N4 is characterized by low mean speed, a high variance of speed in the interior lanes, and high mean volume and variances of volumes. Regime N5 has high mean speed and low speed variances, but high mean volume and high variances of volumes. Finally, Regime N6 is characterized exclusively by very high mean volume, while exhibiting slightly lower than average values of all other variables.

The traffic flow conditions defining the six dry-night Regimes are interpreted in Table 16.

Table 15 Summary of Traffic Flow Variables for Each of the Six Traffic Flow Regimes – Nighttime, Dry Roads

Regime	Mean speed	Var. speed all but right	Var. speed right	Mean volume	Var. volume all but right	Var. volume right
N1	High	High	High	Very low	Low	Low
N2	High	Low	Low	Low	Average	Low
N3	Low	Low	Low	Average	Average	Average
N4	Low	High	Slightly high	High	High	High
N5	High	Low	Low	High	Very high	Very high
N6	Slightly low	Low	Slightly low	Very high	Slightly low	Slightly low

Table 16 Summary of the Six Traffic Flow Regimes, in Order of Average Volume (from lowest to highest) – Nighttime, Dry Roads

Regime	Traffic flow characteristics
N1	**Very low volume free-flow:** Very low average flow and low variances in flow. High average speed and high variances in speed on all lanes.
N2	**Low volume free-flow:** High mean speed and moderately low speed variances. Moderately low flow and low variance of flow in right lane.
N3	**Conservative nighttime driving:** Low average speed. Low variances of speed. Average flow (for periods of darkness) and average variances of flow.
N4	**Sporadically congested flow:** Low average speed. High variances of speed in interior lanes. Moderately high flow (for periods of darkness) and high variances of flow in all lanes.
N5	**Heavy, variable flow:** High flow and very high variances of flow in all lanes. Moderately high mean speed and low variance of speeds.
N6	**Flow near capacity:** Very high volume. Slightly below average mean speed and speed variations. Also slightly below average variations in volumes.

4.3.4 Nighttime, Dry Road Crash Typology Explained by Traffic Flow Regime

Nonlinear canonical correlation analysis (NLCCA) of the 6-category traffic Regime variable versus the three crash characteristics again shows how the traffic flow Regimes are related to patterns of crash characteristics. A two-dimensional NLCCA solution yielded canonical correlations of 0.526 for the first canonical variate and 0.278 for the second variate. The optimally scaled category centroids are plotted in Figures 20 through 22.

We can interpret the two canonical variates (dimensions) based on the positions of the six traffic flow Regimes (which are constant in Figures 20 though 22). The most important canonical variate, the x-dimension, primarily contrasts Regimes N6 and N4 against Regime N1. It is consistent with the flow dimension of the fundamental diagram. The y-dimension, which primarily distinguishes Regime N3 from all other Regimes, is consistent with the density (concentration) dimension of the fundamental diagram. These two dimensions are similar to the canonical variates found for daylight, dry conditions, but they are reversed in terms of explanatory power. Density is more important than flow in explaining the effects of traffic on the types of crashes that occur during the day on dry roads, while flow is more important than density in explaining the effects of traffic on the types of crashes that occur at night on dry roads.

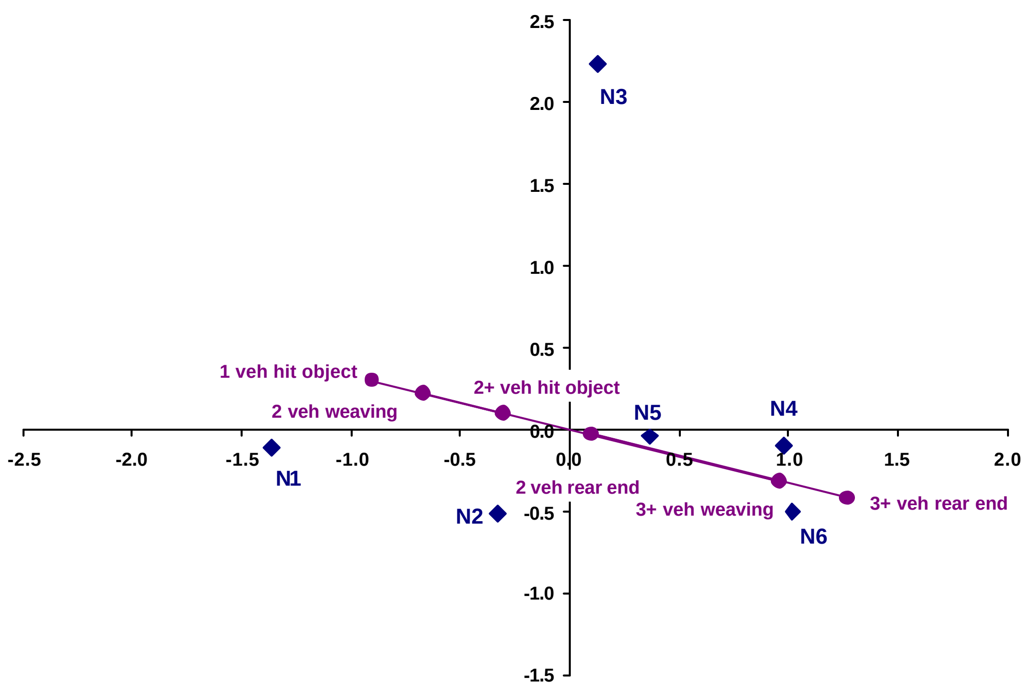

Figure 20 Category Centroids for the Traffic Flow Regime and Collision Type Variables – Nighttime, Dry Roads

The relationship between traffic flow Regime and crash type is depicted in Figure 20. As in the case of dry, daylight conditions (Figure 15), collision type is primarily explained by the first canonical variate, which in this case is consistent with the flow dimension of the fundamental diagram. We found previously that collision type for daylight conditions (Figure 15) was explained more by the canonical variate that was associated with traffic density. The optimal scaling of the collision type variable is also different for day and night. For nighttime conditions, the optimal scaling of the crash type categories contrasts single-vehicle hit-object (low flow) versus three-plus vehicle rear-end and weaving crashes (high flow), with two-vehicle crashes in between (Figure 20). The scaling for nighttime conditions is based more on the number of vehicles involved in the collision. For daylight conditions, the optimal scaling of type categories contrasts hit-object (low density) versus rear-end (high density) crashes, with weaving crashes in-between (Figure 15). The scaling for daylight conditions is based more on kind of the collision, rather than the number of vehicles involved. Three-or-more-vehicle crashes are associated with Regimes N4 and N6, while single-vehicle crashes are associated with Regime N1.

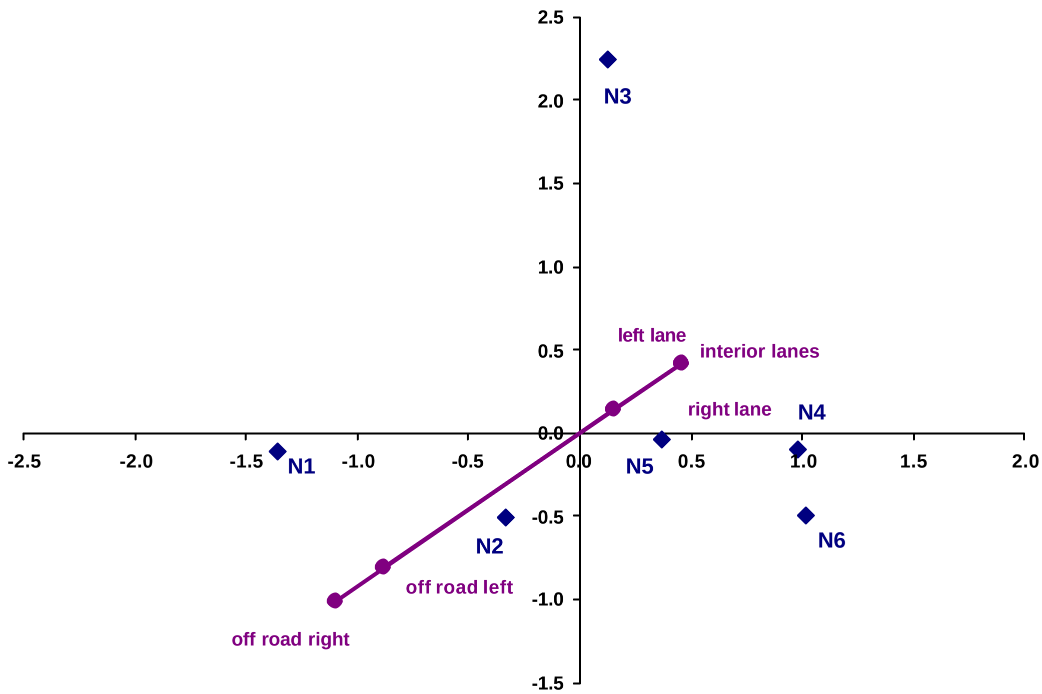

Figure 21 Category Centroids for the Traffic Flow Regime and Collision Location Variables – Nighttime, Dry Roads

Collision location (Figure 21) is explained by both dimensions, with the first canonical variate being stronger. Thus, collision location is primarily a flow phenomenon, and secondarily a density phenomenon. This is the same result found for daylight conditions (Figure 16). The optimal scaling of collision location gives negative values to off-road right, followed by off-road left; left-lane and interior lane locations are scaled with equal positive values, and right-lane locations are close to the middle of the scale. Thus, at night on dry roads, the biggest difference in the effects of traffic flow on crash location is between off-road versus left-lane or interior-lane locations. Off-road crashes are more likely to occur at night in lower density and lower volume conditions. Regimes N1 and N2 are associated with off-road crashes, while Regimes N3 and N4 are associated with left- and interior-lane crashes. For daylight conditions (Figure 16), the optimal scaling of the location variables primarily contrasts left-lane crashes versus all other locations, with interior lane being most different from left-lane.

Finally, crash severity is also explained by both dimensions, and the alignment of the severity vector in Figure 22 is similar to the alignment of the location vector in Figure 21. Thus, the difference between property-damage and injury crashes is both a function of flow and density, in the same way that collision location is. Injury crashes are more likely to occur in lower density and lower flow conditions. Regime N2 is predicted by the NLCCA model to have a higher proportion of injury crashes, while Regime N4 is predicted to have a higher proportion of property-damage-only crashes. This result is different than for daylight conditions (Figure 17), where injury crashes were found to be associated with *higher* flow and lower density conditions.

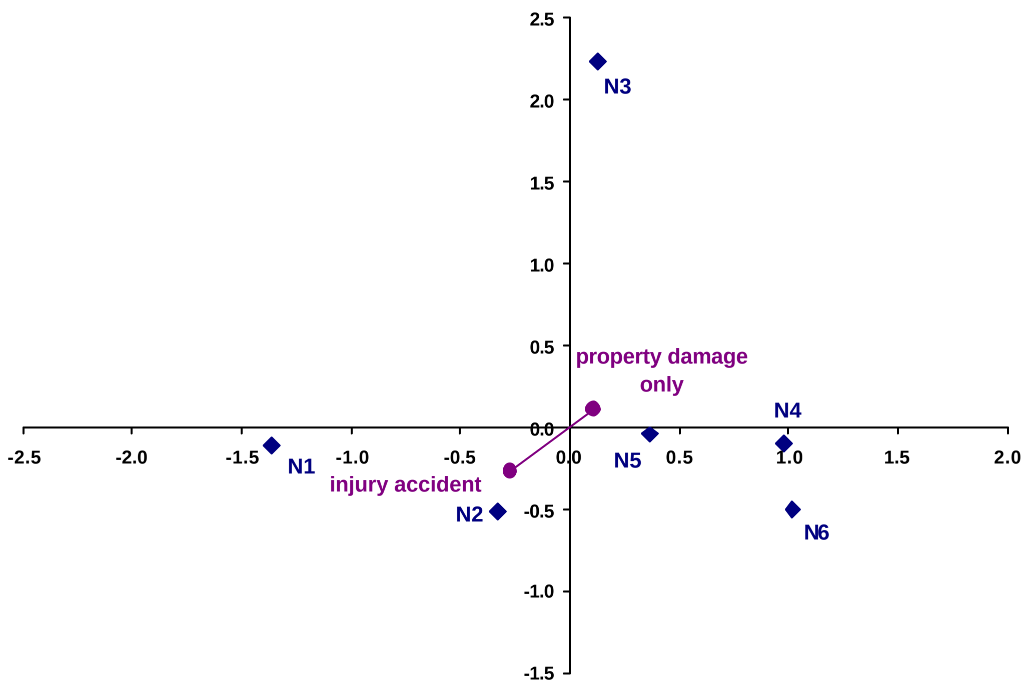

Figure 22 Category Centroids for the Traffic Flow Regime and Crash Severity Variables – Nighttime, Dry Roads

As before, the results of the NLCCA model were verified and refined by cross-tabulating each crash characteristic against the eight-category variable that segments crashes according to traffic flow Regime. The results were once again found to be largely consistent. Table 17 summarizes the main results of the combined NLCCA and cross-tabulation analyses for nighttime, dry road conditions.

Table 17 Typology of Crashes Occurring During Each of the Six Traffic Flow Regimes – Nighttime, Dry Roads

Regime	Collision types		Collision locations		Crash severity	
	More likely	Less likely	More likely	Less likely	More likely	Less likely
N1	**1 veh hit obj**	**any rear end**	**off road right**	**left lane**		
N2			**off road left**		**injury**	**property damage only**
N3	**2 veh weaving 2 veh rear end**	**1 veh hit obj**	**Interior lanes**	**off road**		
N4	**3+ v. rear end 3+ veh weaving**	**1 veh hit obj**	**left lane**	**off road right**	**property damage only**	**injury**
N5						
N6	**3+ v. rear end**	**any hit obj**	**right lane**	**off road right**	**property damage only**	**injury**

4.3.5 Summary of Results for Nighttime, Dry Road Conditions

The traffic flow conditions and associated crash typology for the six traffic flow Regimes for nighttime, dry road conditions are summarized in Table 18.

Table 18 Summary of the Six Traffic Flow Regimes – Nighttime, Dry Roads

	Traffic flow characteristics	Most likely types of crashes
N1	**Very low volume free-flow:** Very low average flow and low variances in flow. High average speed and high variances in speed on all lanes.	**Right-side run-offs**: Single-vehicle hit-object crashes. More off-road right crashes. Fewer left-lane and rear-end crashes.
N2	**Low volume free-flow:** High mean speed and moderately low speed variances. Moderately low flow and low variance of flow in right lane.	**Left-side run-offs:** Off-road left and injury crashes more common.
N3	**Conservative nighttime driving:** Low average speed. Low variances of speed. Average flow (for periods of darkness) and average variances of flow.	**Two-vehicle interior- lane crashes:** Two-vehicle sideswipes and rear-ends in interior lanes more common. Fewer hit-object off-road crashes.
N4	**Sporadically congested flow:** Low average speed. High variances of speed in interior lanes. Moderately high flow (for periods of darkness) and high variances of flow in all lanes.	**Many-vehicle left-lane crashes:** Three-plus-vehicle sideswipe and rear-end crashes. Fewer single-vehicle hit-object off-road-right crashes.
N5	**Heavy, variable flow:** High flow and very high variances of flow in all lanes. Moderately high mean speed and low variance of speeds.	**Mixed:** No prevailing types.
N6	**Flow near capacity:** Very high volume. Slightly below average mean speed and speed variations. Also slightly below average variations in volumes.	**Large right-lane rear-end crashes:** 3-plus-vehicle rear-ends, especially in right lane. Off-road and injury crashes slightly less likely.

4.4 *Wet Road Crashes*

4.4.1 Principal Components Analysis

A third and final series of analyses was performed for all crashes that occurred on wet roads. Results show that the correlation structure among the twelve traffic flow variables is nearly identical for the three weather and lighting conditions. Here six factors account for 87.1% of the variance in the twelve original traffic flow variables, versus 86.8% and 87.7% for dry-daylight (Table 6) and dry-nighttime (Table 13), respectively. The factor loadings and breakdown of the explained variance are listed in Table 19. The factor structure is essentially the same as found previously, so the same six variables were chosen to represent the factors (Table 7); the factor loadings for these variables are underlined and in bold in Table 19.

Table 19 Rotated PCA Loadings for Traffic Flow Variables for Wet Road Crashes (showing only absolute values > 3.0)

		Principal component					
	Traffic flow variable	1	2	3	4	5	6
Percentage of original variance accounted for		21.6%	20.0%	15.4%	11.7%	9.9%	8.5%
Block 1	Median volume/occupancy left lane		0.835				
	Median volume/occupancy interior lane		**0.878**				
	Median volume/occupancy right lane		0.886				
Block 2	Variation in volume/occupancy left lane				0.696	0.457	
	Variation in volume/occupancy interior lane				**0.887**		
	Variation in volume/occupancy right lane					**0.922**	
Block 3	Mean volume left lane	**0.906**					
	Mean volume interior lane	0.919					
	Mean volume right lane	0.815					0.358
Block 4	Variation in volume left lane			**0.907**			
	Variation in volume interior lane			0.845			
	Variation in volume right lane			0.366			**0.869**

4.4.2 Cluster Analysis in the Space of Six Traffic Flow Variables

For crashes on wet roads, the optimal number of clusters in the space of the six variables is seven, based on the internal clustering criteria (Figure 23) and the explanation of crash characteristics (Figure 24). The explanation of crash characteristics peaks at seven clusters (Figure 24), and the seven-level cluster is also consistent with a break in Hotelling's Trace criteria.

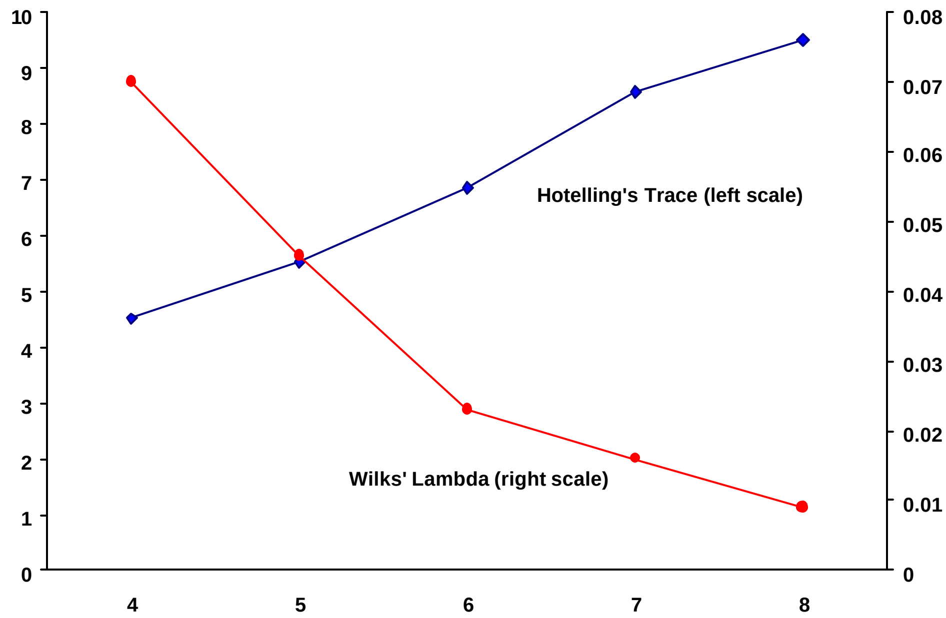

Figure 23 Clustering Criteria by Number of Clusters – Wet Roads

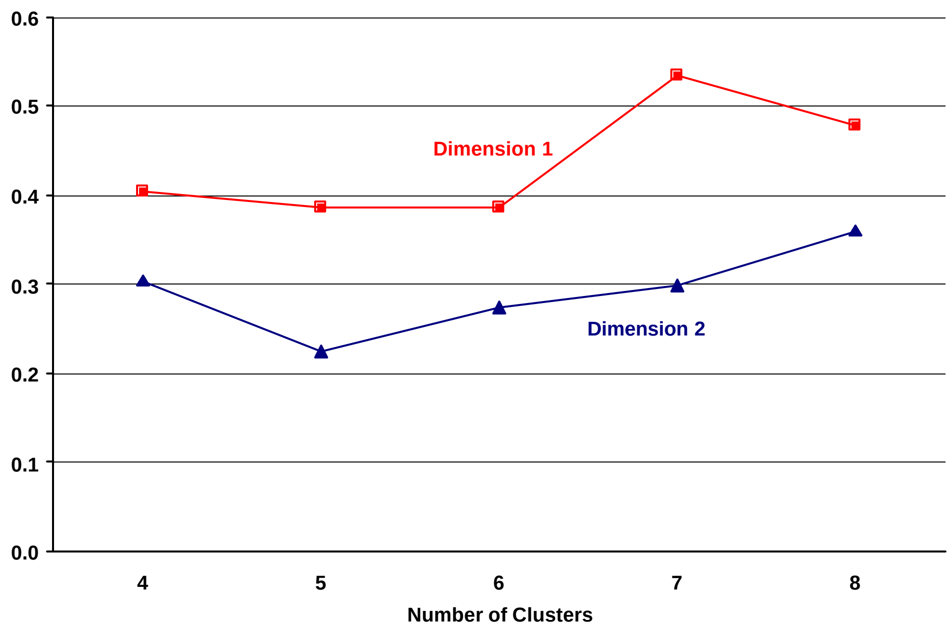

Figure 24 Canonical Correlations for Two-dimensional Nonlinear Discriminant Analysis Solutions for Different Number of Clusters – Wet Roads

The distribution of the 154 wet road collisions over the seven Regimes is shown in Table 20. The flow characteristics of these Regimes are described in the next Section. Once again, the Regimes are labeled in order of increasing average volume.

Table 20 Sample Distribution Among the Seven Traffic Flow Regimes – Wet Roads (N = 154)

Regime	Membership	% of all wet road crashes
W1	26	16.9
W2	22	14.3
W3	27	17.5
W4	26	16.9
W5	13	8.4
W6	15	9.7
W7	25	16.2

4.4.3 Description of the Seven Wet Road Traffic Flow Regimes

Table 21 summarizes the relative levels of the standardized cluster means for the seven traffic flow Regimes for wet road conditions. The Regimes are interpreted in Table 22.

Table 21 Summary of Traffic Flow Variables for Each of the Seven Traffic Flow Regimes – Wet Roads

Regime	Mean speed	Var. speed all but right	Var. speed right	Mean volume	Var. volume all but right	Var. volume right
W1	Slightly low	High	Very high	Very low	Very low	Very low
W2	Slightly high	Average	Slightly low	Low	Slightly low	Low
W3	Slightly high	Slightly low	Average	Average	Slightly high	Average
W4	Average	Average	Low	Average	Slightly high	High
W5	Slightly high	Slightly low	Average	Slightly high	Very high	High
W6	Average	Low	Low	High	High	Slightly high
W7	Low	Slightly low	Slightly low	High	Slightly low	Average

Table 22 Summary of the Seven Traffic Flow Regimes in Order of Average Volume (from lowest to highest) – Wet Road Crashes

Regime	Traffic flow characteristics
W1	**Very low flow, variable speed:** Very low volume and very low variations in volume. Mean speed slightly below average for wet roads. Variations in speed high, especially for right lane.
W2	**Low volume free-flow:** Low average volume and moderately high speed. Low variances in volume and speed in right lane.
W3	**Moderate free-flow:** Moderately high speed and near average volume, volume variances, and speed variances for wet roads.
W4	**Moderate flow with right-lane concentration:** Moderately high speed and near average volume, but high variance of volume and low variance of speed in right lane.
W5	**Heavy, variable flow:** Moderately high speed and volume. Very high variance of volume in left lane and high variance of volume in right lane.
W6	**Very heavy flow:** High volume and average speed with low variances in speed. High volume variances, especially in left lane.
W7	**Flow near capacity:** Low speed and high volume. Average to slightly below average variances of both speeds and volumes.

4.4.4 Wet Road Crash Typology Explained by Traffic Flow Regime

A two-dimensional nonlinear canonical correlation analysis (NLCCA) solution of the seven-category traffic Regime variable for wet roads versus the three crash characteristics yielded canonical correlations between the two sets of variables of 0.532 (first canonical variate) and 0.298 (second canonical variate). The optimally scaled category centroids are plotted in Figures 25 through 27.

The two canonical variates (dimensions) can be interpreted in terms of traffic conditions based on the positions of the seven traffic flow Regimes. The first variate (the x-dimension in Figures 25 through 27) contrasts low volume Regimes (W3, W1 and W2) against the high volume flow Regime with low average speed, W7. Regimes W4, W5 and W6, which have moderate to heavy flows but average to above average mean speeds, score close to zero on this first variate. This variate may be capturing a measure of exposure that is independent of traffic stream speed effects. The second canonical variate contrasts Regime W6, then W2 and W5 (negative scores), against Regime W3, then W7 and W1 (positive). No interpretation for this variate is obvious; perhaps, this is an artifact of the relatively small sample size for this category of environmental conditions. Also, the nature of the environmental conditions is a potentially confounding effect for this particular segmentation. The category is defined by a simple binary variable (wet vs. dry): light rain conditions are indistinguishable from heavy downpours; it also covers the complete spectrum of lighting conditions. So, for example, although it is plausible to interpret that Regime W1 typifies light volume flow conditions under extreme weather (since the ratio of the "very low" volume to occupancy yields inference of only a slightly less-than-average of speed), only the traffic flow variables are directly measurable; any hypothesis regarding causal effects of weather on these variables requires additional environmental information.

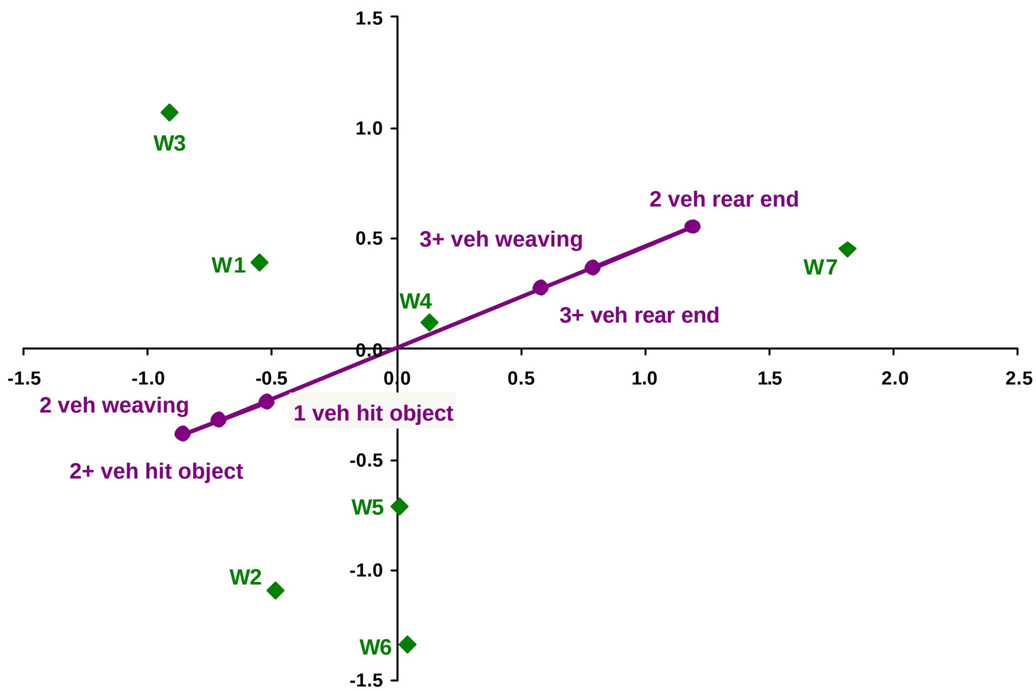

Figure 25 Category Centroids for the Traffic Flow Regime and Crash Type Variables – Wet Road Crashes

The relationship between traffic flow Regime and crash type for wet road conditions is depicted in Figure 25. Crash type is explained by both canonical variates, but more so by the first. The optimal scaling of crash type contrasts hit-object crashes and two-vehicle weaving crashes against two-vehicle rear end crashes. Three-or-more-vehicle collisions are in-between but are more like two-vehicle rear end crashes. Hit-object crashes are more likely in low-volume Regimes W1, W2 and W3. Rear end crashes are more likely in flow conditions approaching capacity, Regime W7

The optimal scaling of crash location is plotted together with the category scores of the Regimes in Figure 26. Location is explained by both canonical variates, and the optimal scaling of the categories is very different from that found for dry conditions. The extreme categories are left lane and off-road left, with off-road right being more similar to left lane than to off-road left.

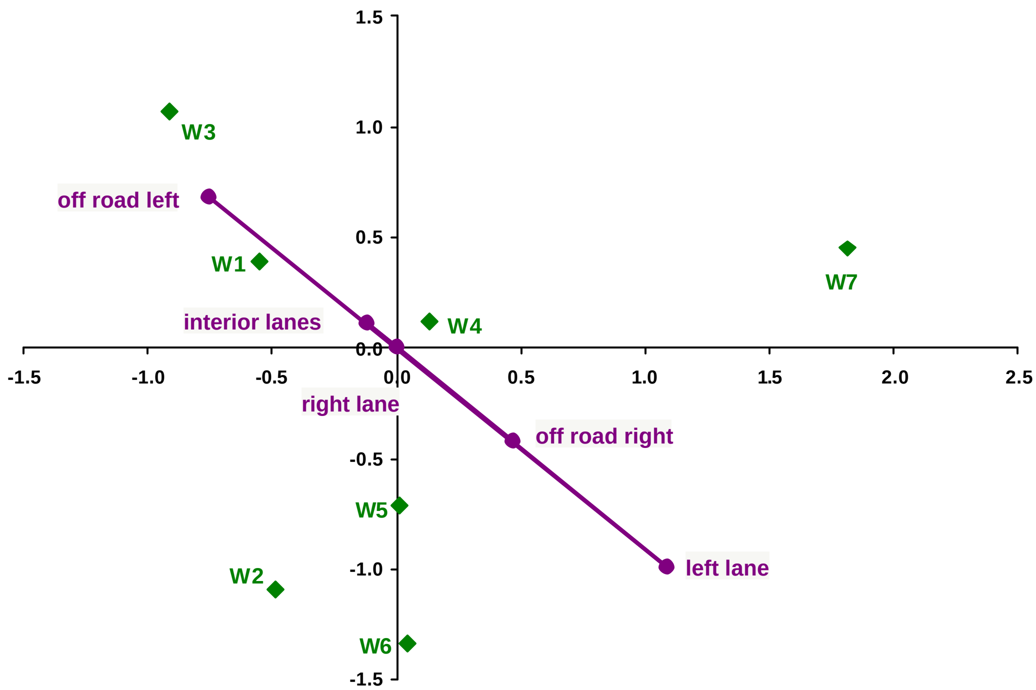

Figure 26 Category Centroids for the Traffic Flow Regime and Crash Location Variables – Wet Road Crashes

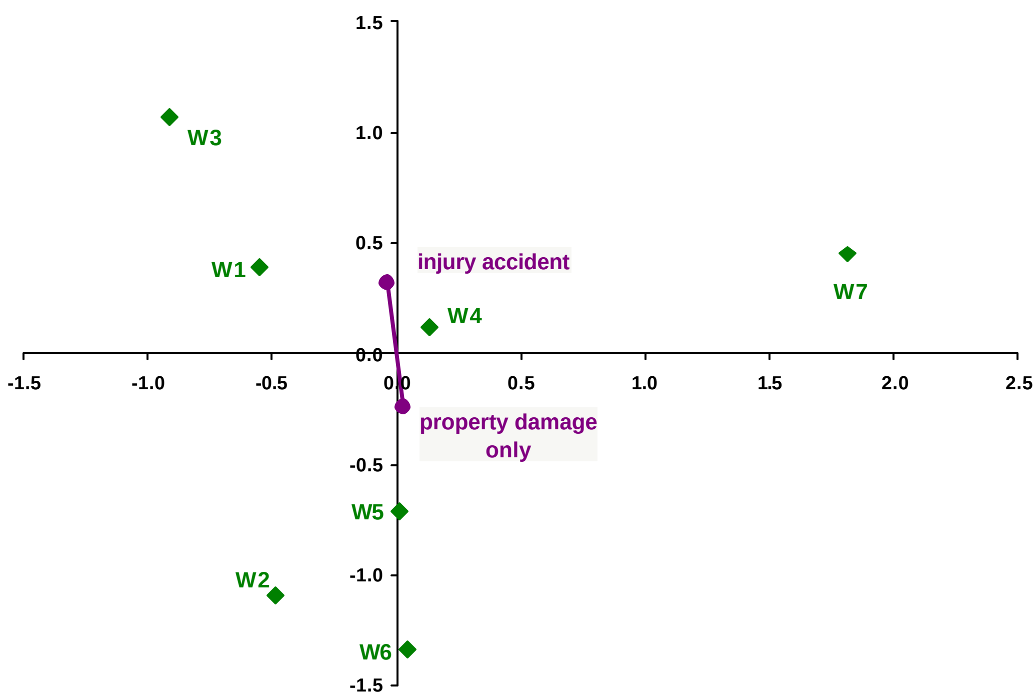

Figure 27 Category Centroids for the Traffic Flow Regime and Crash Severity Variables – Wet Road Crashes

The results of the NLCCA model were verified and refined by cross-tabulating each crash characteristic against the eight-category Regime variable. As in the cases of dry-day and dry-night conditions, the NLCCA and contingency table results were once again found to be consistent. Table 23 summarizes the main results of the combined analyses for wet road conditions.

Table 23 Typology of Crashes Occurring During Each of the xxx Traffic Flow Regimes – Wet Road Crashes

Regime	Collision types		Collision locations		Crash severity	
	More likely	Less likely	More likely	Less likely	More likely	Less likely
W1	1 veh hit object		off road right	left lane right lane		
W2	2+ veh hit object	3+ v. crashes		interior lanes		
W3	1 veh hit object		off road left	left lane		
W4		1 veh hit object		off road left	injury	
W5						
W6						injury
W7	any rear end	1 veh hit object 2 veh weaving	left lane	off road		

4.4.5 Summary of Results for Wet Road Conditions

The traffic flow conditions and associated crash typology for the seven traffic flow Regimes for wet conditions are consolidated in Table 24. We next investigated how the various traffic flow Regimes for all lighting and weather conditions are distributed over time and space on Orange County freeways in calendar year 1998.

Table 24 Summary of the Seven Traffic Flow Regimes – Wet Road Crashes

	Traffic flow characteristics	Most likely types of crashes
W1	**Very low flow, variable speed:** Very low volume and very low variations in volume. Mean speed slightly below average for wet roads. Variations in speed high, especially for right lane.	**Right-side run-offs**: Single-vehicle hit-object crashes. More off-road right. Fewer left- and right- lane crashes.
W2	**Low volume free-flow:** Low average volume and moderately high speed. Low variances in volume and speed in right lane.	**Mixed 1- and 2-vehicle:** All types of collision, with the exception of 3+ vehicle and interior-lane crashes.
W3	**Moderate free-flow:** Moderately high speed and near average volume, volume variances, and speed variances for wet roads.	**Left-side run-offs:** Off-road left crashes more common. Left lane crashes less likely.
W4	**Moderate right-concentrated flow:** Moderately high speed and near average volume, but high variance of volume and low variance of speed in right lane.	**All types of serious multi-vehicle crashes:** Single-vehicle hit-object crashes less likely, especially those off-road left. Injuries more likely.
W5	**Heavy, variable flow:** Moderately high speed and volume. Very high variance of volume in left lane and high variance of volume in right lane.	**Mixed:** No prevailing types.
W6	**Very heavy flow:** High volume and average speed with low variances in speed. High volume variances, especially in left lane.	**Mixed non-injury:** No prevailing types, but Injuries less likely
W7	**Flow near capacity:** Low speed and high volume. Average to slightly below average variances of both speeds and volumes.	**Rear ends:** More rear-end crashes, especially in left lane. Run-offs and two-vehicle sideswipes less likely.

5 Traffic Flow Conditions on Orange County Freeways in 1998

5.1 Drawing of a Random Sample of Locations and Times

To understand where and when the various traffic Regimes were present historically on the six major six Orange County freeways in calendar year 1998, we drew a random sample of 100,000 loop detector stations and times from the complete set of archived VDS 30-second loop detector for the six freeways.[1] Each observation in the sample consisted of 27.5 minutes of 30-second measurements. The date and time of the last observation in the series defines the time signature of each observation.

Of the 100,000 observations, only 11,192 (11.2%) had complete and logically consistent data for volume and occupancy for the right, left, and a selected interior lane for the entire 27.5 minute period (55 30-second intervals). The FITS evaluation tool includes an algorithm to determine the time, for each day of the year, at which dawn begins (0.5 hours prior to sunrise) and the time at which darkness begins (sunset plus 0.5 hours), for the latitude and longitude corresponding to the approximate center of the Orange County freeway network. Also, periods of wet roads were identified from historical records. The breakdown of the 11,193 random observations for 1998 by lighting and weather conditions is as follows. Dry day: 6,615 observations (59.1% of the random sample), dry night: 3,584 observations (32%), and wet roads: 994 observations (8.9%).

Importantly, the missing or suspect loop detector data in the random sample of 100,000 are ***not*** distributed randomly across locations and times. There are systematic patterns in coverage. Certain blocks of time are missing in entirety for the entire Orange County freeway system, and certain loop detector stations are missing for the entire year or for a high proportion of the year. Thus, the patterns described in this Section should not be considered to be representative of all or any one of the six major freeways in Orange County.

First, our sample of useable observations from the random data is easily shown to be spatially biased. There are very few (sometimes none at all) useable data from several adjacent loop detector stations on certain sections of Orange County freeways. Thus the useable data for a specific route do not necessarily represent the entire route because certain sections are underrepresented. Second, there are systematic biases over time. The non-random temporal coverage of the useable data extracted from the archived loop detector data is demonstrated is Figures 27 through 29. If detector or processing system failures and maintenance downtimes were randomly distributed over time, each of the distributions in Figures 27-29 should vary from uniform only by perturbations due to chance.

---

[1] These six freeways are: I-5 (the Santa Ana Freeway and the southern section of the San Diego Freeway in Orange County), SR-22 (Garden Grove Freeway), SR-55 (Costa Mesa Freeway), SR-57 (Orange Freeway), SR-91 (Riverside and Artesia Freeways), and I-405 (the northern section of the San Diego Freeway in Orange County)

The distribution of useable data by month (Figure 27) shows that the probability that randomly selected data would be from a faulty loop detector station increases from January 1998 through August 1998, then decreases until November 1998 and increases again in December 1998. This is presumably due to a maintenance schedule. Consequently, we do not have consistent representation of seasonal and weather effects.

The distribution of useable data by hour of the day (Figure 27) shows that there are less data available for nighttime periods, specifically after 11PM and before 6AM. Finally, the distribution of useable data over days of the week shows that there are relatively less data available on Sundays and Mondays.

Due to our inability to secure a representative random sample of traffic conditions that corresponds to our sample of crashes on Orange County freeways for the calendar year 1998 period, the distributions of traffic Regimes described in the remainder of this Section are used only to illustrate the application of the FITS tool and aid in interpreting the traffic flow Regimes.

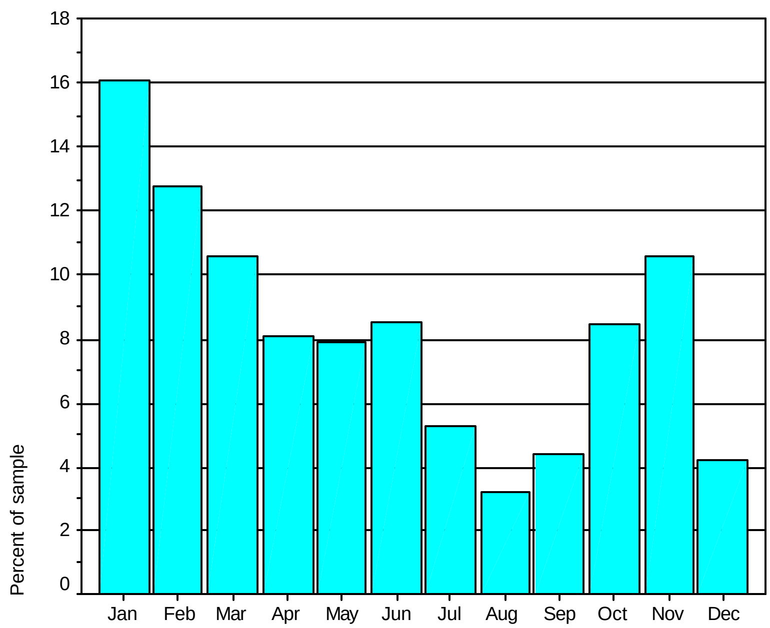

Figure 27 Distribution by Month of 11,192 Useable Observations from the Random Sample of 100,000 Observations of Traffic Conditions on Six Major Orange County Freeways in Calendar Year 1998

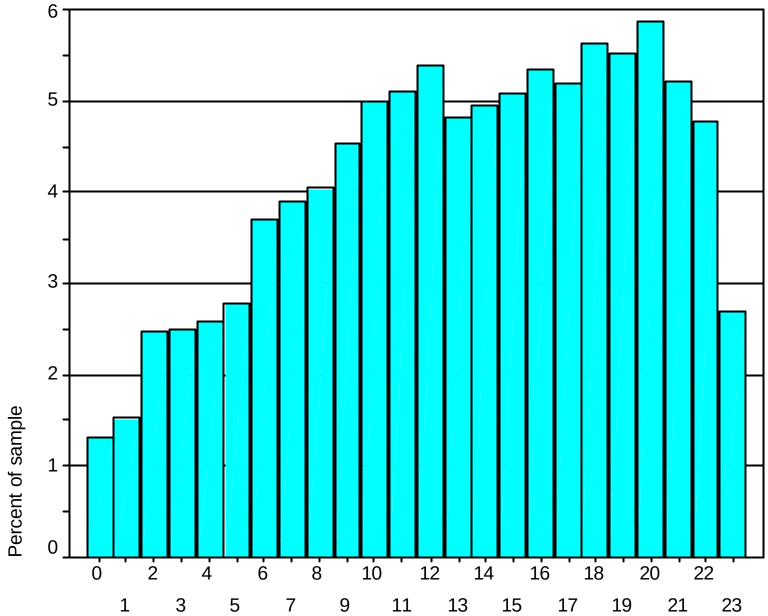

Figure 28 Distribution by Hour of 11,192 Useable Observations from the Random Sample of 100,000 Observations of Traffic Conditions on Six Major Orange County Freeways in Calendar Year 1998

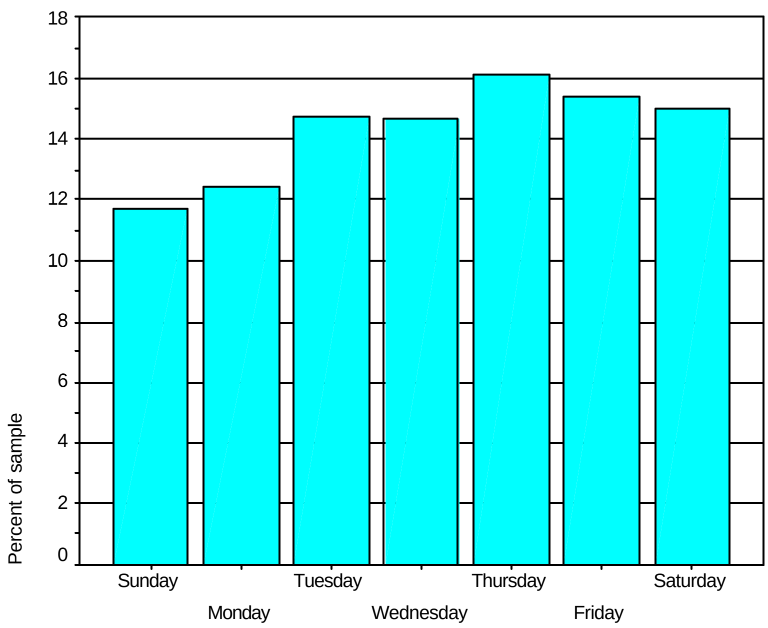

Figure 29 Distribution by Day of 11,192 Useable Observations from the Random Sample of 100,000 Observations of Traffic Conditions on Six Major Orange County Freeways in Calendar Year 1998

5.2 Daylight, Dry Road Conditions on Orange County Freeways in 1998

We first investigated the frequencies of occurrence of the eight dry day traffic Regimes over four weekly time periods: (1) weekday morning peak, 6:00AM to 9AM inclusive, (2) weekday afternoon peak, 3:30 PM to 6:30 PM inclusive, (3) weekday off-peak, and (4) weekends. The classifications, broken down by route, for the AM peak period and the PM peak period, are shown graphically in Figures 30 and 31, respectively. These Figures confirm that all Regimes are possible at any time at a given location. For example, Regime D1 (light free-flow) is possible during peak periods, perhaps due to random perturbations in flow, or as a result of non-recurrent congestion due to an incident or a bottleneck upstream of the detector station, or more likely due to directional flow patterns.

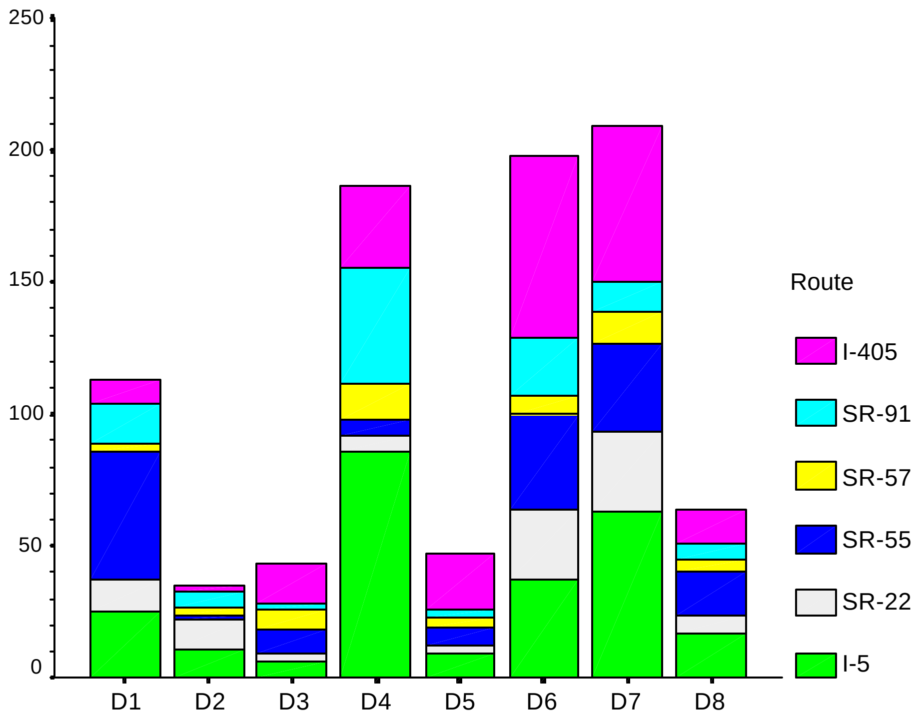

Figure 30 Classification of Traffic Conditions on Each of Six Orange County Freeways During the AM Peak Period (Weekdays 6:00 AM through 9:00 AM) - Daylight, Dry Roads

The most prevalent Regimes during weekday morning peak periods in 1998 were D7 (heavy, steady flow), D6 (heavy variable flow) and D4 (right-concentrated flow). Regime D1 (light free-flow) is also possible at certain times (early in the period) and locations. The two Regimes with heavy flow and the highest speeds – Regimes D7 and D8 – are relatively more common during the afternoon peak, while Regime D6, with slightly lower but more variable flow and Regime D4, another Regime with high variation in flow, are more common during the morning peak.

Investigations of the distributions of the Regimes over specific peak hours of the day (not shown) reveal that Regimes D5 (relatively unstable flow at capacity) and D8 (flow near capacity) occur most predominately between 5PM and 6PM. Each of the other Regimes is distributed rather evenly over the morning and evening peak hours.

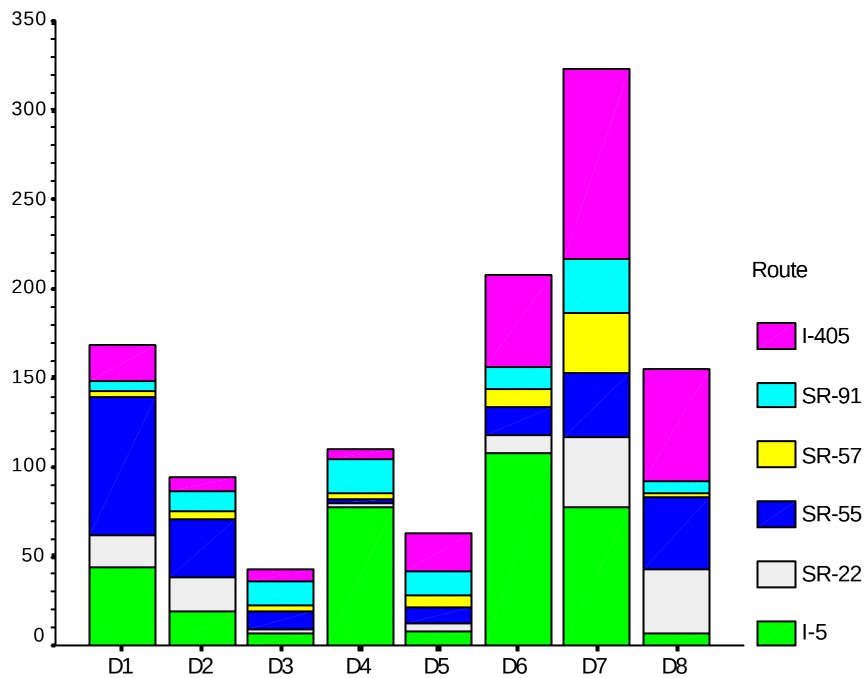

Figure 31 Classification of Traffic Conditions on Each of Six Orange County Freeways During the PM Peak Period (Weekdays 3:30 PM through 6:30 PM) - Daylight, Dry Roads

Were the random sample of these Regimes truly representative of the population of flow regimes occurring on the Orange County freeway system (i.e., were it not for the suspected systematic biases resulting from missing loop data), certain conclusions could be drawn regarding the expected prevalence of specific accident characteristics during the peak periods. For example, Regime D7, the most common Regime during both the morning and evening peak hours, is associated primarily with multi-vehicle crashes precipitated by weaving and single vehicle hit-object crashes (Table 12).

Table 25 lists specific route and Regime combinations that are out of proportion with respect to a random distribution of Regimes over routes. Heavily congested flow (D2) was a more common occurrence on certain segments of the I-405 in the morning and

SR-55 in the afternoon. Congested flow (D3) was more common on certain segments of the SR-57 in the morning and SR-91 in the afternoon. In contrast to the morning peak period, I-405 frequently exhibited high-speed heavy-volume operations (Regimes D7 and D8) in the afternoon peak.

Table 25 Routes with Relatively High or Low Concentrations of Regimes – Morning and Afternoon Weekday Peak Periods – Daylight, Dry Roads

Regime		Morning peak		Afternoon peak	
		More likely	Less likely	More likely	Less likely
D1	Light free-flow	SR-55	I-405	SR-55	I-405
D2	Heavily congested flow	I-405		SR-55	I-405
D3	Congested flow	SR-57		SR-91	
D4	Light, right-variable flow	I-5, SR-91	SR-22, SR-55	I-5	55, 22, 405
D5	Flow at capacity	I-5		SR-91	I-5
D6	Heavy variable flow	I-405	I-5	I-5	SR-55
D7	Heavy steady flow		SR-91	SR-57, I-405	SR-55
D8	Flow near capacity			SR-22, I-405	I-5

Classifications of dry road, daylight off-peak periods according to traffic flow Regimes is depicted in Figure 32 (weekday daylight off peak) and Figure 33 (daylight weekends). For weekday off-peak periods, Regime D4 (light, right-variable flow) is the most common Regime. Light free-flow (D1) is also prevalent, and it was not uncommon to find relatively high speed heavy flows (D6 and D7). With regard to specific weekday off-peak time periods (not shown), Regime D1 (light free-flow) occurs most often (in summer) before the morning and after the evening peak. Regime D4 (right-variable flow) occurs most often in the 10AM to 1PM period, and Regimes D5 (flow at capacity) and D8 (flow near capacity) occur at the margins of the evening peak period (3:00 - 3:30 PM and after 6:30 PM in summer).

Focusing on the breakdown by routes (Table 26), heavy variable flow (D6) was relatively concentrated on segments of the I-405 and SR-57, while heavy, steady flow (D7) was more concentrated only on I-405. The relatively rare, heaviest flow (D8) was more likely to be found on certain segments of the SR-55 and SR-22 weekday off peak.

On weekends in 1998 (Figure 33), light free-flow was most prevalent, but heavy, steady flow (D7) and light, reight-variable flow (D4) were relatively common. Regime D1 (light free-flow) is most predominant before 9AM on Saturday mornings and is rare after 1PM on Sundays. Regime D4 (right-variable flow) is most prevalent 6AM to 10AM on Saturday mornings. The two heavy-flow Regimes that appear on weekends (D6 and

D7) are most commonly observed on Sunday afternoons, from 1PM to 6PM. Regime D6 (heavy, variable flow) was also common 2PM to 4PM on Saturdays.

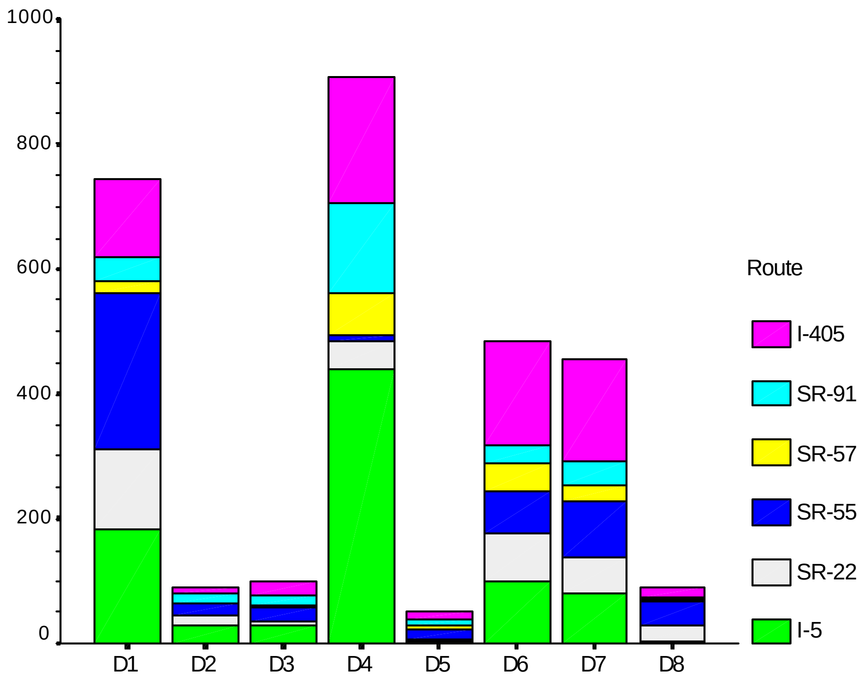

Figure 32 Classification of Traffic Conditions on Each of Six Orange County Freeways During Weekday Off-peak Periods (Weekdays before 6:00 AM, 9:00:30 AM through 3:29:30 PM and after 6:30 PM) - Daylight, Dry Roads

The pattern of Regimes by routes for off-peak times (Table 26) shows that the heavy, steady flow condition on weekends was more likely to be found on SR-91, which carriers heavy traffic back into Orange and L.A. Counties from the eastern mountain and desert regions.

The temporal and spatial patterns of the traffic flow Regimes for Orange County freeways in 1998 in daylight and dry road conditions are summarized in Table 27. Due to the scope of the missing data problem for the historical traffic flow data, the patterns described in Table 27 should not be considered to be a representative depiction of conditions for the entire Orange County freeway system for calendar year 1998.

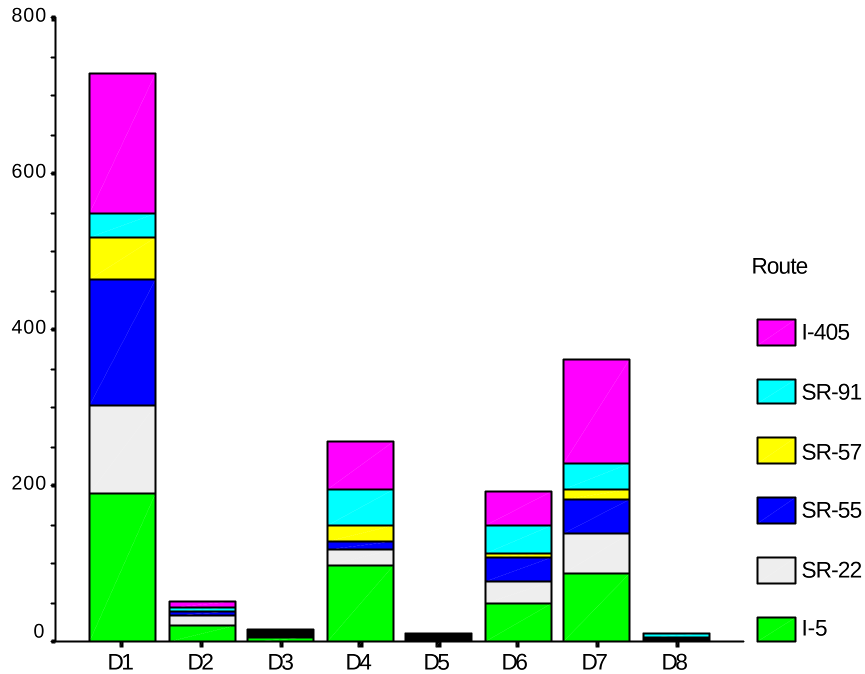

Figure 33 Classification of Traffic Conditions on Each of Six Orange County Freeways During Weekends - Daylight, Dry Roads

Table 26 Routes with Relatively High or Low Concentrations of Regimes – Weekday Off-peak Periods and Weekends – Daylight, Dry Roads

Regime		Weekday off-peak		Weekends	
		More likely	Less likely	More likely	Less likely
D1	Light free-flow	SR-55, SR-22	405, 91, 57, 5	SR-55	SR-91
D2	Heavily congested flow	SR-91	I-405, SR-57		
D3	Congested flow				
D4	Right-variable flow	I-5, SR-91	SR-55, SR-22	SR-91, I-5	SR-55, SR-22
D5	Flow at capacity	SR-55	I-5		
D6	Heavy variable flow	I-405, SR-57	I-5, SR-91	SR-91	
D7	Heavy steady flow	I-405	I-5		I-405
D8	Flow near capacity	SR-55, SR-22	I-5		

Table 27 Summary of the Likely Times and Locations of the Eight Traffic Flow Regimes – Daylight, Dry Roads

	Traffic flow characteristics	Most likely times and places
D1	**Light free-flow:** Very low volume, high average speed, low variance of speed in the right lane and about average variances of speed in the other lanes.	**Weekend and off-peak:** Weekends and other off-peak times, but can also occur during peak periods, presumably upstream of incidents. All freeways, but particularly SR-55.
D2	**Heavily congested flow:** Low volume and very low speeds. Low variances of volume in all lanes. Low variance of speed, particularly in right lane.	**Weekday afternoon peak:** Weekday peak periods, especially afternoons. Most prevalent on I-405 during morning peak, SR-55 during afternoon peak, and SR-91 off-peak.
D3	**Congested flow:** Moderately low average volume and low average speed. High variances in volumes and high variance in speeds except for the right lane.	**Weekday peak periods:** Morning and afternoon peak periods. Most prevalent mornings on SR-57 and afternoons on SR-91. Very rare on weekends.
D4	**Light, right-variable flow:** High mean speeds and moderately low mean volumes. Left and interior lanes high speed; right lane speed variance high and volume variance low.	**Weekday off-peak and mornings:** Most prevalent 10:00 AM to 1:00 PM weekdays and 6AM to 10AM Saturdays. More prevalent on I-5 and SR-91; less prevalent on SR-55.
D5	**Flow at capacity:** Very high variances in speed, average volumes and variances in volume, and moderately low average speeds.	**Weekday peak periods:** Relatively rare off-peak, but possible on SR-55 weekday off-peak. Most prevalent 5:00 PM – 6:00 PM and at the shoulders of the afternoon peak. Very rare on weekends.
D6	**Heavy, variable flow:** Very high volume variances, particularly in the right-lane, and moderately high volumes. High mean speeds and relatively low speed variances.	**All Peak Periods:** Especially the morning peak and on Saturdays 2:00 PM – 4:00 PM and Sundays 3 PM - 6:00 PM. More prevalent AM peak and off-peak on I-405; PM peak on I-5; weekends on SR-91.
D7	**Heavy, steady flow:** High volume and high mean speed, with low temporal variances of speed on all lanes and near-average volume variances.	**All Peak periods:** Both morning and evening peak periods, and the most prevalent afternoon peak period Regime. Also possible off-peak, e.g., Sunday 1-5:00 PM. Most prevalent on I-405.
D8	**Flow near capacity:** High volume, and low volume variances. Speed and speed variations about average to moderately below average.	**Weekday afternoon peak:** Especially 5:00 – 6:00 PM weekdays and at the shoulders of the afternoon peak. More prevalent on SR-22 and I-405 during peak; SR-22 and SR-55 off-peak.

5.3 Nighttime, Dry Road Conditions on Orange County Freeways in 1998

We next investigated the frequencies of occurrence of the six dry nighttime traffic Regimes over two periods: (1) off-peak weekdays, and (2) weekends. The classifications, broken down by route, for these two time periods are graphed in Figures 34 (weekdays) and 35 (weekends). There are too few observations for us to analyze the distribution of night Regimes during peak periods, because darkness during peak hours occurs only in late fall and early winter months.

For both weekdays and weekends, Regimes N2 (low volume free-flow) and N1 (very low volume free-flow) are most common, followed by Regime N5 (heavy, variable flow). Regime N3 (heavily congested flow) is also possible during weekday off-peak periods. Regimes N4 and N6 are both rare on weekends, and Regime N4 is also rare weekday off-peak.

With regard to a weekday distribution over routes (Figure 34), Regime 1 is more likely on I-5, while Regimes N2 and N3 are more likely to be found on SR-22. Regimes N4 and N5 are more likely to be found on SR-91, and N5 on I-405. On weekends (Figure 35), Regime N1 is more likely to be found on SR-55. These results are summarized in Table 28.

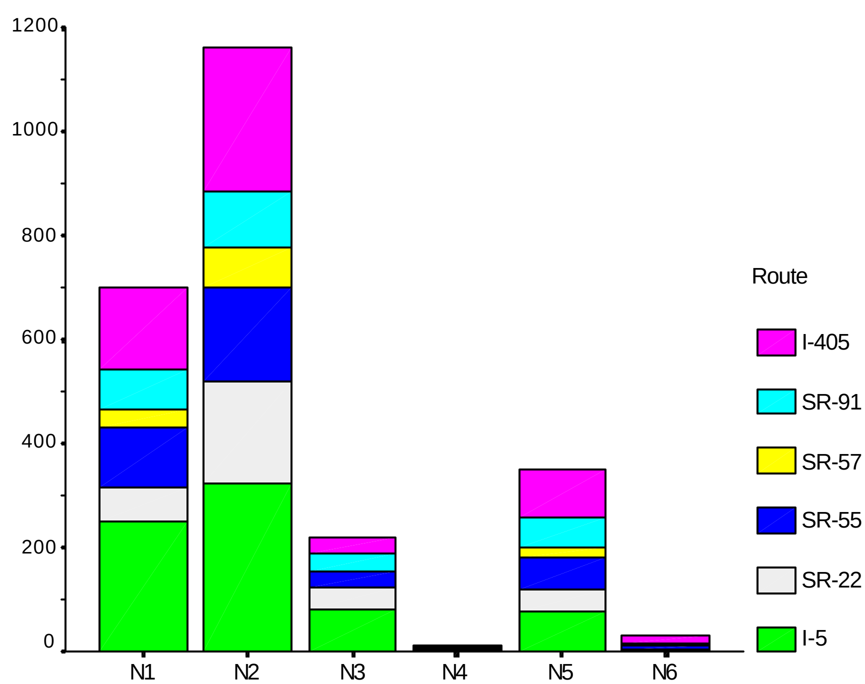

Figure 34 Classification of Traffic Conditions on Each of Six Orange County Freeways During Weekday off-peak periods (Weekdays before 6:00 AM and after 6:30 PM) - Nighttime, Dry Roads

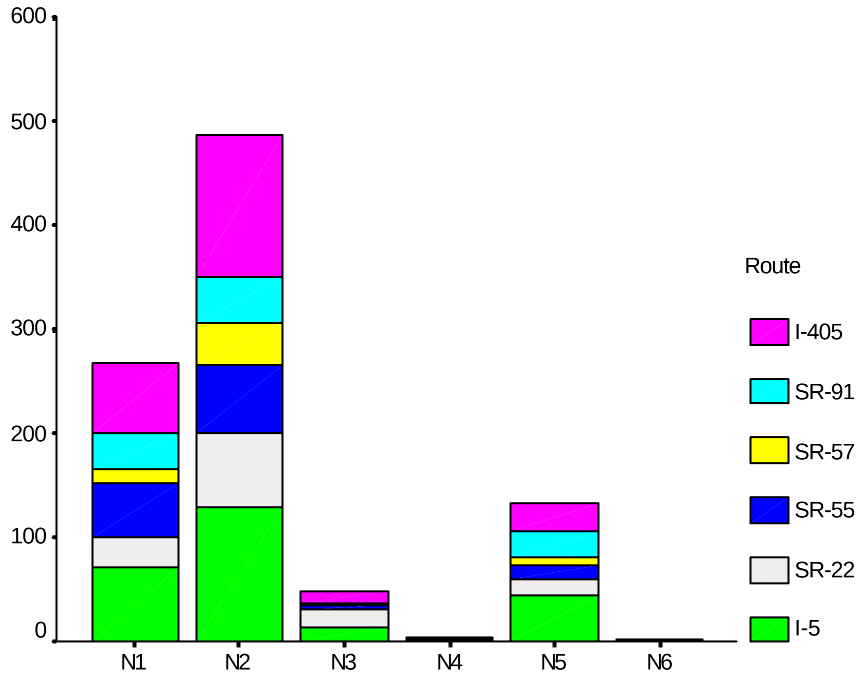

Figure 35 Classification of Traffic Conditions on Each of Six Orange County Freeways During Weekends - Nighttime, Dry Roads

Table 28 Routes with Relatively High or Low Concentrations of Regimes – Weekday Off-peak Periods and Weekends – Nighttime, Dry Roads

Regime		Weekday off-peak		Weekends	
		More likely	Less likely	More likely	Less likely
N1	Very low volume free-flow	I-5	SR-22	SR-55	
N2	Low volume free-flow	SR-22	SR-91		
N3	Conservative driving	SR-22	SR-57, I-405	SR-22	
N4	Sporadically congested flow	SR-91			
N5	Heavy, variable-volume flow	SR-91	I-5	SR-91	
N6	Flow near capacity	I-405	I-5		

A distribution of Regimes by weekday hours (not shown) reveals that Regime N1 typically occurs at night, between midnight and 6AM, while Regime N2 occurs in the evening, between 8PM and midnight. Regime N3 typically occurs 2AM – 5AM, and Regimes N5 and N6 are more likely to occur in the early evening, prior to 9 AM. A similar temporal distribution is found on weekends.

The temporal and spatial patterns of the nighttime dry-road traffic flow Regimes for Orange County freeways in 1998 are summarized in Table 29. Once again, due to the systematic nature of the missing data problem for the historical data, these patterns are not necessarily representative of conditions for Orange County freeways in 1998.

Table 29 Summary of the Likely Times and Locations of the Six Traffic Flow Regimes – Nighttime, Dry Roads

	Traffic flow characteristics	Most likely times and locations
N1	**Very low volume free-flow:** Very low average flow and low variances in flow. High average speed and high variances in speed on all lanes.	**Middle of the night:** More prevalent from midnight – 6:00 AM, very seldom in the early evening. More prevalent on I-5 weekdays, on SR-55 on weekends.
N2	**Low volume free-flow:** High mean speed and moderately low speed variances. Moderately low flow and low variance of flow in right lane.	**Pre-midnight:** More prevalent from 8:00 PM – midnight. More prevalent on SR-22 on weekdays.
N3	**Conservative nighttime driving:** Low average speed. Low variances of speed. Average flow (for periods of darkness) and average variances of flow.	**After closing hours:** 2:00 AM – 5:00 AM. More prevalent on SR-22. Less prevalent on SR-57 and I-405.
N4	**Sporadically congested flow:** Low average speed. High variances of speed in interior lanes. Moderately high flow (for periods of darkness) and high variances of flow all lanes.	**Evening peak:** Most prevalent 5:00 PM – 7:00 PM during autumn and winter periods of darkness, especially on Fridays. More prevalent on SR-91.
N5	**Heavy, variable flow:** High flow and very high variances of flow in all lanes. Moderately high mean speed and low variance of speeds.	**Early evening:** 6:00 PM – 9:00 PM, 5:00 PM – 7:00 PM on Fridays and 5:00 – 9:00 PM on Sundays during periods of early darkness. More prevalent on SR-91 and less prevalent on I-5.
N6	**Flow near capacity:** Very high volume. Slightly below average mean speed and speed variations. Also slightly below average variations in volumes.	**Late afternoon peak:** 6:00 PM - 8:00 PM weekdays, seldom on Weekends. More prevalent on I-405, less on I-5.

5.4 Wet Road Conditions on Orange County Freeways in 1998

The classification of 994 random observations of wet road traffic conditions into the seven wet road Regimes is depicted in Figure 36 – weekdays, 579 observations – and Figure 37 – weekends, 415 observations. The Table 30 lists specific route and Regime combinations that are out of proportion with respect to random distributions of Regimes over routes for weekdays and weekends.

Beginning with the lowest volume Regime W1 (very low volume free-flow), the hourly distributions (not shown) indicate that this Regimes occurs most often 11 PM to 5 AM weekdays on wet roads (midnight to 6AM weekends). The most common wet-roads Regime, W2 (low volume free flow) occurs more often 8PM to midnight both weekdays and weekends. Regime W2 is more commonly found on routes SR-22 and SR-57 on weekends, and less commonly found on SR-91 anytime.

On weekdays, Regime W3 (moderate free-flow) is relatively more common evenings 9PM to 11PM, especially on Routes I-5 and SR-91, and Regime W4 (moderate, right-concentrated) is more likely at the onset of the peak periods, 6AM to 7AM and 3PM to 4PM, especially on SR-55. Regime W5 (very heavy flow) is most likely to occur during both the morning and evening peak periods, and W7 (congested flow) is more likely to occur during the evening peak period only. Regime W6 (very heavy flow) is more commonly found weekdays on I-405.

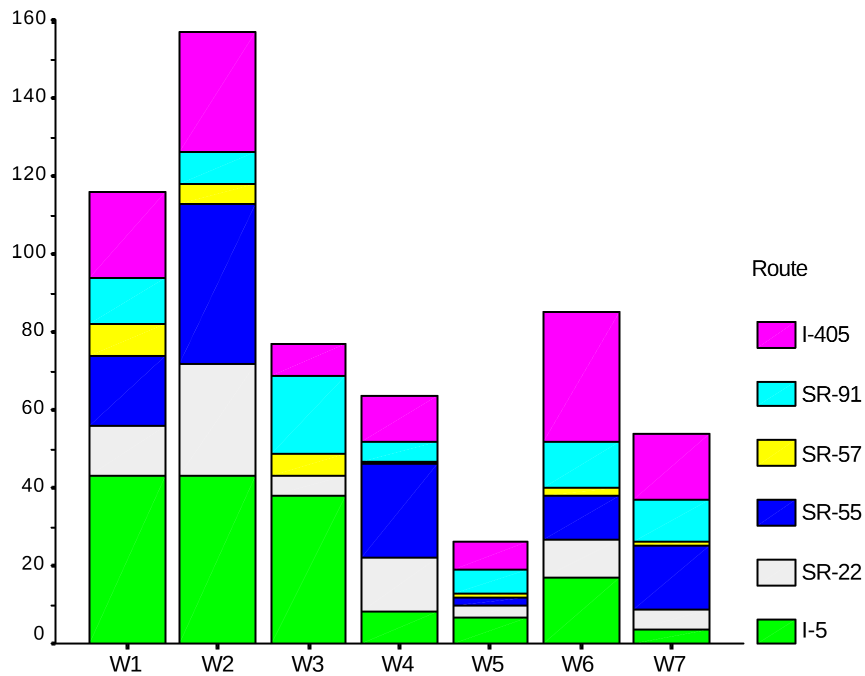

Figure 36 Classification of Traffic Conditions on Each of Six Orange County Freeways During Weekdays - Wet Road Conditions

On weekends, hourly distributions show that Regimes W1 through W4 build up sequentially through the mornings: W1 is more likely to be found before 6AM, W2 from

6AM to 8AM, W3 from 8AM to 10AM, and W4 from 10AM to noon, especially on SR-55. Regime W5 (heavy, variable flow) is also common 10AM – 11AM and on Sr-91. Finally, Regime W6 (very heavy flow) is an afternoon phenomenon, especially on Sundays.

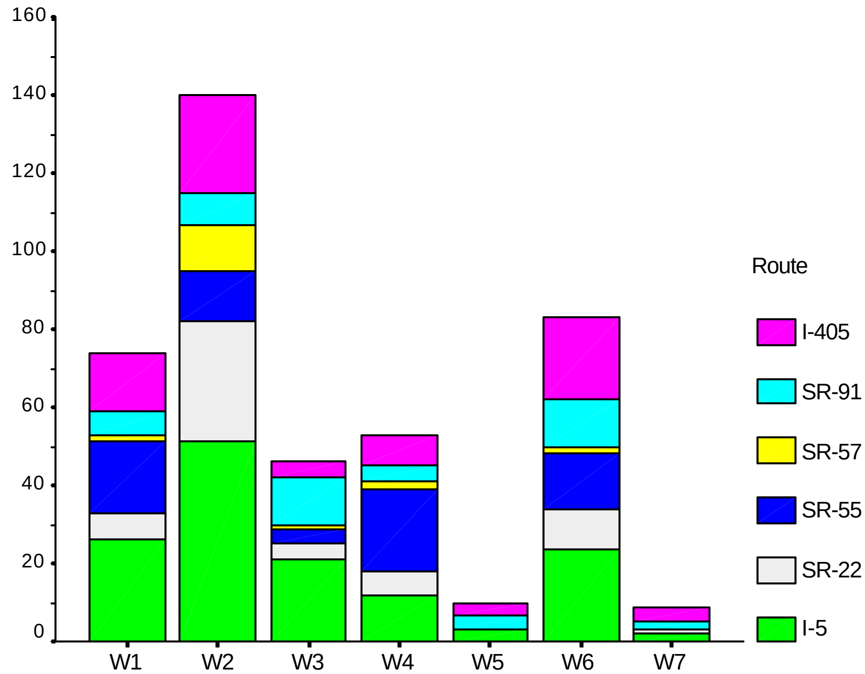

Figure 37 Classification of Traffic Conditions on Each of Six Orange County Freeways During Weekends - Wet Road Conditions

Table 30 Routes with Relatively High or Low Concentrations of Regimes – Wet Roads

Regime		Weekdays		Weekends	
		More likely	Less likely	More likely	Less likely
W1	Very low volume free-flow				
W2	Low volume free-flow		SR-91	SR-22, SR-57	SR-55, SR-91
W3	Moderate free-flow	I-5, SR-91	SR-55, I-405	SR-91	
W4	Moderate, right-concentrated	SR-55	I-5	SR-55	
W5	Heavy, variable flow			SR-91	
W6	Very heavy flow	I-405			
W7	Congested flow		I-5		

The temporal and spatial patterns of the traffic flow Regimes for Orange County freeways in 1998 for wet road conditions are summarized in Table 31. These patterns are not necessarily representative of historical conditions because of the problem with systematic missing loop detector data.

Table 31 Summary of the Likely Times and Locations of the Seven Traffic Flow Regimes – Wet Road Conditions

	Traffic flow characteristics	Most likely times and locations
W1	**Very low flow, variable speed:** Very low volume and very low variations in volume. Mean speed slightly below average for wet roads. Variations in speed high, especially for right lane.	**Nighttime:** Non-peak hours late at night, particularly 10:00 PM to 6:00 AM, and on Sundays. More prevalent on I-5, less prevalent on I-405.
W2	**Low volume free-flow:** Low average volume and moderately high speed. Low variances in volume and speed in right lane.	**Evenings:** Non-peak evening hours. Particularly 7:00 PM – 12:00 AM, especially Saturdays, not Sundays. Less prevalent on I-405.
W3	**Moderate free-flow:** Moderately high speed and near average volume, volume variances, and speed variances for wet roads.	**Post-Peak:** Most common 8:00 – 10:00 AM and 6:00 PM – 9:00 PM., but not on Fridays. Most prevalent on I-5 and SR-91, less on SR-55.
W4	**Moderate right-concentrated flow:** Moderately high speed and near average volume, but high variance of volume and low variance of speed in right lane.	**Evening late-peak:** More prevalent 5:00 PM – 7:00 PM and less prevalent at night or morning peak. More prevalent on SR-55.
W5	**Heavy, variable flow:** Moderately high speed and volume. Very high variance of volume in left lane and high variance of volume in right lane.	**Mid-mornings and afternoons:** Most common 10:00 AM – Noon and 1:00 PM – 3:00 PM on all days and all freeways.
W6	**Very heavy flow:** High volume and average speed with low variances in speed. High volume variances, especially in left lane.	**Morning peak:** Especially 7:00 – 8:00 AM and on Sundays. More prevalent on I-405 and less on I-5.
W7	**Flow near capacity:** Low speed and high volume. Average to slightly below average variances of both speeds and volumes.	**Weekday Peak periods:** More prevalent on Mondays. More prevalent on SR-55, less on I-5.

6 Case Study: One Week at Two Locations on Northbound SR-55

In this Section we demonstrate an application of the FITS Tool by applying it retroactively to streams of loop detector data from two adjacent loop detector stations along northbound SR-55 in the City of Santa Ana, Orange County, California. The time span is the first week of March 1998. The two detector stations, located about 1.25 miles apart between the SR-55/I-405 and SR-55/I-5 interchanges, are depicted in the diagram in Figure 38. FITS output is graphed for consecutive nighttime and daytime periods at these two locations in Figures 39-68. The remainder of this Section contains a brief interpretation of this output.

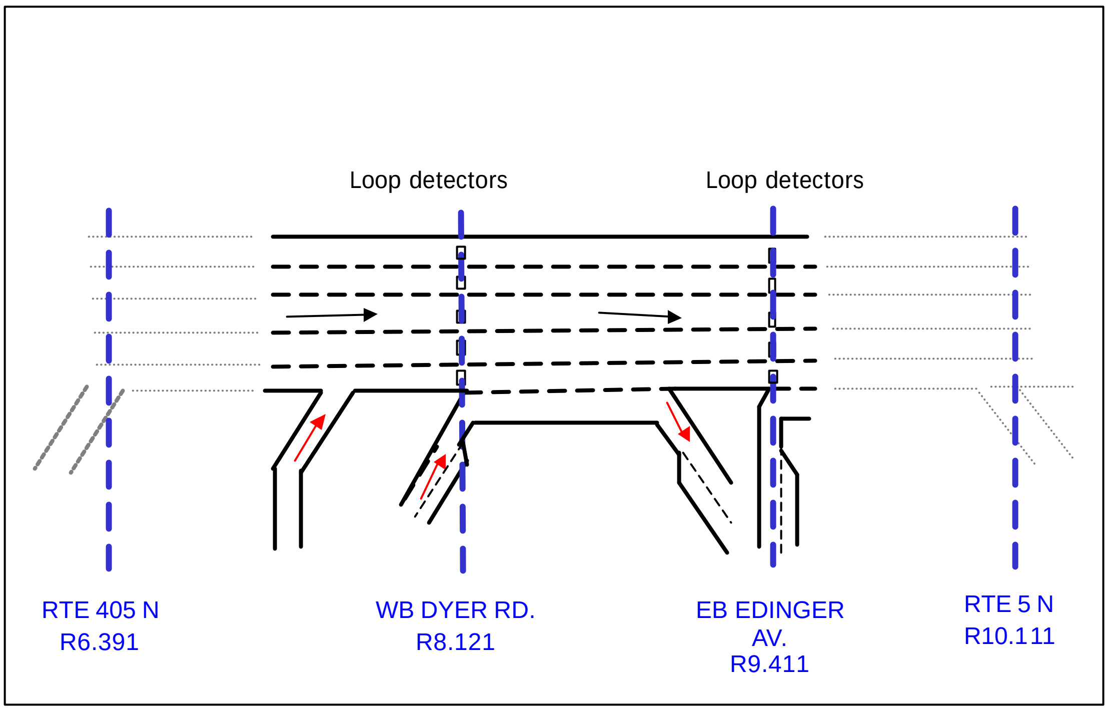

Figure 38 Location of the Two Case Study Loop Detector Stations at Dyer Road and Edinger Avenue on Northbound SR-55 in Orange County

Sunday morning (Figures 39 and 40): The first nighttime period, beginning at approximately midnight Sunday, March 1 sees some heavy, variable-volume flow (Regime N5), with no prevailing type of crash, at the downstream station at about midnight. The downstream location (Figure 36) shifts from Regime N2, low volume free-flow (conducive to left-side run-offs), to Regime N1, very low volume free-flow (conducive to right-side run-offs), at about 3AM. The upstream location (Figure 37) exhibits the same shift earlier, at about 2AM.

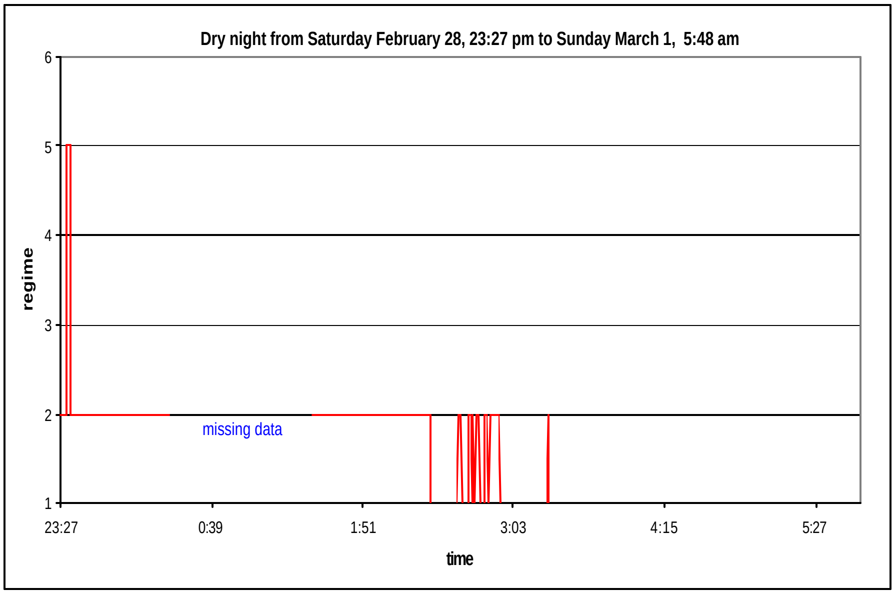

Figure 39 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Early Morning of March 1, 1998 (Regimes N1 – N6)

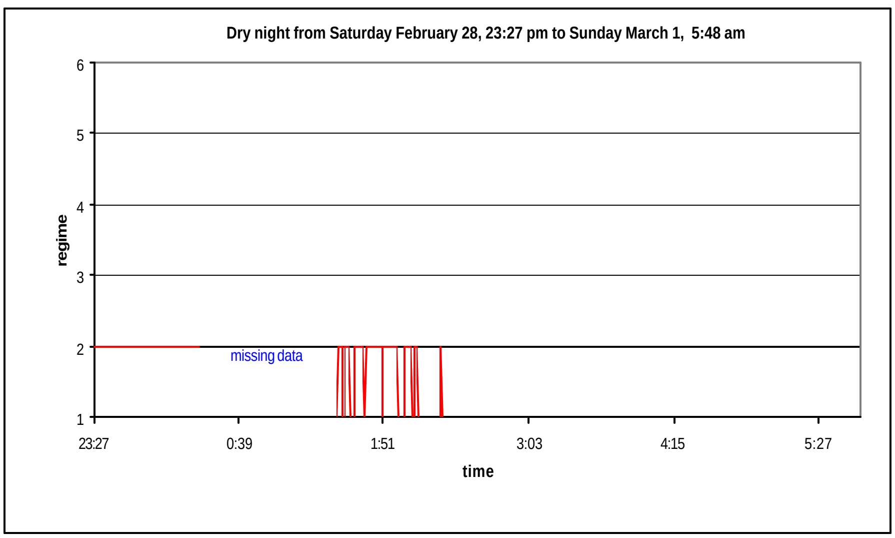

Figure 40 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Early Morning of March 1, 1998 (Regimes N1 – N6)

Daytime Sunday (Figures 41 and 42): At the first location, flow is consistently low (right-side run-offs most likely type of any crash that might occur) until about 10:30AM. Then heavy, variable flow (no predominant type of crash) until about 2:30PM, when the shift is to heavy steady flow (lateral navigation crashes), with a few short periods of the heaviest flow near capacity at about 5PM. At the upstream location (Figure 39), there are periods of light variable free-flow in the morning, but then steady light free-flow until about noon. Flow oscillates between heavy variable flow and light free-flow for about an hour beginning at noon, then settles into heavy steady flow (lateral navigation crashes most likely) for the remainder of the afternoon, with periods of free flow, which indicate periodic downstream traffic bottlenecks.

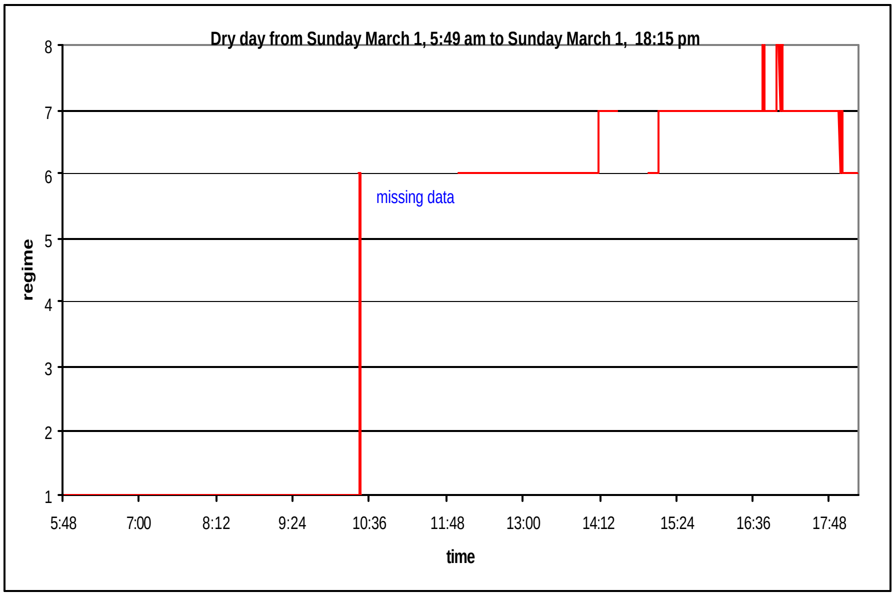

Figure 41 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Daytime Sunday, March 1, 1998 (Regimes D1 – D8)

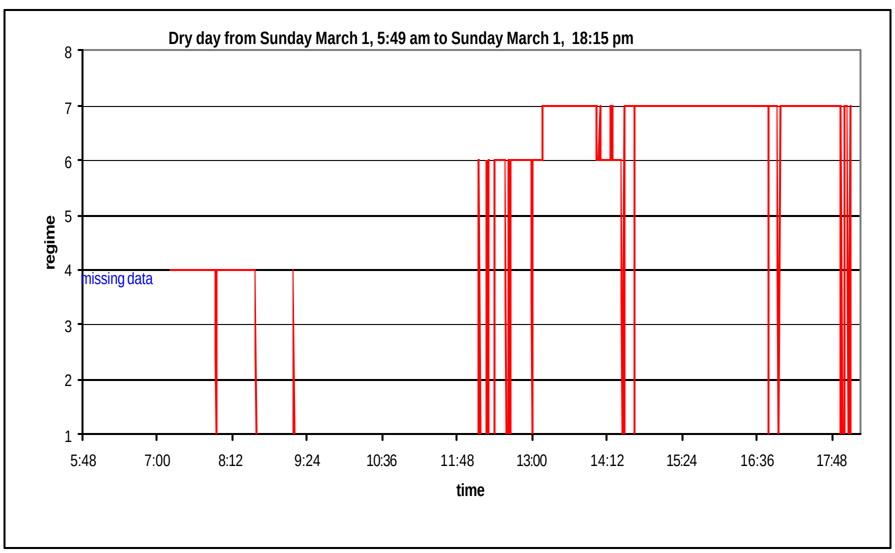

Figure 42 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Daytime Sunday, March 1, 1998 (Regimes D1 – D8)

Sunday night until dawn Monday (Figures 43 and 44): The first location exhibits heavy variable flow (Regime N5), with no prevailing type of crash, until 7:30PM, then again for short bursts around 8:00PM and 9:00PM. The rest of the time, there is low volume free flow (N2), conducive to left-side run-offs, until midnight. At the second location, the transition between heavy variable flow and light free flow occurs earlier, between 6:30 and 7:30. Between midnight and about 5:00AM both locations experience very low volumes (N1), with right-side run-offs the most common type of accident.

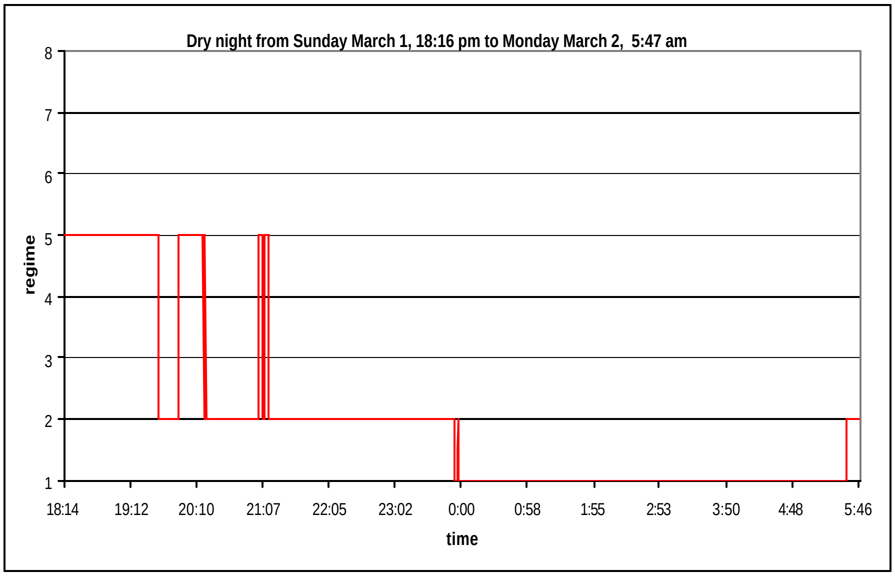

Figure 43 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Nighttime March 1-2, 1998 (Regimes N1 – N6)

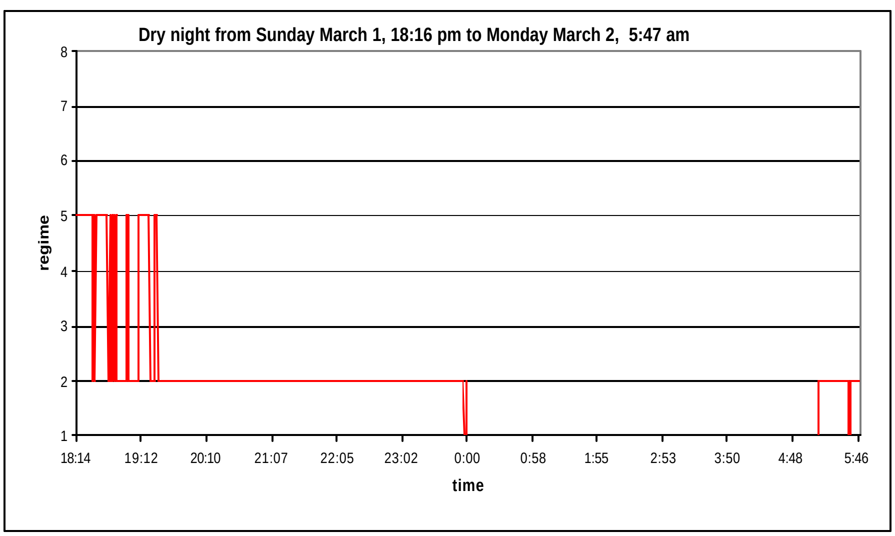

Figure 44 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Nighttime March 1-2, 1998 (Regimes N1 – N6)

Daytime Monday (Figures 45 and 46): Traffic picks up at both locations at about 6:30AM (slightly later at the upstream location). At the first location the sequence in terms of most likely types of crashes is: short period of (D4) serious run-offs and (D6) mixed types, 7:30-9:00 (D7) lateral navigation, 9:00-noon (D6) mixed types, short periods of (D7) and (D8) two-vehicle rear ends, then (D7) again until afternoon peak, In which there are periodic spells of (D8) two-vehicle rear ends. The downstream location operates mainly in Regime (D7) lateral navigation crashes, until about 5:00 PM, when it breaks down into flow at capacity (D5) with left-lane rear ends most likely, and congested flow (D3) which favors two-vehicle rear-end crashes.

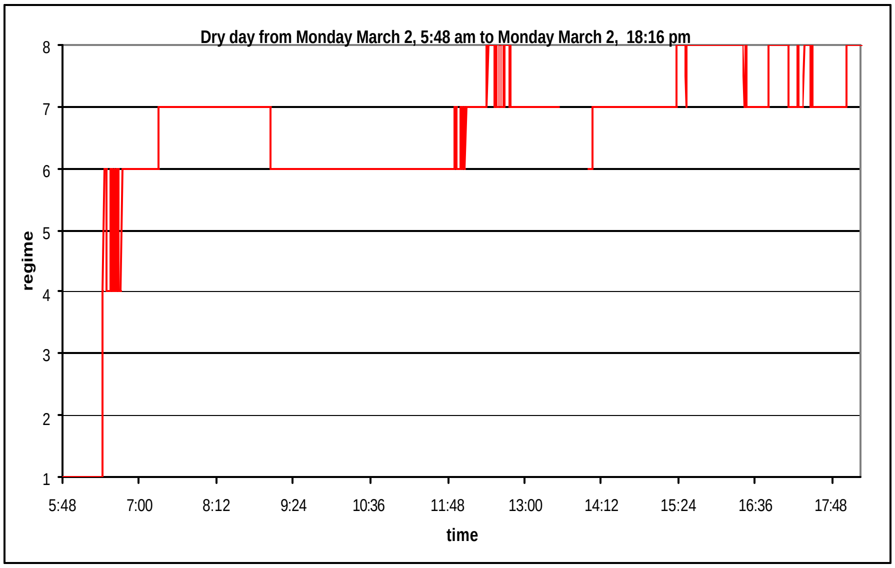

Figure 45 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Daytime Monday, March 2, 1998 (Regimes D1 – D8)

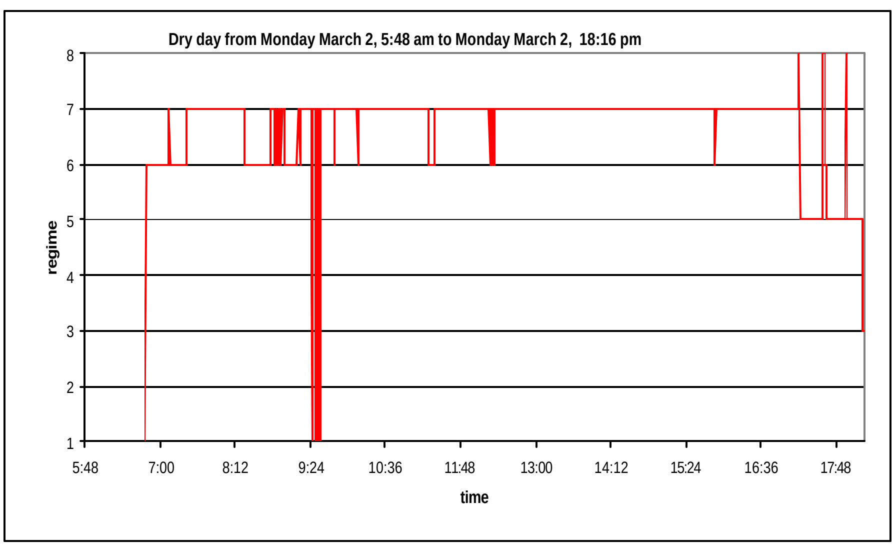

Figure 46 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Daytime Monday, March 2, 1998 (Regimes D1 – D8)

After dark Monday until dawn Tuesday (Figures 47 and 48): Loop detector data Are missing after about 9:00 PM on this night. Prior to 9:00PM, both locations show signs of serious congestion, with the downstream location operating at near capacity (N6), conducive to large right-lane rear ends, then (N5) mixed crash types, with periods of low-volume free flow (N2), when left-side run-offs are more common. The upstream location operates exclusively in (N5) and (N2) modes.

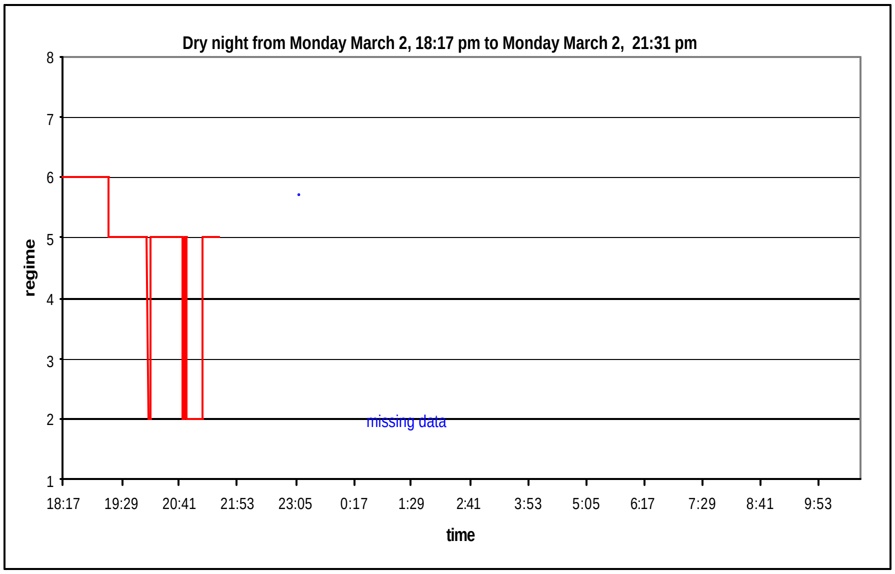

Figure 47 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Nighttime March 2-3, 1998 (Regimes N1 – N6)

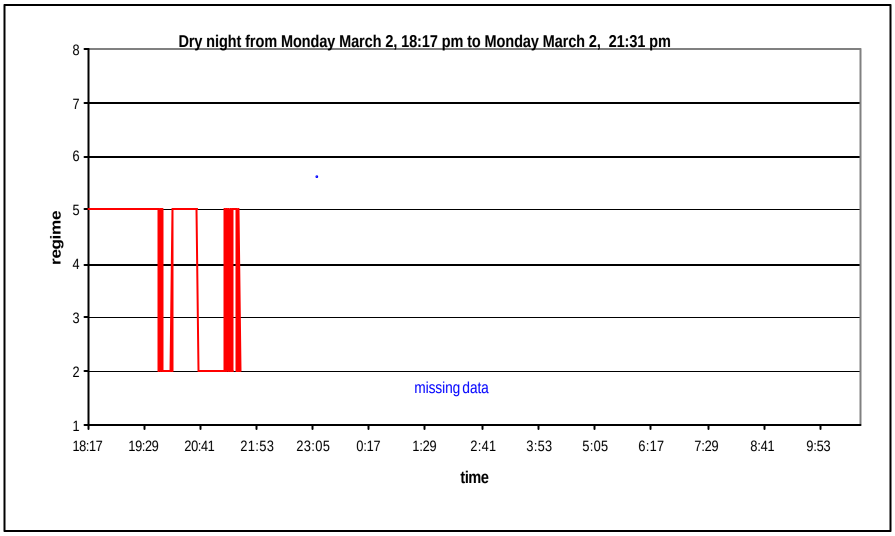

Figure 48 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Nighttime March 2-3, 1998 (Regimes N1 – N6)

Daytime Tuesday (Figures 49 and 50): Loop detector data are missing until 10:46AM. Beginning then, the downstream location oscillates between heavy, variable flow (D6) with mixed crash types, and heavy steady flow (D7) which favors lateral navigation crashes, until about 2:00PM. This location then exhibits periods at near-capacity (D8) with two-vehicle rear ends more likely, intermixed with (D7) conditions. The upstream location operates in heavy, variable flow (D6: mixed crash types) and heavy, steady flow (D7: lateral navigation crashes) conditions all day.

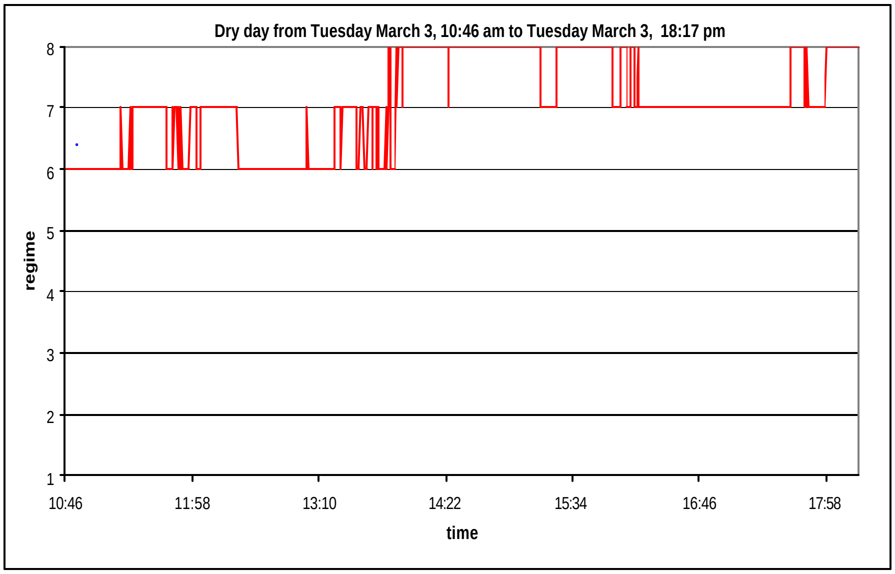

Figure 49 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Daytime Tuesday March 3, 1998 (Regimes D1 – D8)

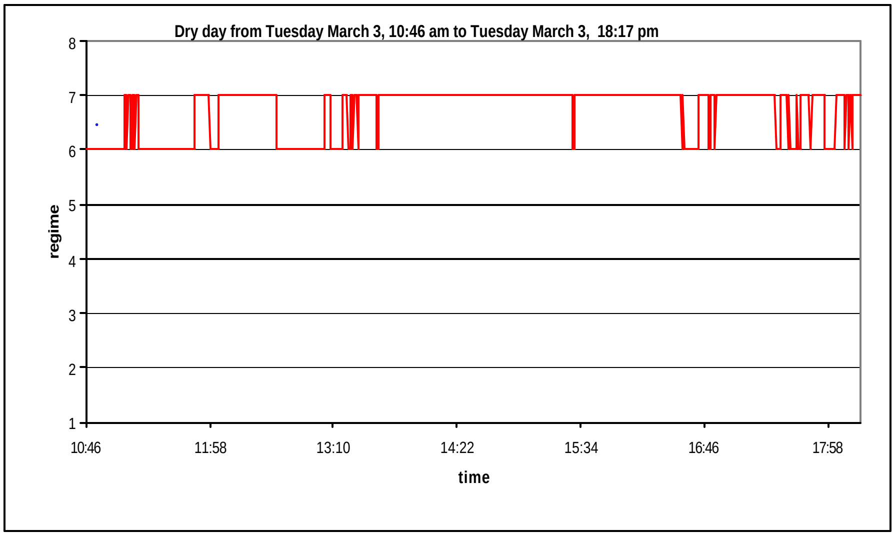

Figure 50 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Daytime Tuesday March 3, 1998 (Regimes D1 – D8)

After dark Tuesday until dawn Wednesday (Figures 51 and 52): The upstream location operates at capacity (N6: associated with large right-lane rear-end crashes), until about 7:30 PM. It then operates mainly in heavy, variable flow (N5: mixed crash types) until about 10:45PM, shifts to low volume (N2: left-side run-offs) a little before 11:00PM, and then to very low volume at about 2:30AM (N1: right-side run-offs). The second location commences in heavy variable flow (N5). The two locations have similar safety profiles from about 8:00PM through the rest of the night.

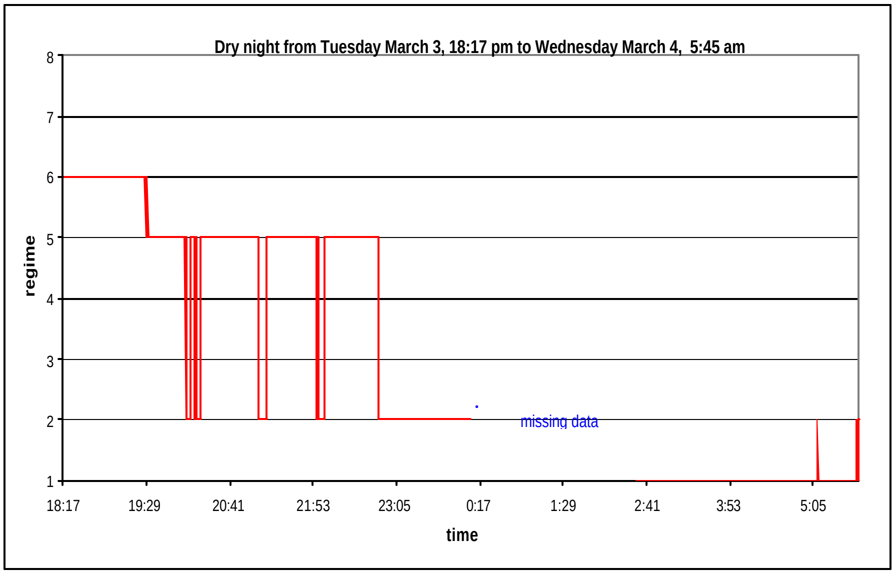

Figure 51 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Nighttime March 3-4, 1998 (Regimes N1 – N6)

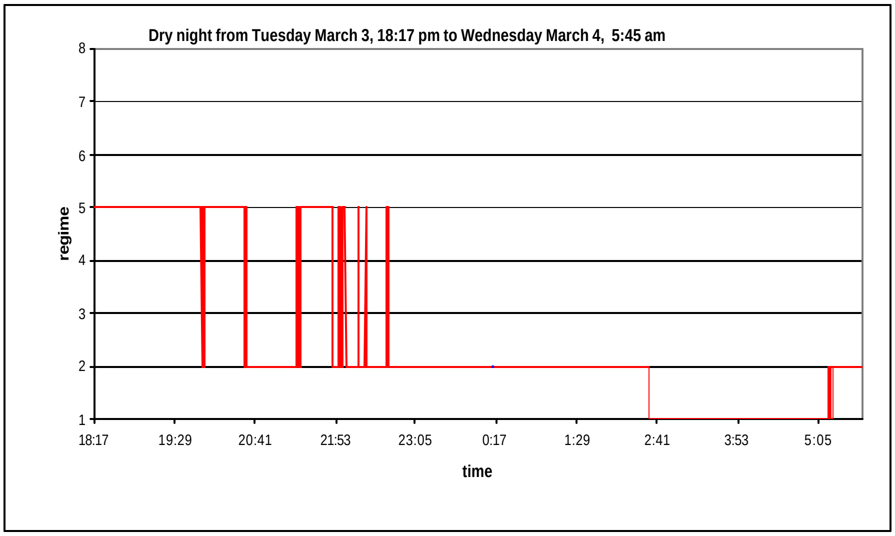

Figure 52 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Nighttime March 3-4, 1998 (Regimes N1 – N6)

Daytime Wednesday (Figures 53 and 54): In the morning, traffic builds up at the first location at about 6:15AM and progresses through three Regimes of increasingly heavier flow (D6: mixed crash types; D7: lateral navigation crashes; and D8: two-vehicle rear ends), finally settling into heavy variable flow (D6) at about 9:20AM. At the second location, traffic builds up a little later, settling mainly into heavy steady flow. In the afternoon, safety conditions are similar to Tuesday (Figures 49 and 50) until about 5:30, when congestion increases at both locations. The downstream locations oscillates among a number of different congested conditions, while the upstream location exhibits a period of heavily congested flow (D2: conducive to mixed multi-vehicle crashes).

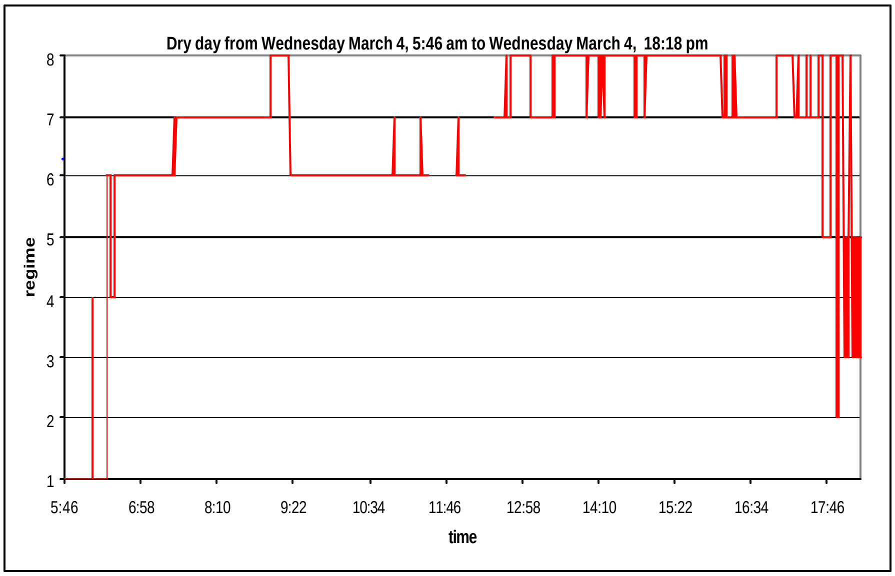

Figure 53 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Daytime Wednesday March 4, 1998 (Regimes D1 – D8)

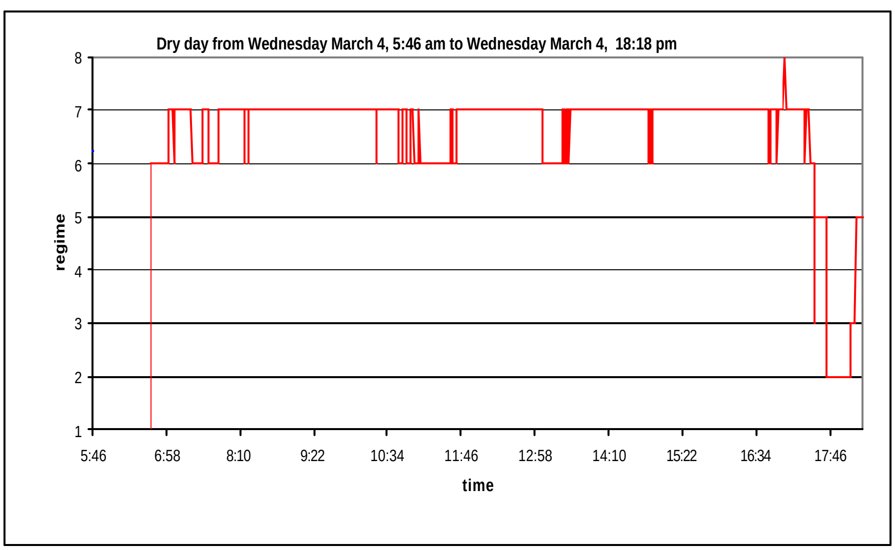

Figure 54 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Daytime Wednesday March 4, 1998 (Regimes D1 – D8)

After dark Wednesday until dawn Thursday (Figures 55 and 56): Safety conditions have roughly the same pattern at both locations on this night as on the previous night (Figures 51 and 52). The only noticeable difference is a slightly more consistent period of heavy, variable flow (N5: mixed crash types) at both locations, up until about 10:45PM.

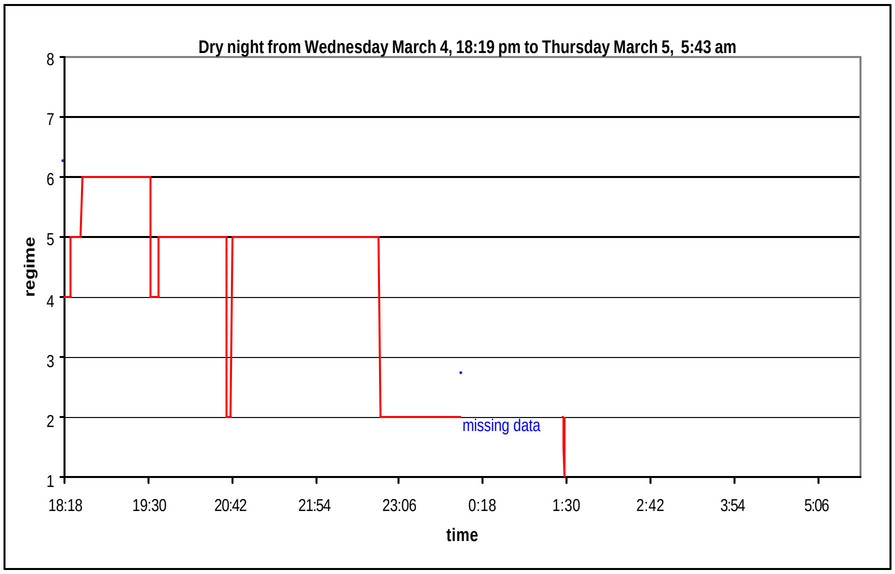

Figure 55 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Nighttime March 4-5, 1998 (Regimes N1 – N6)

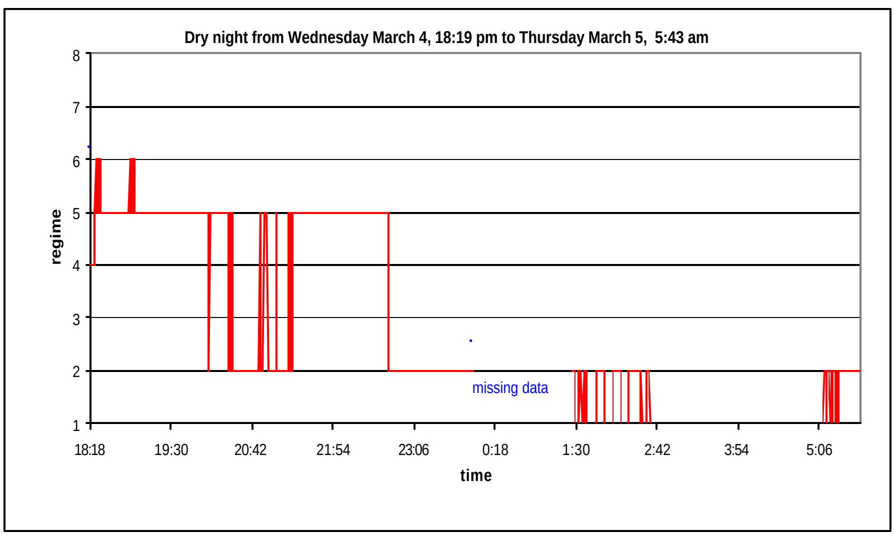

Figure 56 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Nighttime March 4-5, 1998 (Regimes N1 – N6)

Daytime Thursday (Figures 57 and 58): Prior to about 5:00PM, conditions on Thursday are similar to those on the preceding day (Figures 50 and 51). Between 5:00 and 6:00PM, the first location exhibits two fifteen minute periods of heavy congestion (D2: conducive to mixed types of multi-vehicle crashes). This location later experiences intermittent periods relatively unstable flow at capacity (D5: conducive to left-lane rear ends) and flow near capacity (two-vehicle rear ends anywhere on the road). The upstream location is in heavy congestion (D2) from about 5:30 through dusk (6:19PM).

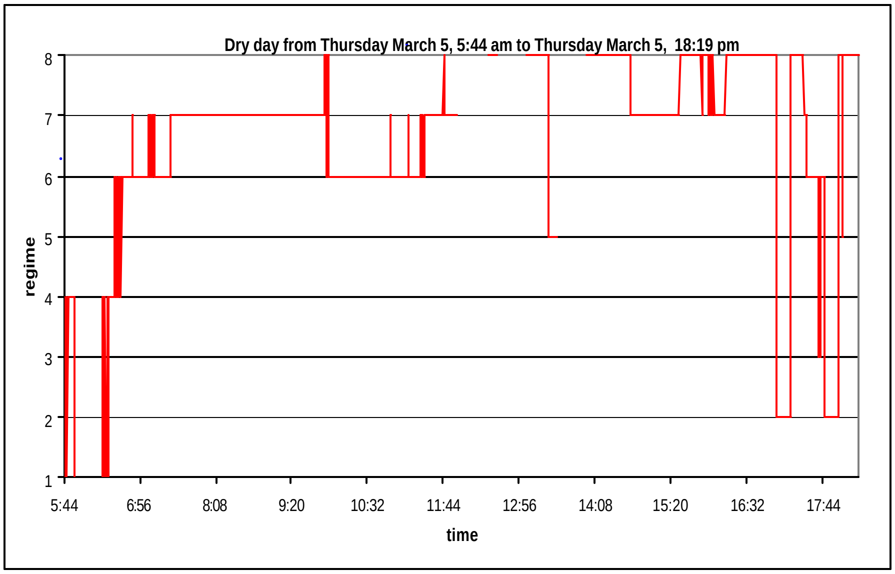

Figure 57 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Daytime Thursday March 5, 1998 (Regimes D1 – D8)

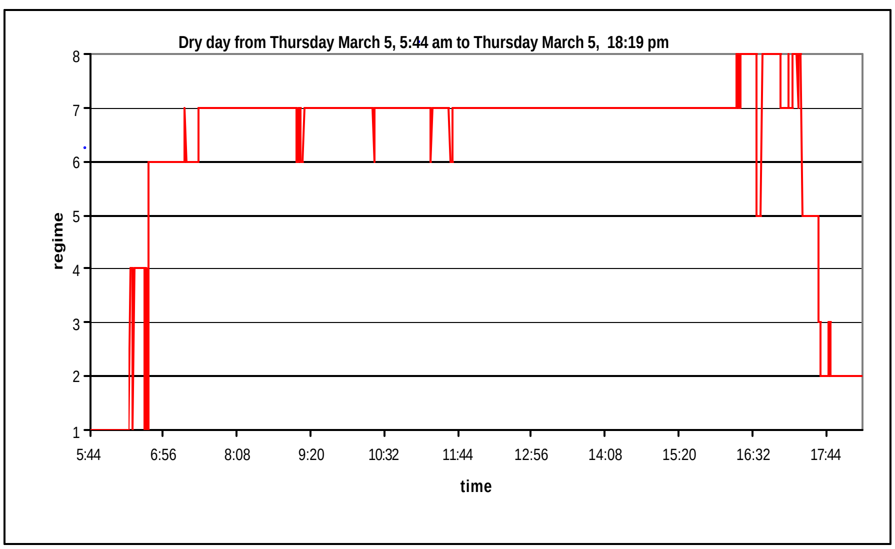

Figure 58 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Daytime Thursday March 5, 1998 (Regimes D1 – D8)

Wet Roads, Thursday evening until 8:30AM Friday (Figures 59 and 60): This rainy period extended congested conditions (W7: conducive to rear end crashes) until about 7:30PM at both locations, after which both locations settled into a brief period of very heavy flows with average speeds (W6: mixed types of non-injury crashes). From 8:40 through midnight, both locations manifest a wide range of wet-road traffic and safety conditions.

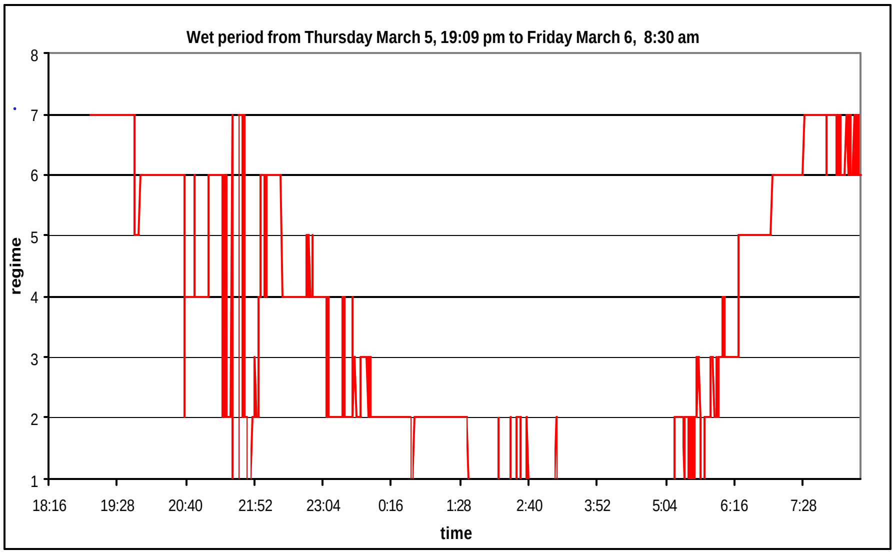

Figure 59 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Period of Wet Roads, March 5-6, 1998 (Regimes **W1** – **W7**)

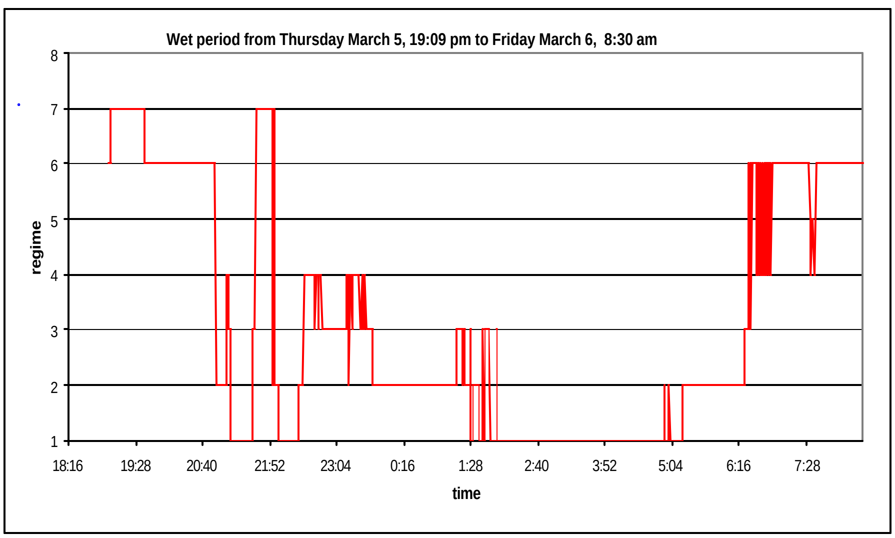

Figure 60 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Period of Wet Roads, March 5-6, 1998 (Regimes **W1** – **W7**)

The downstream location has periods of heavy flow (W6), moderate right-concentrated flow (W4: conducive to serious multi-vehicle crashes) and low volume free-flow (W2: mixed types of 1- and 2-vehicle crashes); the downstream location oscillates between congested flow (W7) and very low volume free floe (W1: right-side run-offs). At the onset of the morning peak period, flow increases at both locations, with the upstream location cycling between very heavy (W6) and congested (W7) flows, and the upstream location settling in to less-congested very heavy flow (W6).

Daytime Friday, dry roads (Figures 61 and 62): Heavy traffic conditions occur at both locations all day Friday, exhibiting heavy steady flow (D7: conducive to lateral navigation crashes), with periods of heavy variable flow (D6) in the morning. The first location also exhibits an extended afternoon periods of very heavy flow near capacity (W8: two-vehicle rear end crashes).

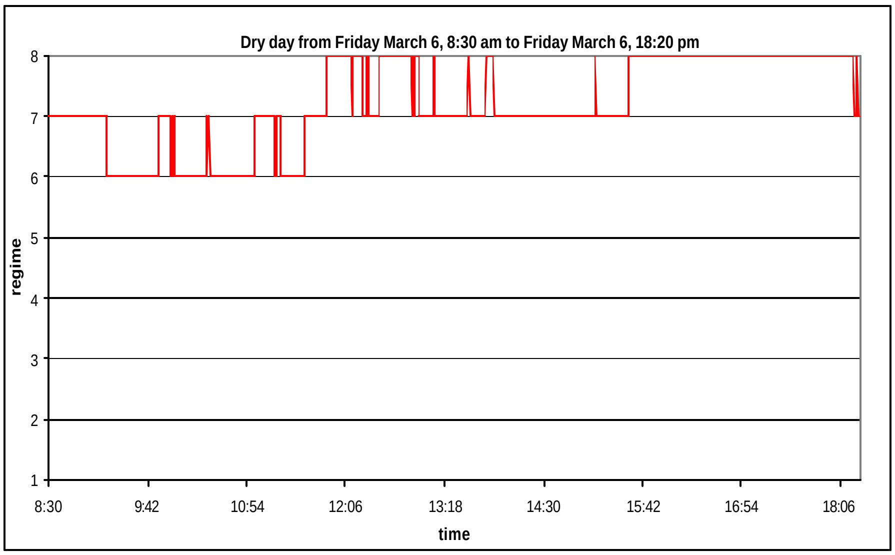

Figure 61 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Dry Daytime, Friday March 6, 1998 (Regimes D1 – D8)

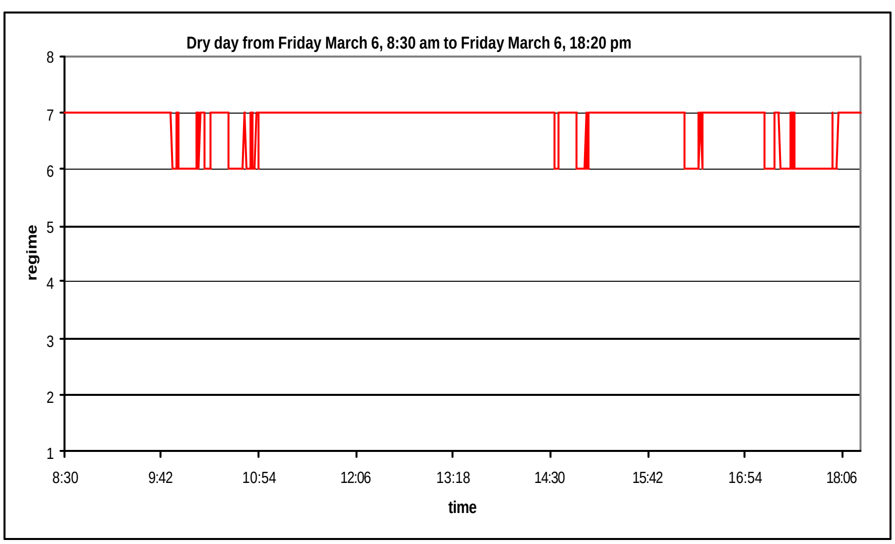

Figure 62 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Dry Daytime Friday March 6, 1998 (Regimes D1 – D8)

After dark Friday until dawn Saturday (Figures 63 and 64): The first location exhibits heavy flow (N6: conducive to large right-lane rear-end crashes) on Friday night until shortly after 7:00PM. Then there are periods of heavy, variable flow (N5: mixed crash types), interspersed with shorter periods of low volume free-flow (N2: left-side run-offs) until about midnight. These periods of heavier flow could correspond to schedules at local entertainment and hopping venues. The upstream location exhibits a similar but more condensed pattern. Very low steady flow (N1: conducive to right-side run-offs) is not reached until after 2:30AM at both locations.

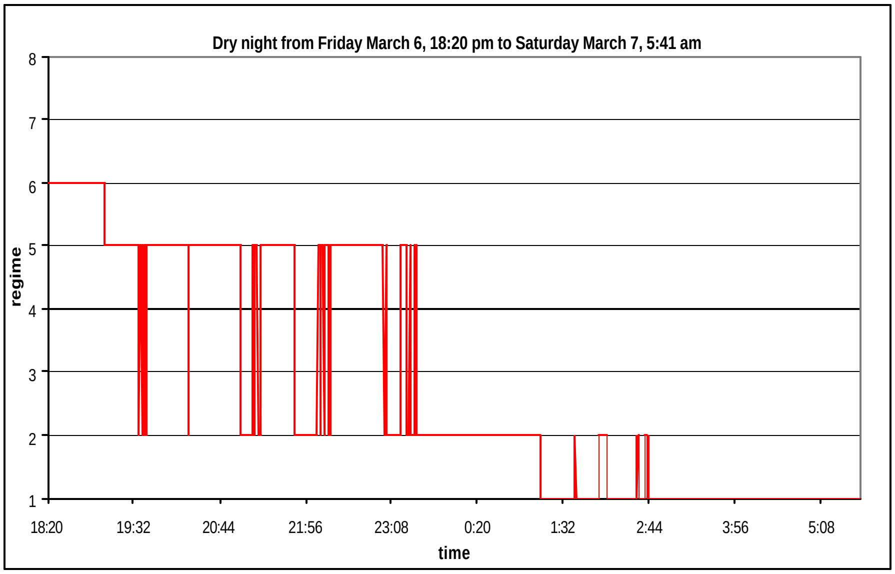

Figure 63 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Nighttime March 6-7, 1998 (Regimes N1 – N6)

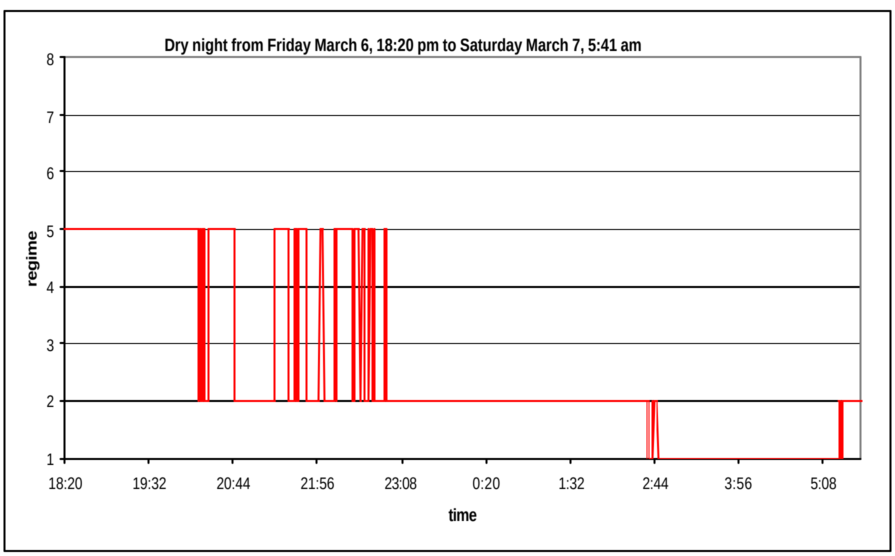

Figure 64 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Nighttime March 6-7, 1998 (Regimes N1 – N6)

Daytime Saturday (Figures 65 and 66): Following early morning free flow (D1 and D4) prior to 8:30AM, the first location shows highly variable flow in the 8:30AM to 9:45AM period, as evidenced by repeated cycles of heavy flow (D6 or D7) and free flow (D1). From 9:45AM until 1:30PM heavy variable flow predominates (D6: mixed crash types), followed by heavy steady flow the remainder of the day (D7: conducive to lateral navigation crashes). The upstream location has low flow (D1 and D4) until about 9:30, then also shifts to heavy steady flow for most of the remainder of the day.

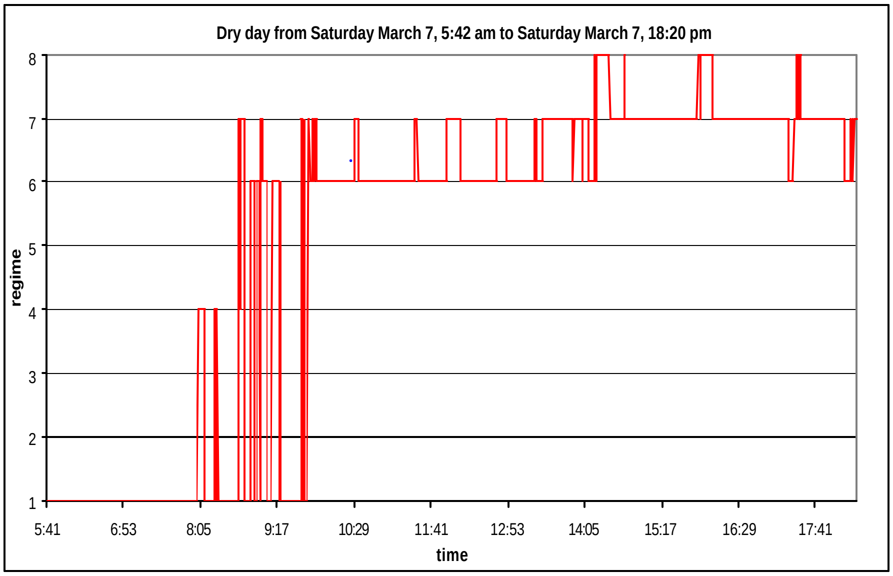

Figure 65 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Daytime Saturday March 7, 1998 (Regimes D1 – D8)

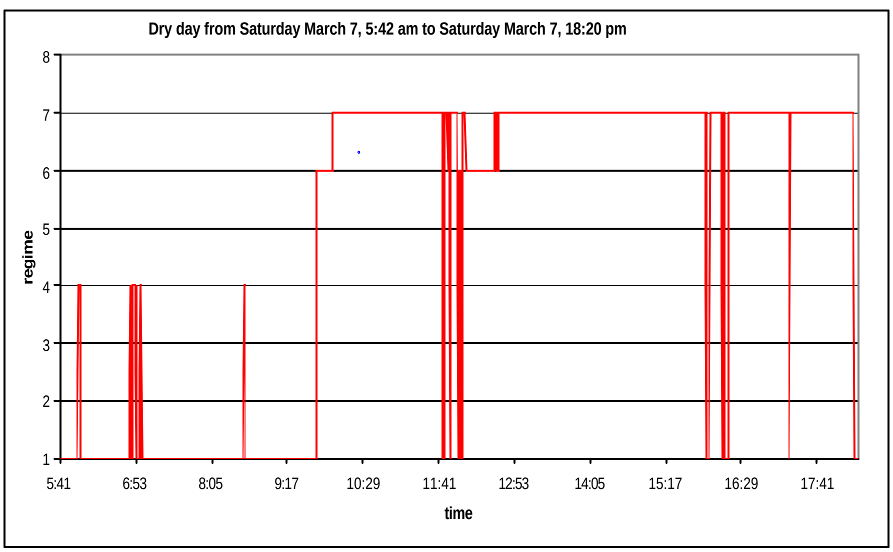

Figure 66 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Daytime March 7, 1998 (Regimes D1 – D8)

Saturday night (Figures 67 and 68): The final case study period is Saturday night through dawn Sunday. Early on Saturday night, both locations have periods of heavy variable flow (N5: no prevailing crash type). The first location is then in a Regime of conservative driving with moderately heavy flow from 8:30PM until about 10:45PM (N3: conducive to two-vehicle interior lane crashes). Low volume free flow is reached at about 11:45PM (N2: left-side run-offs), followed later by very low volume free flow (N1: right-side run-offs). Flow is considerably less at the second location from 7:00PM until 10:30PM, being predominately low volume free flow (N2: left-side run-offs), with a period of heavier flow (N5) from about 10:15 to 10:45PM.

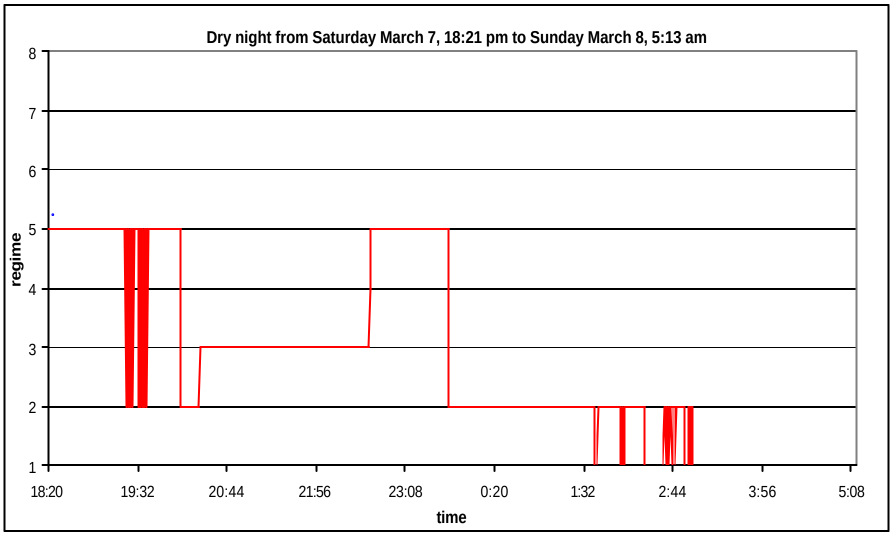

Figure 67 Classification of Traffic Conditions on SR-55 Northbound at Dyer Road, Postmile 8.12, Nighttime March 7-8, 1998 (Regimes N1 – N6)

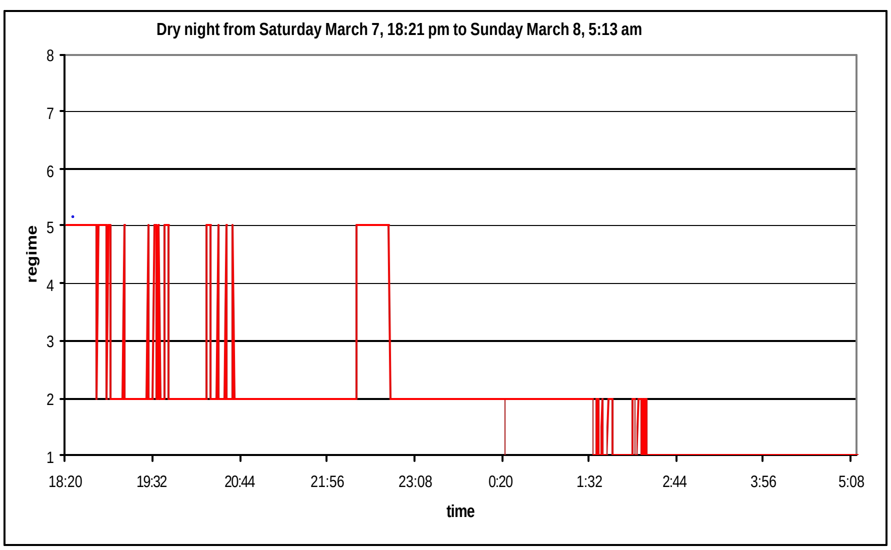

Figure 68 Classification of Traffic Conditions on SR-55 Northbound at Edinger Avenue, Postmile 9.41, Nighttime March 7-8, 1998 (Regimes N1 – N6)

7 Demonstration Application of FITS

In this section, we offer a demonstration of a potential application of the methodology developed in this research. Because of the systematic biases introduced by non-reporting loop stations identified in the previous section (as well as with the sample of crashes used to estimate the models), the following is intended for demonstration purposes only; no claim is made that the results are representative of actual conditions.

Consider a freeway segment, S, during some time interval, T, containing n loop stations, $l_i, i = 1,2,\ldots n$ (Figure 69).

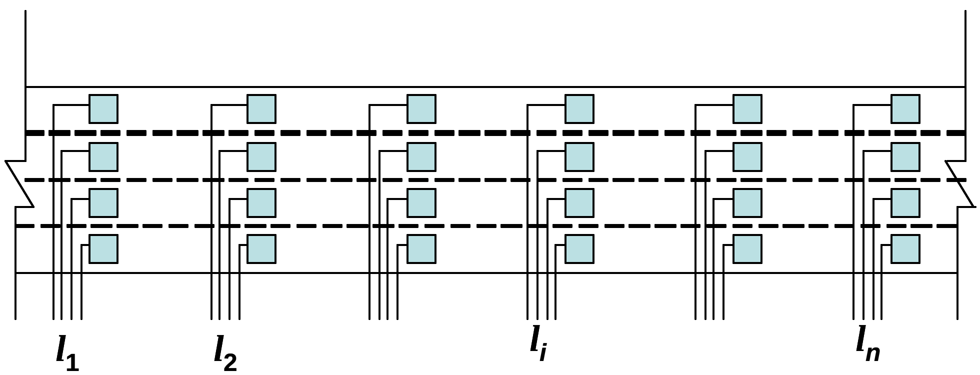

Figure 69 Freeway Segment S

Let R_{it} denote the Regime in the vicinity of loop station $l_i, i = 1,2,\ldots n$, during 30-second time interval $t = 1,2,\ \ldots\ ,T/30\text{sec}$. Ostensibly, each Regime R_{it} defines traffic flow conditions prevailing on a section of freeway extending from the midpoint between loops l_{i-1} and l_i and loops l_i and l_{i+1} during the 30-second time interval t. The FITS program can easily determine R_{it} from 30-second loop count data, based on the membership functions that led to the Regime classifications in Tables 8, 14 and 20 for Dry-Day, Dry-Dark, and Wet crashes, respectively.

The total population of Regimes comprising freeway segment S during T is simply

$$N_{TS} = \frac{nT}{30\text{sec}}$$

where N_{TS} is the total number of Regimes in the population defined by 30-second loop counts on freeway segment S during T. Let

$$n_R = \left|\left\{ R_{it} \mid R_{it} = R, \forall i \in S, \forall t \in T; R \in \mathbf{R} \right\}\right|$$

where n_R is the number of occurrences of any particular Regime R in the population, and $\mathbf{R}$ is the set of Regimes (which may be further disaggregated by particular environmental segmentation, e.g., $\mathbf{R} = \left\{ \mathbf{R}_{Dry-Day} \quad \mathbf{R}_{Dry-Darkness} \quad \mathbf{R}_{Wet} \right\}$). An estimate, $\hat{n}_R$, of n_R can be obtained as follows:

1. Draw a random sample of N^{Sample} 30-second Regimes. Each such sample requires 27.5 minutes of preceding loop data in order to calculate Regime membership.
2. Compute $n_R^{Sample} = \left|\left\{ R_l \mid R_l = R, \forall l \in N^{Sample}; R \in \mathbf{R} \right\}\right|$, We note $\sum_{R \in \mathbf{R}} n_R^{Sample} = N^{Sample}$.
3. Compute the frequency of occurrence of Regime R in the sample, $f_R^{Sample} = n_R^{Sample} \big/ N^{Sample}$.
4. Compute an estimate of n_R as $\hat{n}_R = f_R^{Sample} \cdot N_{TS} = n_R^{Sample} \cdot N_{TS} \big/ N^{Sample}$.

Now, from the analysis provided in the previous sections, an output of FITS is the distribution of accident typologies (for accidents contained in the database on which the analysis was performed) relative to the various Regimes that were identified by the analysis. Specifically, it is possible to assign each of the specific accident typologies (e.g., type, location severity) of each of the accidents contained in the database to a particular Regime. So, for example, we can compute from the accident database and the analysis results:

$$f_{CR}^{base} = \frac{N_{CR}^{base}}{N_C^{base}}$$

where

$f_{CR}^{base} =$ frequency distribution of database accidents of typology C relative to Regime R,

$N_{CR}^{base} =$ Total number of database accidents of typology C assigned to Regime R by FITS, and

$N_C^{base} =$ Total number of database accidents of typology C.

From the TASAS database, it is possible to identify the total number of accidents of typology *C* that have occurred on any freeway segment *S* during a specified time interval *T* (e.g., number of fatal collisions on I-5 in Orange County during the morning peak period of the year 1998), say N_{CTS}. Then, $f_{CTS} = N_{CTS} \big/ N_{TS}$ is the frequency distribution of accidents of typology *C* per 30-second loop count occurring on freeway segment *S* during time *T*. And, $r_R^C = f_{CR}^{base} \cdot N_{CTS} / \hat{n}_R = f_{CR}^{base} \cdot f_{CTS} \cdot N_{TS} / \hat{n}_R$ is an estimate of the expected number of accidents of typology *C* per occurrence of Regime *R* on freeway segment *S* during time *T*. Finally, an estimate of the expected number of accidents of typology *C*, $\hat{N}^{C}_{accident}$, is given by:

$$\hat{N}^{C}_{accident} = \sum_{R} r^{C}_{R} \cdot \hat{n}_{R}$$

7.1 Demonstration Example

As a demonstration of the procedure outlined above, we consider accidents occurring during the morning peak hours on the six major freeways in Orange County, CA, using the year 1998 as a base. There are a total of 551 loop stations on these freeways, distributed as shown in Table 32.

Table 32 Loop Stations on Six Orange County Freeways

Freeway	# loops
SR-22E	24
SR-22W	25
SR-55N	31
SR-55S	26
SR-57N	21
SR-57S	18
SR-91E	43
SR-91W	41
I-405N	55
I-405S	53
I-5N	110
I-5S	104
Total	551

As defined in Section 6, the weekday morning peak comprises 6:00AM to 9AM inclusive (a total of 10,800 sec/day). Hence, for the year, there are

$$N_{TS} = \frac{nT}{30\text{sec}} = \frac{(551\text{ loops})\cdot(10{,}800\text{sec})\cdot(260\text{ weekdays})}{30\text{sec}} = 51{,}573{,}600 \text{ Regime occurrences.}$$

For purposes of this example, we make the simplifying assumption that all of these occurrences correspond to dry conditions. A total of $N^{Sample} = 895$ of the random sample of 30-second Regimes occurred during the dry weekday morning peak hours. The distribution of these among the eight Dry-Day Regimes is given in Table 33.

Table 33 Distribution of Dry-Day Regimes in the Random Sample

R	n_R^{Sample}	$\hat{n}_R = n_R^{Sample} \cdot N_{TS} / N^{Sample}$
D1	113	6,511,527
D2	35	2,016,845
D3	43	2,477,838
D4	186	10,718,089
D5	47	2,708,334
D6	198	11,409,579
D7	209	12,043,444
D8	64	3,687,945
N^{Sample}	895	N_{TS} = 51,573,600

The distribution of crash types in the analysis database with respect to the eight Dry-day Regimes is given in Table 34.

Table 34 Distribution of Crash Type with respect to the eight Dry-Day Regimes

Crash Type	Dry Day Regimes								Total	%
	D1	D2	D3	D4	D5	D6	D7	D8		
single veh hit object	29	15	24	46	56	72	49	22	313	38.2
2 +veh hit object	18	9	18	28	31	50	24	8	186	22.7
2 veh lane-change	16	17	12	32	13	21	17	18	146	17.8
3 +veh lane-change	3	7	7	10	4	3	4	4	42	5.13
2 veh rear-end	1	23	15	23	2	6	2	14	86	10.5
3 + veh rear-end	1	14	5	9	2	7	3	5	46	5.62
Total	68	85	81	148	108	159	99	71	819	
%	8.3	10.3	9.89	18.0	13.1	19.4	12.0	8.67		100

The distribution of crash severity in the analysis database with respect to the eight Dry-day Regimes is given in Table 35.

Table 35 Distribution of Crash Severity with respect to the eight Dry-Day Regimes

Crash Severity	Dry Day Regimes								Total	%
	D1	D2	D3	D4	D5	D6	D7	D8		
Property damage	59	53	61	105	85	124	75	52	614	74.9
Injury	9	32	20	43	23	35	24	19	205	25.0
Total	68	85	81	148	108	159	99	71	819	
%	8.30	10.3	9.89	18.0	13.1	19.4	12.0	8.67		100

Calculations of f_{CR}^{base} (Table 36 and 37) may be obtained directly from Tables 34 and 35.

Table 36 $f_{CrashType,R}^{base}$ for Crash Type for the eight Dry-day Regimes

Crash Type	Dry Day Regimes							
	D1	D2	D3	D4	D5	D6	D7	D8
single veh hit object	0.093	0.048	0.077	0.147	0.179	0.230	0.157	0.070
2 +veh hit object	0.097	0.048	0.097	0.151	0.167	0.269	0.129	0.043
2 veh lane-change	0.110	0.116	0.082	0.219	0.089	0.144	0.116	0.123
3 +veh lane-change	0.071	0.167	0.167	0.238	0.095	0.071	0.095	0.095
2 veh rear-end	0.012	0.267	0.174	0.267	0.023	0.070	0.023	0.163
3 + veh rear-end	0.022	0.304	0.109	0.196	0.043	0.152	0.065	0.109

Table 37 $f_{CrashSeverity,R}^{base}$ for Crash Severity for the eight Dry-Day Regimes

Crash Severity	Dry Day Regimes							
	D1	D2	D3	D4	D5	D6	D7	D8
Property damage	0.096	0.086	0.099	0.171	0.138	0.202	0.122	0.085
Injury	0.044	0.156	0.098	0.210	0.112	0.171	0.117	0.093

There were a total of $N_{CTS} = 9{,}341$ reported crashes on the six major Orange County freeways during 1998. Of these, 1,639 occurred during the AM weekday peak hours between 6:00 am and 9:00 am. The distributions of these crashes by typology (for accident type and accident severity) are given below in Tables 38 and 39.

Table 38 $f_{CTS} = N_{CTS}/N_{TS}$ for Crash Type for the eight Dry-Day Regimes

Crash Type	Frequency	$f_{CTS} = N_{CTS}/N_{TS}$
single veh hit object	102	1.97776E-06
2 +veh hit object	47	9.11319E-07
2 veh lane-change	310	6.01083E-06
3 +veh lane-change	90	1.74508E-06
2 veh rear-end	671	1.30105E-05
3 + veh rear-end	419	8.12431E-06
Total	1,639	

Table 39 $f_{CTS} = N_{CTS}/N_{TS}$ for Crash Severity for the eight Dry-Day Regimes

Crash Severity	Frequency	$f_{CTS} = N_{CTS}/N_{TS}$
Property damage	1289	2.49934E-05
Injury	350	6.78642E-06
Total	1639	

From Tables 33, 36, 37, 38 and 39, we can calculate the respective $r_R^C = f_{CR}^{base} \cdot f_{CTS} \cdot N_{TS} / \hat{n}_R$. These probabilities are listed in Tables 40 and 41, for crash type and crash severity, respectively.

Finally, from Tables 40 and 41, we calculate $\hat{N}_{accident}^{C} = \sum_R r_R^C \cdot \hat{n}_R$ and their expected distribution across the various Regimes. These distributions are listed in Tables 42 and 43.

Table 40 $r_R^C = f_{CR}^{base} \cdot f_{CTS} \cdot N_{TS} / \hat{n}_R$ for Crash Type for the eight Dry-Day Regimes

Crash Type	Dry day Regimes							
	D1	D2	D3	D4	D5	D6	D7	D8
single veh hit object	1.46E-06	2.43E-06	3.17E-06	1.40E-06	6.74E-06	2.06E-06	1.33E-06	1.94E-06
2 +veh hit object	7.00E-07	1.12E-06	1.84E-06	6.62E-07	2.90E-06	1.11E-06	5.03E-07	5.48E-07
2 veh lane-change	5.24E-06	1.78E-05	1.03E-05	6.33E-06	1.02E-05	3.91E-06	2.99E-06	1.03E-05
3 +veh lane-change	9.81E-07	7.45E-06	6.07E-06	2.00E-06	3.16E-06	5.60E-07	7.10E-07	2.32E-06
2 veh rear-end	1.24E-06	8.88E-05	4.71E-05	1.67E-05	5.70E-06	4.12E-06	1.28E-06	2.97E-05
3 + veh rear-end	1.42E-06	6.32E-05	1.84E-05	7.66E-06	6.65E-06	5.58E-06	2.26E-06	1.24E-05

Table 41 $r_R^C = f_{CR}^{base} \cdot f_{CTS} \cdot N_{TS} / \hat{n}_R$ for Crash Severity for the eight Dry-Day Regimes

Crash Severity	Dry day Regimes							
	D1	D2	D3	D4	D5	D6	D7	D8
Property damage	1.90E-05	5.50E-05	5.15E-05	2.06E-05	6.57E-05	2.28E-05	1.31E-05	2.97E-05
Injury	2.37E-06	2.71E-05	1.38E-05	6.86E-06	1.45E-05	5.25E-06	3.40E-06	8.83E-06

Table 42 $\hat{N}_{accident}^C = \sum_R r_R^C \cdot \hat{n}_R$ for Crash Type for the eight Dry-Day Regimes

Crash Type	Dry day Regimes								Total
	D1	D2	D3	D4	D5	D6	D7	D8	
single veh hit object	9	5	8	15	18	23	16	7	102
2 +veh hit object	5	2	5	7	8	13	6	2	47
2 veh lane-change	34	36	25	68	28	45	36	38	310
3 +veh lane-change	6	15	15	21	9	6	9	9	90
2 veh rear-end	8	179	117	179	15	47	15	109	670
3 + veh rear-end	9	127	46	82	18	64	27	46	419
Total	72	365	215	373	96	198	109	211	1,638

Table 43 $\hat{N}^{C}_{accident} = \sum_{R} r^{C}_{R} \cdot \hat{n}_{R}$ for Crash Severity for the eight Dry-Day Regimes

Crash Severity	Dry day Regimes								Total
	D1	D2	D3	D4	D5	D6	D7	D8	
Property damage	124	111	128	220	178	260	157	110	1,288
Injury	15	55	34	74	39	60	41	33	350
Total	139	165	162	294	217	320	198	142	1,638

The row totals here, by definition, match the observed values; the categorizations by Regime are products of FITS. However, the model may also be used in a forecasting mode to estimate expected modifications in safety outcomes accrued from changes in flow patterns, say through reducing congestion by ramp metering.

7.2 Hypothetical Scenario

Let us suppose that, through traffic control measures, we were able to virtually eliminate the two "congested flow" Regimes (D2 and D3), transferring these previously congested periods to the "heavy, steady flow" Regime D7. The expected distribution of Dry-Day Regimes under this scenario is shown in the third column of Table 44.

Table 44 Forecast Distribution of Dry-Day Regimes

R	$n_R^{Existing}$	$\hat{n}_R^{Forecast}$
D1	6,511,527	6,511,527
D2	2,016,845	0
D3	2,477,838	0
D4	10,718,089	10,718,089
D5	2,708,334	2,708,334
D6	11,409,579	11,409,579
D7	12,043,444	16,538,127
D8	3,687,945	3,687,945
N_{TS}	51,573,600	51,573,600

The corresponding forecast expected accident distributions under these new traffic flow conditions (i.e., the revised Tables 42 and 43) are shown in Tables 45 and 46.

Table 45 Forecast $\hat{N}^{C}_{accident} = \sum_{R} r^{C}_{R} \cdot \hat{n}_{R}$ for Crash Type for the eight Dry-Day Regimes

Crash Type	Dry day Regimes								Total
	D1	D2	D3	D4	D5	D6	D7	D8	
single veh hit object	9	0	0	15	18	23	22	7	95
2 +veh hit object	5	0	0	7	8	13	8	2	42
2 veh lane-change	34	0	0	68	28	45	49	38	262
3 +veh lane-change	6	0	0	21	9	6	12	9	63
2 veh rear-end	8	0	0	179	15	47	21	109	380
3 + veh rear-end	9	0	0	82	18	64	37	46	256
Total	72	0	0	373	96	198	150	211	1,099

Table 46 Forecast $\hat{N}^{C}_{accident} = \sum_{R} r^{C}_{R} \cdot \hat{n}_{R}$ for Crash Severity for the eight Dry-Day Regimes

Crash Severity	Dry day Regimes								Total
	D1	D2	D3	D4	D5	D6	D7	D8	
Property damage	124	0	0	220	178	260	216	110	1,108
Injury	15	0	0	74	39	60	56	33	277
Total	139	0	0	294	217	320	272	142	1,385

We note that, when applied in a forecast mode, FITS does not guarantee consistency between typologies (e.g., the total number of crashes forecast by type of crash typology, 1,099, does not match the corresponding forecast by crash severity typology, 1,385). This is because the membership functions for each typology were determined independently; resolving such inconsistency through a combined analysis (e.g., by a two-dimensional classification scheme, such as crash type **and** severity) both presents methodological difficulties and could not be supported by the sample data that was available for the present study. However, the current hypothetical example provides a first-order, albeit crude, approximation in the form of a demonstration of the potential application of FITS. We conclude this hypothetical example by displaying, in Tables 47 and 48, the summaries of improvements in safety that would be expected under the above scenario:

Table 47 Expected Change in Accident Occurrences by Crash Type

Crash Type	Existing	Forecast	Change
single veh hit object	102	95	-7
2 +veh hit object	47	42	-5
2 veh lane-change	310	262	-48
3 +veh lane-change	90	63	-27
2 veh rear-end	670	380	-290
3 + veh rear-end	419	256	-163
Total	1,638	1,099	-539

Table 48 Expected Change in Accident Occurrences by Crash Severity

Crash Severity	Existing	Forecast	Change
Property damage	1,288	1,108	-180
Injury	350	277	-74
Total	1,638	1,385	-253

8 Conclusions and Directions for Further Research

We have developed a tool, called FITS (Flow Impacts on Traffic Safety), that can be used to assess the changes in traffic safety that result from changes in traffic flow. The only input that FITS requires is a stream of 30-second observations from single inductive loop detectors. FITS can be used as part of any evaluation that compares before and after traffic flow data, as measured by single loop detectors. Such an evaluation might involve assessing the benefits of ATMS operations.

FITS applies only to urban freeways with at least three lanes in each direction. In particular, the statistical models that underlie the tool have been estimated using historical data for freeways in Orange County, California. We presume that the relationships uncovered are indicative of all California urban freeways, particularly those in the San Francisco Bay, San Diego, and Sacramento Metropolitan Areas, but validation has not yet been conducted, so we cannot confirm the degree of spatial transferability.

FITS has its limitations. First, due to the quality of the historical loop detector data that were used in calibrating the tool, we were unable to include crash rates as a function of vehicle miles of travel, as originally intended. The historical traffic flow data were not sufficiently representative of Orange County for an entire year, because there were systematic patterns in missing data as a function of freeway route, location along each route, day of week, and week of the year. Thus, we were unable to accurately calculate the rates, in terms of vehicle miles of travel, for crashes that happened to vehicles that were exposed to different traffic flow conditions. Consequently, FITS provides information as to which types of crashes are more likely under different types of traffic flow, but does not forecast crash rates. The enhancement of FITS to include crash rates as well as types is an important subject for future research.

In spite of these limitations, we believe that we have demonstrated that FITS can be used to gain insight into how changing traffic flow conditions affect traffic safety. To the extent that changed conditions are due to ATMS operations, or other projects that influence traffic operations, FITS can be used in evaluating the effectiveness of such projects. FITS can also be used as a forecasting tool combined with simulation studies of the likely future conditions; FITS can be used to evaluate the safety conditions of alternative scenarios of operations with different ATMS or infrastructure treatments. Due to the problem with missing traffic flow data for 1998, it is strongly recommended that FITS be re-calibrated with more recent accident and traffic flow data before any large-scale deployment of this tool.

References

Aljanahi, A.A.M., A.H. Rhodes and A.V. Metcalfe (1999). Speed, speed limits and road traffic accidents under free flow conditions. *Accident Analysis and Prevention*, 31: 161-168.

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

Cedar, A. and L. Livneh (1982). Relationship between road accidents and hourly traffic flow. *Accident Analysis and Prevention*, 14: 19-44.

De Leeuw, J. (1985). The Gifi system of nonlinear multivariate analysis. In E. Diday, et al., eds., *Data Analysis and Informatics, IV: Proceedings of the Fourth International Symposium.* North Holland, Amsterdam.

Dickerson, A., J. Peirson and R. Vickerman (2000). Road accidents and traffic flows: An econometric investigation. Economica, 67: 101-121.

Frantzeskakis, J.M. and D.I. Iordanis (1987). Volume-to-capacity ratio and traffic accidents on interurban four-lane highways in Greece. *Transportation Research Record,* No. 1112: 29-38.

Fridstrøm, L., J. Ifver, S. Ingebrigtsen, R. Kulmala and L.K. Thomsen (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis and Prevention*, 27: 1-20.

Garber, N.J. and R. Gadiraju (1990). Factors influencing speed variance and its influence on accidents. *Transportation Research Record* 1213: 64-71.

Gifi, A. (1990). *Nonlinear Multivariate Analysis*. Wiley, Chichester.

Gwynn, D.W. (1967). Relationship of accident rates and accident involvement with hourly volumes. *Traffic Quarterly*, 21: 407-418.

Hall, J.W. and O.J. Pendleton (1989). Relationship between V/C ratios and Accident Rates. Report FHWA-HPR-NM-88-02, U.S. Department of Transportation, Washington, DC.

Israëls, .Z. (1987). *Eigenvalue Techniques for Qualitative DATA.* DSWO Press, Leiden.

Jansson, J.O. (1994). Accident externality charges. Journal of transport Economics and Policy, 28: 31-43.

Johansson, P. (1996). Speed limitation and motorway casualties: A time series count data regression approach. *Accident Analysis and Prevention*, 28: 73-87.

Jones-Lee, M.W. (1990). The value of transport safety. Oxford Review of Economic Policy, 6: 39-60.

Maher, M.J. and I. Summersgill (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis and Prevention*, 28: 281-296.

Mensah, A. and E. Hauer (1998). Two problems of averaging arising from the estimation of the relationship between accidents and traffic flow. *Transportation Research Record* 1635: 37-43.

Michailidis, G. and J. de Leeuw (1998). The GIFI system of descriptive multivariate analysis. Statistical Science, 13: 307-336.

Newberry, D. (1988). Road user charges in Britain. *Economic Journal*, 98: 161-176.

O'Reilly, D., J. Hopkin, G. Loomes, M. Jones-Lee, P. Philips, K. McMahon, D. Ives, B. Sobey, D. Ball and R. Kemp (1994). The value of road safety: UK Research on the valuation of preventing on-fatal injuries. *Journal of Transport Economics and Policy,* 28: 45-59.

Sandhu, B. and J. Al-Kazily (1996). Safety impacts of freeway traffic congestion. Presented at Annual Meeting of Transportation Research Board, January 7-11, Washington, DC.

Shefer, D. and P. Rietveld (1997). Congestion and safety on highways: Towards an analytical model. Urban Studies, 34: 679-692.

Stokes, R.W. and M.I. Mutabazi (1996). Rate-quality control method of identifying hazardous road locations. *Transportation Research Record,* No. 1542: 44-48.

Sullivan, E.C. (1990). Estimating accident benefits of reduced freeway congestion. *Journal of Transportation Engineering*, 116: 167-180.

Sullivan, E.C. and C-I. Hsu (1988). *Accident rates along congested Freeways*. Research Report UCB-ITS-RR-88-6, Institute of Transportation Studies, University of California, Berkeley.

Ter Braak, C.J.F. (1990). Interpreting canonical correlation analysis through biplots of structure correlations and weights. *Psychometrika*, 55: 519-531.

van Buren, S. and W.J. Heiser (1989). Clustering N-objects into K-groups under optimal scaling of variables. *Psychometrika*, 54: 699-706.

van de Geer, J.P. (1986). Relationships among *k* sets of variables, with geometrical representation, and applications to categorical variables. In J. de Leeuw, *et al.*, (Eds.), *Multidimensional Data Analysis,* DSWO Press, Leiden.

van der Burg, E. and J. de Leeuw (1983). Non-linear canonical correlation. *British Journal of Mathematical and Statistical Psychology,* 36: 54-80.

ver Boon, P. (1996). *A Robust Approach to Nonlinear Multivariate Analysis*. DSWO Press, Leiden.

Vickrey, W. (1969). Congestion theory and transport investment. *American Economic Review,* 59: 251-260.

Vitaliano, D.F. and J. Held (1991). Road accident external effects: An empirical assessment. *Applied Economics*, 23: 373-378.

Zhou, M. and V.P. Sisiopiku (1997). Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record,* No. 1581: 47-52.