UMTRI - 97-39

Modeling Intersection Crash Counts and Traffic Volume

FINAL REPORT

prepared for the
Federal Highway Administration
under

Evaluation of Exposure Data Sources for Highway Safety
Contract No. DTFH61-93-C00123

Hans C. Joksch
Lidia P. Kostyniuk

September 1997

University of Michigan
Transportation Research Institute
2901 Baxter Road
Ann Arbor, Michigan 48109-2150

FOREWORD

A. George Ostensen
Director, Office of Safety and
Traffic Operations Research
and Development

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturers' names appear in this report only because they are considered essential to the object of this document.

031065

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-98-096		

4. Title and Subtitle	5. Report Date
MODELING INTERSECTION CRASH COUNTS AND TRAFFIC VOLUME	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Hans C. Joksch and Lidia P. Kostyniuk	UMTRI-97-39

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
The University of Michigan Transportation Research Institute 2901 Baxter Road Ann Arbor, MI 48109	NCP:3A3B
	11. Contract or Grant No. DTFH61-93-C-00123

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Safety and Traffic Operations R&D Federal Highway Administration 6300 Georgetown Pike McLean, Virginia 22101-2296	November 1995 - March 1997
	14. Sponsoring Agency Code

15. Supplementary Notes

Contracting Officers Technical Representative (COTR): Joe Bared, HSR-20

16. Abstract

This research explored the feasibility of modeling crash counts at intersections with use of available exposure measures. The basic purpose of "exposure" is to serve as a size factor to allow comparison of crash counts among populations of different sizes. In the context of highway crash studies, at first glance, vehicle miles of travel (VMT) appears to be a natural exposure measure. However, VMT is closely related to traffic density and this raises doubts if it can serve the intended purpose of an exposure measure. Data from four-leg signalized intersections in Washtenaw County, Michigan, and the states of California and Minnesota were used in this study. Traffic volumes on the approaches are the routinely available exposure measure. It was noted that in these data sets the same values of traffic volume were often "carried over" several intersections. Using such values of traffic volume as measures of exposure results in correlations between errors of the independent variables, which violates the requirements of standard statistical procedures.

It was found that the relationships between crash counts and traffic volumes could not be adequately represented by the standard loglinear model that is also the basis for more sophisticated models. Therefore, nonparametric regression in the form of kernel smoothing was used. This allowed a realistic representation of complex relationships. The relationships found differed among the three data sets, with California showing dramatic deviations from the loglinear model. The relationship between crash counts and traffic volumes on the approaches to the intersection and those within the intersection were found to be very different. This makes it unlikely that realistic models for all intersection-related crashes can be developed. Turning counts are plausible candidates for exposure measures for turn-related intersection crashes. However, since turning counts are not routinely available, the possibility of using proportions of crashes involving turns was explored. The results were negative, but because of the small case number, no definite conclusions could be drawn.

17. Key Words	18. Distribution Statement
Intersection crashes, exposure, intersection characteristics, smoothing	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	174	

Form DOT F 1700.7 Reproduction of completed page authorized

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
	LENGTH			
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
	AREA			
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^3	square miles	2 59	square kilometers	km^2
	VOLUME			
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3 785	liters	L
ft^3	cubic feet	0 028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
	MASS			
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
	TEMPERATURE (exact)			
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	°C
	ILLUMINATION			
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
	FORCE and PRESSURE or STRESS			
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
	LENGTH			
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
	AREA			
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
	VOLUME			
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
	MASS			
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
	TEMPERATURE (exact)			
°C	Celcius temperature	1.8C + 32	Fahrenheit temperature	°F
	ILLUMINATION			
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
	FORCE and PRESSURE or STRESS			
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised September 1993)

Table of Contents

Tables

Figures

1. Modeling intersection crash counts in relation to exposure

1.1 The concept of exposure

Comparing the annual number of crashes in California with that in Rhode Island simply shows the obvious consequence of California being much larger than Rhode Island, with many more motor vehicles that together travel many more miles. A comparison that reveals something more has to account for the difference in scale. "Exposure" is such a scale factor. Dividing crash counts by an exposure measure gives an indication of crash risk relative to that exposure measure.

Sometimes registered vehicle-years or insured vehicle-years are used as exposure measures. Such measures reduce the effect of large discrepancies in "size" between states, or between various populations of vehicles, such as vehicle types, or vehicle makes and models. However, they are not fully satisfactory measures of exposure because they do not control for differences in the annual miles traveled by different vehicles. Therefore, vehicle miles of travel (VMT) is a preferred exposure measure.

VMT is a plausible exposure measure for studies of some crash types, such as running off the road or collisions with a roadside object. However, for other types of crashes, such as head-on collisions, VMT is not a good exposure measure. In the case of head-on crashes, an "exposure" to a collision is present only when a vehicle encounters an oncoming vehicle. The number of such encounters on a segment of highway will change proportionately to the square of VMT, not proportionately to VMT. On the other hand, if we compare several states with the same ratio of VMT to highway miles, then the number of such encounters will be proportional to the numbers of VMT in the states.

The situation is different if specific locations on highways, such as intersections rather than aggregations of highways or larger areas, are considered. In principle, VMT within the intersection could be defined as an exposure measure. This would imply that the expected number of crashes should increase proportionately to the width of the intersecting highways. An alternative to this assumption is to use the number of vehicles entering an intersection as the exposure measure, and to include the width (or the number of lanes as an indicator of width) among the other factors to be studied. Thus, for a n-leg intersection, the exposure measure would have n components.

First consider only intersections with crossing traffic streams and no turning maneuvers. In this case, the two traffic volumes on the crossing roads would suffice as exposure measures. If the intersection is uncontrolled, then the number of "encounters" between vehicles, where a crash could occur, is proportional to the product of the two volumes.

However, in reality, vehicles turn at intersections and the numbers of the different types of turns determine the encounters in which certain types of crashes can occur. The situation is even more complicated when traffic at the intersection is controlled by signs or a traffic signal. In that case the number of encounters within the intersection is reduced, but the number of different types of encounters during which rear-end collisions can occur just outside the intersection is increased. The magnitude of this shift depends on the lengths of the phases of the traffic signal. Even in the simplest situation, it cannot be expected that the number of encounters, which represent exposure to the possibility of a crash, can be represented by relatively simple mathematical functions of the entering traffic volumes.

Thus, ideally, exposure measures should be related to the potential conflicts between vehicles in the intersection, i.e., situations where more than one vehicle can occupy the same space at the same time.[1] In this study, we explore what can be achieved in modeling intersection crash counts using the readily available "exposure" measures of traffic volumes on the crossing roads.

1.2 The purposes of modeling intersection crash counts

The two major purposes for using models of intersection crash counts as functions of traffic volumes and intersection characteristics are:

- to determine how the crash risk at intersections depends on various intersection characteristics, especially those that can be modified.

---

[1]W.D. Glauz, D.I. Migletz, *Application of Traffic Conflict Analysis at Intersections*. Transportation Research Board, NCHRP Report 219, 1980

- to determine whether a specific intersection has an "unusually" bad crash experience.

For the first purpose, a population of intersections is selected and each intersection is treated as an "observation." The dependent variable is the crash count for some period of time, usually a year, and the independent variables consist of intersection characteristics, including traffic volumes on the approaches to the intersection. Of special interest are those characteristics that can be modified to reduce the crash risk at the intersection. Statistical methods are used to fit a model approximating the crash counts by a function of the independent variables. The form of the function is usually assumed on the basis of mathematical convenience, and typically is not empirically determined.

For the second purpose, a model is applied to an individual intersection and the expected number of crashes is predicted and compared against the actual number of crashes at the intersection. The difference between them is examined to determine if it is greater than that expected from random variations. If so, the intersection is studied to identify factors responsible for the elevated risk.

When applying the model to determine if the intersection has an "unusually" bad crash experience, a distinction has to be made whether this intersection was or was not used to develop the model.

If the intersection was used in the development of the model and "all" relevant factors were included in the model, and the mathematical form of the model is correct, then the crash counts at any individual intersection should not differ from the modeled value by more than random variation. In this case any differences between modeled and the actual values cannot be attributed to the factors included in the model. Thus, if an intersection's crash count differs from the modeled value by more than random variation, either the model is incomplete or it is mathematically incorrect. If this occurs the model should be revised. Factors other than those included in the original model must be sought to explain the discrepancy, or a better mathematical form of the model must be found.

Unfortunately, there can be situations where influential factors are not included in the model or where the mathematical form is not correct, yet they are accepted as correct.

For example, the data may be configured in a way that allows the model fitting process to include data points that should be outliers. This will bias, possibly dramatically, other coefficients of the model.

These problems do not arise if the model is applied to an intersection that was not in the population used to develop the model. However, different problems can arise. Consider the situation where the assumed mathematical form of the model, while not correct, may still be good enough to represent the data over the range of the independent variables in the data set. If the "new" intersection is outside, or possibly just within the range, of the independent variables used in the development of the model, the actual crash count can deviate substantially from the modeled value. This occurs because model errors tend to increase toward the limits of the range over which it was calibrated, and may "explode" beyond it. Trying to explain such differences between new cases and an apparently satisfactory model can give completely wrong results.

These examples should serve as a warning that it may not be possible to achieve the goal of modeling intersection crashes.

1. 3 Some critical assumptions

The difficulties discussed in the previous section can arise when applying models. However, technical problems also can arise when basic assumptions of the modeling process are not satisfied in the development of models.

One important point to remember is that safety features often are installed in response to a perceived "crash problem" that has been identified by high crash counts or by large deviations of crash counts from a model. Such safety features should be included among the variables describing intersection characteristics. Depending on the overall quality of the model, in terms of completeness and realism of its mathematical form, this can distort the model coefficients, and can sometimes show a crash increase effect from the safety feature.

Other problems result from aggregation of data. The mathematical relationship between crashes within the intersection and traffic volumes and other intersection characteristics may be very different from the mathematical relationship between

crashes on the approaches to the intersection and traffic volumes and intersection characteristics. However, crashes within the intersection and on its approaches are usually aggregated together as intersection crash counts because both are "intersection related." Even if the intersection and intersection approach crashes followed simple relationships such as

$$a*x^{b}*y^{c} \qquad (1\text{-}1)$$

where x and y are the volumes on the intersecting roads, their sum will usually not follow such a relationship. The same holds for aggregation across different crash types.

The daily (and possibly weekly and seasonal) variation of traffic volumes causes another problem. For instance, let $x_i ... x_n$ and $y_i ... y_n$ be the traffic volumes during n time intervals, e.g., of the day, and let x and y be their totals. Assume that for each time interval a relationship

$$z_i = a * x_i * y_i \qquad (1\text{-}2)$$

holds. Generally, a corresponding relationship

$$\sum z_i = A * \sum x_i * \sum y_i \qquad (1\text{-}3)$$

will not hold for the sums over the time intervals. Rather, the relationship for the total is

$$\sum z_i = a*(\sum x_i * \sum \frac{y_i}{n} + R * SQRT\,(\sum (x_i - \bar{x})^2 * \sum (y_i - \overline{y^2}))) \qquad (1\text{-}4)$$

where R is the correlation coefficient between the x_i and y_i. Usually, since traffic tends to be high, or low, on all approaches at the same time, R will be high, and the second term will not vanish. Thus, even if a relationship

$$z = a * x * y \qquad (1\text{-}5)$$

or a more complicated, similar relationship holds at any time, a similar one will not hold for aggregated traffic volumes if they vary and are correlated.

1.4 The conventional statistical approach

The conventional approach uses statistical techniques to fit an analytical model to the data. Examples of such models are:

$$a*\exp\{b_0+b_1*x_1+b_2*x_2+b_3*x_3\cdots\} \qquad (1\text{-}6)$$

or

$$a*x_1^{b_1}*x_2^{b_2}*x_3^{b_3}\cdots \qquad (1\text{-}7)$$

where x_1, x_2 and x_3 are variables describing intersection characteristics, including traffic volumes on the approaches and possibly volumes of turning traffic. For qualitative characteristics, "dummy" variables with values of 0 or 1 are used. Statistical techniques to fit such models include regression on transformed variables and maximum likelihood estimates.

If all assumptions underlying these statistical techniques are satisfied, valid estimates of the model parameters can be obtained and the effects of intersection characteristics on crash risk can be determined. However, some of the assumptions are often, if not always, violated. Basic assumptions are that the deviations between the model and the crash counts have expected values of zero and that they are independent. Both assumptions are violated if the model does not reflect the relationship correctly, which is likely, as discussed above. The second assumption is violated if the traffic passing through several intersections in the population is largely the same. Factors such as driver-age distribution, vehicle mix, and possibly trip purpose cause deviations from the average crash risk that are treated as random variations by the model. These deviations will be correlated over the intersections that are passed by largely the same traffic.

Systematic deviations between the assumed mathematical forms of the model and the actual relationship between crash counts and intersection parameters, especially volumes, can have a serious biasing effect. This can occur if the majority of intersections has medium volumes, but a few intersections have high volumes. A similar effect also can occur if there are a few rare intersections with low volumes. In such cases, the model parameters for the volumes will be determined primarily by the majority of the intersections and will provide a good fit in the area covered by them. If the assumed model is not correct, the predicted crash counts for high-volume intersections would show large, systematic deviations. However, if there are characteristics that are mainly present at high-volume intersections, then the statistical algorithms may use the coefficients of these characteristics to reduce the systematic deviation. In extreme cases, a very significant parameter coefficient may appear that depends on a single intersection. To what extent this actually occurs has to be established in each case by a very detailed analysis.

Standard statistical techniques assume that the independent variables are not subject to error. This does not hold for traffic volumes. Traffic volume data can be subject to large errors if special counts are not available. Often, values of average daily traffic (ADT) are carried over long distances of roadway and used for several adjacent intersections. In that case, the errors of the independent variables also will be correlated, complicating an already complex situation even further. For linear regression, the problem has been studied and suitable approaches have been developed. However, this still has to be done for the nonlinear models used for intersection modeling.

The simplest way of assessing the model fit is to test correlation coefficients or similar aggregate measures. This is wholly inadequate. First, the explicit or implicit null hypothesis of such tests is that there is no relationship between the parameters and crash counts. However, even if one of the included factors has a relationship with crash counts, the test will show a significant relationship, even though many terms of the model may contribute nothing but noise. To recognize this, more sophisticated tests, which separate the contributions of the independent variables, have to be applied. A second problem is that, even if a test shows a high significance level, the modeled relationship may have nonnegligible systematic errors.

Many of these problems can be reduced and sometimes even avoided by using better than run-of-the-mill statistical techniques. Serious difficulties, however, can remain if the actual relationship between crash counts and intersection characteristics cannot be expressed by manageable mathematical functions.

1.5 Smoothing techniques

The difficulties arising from relationships that cannot be described by simple mathematical functions can be avoided by using smoothing methods. Smoothing techniques are based on the mapping of the data onto an n+1 dimensional space (where n is the number of independent variables), selecting a grid of appropriate spacing and fitting relatively simple "local" functions to the data points "near" each grid point. The values of these local functions at the grid point or at actual data points are the smoothed values. Smoothing techniques do not provide a relationship in mathematical form and work only for continuous relationships, such as those between crash counts and traffic volumes. Categorical variables must be treated as additive or multiplicative terms, or by splitting the data into subsets.

Smoothing techniques are simple and the results can be easily interpreted if there are one or two continuous independent variables. Fitting the model is only slightly more difficult with more variables, but presenting and interpreting the results becomes complex. The results can be presented in a simple form only if the relations are additive or multiplicative with respect to the continuous variables.

Estimating errors for a smoothed model is more laborious than for traditional analytical models. One approach is to use the error estimate obtained when calculating each smoothed value. Another approach is to split the data set and derive separate models for each part, with the differences between them providing error estimates. Bootstrapping is a similar technique, with an additional feature that allows the incorporation of the effect of "influential" observations into the errors. These approaches give error estimates for each grid point, or each data point. This makes these error estimates more realistic, but more cumbersome, than those obtained from analytical models. It also is possible to define overall error estimates for a smoothed model. Significance testing, however, if at all possible, is very complex.

2. The smoothing technique used

The purpose of smoothing is to describe empirical data that cannot be adequately approximated by simple analytical functions of the independent variables. Although the technique can be used with any number of independent variables, in this study it was applied to the case of only two variables.

Let z_i be the value of the dependent variable and x_i and y_i be the values of the independent variables at data point i. The process of smoothing calculates the smoothed values ζ_j for a number of grid points j with coordinates ξ_j and η_j. The technique used is "kernel smoothing," which uses a weight function $w(d_{ij})$, where d_{ij} is a measure of a distance between an observation i and grid point j. The weight function is so defined that it is largest for a distance zero between grid point and observation, and declines with increasing distance.

The procedure fits a separate linear regression to each grid point, using the weight functions. This means that the model fits the data points near the grid points very well. while the fit at remote data points may not be as good. The value of this linear function at the grid point is the smoothed value. This procedure is repeated for each grid point, resulting in smoothed values at each grid point. In the case of two independent variables, the resulting smoothed function can be represented as a surface in a three-dimensional space by connecting the smoothed values with grid lines.

Using a linear regression at each point has the disadvantage that real nonlinearities, such as maxima, edges, valleys, etc., also are smoothed and flattened. This can be avoided, to some extent, by using a quadratic local fit. We experimented with quadratic fits, but the results were unsatisfactory. If the fit represented larger nonlinearities, it also was overly influenced by individual points with very large or very small values. However, if the influence of such points was reduced, the nonlinear features were obscured by excessive flattening.

We used the following Gaussian kernel as the weight function

$$w_{i,j}=\exp\left(-\left(\frac{x_i-\xi_j}{a}\right)^2-\left(\frac{y_i-\eta_j}{b}\right)^2\right) \qquad (2\text{-}1)$$

where we call (a x b) the size of the "window." The weight of a data point depends on the values of a and b; the larger a or b, the wider the range in x or y over which the data points have a large weight. The choice of the term "window" becomes clearer if we consider a generalization of the Gaussian kernel, where the exponent value of 2 is replaced by another number. Figure 1 shows the value of w_{ij} in equation (2-1) as a function of x_i and y_i, for fixed values of ξ_j and $\eta_{\cdot j}$, which represent the coordinates of the top of the surface. The values of a and b correspond to 4 and 2, respectively, the distances between grid lines. Figure 2 shows the values of w_{ij} with the exponent 2 replaced by 10. The weight function shows a very steep drop at distances a in the x-direction and b in the y-direction, with approximately the value 1 within the "frame" and zero outside the "frame." For larger values of the exponent, the drop becomes steeper, the frame narrower, and a moving average is approached. Therefore, (a x b) is called the size of the window. This is done for convenience since the actual size of the window is really ((2a) x (2b)).

Several choices have to be made when fitting a model. These include the density of the grid, the exponent, the size of the window, and the option of varying the orientation of the window. We experimented to determine which values of grid density and window size and orientation eliminated irregular waves and patterns, likely to be noise, but still retained features of the surface that may reflect real effects. Though statistical criteria can be developed, they are cumbersome, and we did not use them. We varied the grid size until a clear picture of a continuous surface was obtained visually. We also varied the window size until the surface appeared to be smooth, with only low local "waves." If the surface showed a "ridge" or similar pattern, we tried varying the orientation of the window. This, however, did little to reveal a pattern and we decided not to vary the orientation of the windows in this study.

The results of the smoothing process are shown as surfaces, formed by connecting the smoothed values at the grid points along the grid lines. The width of the lines is proportional to the number of cases in each cell. If the line was too narrow to be printable, it was printed as a dotted line. No grid lines are shown where there are no data points. Sometimes, a cell with no data points is surrounded by cells with data points. In that case two "sides" of adjacent cells will be seen through the "hole" in the surface at the location of the cell with no data points.

Although there is some theoretical work on smoothing and a body of experimental work exists,[2] additional development is desirable. For example, data points are usually not uniformly distributed over a surface. In areas covered densely with data points a small window suffices to average out the random variations, and reveal more details of the surface than a large window. However, in areas with sparse observations a large windows shows the general trend of the surface, but less detail. A procedure with adaptive window size, which would allow an analyst to make the best use of available information, still needs to be developed.

3. The data

3.1 Washtenaw County, Michigan

Data from intersections in Washtenaw County, Michigan, were taken from a report by the Ann Arbor - Ypsilanti Urban Area Transportation Study Committee, *Washtenaw County Intersection Crash Analysis,* dated April 1991. It contains data for 134 signalized urban or suburban intersections consisting of traffic volumes on the crossing streets, and crash counts for each of the years 1986-89. Only intersections with 10 or more crashes in at least one year were included. This biases the models derived from the data, and the findings should be interpreted with caution. In a few cases, data were not available for all four years. These counts were expanded by the appropriate factor to estimate a total for the four years.

A careful inspection of the traffic volumes showed that they often remained constant over several adjacent intersections on the same road. While it may be the case that traffic volume varies little over longer stretches of certain roads, this has to be established. Simply carrying data from counting sites over several intersections not only biases the traffic volumes used, but also introduces a correlation between the errors of the volumes of these intersections. This invalidates error estimates obtained by standard statistical methods.

Despite the limitations, the Washtenaw County data were studied because they were readily available early during the study.

---

[2]T.J. Hastie, R.J. Tibshirani, *Generalized Additive Models.* Chapman & Hall, 1990. J.S. Simonoff, *Smoothing Methods in Statistics.* Springer 1996

3.2 California data

Data for the four-leg signalized urban intersections in California were those used by the Mid-West Research Institute in its study, *Statistical Models of At-grade Intersection Crashes,* by K.M. Bauer and D.W. Harwood. The data were not modified or processed in any way.

3.3 Minnesota data

Minnesota data were obtained from the Highway Information System (HSIS) files. The Highway Safety Research Center of the University of North Carolina prepared the following files from the complete data set:

- An intersection route file, which contained one record for each route passing through a given intersection. This record contained information for both the incoming and outgoing legs. It also contained an intersection identification number that allowed matching with the record of the other route crossing the same intersection.

- A segment file, which contained cross-section data for each leg of each intersection. Legs were eliminated if they contained an extra intersection, including interchanges, grade crossings, etc. This was done to ensure that crashes linked with the legs are not related to any other intersection or interchange.

- A crash file, which contained records for all crashes anywhere on the legs retained, and corresponding vehicle and occupant files. Crash data for the year 1993 were used.

Route and segment files were processed to combine all data pertaining to a specific intersection into one record. Data for a total of 3,288 four-leg intersections were obtained. After selecting signal-controlled urban intersections that had volume information for all four legs, 71 intersections remained. Volume information was coded on both the route and segment files. The route file contained values of annual average daily traffic (AADT) for each of the two legs for up to five years. The segment file

contained one value of AADT and the calendar year to which it pertained. Review of these values showed that the information on the route file was often old or incomplete. Therefore, information from the segment file was used when available. Otherwise volume information from the route file was used.

A review of the volume data showed that the same values of volume were often carried over long distances and over several intersections. Furthermore, often the four volumes were so different that it appeared questionable whether they could result from physical maneuvers. Therefore, the analysis described in the next section was performed.

3.4 Identifying unrealizable intersection approach volumes

AADT values for all legs of many intersections were available in the Minnesota HSIS file. The question arose: Could all the given values have arisen from real traffic flows?

As an example, we considered an intersection with the AADT values 13247, 13247, 9488, and 13046 vehicles per day (vpd), shown in Figure 3A.1. At first glance it appeared questionable whether these values could have resulted from actual traffic: 13247 vpd would have been the traffic between the first and second leg, 9488 vpd could have been traffic from the third to the fourth leg, leaving 3556 vpd on the fourth leg, which could be explained only by 1779 vpd entering the intersection from the fourth leg, making a U-turn in the intersection, and leaving it on the fourth leg. This pattern is shown in Figure 3A.2 and is very implausible.

A slightly more complex traffic pattern, shown in Figure 3A.3, could explain the observed volume values. We assumed that 9488 vpd is the through traffic between the third and fourth leg. We interpreted the remaining volume of 3556 vpd on the fourth leg as a flow of 1778 vpd between the first and the fourth legs, and a flow of 1778 vpd between the second and fourth legs. A flow of 11469 vpd = 13247 vpd -1778 vpd between the first and the second leg completed the pattern. Thus, the observed volume values could have originated from actual traffic, though the actual traffic pattern is likely to be different from that used to demonstrate the feasibility.

Since it is not always obvious when a combination of volumes is realizable, a simple mathematical test was derived. We assume that only total volumes, without

distinguishing directions are given, that no legs are one-way, and that there are no turn restrictions. It also is possible to derive tests for such conditions, but they would be more complicated.

Figure 3 shows the flows we distinguish at a four-leg intersection. They must add up to the volumes u, v, w, and x on the legs:

$$f_1 + f_4 + f_5 = u \quad (3\text{-}1)$$

$$f_1 + f_2 + f_5 = v \quad (3\text{-}2)$$

$$f_2 + f_3 + f_6 = w \quad (3\text{-}3)$$

$$f_3 + f_4 + f_6 = x \quad (3\text{-}4)$$

For the four volumes u, v, w, and x to be realizable by actual traffic flows, the system of equations (3-1) through (3-4) must be solvable by nonnegative values f_i, (i = 1, 2, ..,6).

To derive a condition for that, we represent the pattern of Figure 3 in a different manner in Figure 4. We assume that a realizable solution exists, and that x is the smallest of the four volumes.

Figure 5 shows how the flows $f_1 ... f_4$ can be changed, so that a realizable solution remains. One can choose d so large that either f_1-d or f_3-d becomes zero, or that either f_4-d or f_1-d becomes zero. Since x was assumed to be the smallest of the four volumes, it can be shown that either f_3-d=0, or f_4-d=0 can be obtained. This gives a new flow pattern. The case where f_3 is reduced to zero is shown in Figure 6.

Again, the new flow pattern can be changed by adding and subtracting the same amount d to certain flows. This results in reducing f_4-d or f_5-d to zero, again because x is defined to be the lowest or one of the lowest volumes. Similarly, the other modifications of Figure 5 can be modified. All possible resulting flow patterns are shown in Figure 7.

These three patterns can be transformed into a single pattern. In the first case, one sets f_5=x, and replaces u by u_1=v=x. In the second case, one sets f_4=x and replaces u by u^1=u=x, and in the third case one sets f_3=x and replaces u by w^1=w-x Thus, all three cases can be represented by the simple problems shown in Figure 8, where one of the u^1,v^1,w^1 is the modified one, and the other two have the original values.

The conditions to be satisfied are:

$$f_1 + f_6 = u^1 \quad (3\text{-}5)$$
$$f_1 + f_2 = v^1 \quad (3\text{-}6)$$
$$f_2 + f_6 = w^1 \quad (3\text{-}7)$$

If they can be satisfied by nonnegative values f_1, f_2 and f_6, the original u, v, w, and x can result from real traffic flows. To obtain conditions for satisfying conditions (3-5)...(3-7) with nonnegative numbers, we solve for f_1, f_2, and f_6.

$$2f_1 = u^1 + v^1 - w^1 \quad (3\text{-}8)$$
$$2f_2 = u^1 + v^1 - w^1 \quad (3\text{-}9)$$
$$2f_6 = u^1 + v^1 - w^1 \quad (3\text{-}10)$$

Since the f_1 must be nonnegative, we have the conditions

$$u^1 + v^1 - w^1 \geq 0 \quad (3\text{-}11)$$
$$-u^1 + v^1 + w^1 \geq 0 \quad (3\text{-}12)$$
$$u^1 - v^1 + w^1 \geq 0 \quad (3\text{-}13)$$

The u^1, v^1, w^1 could have originated in three ways: by subtracting x from u, from v, or from w. Thus, we actually have three sets of conditions:

$$u + v - w \geq x \quad (3\text{-}14)$$
$$-u + v + w \geq -x \quad (3\text{-}15)$$
$$u - v + w \geq x \quad (3\text{-}16)$$

or

$$u + v - w \geq x \quad (3\text{-}17)$$
$$-u + v + w \geq x \quad (3\text{-}18)$$
$$u - v + w \geq -x \quad (3\text{-}19)$$

or

$$u + v - w \geq -x \quad (3\text{-}20)$$

$$-u+v+w \geq x \quad (3\text{-}21)$$

$$u-v+w \geq x \quad (3\text{-}22)$$

only one of which needs to be satisfied to ensure realizability of *u, v, w,* and *x.* It is possible to rename the variables so that

$$u \geq v \geq w \geq x \quad (3\text{-}23)$$

Then, $u+v \geq w+x$ holds. Therefore conditions (3-14) and (3-17) are satisfied, and since w+x ≥ -v, condition (3-20) also is satisfied. Since $u+w \geq v+x$, conditions (3-16) and (3-22) hold, and because $v+x \geq v-x$, condition (3-19) is satisfied. What remains is that one of the conditions (3-15), (3-18), or (3-21) must be satisfied. Since the last two are identical,

$$v+w \geq u-x \quad (3\text{-}24)$$

and

$$v+w \geq u+x \quad (3\text{-}25)$$

remain. It is sufficient that one of the two is satisfied. If condition (3-24) is satisfied, a realizable flow pattern exists. If condition (3-24) is not satisfied, then condition (3-25) cannot be satisfied. Thus, the only effective condition for reliability of volume by flow patterns is that

$$v+w+x \geq u \quad (3\text{-}26)$$

or that the largest volume on a leg must not be larger than the sum of the other three volumes.

The condition for realizable traffic volumes for three-leg intersections, obtained in the same manner, requires that the largest volume cannot be larger than the sum of the two smaller volumes.

The test was applied to four-leg intersections in the Minnesota data file. For signalized intersections, no unrealizable volume combinations were found. Among the 139 urban stop-controlled intersections with complete volume information, 2, or 1.4 percent had unrealizable volume combinations. Among the 438 rural stop-controlled intersections,

17, or 3.9 percent had unrealizable volume combinations. The data for these intersections are shown in Tables 3.4-1 and 3.4-2.

Table 3.4-1. Four-Leg Urban Stop-Controlled Intersections Violating Condition (3-26)				
Volume (vpd) on legs in decreasing order				Sum of volumes on all legs, h
u	v	w	x	h=u+v+w+x
3925	1847	1324	371	3542
2719	882	855	771	2508

Table 3.4-2. Four-Leg Rural Stop-Controlled Intersections Violating Condition (3-26)				
Volume (vpd) on legs in decreasing order				Sum of volumes on all legs, h
u	v	w	x	h=u+v+w+x
4720	2770	498	498	3766
14364	11902	1386	5	13293
5438	4566	351	80	4997
1077	739	177	134	1050
3642	1436	1125	940	3501
2514	1724	427	353	2504
3591	2308	661	605	3574
1488	811	296	35	1142
34136	13247	2641	1402	17290
2360	1539	567	144	2250
4617	3796	554	167	451
2719	253	229	150	632
1962	1539	129	45	1713
3386	1642	1081	559	3282
1744	1129	378	55	1562
22927	20380	1294	776	22450
1744	616	487	91	1194

A closer look at the tables shows that, in some cases, condition (3-26) is nearly satisfied, whereas in others the discrepancy is very large. If the AADT values were obtained from a single 24-hour count, a plausible first approximation is to assume them to be Poisson - distribution, for which an estimate of the variance is equal to the actual count. If a seasonal adjustment factor is applied, its square must be applied to the variance of the actual count. If the AADT is calculated from the average of n 24-hour

counts, then its variance is reduced by a factor of n compared with that obtained by a single 24 hour count. Therefore, without knowing the seasonal factor and the number of days of counting, we cannot realistically estimate the variance resulting from the random character of traffic. However, if one is willing to use the Poisson assumption for estimating the standard errors of an AADT figure, one can make some comparisons. In Table 3.4-1, the discrepancies are four and six times the standard deviation, which is too much to be acceptable as resulting from random variability.

In Table 3.4-2, the discrepancies range from 0.2 to 128 times the assumed standard deviation. In five cases the discrepancies are less than twice the assumed standard deviation, indicating that least some of them may be due to random fluctuations of the traffic count. However, in the remaining 12 cases, the discrepancies are unlikely to have resulted from random variations.

4. Selected Intersections in Washtenaw County, Michigan

The list of Washtenaw County intersections used for this analysis contained only intersections that had at least 10 crashes in any one of the years 1986-89. The exclusion of intersections with no or few crashes biases our findings. Therefore, this analysis should be considered as exploratory only. We used data that give a fairly smooth surface, primarily to illustrate some statistical aspects.

Figure 9 shows the distribution of traffic volumes for the 134 signalized four-leg intersections in Washtenaw County. Most are in the city of Ann Arbor, some in the city of Ypsilanti, and the others in essentially suburban areas. Two intersections with middle values of traffic volumes but extreme crash counts were omitted from the analysis.

4.1 Smoothing for signalized four-leg intersections

Figure 10 shows the results of smoothing with a (4,000 x 4,000) window. The surface shows some waves. The surface obtained with a wider window of (6,000 x 6,000), shown in Figure 11, appears smoother. Using a wider smoothing window resulted in a practically plane surface, which indicates that, over a wide range of traffic volumes, crashes vary essentially linearly with both the major and minor road volumes.

4.2 Analytical models

Because the relation in Figure 11 appeared so smooth and simple, attempts were made to fit analytical models to the data. The simplest model is linear in both volumes. A standard linear regression analysis resulted in

$$z=-2+2.07*x+3.78*y \qquad (4\text{-}1)$$
$$(7)\ (0.34)\ (0.51)$$

where standard errors are shown in parentheses below the coefficients, and the major volume x and minor volume y are in thousands. The intercept is so small and uncertain that for all practical purposes the function is a plane though the origin, that is, crashes increase independently and proportionally with x and with y. Figure 12 shows this plane.

The simple linear model is not commonly used because it does not reflect an interaction of two traffic streams and is therefore implausible. Most models used are refinements of the model

$$z=a*x^b y^c, \qquad (4\text{-}2)$$

which is often fitted to the data in the log-linear form such as,

$$log(z) = a+b*log(x)+c*log(y) \qquad (4\text{-}3)$$

Analytically, these models are equivalent. Statistically, however, they are different.

Fitting a model of type (4-3) by linear regression to the logarithm of the data gives

$$z = 11.6*x^{0.352}*y^{0.343} \qquad (4\text{-}4)$$
$$(9.4)(0.082)\ (0.052)$$

It is noteworthy that both exponents agree well within their errors. This means that the function is symmetric in both volumes. Figure 13 shows the corresponding surface.

Comparison with Figure 12 shows that the log-linear model "predicts" more crashes than the linear model for low volumes, and fewer crashes for high volumes.

Fitting a model of type (4-2) using a nonlinear regression routine results in the following:

$$z = 3.4 * x^{0.737} * y^{0.405} \qquad (4.5)$$
$$(1.1)\ (0.101)\ (0.057)$$

Here, the exponents of x and y differ well beyond their standard errors. The exponent of x differs very noticeably from that in model (4-4), whereas the exponent of y differs only by about the standard error of their difference, 0.077.

Figure 14 shows the surface representing model (4-5). Comparison with Figure 13 shows that for low volumes the log-linear model "predicts" more crashes, whereas it "predicts" fewer crashes for higher volumes. For actual intersections, the differences are large at the extreme values. In areas without intersections where the model truly "predicts," the differences are dramatic. This raises the question of what is the reason for these discrepancies between models of the same analytical form?

Assume that the value z_i of observation i has a standard error of σ_i. Then, the standard error of $u_i = \log(z_i)$ is $\tau = \sigma_i/z_i$. The corresponding weights to be used in the analyses are $v_i = 1/\sigma_i^2$ in the nonlinear regression and $w_i = 1/\tau_i = z_i^2/\sigma_i^2$ in the log-linear case. In the simple nonlinear regression, the standard errors of all observations are assumed to be equal. Similarly, in simple log-linear modeling one assumes that the standard errors of the logarithms of the observations are equal. If, in fact, the standard errors of the observations were equal, then the nonlinear regression techniques would provide the correct fit. If simple log-linear models were used, the result could be wrong, because wrong weights were used. The correct weights would have to be proportional to z^2, which means observations with high values of z are underweighted. The correct model could be obtained if a weight proportional to z^2 was used in the log-linear analysis. The reverse holds if, in fact, the variances of the logarithms of the observations are constant.

How does this relate to the well-known theorem that coefficients of weighted and unweighted models have the same expected values under fairly general conditions?

First, though the expected values may be the same, the actually computed values from a finite sample are usually different. Second, the theorem assumes implicitly that the model is "correct," i.e., that it has no systematic errors. In our situation, it is very likely that the arbitrarily selected simple analytical model has some systematic errors.

A practical question then is: What are the "correct" or at least "good" weights to be used in model fitting? The simplest model for crash counts is the Poisson distribution, which assumes that if a crash count z has an expected value m, then its variance is also m. We used this in the fourth model where we fitted a nonlinear model (4-2), but used weights proportional to 1/m. Since the value of m is not known, we proceeded interactively by starting with equal weights and fitting a model. We then used the fitted values of this model as weights for fitting a second model, and so on until the coefficients no longer changed. This procedure is asymptotically equivalent to fitting a maximum likelihood model to the data, using likelihood functions derived from the Poisson distribution.

The result of our iteratively weighted nonlinear modeling is

$$z = 7.0 * x^{0.517} * y^{0.309} \qquad (4\text{-}6)$$
$$(1.9)\ (0.091)\ (0.055)$$

The coefficients are between those for models (4-4) and 4-5). Figure 15 shows the corresponding surface.

This model should not be interpreted as being "better" or more "realistic" than the other models. The reason is that the Poisson assumption accounts for only the random component of the residual variance. Even here probably only part of the variation is accounted for because empirical distributions of crash counts tend to be broader than the Poisson - distribution. Very likely there may be systematic-error components, resulting from factors not included in the simplistic models, and from differences between the assumed simple relationship between volumes and crash counts and any "real" relationship. Such potential differences are the reasons why comparing overall measures of goodness of fit, such as R^2, or likelihood ratio are not sufficient. They rely heavily on the large number of cases concentrated in certain areas of the x-y plane, and relatively little on the few points outside of these areas. These points, however, may be of greatest interest because the differences between the various model are

greatest there. Therefore, in the following section we will visually look at the differences between the models, and between the models and actual data.

4.3 Visual comparison of actual data and models

Figures 16 and 17 show cross-sections through the surfaces shown in Figures 11, 12, 13, and 14, at minor volumes of 4,000 vpd and 14,000 vpd, and at major volumes of 16,000 vpd and 30,000 vpd. The locations for these cross-sections were chosen because they touched a wide range of cells that contained intersection data.

Figure 16 shows cross-sections across major volumes. There is a consistent discrepancy between the smoothed curve, approximating the actual values, and the analytical models, except the linear model, at a minor volume of 4,000 vpd. The smoothed curve is convex from below and the models are convex from above. For low values of the major volume, most models predictions are below the actual values; for middle values all are above the actual values; and for the highest values of major volume, all models are below the actual values. None of the models appear clearly superior. At a minor volume of 14,000 vpd, the smoothed curve is convex from below (except for a short piece at high volumes on the major road). All models show nearly straight lines. Again, none of them is consistently better than the others, and none is really "good."

Figure 17 shows cross-sections at a major volume of 16,000 vpd across minor volumes. The smoothed relationship is close to a straight line, the linear model is a nearly parallel line, and the nonlinear models are slightly convex from above. The portion of the line below 2,000 vpd should be ignored because there are only 10 intersections with minor volumes between 1,000 vpd and 2,000 vpd, and the curves rely heavily on extrapolation or are forced to zero at zero volume. Aside from this area, the modeled curves are above the smoothed curve nearly everywhere. In this case, the simple unweighted nonlinear model tends to be closest to the smoothed values. At a major volume of 30,000 vpd, the smoothed relationship shows a nearly linear increase with minor volume up to 14,000 vpd. It then levels off and increases above 20,000 vpd. If we again ignore minor volumes below 2,000 vpd, the log-linear model is always much too low, and the unweighted nonlinear model nearly always too high. Overall, the iteratively weighted nonlinear model appears to be the best.

No clear picture emerges from these comparisons, except that the predictions of crash counts from the simple log-linear model are consistently too low.

4.4 Stop-controlled intersections

The range of volume on the major road for stop-controlled intersections was the same as that for signal-controlled intersections. Volumes on the minor road ranged up to the volume of the major road up to 20,000 vpd. At volumes of the major road higher than 20,000 vpd, volumes on the minor road tended to be much lower.

Figure 18 shows the smoothed crash counts using a window of (3,000 x 3,000). Except for a "spike" at very high major, and very low minor volumes, the surface shows no trend. Figure 19 shows the data smoothed with a (6,000 x 6,000) window. Aside from the "spike" that remains, there is no apparent trend of crashes with major volume. At low values of major volume, crashes show a slightly increasing trend with minor volume. This trend disappears at higher values of the major volume.

The "spike" is due to a single intersection with 100 crashes, which is not balanced by other intersections with low crash counts. If the effect of this intersection is ignored, the data show such a pattern of variation over a relatively narrow range of crash counts that cannot be represented by the conventional models of the form $z = a^*x^b y^c$.

5. Signalized four-leg urban intersections, California

5.1 Distribution of intersections by traffic volumes

Figures 20 and 21 show the distribution of signalized four-leg urban intersections by volume on the major and minor roads. Figure 21 shows the same distribution in a form that can be more easily compared with the figures showing smoothed crash counts and other variables. The height of the blocks is always the same, but the width of the lines indicates the number of cases in each cell. If there are no cases, no block is shown. A few isolated cases are outside the range of this diagram.

5.2 Total crash counts

Figure 22 shows total crash counts, as given in the California intersection file, moderately smoothed with a window size of (5,000 x 5,000). While the surface is somewhat "wavy," a distinct though not simple pattern is recognizable. This pattern is clearer in Figure 23 where a greater window size of (10,000 x 5,000) was used. This surface is more smooth and the waves essentially disappear.

The most obvious feature of the surface is the "ridge," at minor volumes of about 20,000 vpd, which becomes a "hump" beyond a major volume of about 50,000 vpd. Up to this ridge, crash counts increase with minor volume. Beyond it they decline and at even higher volumes they tend to rise again, but not to the height of the ridge.

Crash counts increase with increasing major volume. However, above about 20,000 vpd, the increase is much slower than for lower volumes. Indeed, there is very little change for low values of the minor volume. For major volumes above 60,000 vpd, data become scarce and the pattern becomes complicated. In some areas crash counts are higher and in other areas much lower than for lower volumes.

This complicated surface cannot be represented, or even reasonably approximated, by a simple mathematical function. This contradicts the intuitive expectation that crash counts should vary in a relatively simple way with traffic volumes, and that intersection crash counts should deviate from it only because of their individual characteristics, including safety features. If these characteristics were randomly distributed over the intersections, then the shape of the relationship between crash counts and volumes should be only slightly affected. If, however, the presence of such features was related to the two volumes, then the combined effects of these features and of the pure volume effect could result in an apparent relationship between crashes and volumes that is very different from the pure crash-volumes relationship.

In our case, the complicated surface could be the result of a combination of a volume effect, and the effects of features that become increasingly common with minor volume once a value of 20,000 vpd is exceeded. There might also be intersection features that are more common for major volumes above 60,000 vpd. Another potential explanation is that different crash types may depend in different ways on the

two traffic volumes. Adding several simple functions can easily result in a much more complex function.

Before we explore these possibilities, we want to make sure that the overall shape shown in Figure 23 is "real" and not an artifact of the smoothing procedure. It must be emphasized that artifacts can also be produced by analytical models. We separated the intersections into three groups. The first group included all intersections with major volumes less than or equal to 60,000 vpd and minor volumes less than or equal to 20,000 vpd, which represents most intersections. The second group included intersections where the major volume exceeded 60,000 vpd, which has no intersections with minor volume greater than 32,500 vpd. The third group included all intersections where the major volume was less than or equal to 60,000 vpd and the minor volume exceeded 20,000 vpd, which is beyond the ridge. The data for each group were smoothed separately.

Figure 24 shows the smoothed surface for the first and largest group of intersections. Comparing it with the corresponding part of Figure 23 shows that the pattern is qualitatively very similar, though quantitatively different. Note that the vertical scale is different, to accommodate the higher peak. The increase of crash counts toward the ridge is steeper, and the "hump" in Figure 23 is now a much higher "peak." It is not surprising that such features are softened in Figure 23 since this is the normal effect of smoothing. This effect is usually desired, except if there are theoretical or empirical reasons to assume that there are lines where the function is discontinuous, or has discontinuous derivatives.

Figures 25 and 26 show the smoothed surfaces for the other two groups of intersections. Figure 27 shows the smoothed surface for all intersections with the part for the first group "cut out" to make the comparison of the second and third groups to all intersections easier. The overall shape of the surfaces for both groups is very similar to that for all intersections, but there are differences in the details. For instance, the surface for the third group shows a steeper increase for high values of the major volume than does the smoothed surface for all intersections.

5.3 Crashes within and near the intersection

The crash data file shows whether a crash occurred within the curb lines of the intersection, on a major approach leg, or on a minor approach leg.

Figure 28 shows the smoothed surface for crashes within the intersection. Comparison with Figure 23, representing total crashes, shows that crashes within the intersection proper are only a relatively small proportion of all crashes. The comparison also shows that the relationship between crash counts and volumes appears much weaker. There are more "local" patterns. One is that intersections with both volumes around 40,000 to 50,000 vpd have many more crashes than most other intersections, and that there are a few intersections with very high crash counts at major volumes of about 77,500 vpd and minor volumes of about 25,000 vpd. Furthermore, the hump at a major volume of 60,000 vpd and a minor volume of 20,000 vpd is more pronounced. This suggests that individual intersection characteristics have a strong influence on crash counts.

Figure 29 shows the smoothed surface of crashes on the major approaches. Comparison with Figure 23 shows that these crashes account for a high percentage of all crashes, and that the shape of the surface is very similar to the surface for all crashes.

Figures 30, 31, and 32 also show the same surface as Figure 29, but cut off at major volumes of 20,000, 40,000, and 60,000 vpd, respectively. These cuts show the nature of the ridge clearly. For low major volumes the ridge is near a minor volume of 15,000 vpd and is a soft maximum. For roads with higher major volumes the ridge occurs at minor volumes of about 20,000 vpd and is much sharper.

Figure 33 shows smoothed crash counts on the minor approaches. There is a clear pattern for major volume up to 60,000 vpd and minor volume up to 20,000 vpd, which consists of a nearly plane surface, increasing in a roughly proportional rate with minor volume, and with relatively little or no increase with major volume. The cross - sections shown in Figures 34, 35, and 36 clearly show the initial nearly proportional increase with minor volume. For higher major road volumes, however, the nearly proportional increase stops abruptly and reverses.

5.4 Crash types within the intersection

Crashes that occurred within the curb lines of an intersection were further classified. Those involving a single vehicle were excluded, because they do not directly result from the intersection of two traffic streams. Crashes involving pedestrians, bicycles, and animals also were excluded. Although some of these crashes result from crossings of the road at an intersection, and thus are typical intersection crashes, their study requires additional exposure measures for pedestrians, bicyclists, and even animals.

We distinguished the following crash types indicated in the crash data file:

- rear-end collision
- angle collision
- right-turn collision
- left-turn collision (or U-turn)

and combined head-on, sideswipe, and other multiple-vehicle collision into "other" collisions.

Figure 37 shows the smoothed surface for left-turn collisions within the intersection. This surface is very similar to that for all crashes within the intersection, shown in Figure 28, but the dependence on the major volume appears to be slightly weaker.

Figure 38 shows the smoothed surface for right-turn collisions. The pattern also is similar to that of all crashes though the isolated "peaks" at the right side, outstanding in the surfaces for all within-intersection and left-turn crashes, are missing.

The smoothed surface for rear-end collisions, Figure 39, shows an unusually simple pattern, consisting of a roughly linear increase with the major volume up to about 60,000 vpd, and a general tendency to increase with minor volume, albeit with some waves.

Figure 40 shows the smoothed surface for angle collisions. There is no clear pattern, except for the hump similar to the one seen for all within-intersection crashes, and the "spike" nearby, which also appears for all within-intersection crashes and left-turn crashes.

The smoothed surface for "other" collisions, shown in Figure 41, is the only surface that has a ridge, similar to the surface for all intersection crashes, but in a slightly different position.

These surfaces were smoothed with the same windows as those shown in the earlier figures, so that potential similarities could be recognized. However, because of the smaller number of cases, the surfaces are more "wavy." Therefore, in Figures 42, 43, 44, 45, and 46 we also show the surfaces for the various crash types, smoothed with a larger window size of (15,000 x 10,000). The general patterns are more easily recognizable in this set of figures.

As seen in Figure 42, the number of left-turn crashes at low volumes on the minor road is practically constant and independent of the major volume. That the surface shows two collisions for no traffic is a consequence of smoothing, which extrapolates to this point. The same can occur with an analytical model if it does not force the number of crashes to be zero for major and minor volumes equal to zero. For low major volumes, left-turn crashes vary very little with minor volume. For larger values of the major volume they increase only slowly with minor volume, except for high values of the minor volume, where they increase rapidly.

The smoothed surface for right-turn collisions in Figure 43 shows a soft ridge at a minor volume of about 12,500 vpd, and a relatively simple surface, except for intersections with middle values for both the major and minor volumes.

The smoothed surface for rear-end collisions in Figure 44 is nearly a plane up to major volumes of about 60,000 vpd, and minor volumes of about 20,000 vpd. The surface shows that rear-end crashes decline rapidly for larger values of the major volume. For

larger values of the minor volume they level off and even decline somewhat if the minor volume becomes even larger.

The smoothed surface for angle collisions in Figure 45 is smooth, but not simple.

The familiar ridge is evident in the smoothed surface for "other" collisions shown in Figure 46. The ridge is located at a minor volume of 20,000 vpd for low values of the major volume, and at a minor volume of 25,000 vpd for high values of the major volume. Up to the ridge, the increase with minor volume is roughly linear.

The patterns that appear in the proportions of the different types of crashes may be missed if only the smoothed surfaces of crash counts are examined. Since the proportions of crash type showed greater variations than crash counts, a larger smoothing window size of (15,000 x 15,000) was used. Figures 47, 48, 49, 50, and 51 show the smoothed surfaces for proportions of different crash types. The smoothed surface for the proportion of left-turn collisions in Figure 47 initially declines with increasing volumes, but changes relatively little when volumes are greater, except when the volumes on the major and minor roads are in the middle range. The surface for the proportion of right-turn collisions increases rapidly with both volumes, but levels off later and shows relatively little change, except at the highest volumes on the major road. The smoothed surface for the proportion of rear-end collisions shows an initial increase with both volumes. For higher volumes on the major road there is relatively little change, and there is a slight decline with increasing volume on the minor road. There is a decrease in the proportion of rear-end collisions at the highest volumes on the major road. Figure 50 shows that the proportion of angle crashes is fairly constant, although there is a slight increase when both the major and minor volumes are in the middle range. The smoothed surface for the proportion of "other" crashes shown in Figure 51 also increases initially with both volumes, but then remains fairly constant, except for a group of intersections with the highest volumes on the major road.

Care must be taken when trying to interpret such figures. For instance, left-turn collisions usually occur with oncoming vehicles on the same road. Therefore, it is expected that their occurrence depends primarily on the volume of the road the vehicles are traveling on and not on the volume of the crossroad. However, the same holds for vehicles on the crossroads. Thus, the total number of left-turn collisions is the sum of the left turns on both roads, each separately determined by its own volume.

Similar arguments hold for rear-end collisions. However, the occurrence of angle and right-turn collisions is expected to be determined by an interaction of both volumes.

5.5 Intersection characteristics

Intersection characteristics, especially those reflecting crash countermeasures are expected to have an effect on the occurrence of crashes. Then, intersections with a certain, safety related feature would have fewer crashes than otherwise comparable intersections without this feature or characteristic. Comparing otherwise comparable intersections with and without the feature should allow the effect of the feature on crashes to be estimated.

The common way to overcome this problem is to use all available intersections to develop an analytic model. Such a model would have the crash counts as the dependent variable, the traffic volume or continuous independent variables, and other intersection characteristics, including the presence or absence of certain countermeasures, or categorical independent variables. Because such a model is based on much greater case numbers than those available for "matched" comparisons, the standard error of the coefficient, including those reflecting countermeasures, tend to be lower.

However, even this may not resolve the problem of lacking comparable intersections. For instance, many features reducing the crash risk will also improve traffic flow. Therefore, they are more likely to be installed at high-volume intersections. This means that there is a correlation between the independent variables volumes and presence of the feature. If the correlation is high enough, no statistical technique can separate the two effects. Furthermore, the estimate effect of a feature depends completely on the assumed mathematical form of the relation with the volumes. There is no theoretical basis for the mathematical form of this function. Those mathematical functions used are arbitrary, though often plausible.

Therefore, we explored whether the unexpected complicated relation between crashes and traffic volumes could be explained by the presence of intersection characteristics related to the two traffic volumes.

The California intersection file contains a number of intersection characteristics that may have an effect on crashes. These include:

- design speed
- traffic control type
- lighting type
- major road left-turn channelization
- major road right-turn channelization
- major road traffic regulation
- major road number of lanes
- crossroad left-turn channelization
- crossroad right-turn channelization
- crossroad traffic regulations
- crossroad number of lanes
- median on major road
- information on shoulders
- details on the median
- road bed dimensions

We selected the following characteristics that are most likely to have an effect on crashes:

- design speed
- traffic control type:
  2-phase signal or multi phase signal
- major road left-turn channelization
- major road provision for free right turn
- minor road left turn channelization
- minor road provisions for free right turn
- major road number of lanes
- minor road number of lanes
- median on major road

We also considered using the presence of turn restrictions, distinguishing intersections where left turns were not permitted or were restricted during peak hours. However,

only 13 intersections had such restrictions on the main road, and nine intersections had such restrictions on the crossroads.

Figure 52 shows the average design speed in relation to the two volumes. The design speeds are mostly between 80 and 100 km/h and no clear pattern is seen. It should be noted that this is the design speed and not the speed limit or actual travel speed. Thus, no conclusions on relations between speed and traffic volume and indirectly on relations between design speed and crashes can be drawn from our data.

Figure 53 shows that the proportion of intersections with multi-phase signals is very high with no simple pattern. The figure clearly shows that there is no consistent increase in multi-phase signals with traffic volumes.

The proportion of intersections with left-turn channelization on the main road is shown in Figure 54. It is very high and shows only minor variations with volume.

The corresponding proportion of intersections with channelization on the minor road is shown in Figure 55. The proportion varies only slightly with volume of the major road, but strongly with the volume of the minor road. This variation is roughly linear up to minor volumes of around 20,000 to 25,000 vpd. Beyond that it levels off and even reverses in one part of the diagram.

This surface bears some similarity to the smoothed surface of crash counts on the minor approaches shown in Figure 33. Both crashes and the presence of left-turn lanes increase with increasing volume. Thus, while left-turn lanes may possibly reduce collisions within the intersection, they may increase crashes on approaches. This is well known to traffic engineers. However, the pattern in Figure 37 gives no indication that left-turn crashes on any approach are inversely related to left-turn channelization on the minor approach.

Figures 56 and 57 show the proportions of intersections with provisions for free right turns on the major and minor roads, respectively. There is some similarity between the surfaces, especially since free right turns become rarer with increasing volume on the minor road. However, there are local deviations from these patterns. One of these is a slight local "hump" near major volumes of 60,000 vpd and minor volumes of 20,000 vpd, which appears in several graphs.

These two surfaces show no similarity to any of the crash surfaces. Figure 58 shows the number of lanes on the major road. As expected, it increases from an average of three lanes for the lowest-volume roads to six lanes for the highest. There is relatively little variation with the volume of the minor road, except for an area where both roads have middle-range volumes.

Figure 59, showing the number of lanes on the minor approach, reveals the complementary pattern with little variation with the volume of the major road and a roughly linear increase with minor road volume. These is no similarity to any crash pattern.

Figure 60, showing the presence of a median on the main road, yields an interesting pattern. The main feature is that the presence of medians increases rapidly with volumes on the main road up to about 30,000 vpd. Beyond 30,000 vpd, medians are nearly always present. However, for high values a "valley" appears near a minor volume of about 20,000 vpd. This is more recognizable in Figures 61 through 63. That means that at these volumes a median is less common than at other volumes. One might speculate that this could allow relatively more collisions on the main approach, accounting for the "ridge" that appears for crashes on this approach. However, it would not explain the less sharp ridge that appears for crashes on the minor approach except if presence of a median on the major road is related to presence of a median on the minor road.

To determine whether a high number of crashes on the relatively many intersections with undivided main approaches creates the ridge and/or hump, we separated intersections by divided and undivided main approach. We had to use the term "main" approach, rather than the "major," because division was given for only one approach. Thus, these surfaces are not strictly comparable with the other surfaces in this study. Figure 64 shows the smoothed crash counts for the major approaches for divided intersections. Comparing this figure with Figure 29 shows that the overall patterns, with the ridge and to some extent the hump, are similar. Figure 65 shows the smoothed surface of crash counts for undivided highways. There are indeed a few intersections with very high crash counts in the area with main volumes of 50,000 to 55,000 vpd and crossroad volumes of about 25,000 vpd. However, they are already beyond the ridge and do not substantially contribute to it. Consideration of these two figures and Figure 33 indicates that division of the major highway, or lack of it, cannot explain the ridge.

Comparing Figure 64 with Figure 65 shows another interesting feature. In the area of low volumes where there are sufficient numbers of intersections with divided and undivided main approaches, the surface for undivided approaches tends to be below that for divided approaches. This contradicts the intuitive expectation that divided highways should have fewer crashes than undivided highways. Other factors must have a stronger effect than the separation of traffic streams, and the divisions may have been made in response to high crash counts without reducing them to the level of those intersections where the approach remained undivided.

5.6 The length of the influence zone

Crash counts on intersection approaches depend not only on intersection characteristics and traffic - flows, but also on the length of the "influence zones." These zones are defined by convention or judgment and are not a specific function of other intersection characteristics. Therefore, they have to be considered when studying crashes on intersection approaches, as well as intersection crashes that include those within the influence zones.

The California data base always had values for the length of an influence zone for the main approaches. The value for the length of influence zone on the crossroad approaches was usually zero and only had positive values for a few cases. Fifty-three percent of the intersections have influence zones of 75 meters. The next most frequent value is 45 meters for 8 percent of the zones, followed by 60 meters and 30 meters for 4 percent of the zones. The maximum length of the influence zone is 350 feet and the minimum is 7 feet. Nine percent of the influence zones are less than 100 feet. While one would not expect the crash count to be proportional to the length of the influence zone, one would expect it to increase with the length. Therefore, if the length of the influence zone is related to traffic volumes, which is not implausible, a relationship between crash counts and traffic volumes is expected. Indeed, the length of the influence zone determines the "exposure" on the approaches.

A simple relationship between the length of the influence zone and traffic volumes should be recognizable in Figure 66, which shows the smoothed average length of the influence zone on the main approach in relation to the two traffic volumes. There could

be more subtle relations, e.g., in some areas of the diagram the lengths might be equal for all intersections, whereas in other areas the average could still be the same, but the individual values vary widely.

The surface shows no overall trend and an appreciable variation. Of special interest is the slight local hump that appears in the area where we have previously found a hump with a ridge. However, the ridge appears for crossroad volumes of 15,000 vpd, whereas at 20,000 vpd there is a valley, where previously there was a ridge.

Because Figure 66 is based on main and cross roads, and the previous figures were based on major and minor roads, they are not strictly comparable. Therefore, Figure 67 shows crashes on the main approach by volumes of main and cross roads. It corresponds to Figure 29. While there are some differences in detail, the overall pattern is the same.

Thus, we conclude that the peculiar surface for crashes on the main (and also the major) approaches cannot be explained by the length of the influence zone as a scale factor.

5.7 Analytical modeling

We also developed a simple analytical model. Its purpose is not to compete with very detailed analytical models, such as those developed by Bauer and Harwood,[3] but to see to what extent simple analytical models can approximate the empirical representation of the data by smoothing techniques.

One important point to keep in mind is that there is no theoretical basis for a specific mathematical form. Commonly used models are based on very simple considerations of plausibility and mathematical convenience. A commonly used model is

$$z=a^{*}x^{b}\,y^{c} \qquad (5\text{-}1)$$

[3] K.M. Bauer, D.W. Harwood, *Statistical Models of At-grade Intersections Crashes.* Mid-West Research Institute, November 1961.

The current state - of - the - art is to specify (5-1) as the model for the expected value of a Poisson distribution for the crash counts, and then get a maximum-likelihood estimate of the parameters. A negative binomial distribution with an additional parameter could also be used. We used a mathematically simpler process to obtain practically the same result. We assumed (5-1) as the model for the crash counts, but weighted each observation with the inverse of the modeled crash count, which is the variance of a Poisson distributed variable. The estimates were obtained using the SAS procedures NLIN repeatedly, re-weighting each time. Though not statistically rigorous, these estimates are usually close to true maximum likelihood estimates. Considering the only modest overall fit of the resulting model, a more sophisticated approach would not have been justified.

The result was:

$$z = 2.60 * x^{0.522} y^{0.219} \qquad (5\text{-}2)$$
$$(0.38)\ \ (0.042)\ (0.017)$$

with standard error shown in parentheses.

The correlation between b and c is -0.20 and between a and c it is -0.02. These values are negligible. The correlation between a and b is -0.97, which is large, as is usual if the averages of the variables are far from zero. This means that despite their low relative standard errors, a and b can be varied considerably, as long as it is done according to the correlation, without affecting the model fit too much.

Figure 68 shows the surface resulting from the model, and Figure 69 shows the same surface but cut at x = 20,000 vpd and y = 62,500 vpd. Comparison with Figures 23 and 27 shows that the analytical model is only a rough approximation of the smoothed surface. In some areas the trend of the analytical model contravenes the trend of the data as reflected by the smoothed surface.

Figures 70 and 71 show this even more clearly. However, they also show that our overall smoothed surfaces still are not satisfactory representations of the crash patterns. If separate surfaces are constructed for three separate blocks of data, bounded by the apparent ridges ($x \leq 60{,}000$, $y \leq 20{,}000$, $x > 60{,}000$, $y > 20{,}000$), they do not meet at the boundaries. This means that more sophisticated smoothing

techniques have to be used that distinguish "real" ridges from "noise," and avoid "smoothing out" of such ridges. Such techniques are not yet available for routine work.

5.8 Conclusion regarding the four-leg signalized intersections in California

We have found that crashes at four-leg signalized urban intersections in California have complicated relationships with the volumes of the two roads, which cannot be expressed by mathematical functions of the form $z = a*x^b*y^c$, which is equivalent to a log-linear model. It is still possible that such a model is tenable, if it is expanded to include certain intersection or traffic characteristics. However, the intersection characteristics available in the data file show no patterns that can explain the deviation of the actual crash patterns from the analytical models. Traffic characteristics beyond the volumes of the two roads are not available in the file. We tried to infer some traffic characteristics by separating certain classes of crashes. Again, these classes of crashes showed no patterns likely to provide an explanation for the deviations. Of course, patterns of specific crash classes represent a combination of risk patterns and specific exposure patterns. Thus, it could be possible that very specific exposure measures exist that can explain the deviations.

Another important finding is that crashes on the major and minor approaches showed very different patterns in relation to the volumes, and that crashes within the intersection itself showed yet another different pattern. Different crash types within the intersection also showed different patterns.

While it might be possible to develop useable analytical model for specific crash types with the use of exposure measures specific to the type of crashes, it is very unlikely that a simple analytical model can be found that adequately represents the sum of many different crash types, the proportions of which vary across the intersections.

6. Minnesota intersections

6.1 Distribution of intersections by traffic volumes

Figure 72 shows the distribution of traffic volumes for the 71 useable signalized urban four-leg intersections found in the Minnesota data files. Volumes for the four approaches may all be different. However, it appeared questionable whether actual

volumes were as different as the volumes shown in the file. Also, using the four approach volumes would have required a complicated analysis that would not have been justified with 71 data points. Therefore, the volumes of the two approaches on each road were averaged, and only the resulting average volumes for the major and the minor road were used.

Figure 73 shows the distribution of the two volumes in a form easily comparable with the following figures. Comparing this distribution with the distribution for California intersections shows that there are relatively few intersections with low volumes on the minor road.

6.2 Smoothed crash counts

Figure 74 shows the smoothed values for the crash counts found in the data file, using a window of (4,000 x 4,000). There is a tendency for crash counts to increase roughly linearly with the minor volume, at least up to certain values. The relationship with the major volume, however, is more complicated. There is relatively little variation up to 16,000 vpd, and again from 22,000 vpd upward, except for some isolated points at high volumes. Between 16,000 vpd and 22,000 vpd, crash counts increase quite rapidly with the major volume.

Figures 75 and 76 show this pattern more clearly. Figure 75 shows the surface cut off at a minor volume of 6,000 vpd, where the pattern, described above, is very pronounced. Figure 76 shows a cut-off at a major volume of 10,000 vpd. Here, the initial level part is no longer present, but there is a linear increase up to a major volume of 22,000 vpd, above which crash counts also remain stable.

While this pattern appears to be very clear, there is still a possibility that it is due to only a few intersections. Therefore, we also generated the surface using a larger smoothing window size of (8,000 x 4,000). The resulting surface is shown in Figure 77. The fairly steep increase between 16,000 vpd and 22,000 vpd has been smoothed out to a more uniform increase up to 24,000 vpd. Beyond that, however, crash counts still vary only slightly with the major volume, except for a few points with very high volumes.

6.3 An analytical model

We fitted the data to a standard log-linear model, using a nonlinear regression technique with the following result:

$$z = 2.18 * x^{0.205} y^{0.388} \qquad (6\text{-}1)$$
$$(0.63) \quad (0.102)(0.090)$$

Figure 78 shows the surface representing this model. Like the smoothed surface, It shows only a weak increase of crash counts with major volume. Obviously, it cannot represent the different pattern for lower volumes. This becomes clear when Figure 79, which shows the surface cut-off at a minor volume of 6,000 vpd, is compared with Figure 75, which shows the actual surface cut-off at the same place. If Figure 80 is similarly compared with Figure 76, it is clear either that log-linear models are inadequate, or that intersections with low major volume differ from those with high major volume by some important feature that starts appearing at volumes between 16,000 vpd and 22,000 vpd.

6.4 Crash types

Since total crash counts aggregate very different types of crashes, studying different crash types separately may provide insights that cannot be found in aggregate crash patterns. The simplest distinction between crash types is by location, that is, whether the crash occurred within the intersection proper or on an approach to the intersection.

Figures 81 and 82 show the smoothed relation of all crashes in the intersections proper and within 60 meters of the intersections. There is a general similarity between Figure 81 and Figure 74 which represents only crashes within the intersection proper. A closer look, however, shows some differences that are reflected in Figures 83 and 84, which show crashes that occurred on the approaches within 60 meters of the intersection. The less smoothed surface in Figure 83 shows no clear pattern. The more smoothed surface in Figure 84 shows a weak but steady increase with the major volume, and relatively little variation with the minor volume. This simple pattern is very different from the pattern seen in Figure 76, where the increase with the major volume is less consistent, but the increase with the minor volume is strong.

Even crashes within the intersection itself are not homogeneous. They can result from very different pre-crash situations and intended maneuvers. The occurrence of certain crash configurations depends on both the frequencies of the underlying maneuvers and on their inherent crash risk. One would expect that intersections with a high proportion of "risky" maneuvers also have higher crash counts.

The crash file contains some information on crash configuration. We excluded crashes characterized as "other" or "unknown." We also excluded crashes where the vehicle ran off the road, because they are not typical for an intersection. The other "typical intersection crashes" were grouped as follows:

- rear end
- left turn
- right turn
- angle
- other, consisting of side-swipes and head-on collisions

A number of intersections had no typical intersection crashes. Figure 85 shows the distribution of volumes for intersections with typical intersection crashes. A comparison with Figure 73 shows relatively few differences.

Figures 86 and 87 show the surfaces for the typical intersection crashes. The less smoothed surface in Figure 86 shows some irregular variation, aside from a few unusually high cases. If we again disregard the few high-volume cases, the more smoothed surface in Figure 87, shows practically no variation with the major volumes and only little variation with the minor volumes. This is an interesting observation for which we do not have even a speculative explanation. Counts for specific crash types are even lower and, therefore, show more random variations. The proportions of the crash types tended to vary less and are therefore shown.

Figure 88 shows this proportion of crashes involving left turns. This proportion tends to decrease with increasing major volumes, though not uniformly. The pattern changes in a systematic way with regard to minor volume. For low values of the major volume the proportion of left-turn crashes decreases with the minor volume, for medium volumes it stays nearly constant, and for high values of the major volume it increases with the minor volume. The surface pattern can be roughly characterized as a twisted sheet.

Figure 89 represents angle collisions. Again it is complicated but can be simply described as an elongated saddle. Figure 90 shows the surface for rear-end collisions, which is somewhat complementary to that for the angle collisions. A different pattern is seen for "other" collisions in Figure 91. However, there is a ridge, which is approximately at the same location as the ridge in Figure 89 and the trough in Figure 90. Together, these three figures suggest that around minor road volumes of 4,000 vpd to 6,000 vpd there is a change in the crash pattern. There were too few right-turn crashes to allow a meaningful representation by smoothing.

6.5 Relating proportions of crash types to number of intersection crashes

The idea that led to this analysis was that an intersection with a high proportion of left turns would also have a high proportion of left-turn crashes. Because of the high risk in left turns, such an intersection would have a higher crash count than otherwise similar intersections with few left turns. If this turned out to be the case, the proportion of left-turn crashes could be used as a proxy for exposure.

For the following analyses, we used only those intersections that had crashes of the selected types within the intersection and had a minor volume of less than 16,000 vpd, because the few intersections with greater volumes had much higher crash counts than others and showed only very weak relationships between volumes and crash counts. Fifty-seven intersections were selected.

As a first step, we looked at relationships between the number of crashes within the intersections and the proportions of the four crash types we distinguished. These data are plotted in Figures 91 to 95. Also shown are crash counts smoothed over the proportions. A common pattern is that the smoothed values are always low for proportions of 0 and 1, and are at their highest value near 0.25 to 0.3. Interpreted literally, this means that total crashes would be the highest if the four crash types were about equally present, that is, each accounted for about 25 percent of the total crashes. Furthermore, the total crashes would be lowest if all the crashes were of one type.

However, it is easy to see that this is a statistical artifact. Intersections with only one crash can have only the proportions 0 or 1 for each crash type. For instance, in Figure 91 there are eight intersections with one crash that have no angle collisions, and two that have one angle collision. For intersections with two crashes the proportions of 0,

0.5, and 1 are possible. In Figure 92 there are four intersections with no angle collisions, six with one, and one with two. More possibilities appear with increasing number of collisions. For the one intersection with 17 crashes, the proportions 0, 0.059, 0.118, ...0.941, and 1.0 are possible. Actually this intersection has proportions of 0.294 for angle collisions, 0.412 for left-turn collisions, 0.235 for rear-end collisions, and 0.059 for other collisions. It can be concluded from this observation that, if there were no relationships between crash counts and proportions of crashes, and the probabilities for the four types are equal, then most intersections with many crashes would have proportions near 0.25, and very few intersections, if any, would have proportions near 0 or 1. The lower the number of crashes, the more often intersections with proportions of 0 or 1 must appear, until finally intersections with one crash can have only zero or one crashes of each type.

A consequence of the above is that primarily intersections with low crash count are represented at the endpoints 0 and 1 and that the average crash counts at these points must be low. Other proportions can be realized only at intersections with more crashes and, therefore, for proportions between 0 and 1, the average crash count must be higher.

Correcting for this confounding factor is very difficult. However, it can be avoided if one looks at the graph "the other way," that is, by looking at proportions of crashes versus crash counts. Figures 96 to 99 show the data this way. For each crash count (intersections with the same crash count are combined) the proportions of crashes of one of the four types are shown, together with the range of one standard error. For proportion 0, a standard error of 0.04 was assumed.

The smoothed curves also are shown. They use a generalized Gaussian kernel and weight the points according to their standard errors. Figure 98, which shows the proportion of rear-end crashes, is the only figure that shows a potential relationship between total crashes and the proportion of crashes of one type. To explore this further, we used our original idea that the proportions of crash types might possibly explain deviations of individual intersection counts from the average surface for all intersections, which should be determined largely by volume effects. Therefore, we calculated the difference between the intersection crash count and the value from the smoothed surface for each of the 57 intersections. These differences are used in Figure 100. Comparing Figure 100 with Figure 98, which shows the smoothed surface

for the actual crash counts, shows that although the two figures initially look very different, there are some similarities. A comparison with Figure 94 shows even greater similarities (note that abscissa and ordinates are exchanged). The clusters of data points that appear in Figure 94 are spread out, but not much. The reason is that the differences against the smoothed value do not differ much from the difference against the overall average, because the smoothed surface does not differ much from the overall average.

The straight line shows a linear regression fitted to the data points. Its slope does not differ significantly from zero. This can be deceiving, since the linear specification of the model may be wrong. A quadratic regression fit gave the bold curved line. None of the coefficients, or both coefficients together, were significantly different from zero. Again, the quadratic specification may not be adequate. Smoothing with a wide window gave the broken line close to the quadratic model. Smoothing with an intermediate window gave the dash-dotted line, and smoothing with a narrow window gave the light solid line, which, aside from local variations, suggests a step somewhere between crash differences of 6 to 9.

The smoothed functions do not provide simple error estimates as regression models do. However, some estimates can be obtained. One approach utilizes the concept that each point of the smoothed curve is an estimated point of a local regression. The error of this estimate can be used as the error of the smoothed value. An objection to this approach is that both the proportion and the crash difference are random variables. The standard error estimates of regression analysis, however, assume that only the dependent variable is subject to random errors. An alternative, nonparametric approach is bootstrapping. One way to bootstrap is to fit a smooth model, calculate the residual of the data points against the model, repeatedly add random samples (with replacement) of the residuals to the smoothed values, and smooth these artificial values. If one repeats this often enough, one obtains a visual "confidence band" where only a certain specific percentage of the smoothed points lie outside this band. This approach, however, makes the same unrealistic assumption as the classical regression approach, i.e., that the values of the independent variables have no errors. This assumption can be avoided by a different, even simpler, application of the bootstrapping principle. Samples (with replacement) of 57 from the given 57 intersections are drawn repeatedly and a smoothed curve is fitted to each sample.

Figure 101 shows the results of 10 smoothing fits using the wide window. Beyond a crash difference of 7, the fit becomes very uncertain and often gives impossible negative proportions. Figure 102 shows the same curves, with the range of the ordinate between 0 and 1. Obviously, the smoothed relationship between the proportion of rear-end crashes and the crash difference is very uncertain, even where the proportions are not negative.

This high uncertainty is not a weakness of the smoothing technique. Rather, it is a result of the wide scatter of the data. This is confirmed in Figures 103 and 104. They show quadratic regressions fitted to 10 bootstrap samples of intersections. Except for one sample, the range of variability is as great as with the smoothed relations. If the comparison is limited to the range of feasible proportions between 0 and 1, the quadratic model is much less certain than the smoothed model, as shown in Figures 102 and 104. The reason is that the two data points with the highest values of crash difference are very "influential." To some extent smoothing can accommodate such influential points and reduce their influence on other parts of the curve. This is difficult with analytical models, and models with many parameters must be used.

Figure 105 shows the result of fitting linear regressions to the bootstrapped samples. Here the variability is much less than with the quadratic model. It should be noted that the envelope of the lines is very close to the typical 90 percent confidence "trumpet" derived by analytic models.

6.6 Conclusion

The overall conclusion is that only two intersections are responsible for the apparent, but not significant, relationship between the proportion of rear-end collisions and the crash difference. To determine whether such a relationship might be real, a much larger number of intersections would be needed. No relationship between their proportion and the deviation of crash counts from the smoothed model was apparent for the other crash types.

7. Conclusions on modeling intersection crashes in relation to traffic volumes as exposure measures

7.1 Relations between crash counts and traffic volumes at four-leg signalized intersections

The currently readily available exposure measures at intersections are the average traffic volumes on the intersecting roads. Intuitively, one expects crashes in intersections to result from the interaction of the two traffic streams. The simplest mathematical function expressing an interaction is $z=a*x*y$. This expression is too rigid because it has only one parameter, a. A simple generalization is $z=a*x^b y^c$. It is often used, usually in logarithmic from as a log-linear model.

A critical question is, is this form adequate to represent the actual relation between crashes and traffic volumes, or are there better relationships? The simplest way to obtain an approximation to the "true" relationship between crashes and traffic volumes is to smooth crash counts over the two traffic volumes. We did this for three data sets: four-leg signalized and stop-controlled intersections in Washtenaw County, Michigan; four-leg signalized intersections in California; and four-leg signalized intersections in Minnesota.

The Washtenaw County data suffered from a selection bias because only intersections with more than a certain number of crashes were included. Such data are rarely selected in practice. For this data set, crash counts increased with increasing major volume as well as with increasing minor volume. The conventional log-linear model appeared to be an acceptable qualitative representation. Quantitatively, however, there were complex systematic deviations between the data and the model. There was practically no relationship between crash counts and traffic volumes for crashes at stop-controlled intersections in this data set.

The large number of intersections in the California data set allowed detailed analysis. A fairly simple visual, but analytically complex, relationship was apparent. The most obvious feature of the surface was a "ridge" at minor road volumes of 20,000 vpd. Up to that value, crash counts increased nearly linearly with the minor volume. Beyond that volume, they initially dropped rapidly, and then leveled off. The relationship with the major volume was not that pronounced but showed a fairly strong increase at low

volumes, no or moderate increase at middle volumes, and an irregular decrease at high volumes.

Crashes within the intersection itself showed a similar but much less pronounced pattern with a weaker variation with the volumes. Crashes on the major approach showed a pattern very similar to that of all crashes, while those on the minor approach showed a definitely different pattern. There was, however, a "ridge" beyond which again crashes declined, a very strong increase with minor volume, and relatively little variation with major volume. It was obvious that a log-linear model could not even roughly approximate the actual surface.

One possible reason for a deviation from the expected pattern is that intersections that otherwise would have very high crash counts have been "improved" so as to reduce the crash risk. However, none of the intersection characteristics given in the data file that might reduce the crash risk appeared more frequently in the areas of the diagrams where the unexpected decline of the crash counts occurred. Thus, there must be either other intersection features not available in our data file, or the relationship between crash counts and traffic volumes must be far from log-linear.

The analysis of Minnesota intersections was limited by their low number. They showed a complex pattern. The relationship of major volume to crash counts was nearly a step function, which was approximately constant for low volumes, even more so for high volumes, with a "ramp" connecting the two levels. Strong smoothing that came close to fitting a plane to the data points resulted in a surface that increased with both volumes.

Crashes within the intersection itself showed a slightly different pattern. The pattern for the major volume again had two levels connected by a ramp, but there was a fairly strong increase with minor volume. Crashes on the approaches showed no clear pattern. Strong smoothing revealed only a weak increase with major volume and a stronger increase with minor volume.

A log-linear function was qualitatively similar to the more strongly smoothed surface, but deviated quantitatively. It could not represent the less strongly smoothed surface showing two levels and a connecting ramp.

If the crash risk in specific intersection maneuvers, such as turning left, turning right, going straight, etc., were the same across and independent of the volumes, but not across maneuvers, the frequencies of crashes reflecting such maneuvers would be proportional to the frequencies of the maneuvers. If an intersection had many high-risk maneuvers, one would expect more crashes than at intersections with comparable volumes but with fewer such maneuvers. Therefore, we also explored possible relationships between the frequencies of crash types, and of total crashes, in the Minnesota and California data sets. We found none.

Our three data sets showed very different relationships between crash counts at four-leg signalized intersections and the traffic volumes on the intersecting roads. None of them could be adequately represented by the conventional log-linear model. Either other intersection characteristics that were not readily available in our data sets had a strong influence on crash counts, or average annual daily traffic is not an adequate exposure measure.

7.2 What can currently be done?

Considering our negative conclusions about the usefulness of using conventional mathematical models to represent relationships between crash counts at signalized intersections and traffic volumes, what can be done? Smoothing is a promising alternative because it allows the fitting of even complicated surfaces by a simple process and avoids arbitrary assumptions. Using a function of two volumes, as done in this study, is a relatively simple matter. If more than two volumes, or other variables, especially categorical variables, are desired, the procedures have to be extended and refined, as discussed in subsequent sections.

We cannot rule out the possibility that someone may find manageable and not too ad-hoc mathematical expression for the relationship between traffic volumes and crash counts. By ad hoc we mean mathematical expressions selected specifically to fit the data sets studied, without consideration whether they could plausibly be extended to other data sets. However, such mathematical expressions have to be validated by more detailed criteria than correlative coefficients, likelihood ratios, or similar aggregate measures to be acceptable substitutes for smoothed surfaces.

What can be done in practice with such smoothed (or validated analytical) relationships? They can be used to compare the experience of an individual intersection with that expected from the relationship. If the difference is sufficiently large then the crash experience of that intersection should be studied in detail. Criteria for what is considered sufficiently large still have to be selected. Possibly, an explanation that suggests either which countermeasure should be applied to that intersection or which features of that intersection might have the beneficial effect of a crash countermeasure could be found.

This approach has its limitations. It can work only for intersections in an "area," defined by combinations of the two volumes, with enough data points in the "area" so that the individual peculiarities of the intersections will average out. It will not work for relatively isolated intersections near the boundary of the area covered by intersections. There, the smoothed surface will be "pulled" toward the value of each individual intersection. Even if the actual crash count for an intersection may be much higher than to be expected from the "true" relationship between crashes and volumes, this deviation may not be recognizable.

7.3 Substantive research needed

Before one can realistically think about modeling intersection crash counts, one needs to develop a more realistic logical and functional structure for such models. A first step is to re-think the concept of exposure. As already discussed, traffic volumes on the intersecting roads are conceptually unsatisfactory. Only in the simplest case of uncontrolled intersections can one expect crash counts to be log-linear or similar functions of the volumes. An exposure measure should count the opportunities for collisions. These depend heavily on the type and characteristics of traffic control provided. Promising steps to develop more meaningful exposure measures have been taken.[4] However, much more work on the problem using different perspectives is needed.

---

[4]F.M. Council, J.R. Stewart, D.W. Reinfurt, W.W. Hunter, *Exposure Measures for Evaluating Highway Safety Issues*. University of North Carolina Highway safety Research Center, 1983. F.M. Council, J.R. Stewart, E.A. Rodgman, *Development of Exposure Measures for Highway Safety Analysis*. University of North Carolina Highway Safety Research Center, 1987.

Another aspect is that intersection crashes, and even more so, intersection-related crashes, are very inhomogeneous. It cannot be expected that a single model will describe their frequency in a manner reflecting crash causation. Therefore, a closer examination of intersection crash types should be made and classes for meaningful modeling must be identified. This might require performing nearly a "clinical" analysis of individual crashes.

In reality, many intersections have certain features exactly because they had much crash experience. This creates relationships that make the standard statistical models uninterpretable. Either much more sophisticated models have to be developed, or different techniques used.

One alternative to the conventional approach, that of using a large set of intersections and including many variables in a complicated model, is to select intersections that are matched in many respects as closely as practical, and differ only in one or very few characteristics to be studied. This is much more likely to isolate any effect of such characteristics. If this is done, in turn, with many different subsets of intersections, a realistic model may be built in a stepwise fashion.

7.4 Methodological research needs

Before smoothing can be used routinely to model relationships between crash counts and exposure measures and other intersection characteristics, additional research needs to be done.

A realistic model will contain one or several exposure measures that are continuous variables (or counts that can be treated as continuous variables), intersection characteristics that will usually be described by 0/1 categorical variables, and possibly other continuous variables, such as travel speeds. In principle, one can smooth over all continuous variables simultaneously, but one cannot smooth over the categorical variables. They have to be accommodated by either additive or multiplicative terms, or the entire data set may have to be split according to a categorical variable, or combinations of several categorical variables, and each part modeled separately. Criteria have to be developed to decide when each of these treatments is appropriate.

Though it is possible to smooth data sets with a large number of continuous independent variables, it is not very useful. The data can be stored in a computer, or in hardcopy tables, and the smoothed value can be calculated for any combination of the independent variables. However, if the number of variables is greater than two, or at most three, the smoothed surface cannot be visualized or intuitively assessed for overall shape and smoothness. As a practical matter, one wants to separate the model into additive or multiplicative components, each of which can be studied and assessed separately. Indeed, this is the same approach used in analytical modeling, where one uses additive or multiplicative terms. If interactions have to be considered, they are introduced as additional additive or multiplicative terms. Since one can easily visualize a surface smoothed over two variables, one only needs to determine how to separate a model into components representing main effects, or interactions of any two variables. Research is needed to learn how to do this best and how to assess the adequacy of such additive or multiplicative models.

If one deals with experimental data where by design the data points can cover the range of the variables of interest more or less uniformly, smoothing by standard methods can give a good representation of the relationship, and the deviations of the individual points from the surface can give a good idea of the random variability of the data points.

In the case of intersections, and probably also other highway locations, the situation is different. Most observations are concentrated in only part of the entire area covered with observations. Toward the edges of this area, observations become more sparse and may be isolated. This poses a dilemma for smoothing. In the areas densely covered with observations, a narrow smoothing window may be appropriate, because it can well represent a complex relationship and still provide adequate smoothing. Where the points are more isolated, such a narrow window is no longer appropriate, because in extreme cases it may result in a perfect or at least very good fit to any single point or to a combination of only a few points. This can result in erratic behavior of the smoothed surface toward its boundaries. To avoid this, one might enlarge the smoothing window. While this has the effect of giving a smoothed surface near the boundaries of the covered area, it can result in smoothing out important details in the area well covered with points. Techniques should be developed that avoid this, for instance, by using a window with adoptive size, or by identifying parts of the smoothed surface that depend on only a few data points.

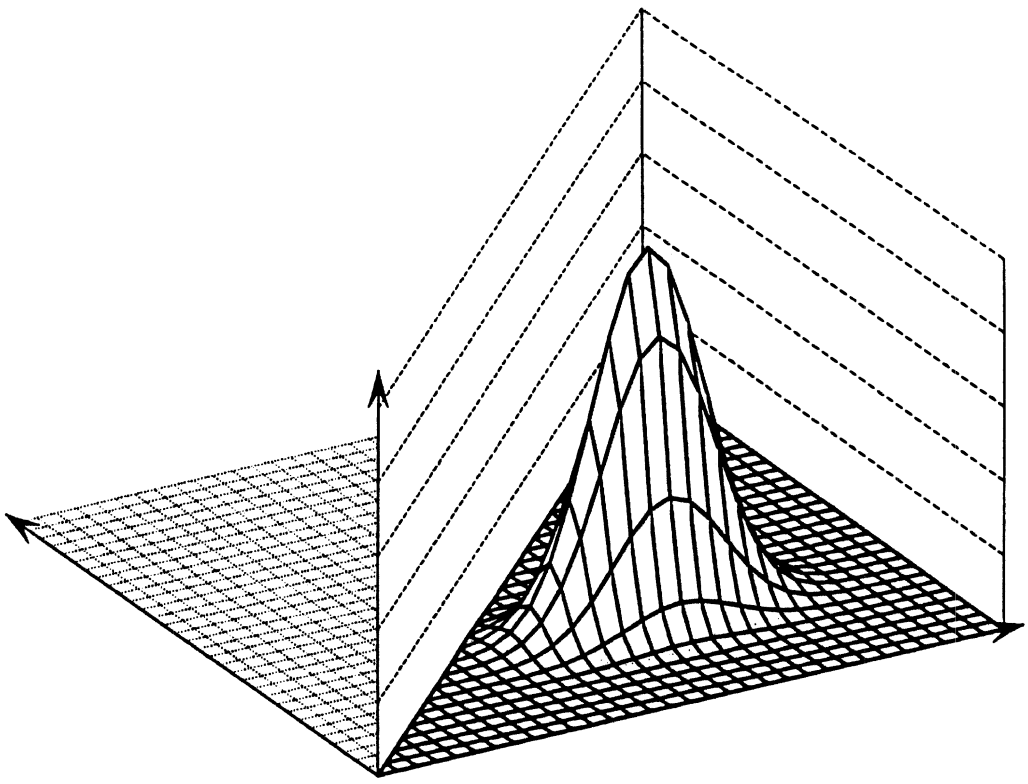


Figure 1. Representation of a Gaussian kernel, as represented by (2-1).

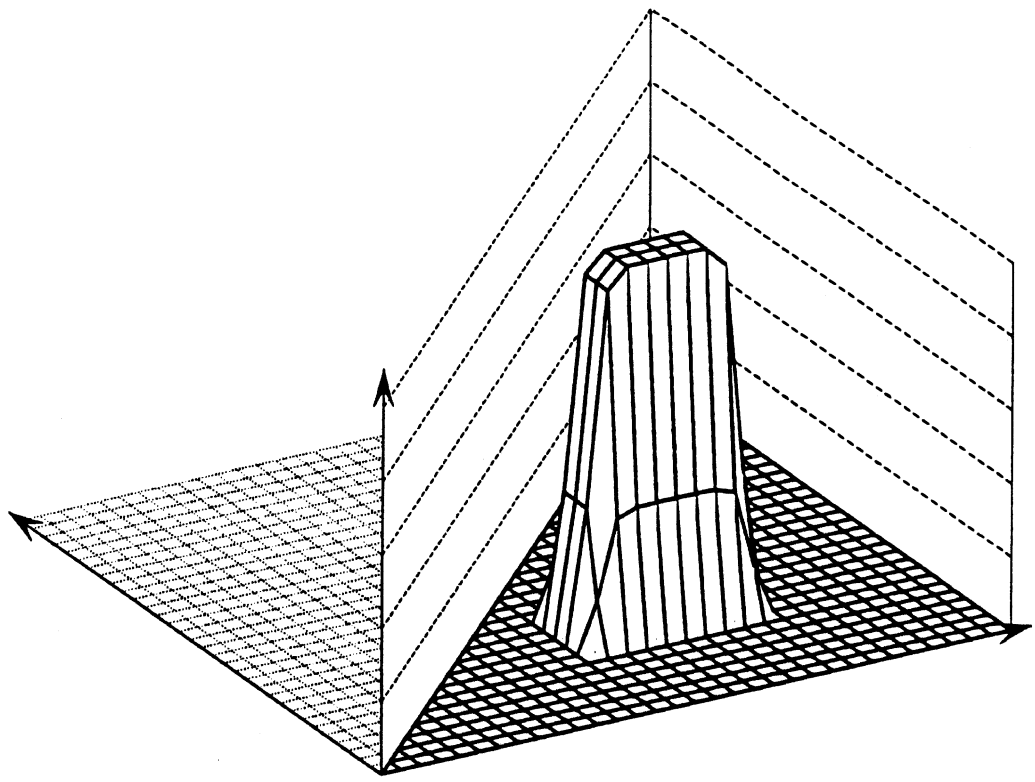


Figure 2. Representation of a Gaussian kernel with an exponent of 10.

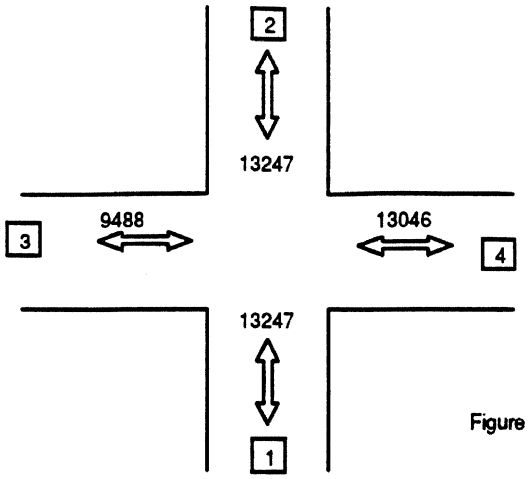


Figure 3A.1 Example Volumes

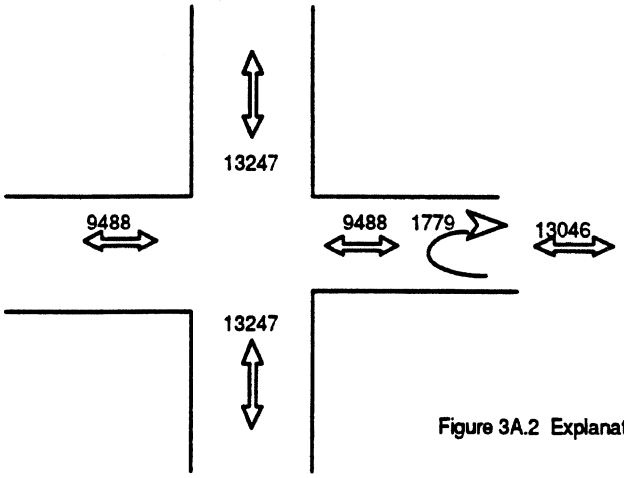


Figure 3A.2 Explanation 1

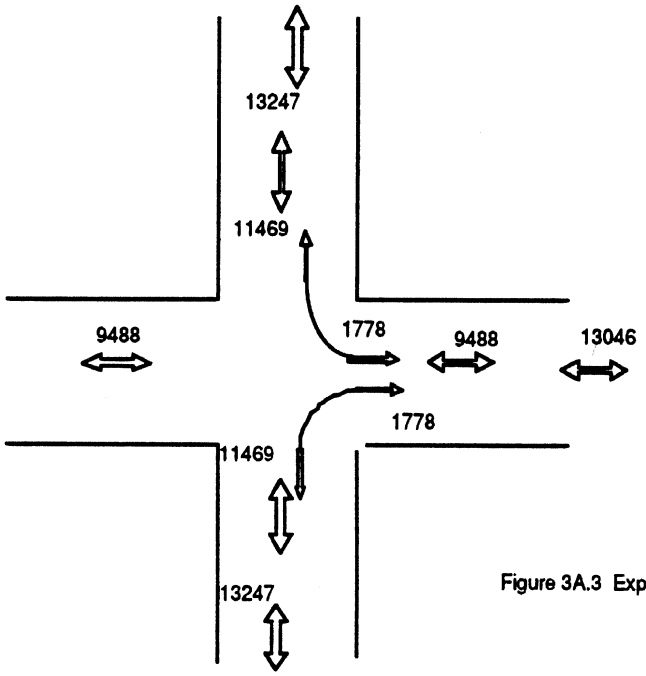


Figure 3A.3 Explanation 2

Figure 3A. Example of two-way volumes and two possible explanations.

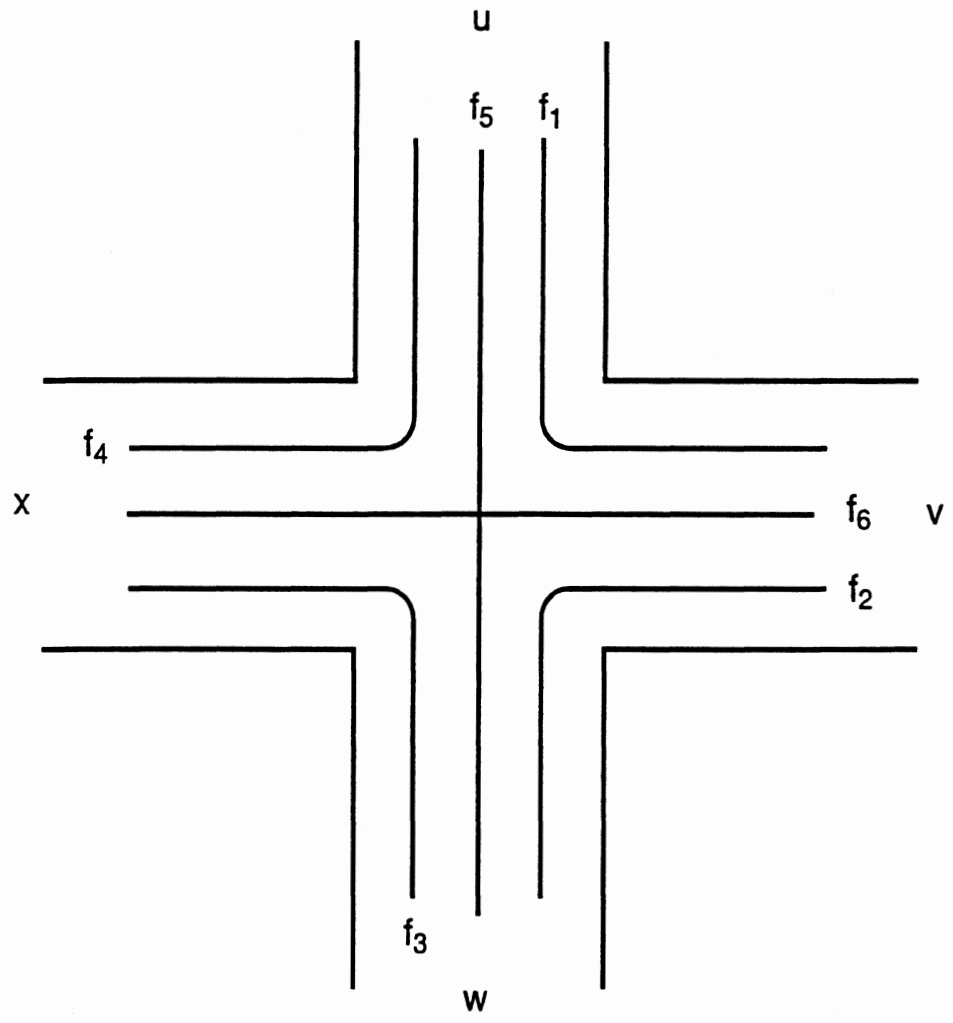


Figure 3. Flows distinguished at an intersection.

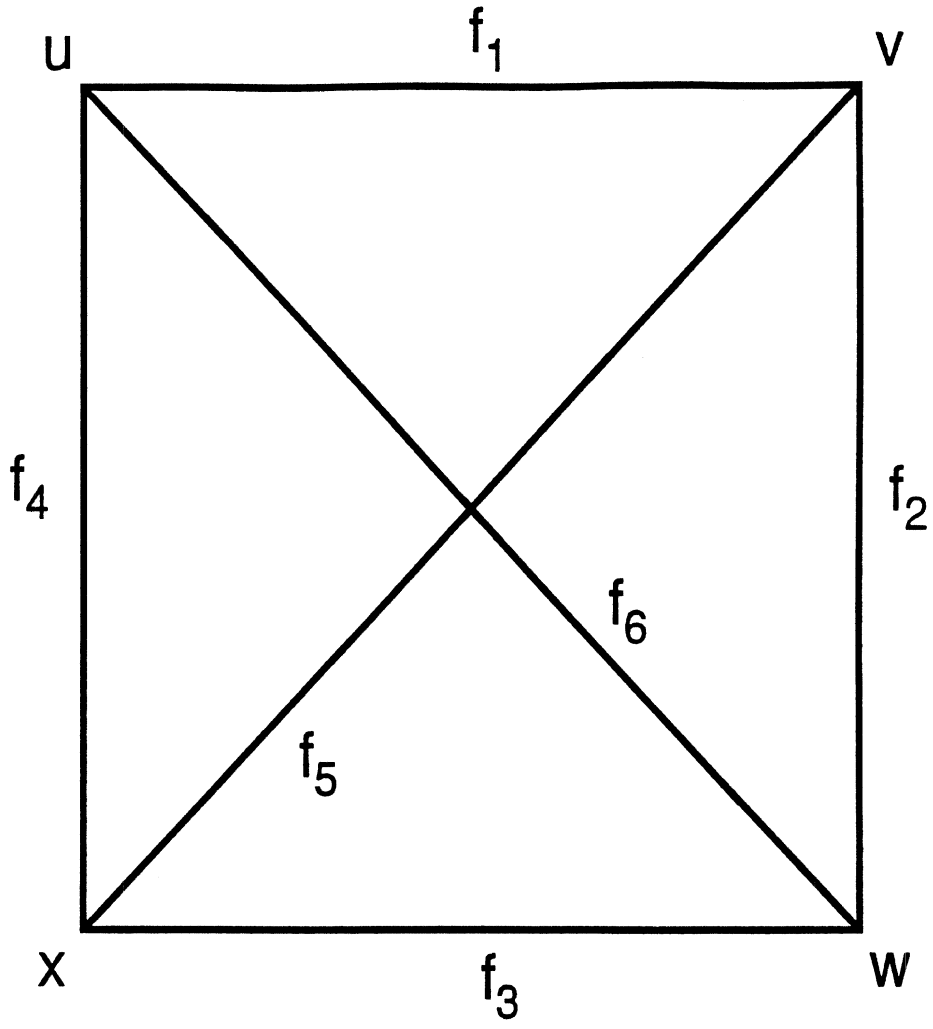


Figure 4. Different representation of the traffic flows shown in Figure 3.

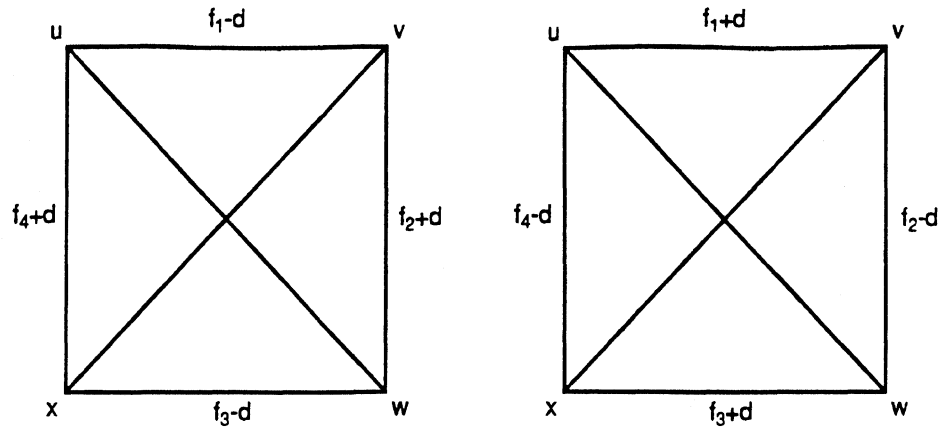


Figure 5. Deriving other realizable solutions from a given realizable solution. d is the value by which the original flows are changed.

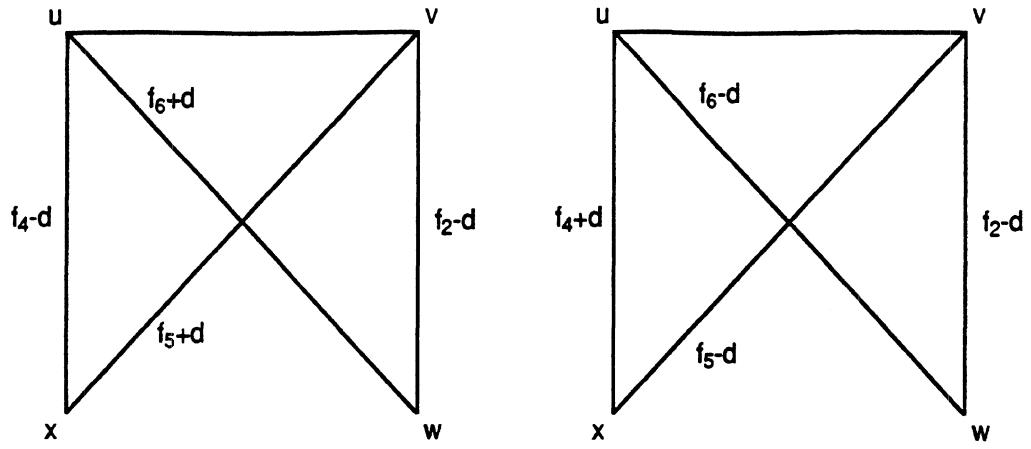


Figure 6. A new flow pattern, resulting from a modification shown in Figure 5, and possible further modifications of the flow pattern.

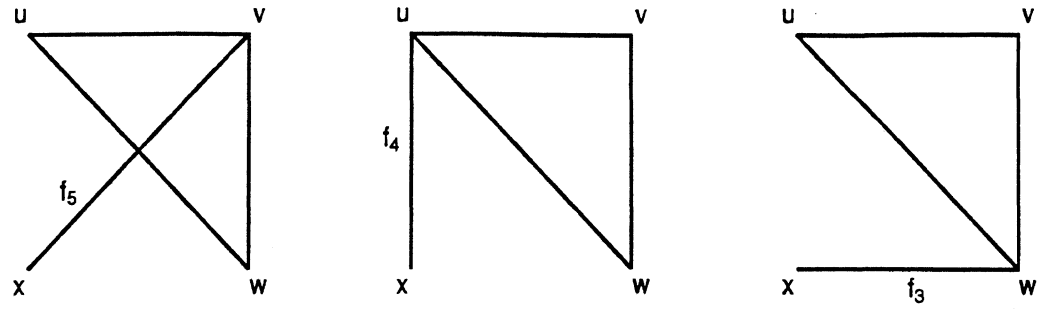


Figure 7. The simplest flow patterns obtainable if x is a minimal volume on the legs.

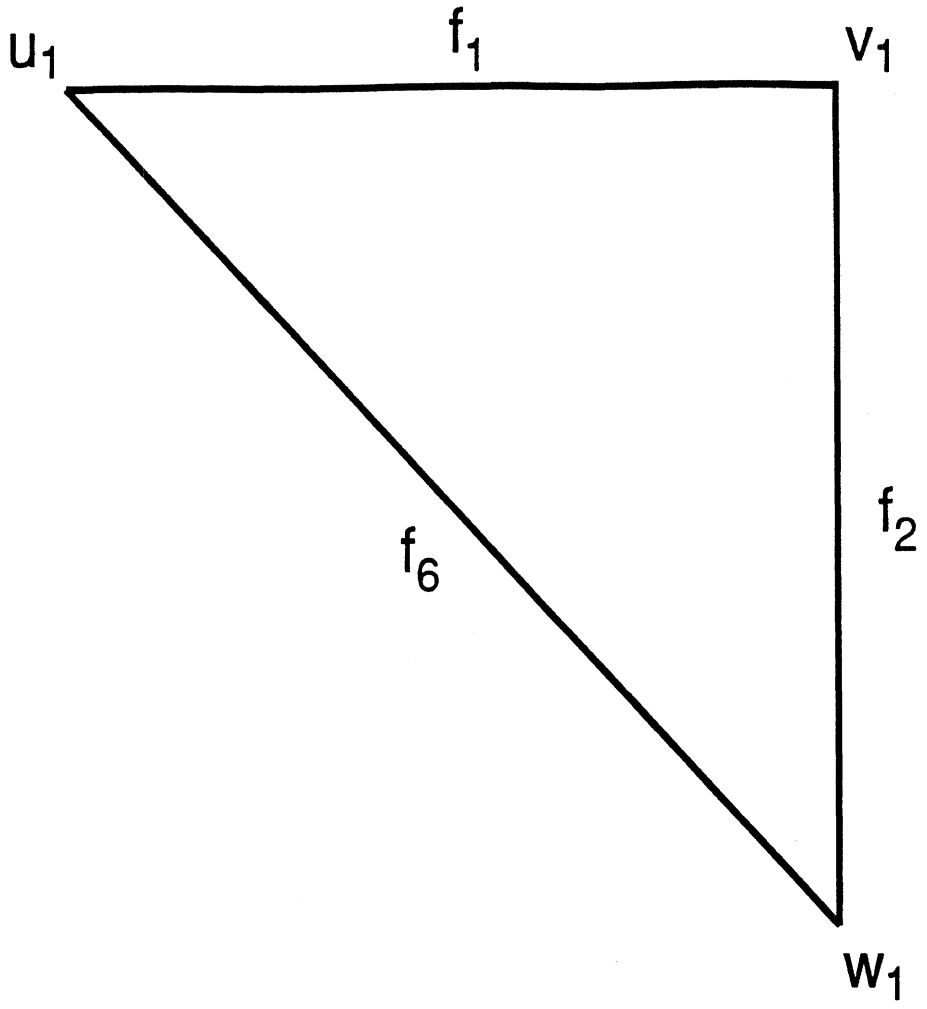


Figure 8. Reduced flow pattern to derive conditions for reliability of leg volumes.

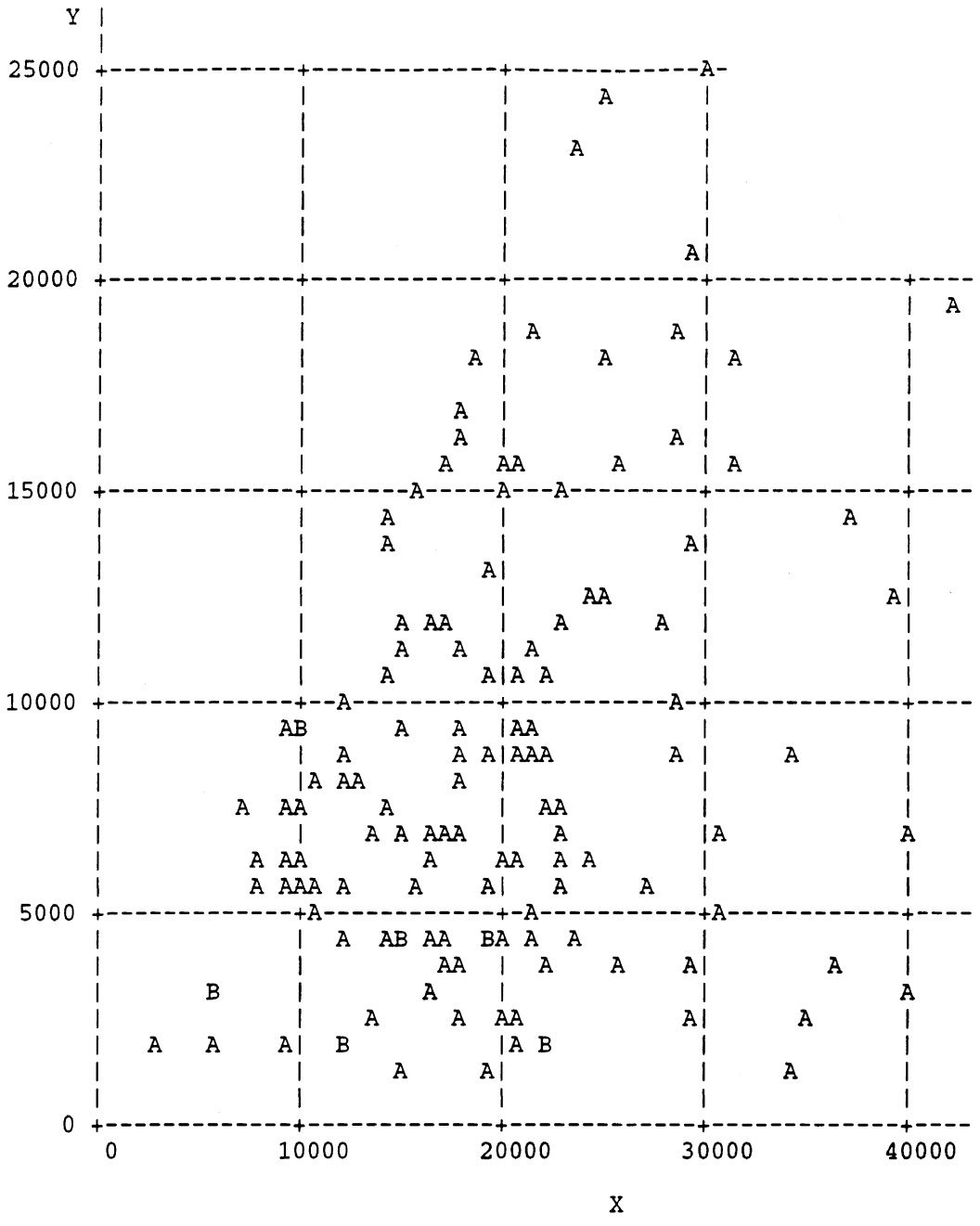


Figure 9. Distribution of traffic volumes at signalized four-leg intersections in Washtenaw County, Michigan. X=volume on major, Y= volume on minor road.

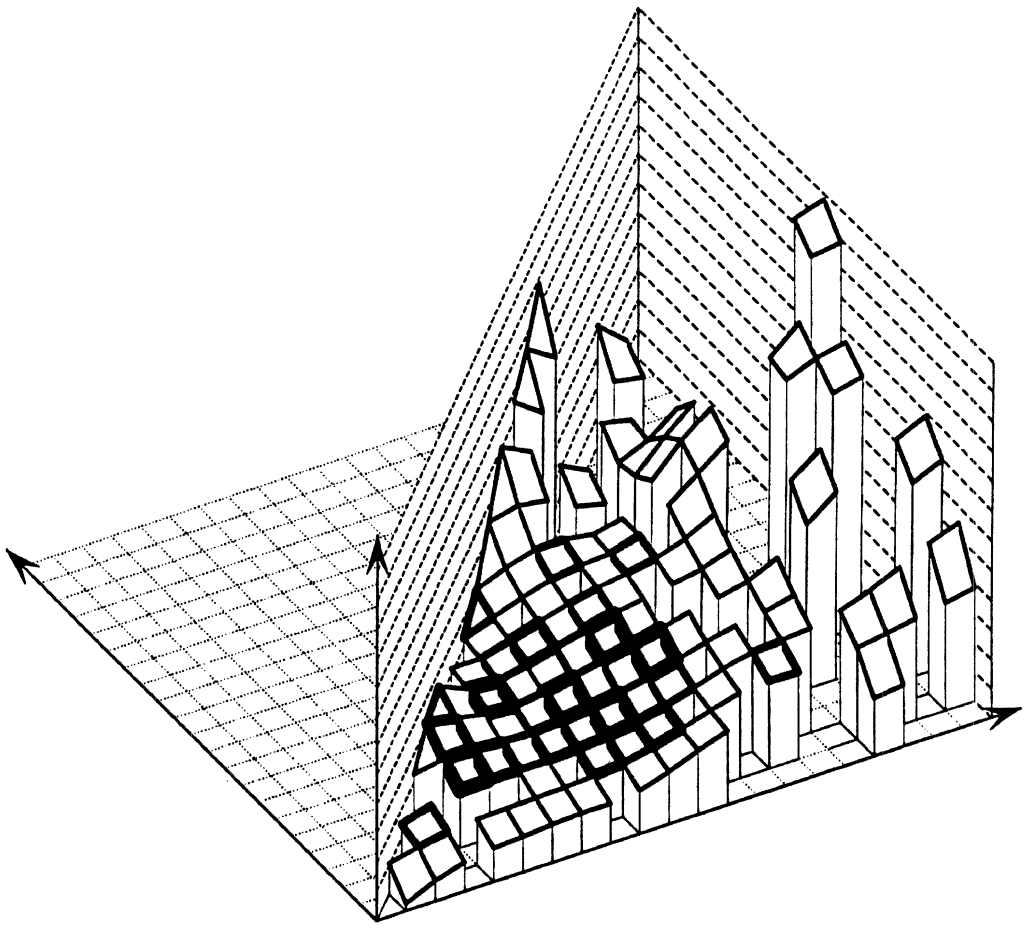


Figure 10. Signalized four-leg intersections in Washtenaw County, Michigan. Accident counts smoothed with a 4,000 x 4,000 window.

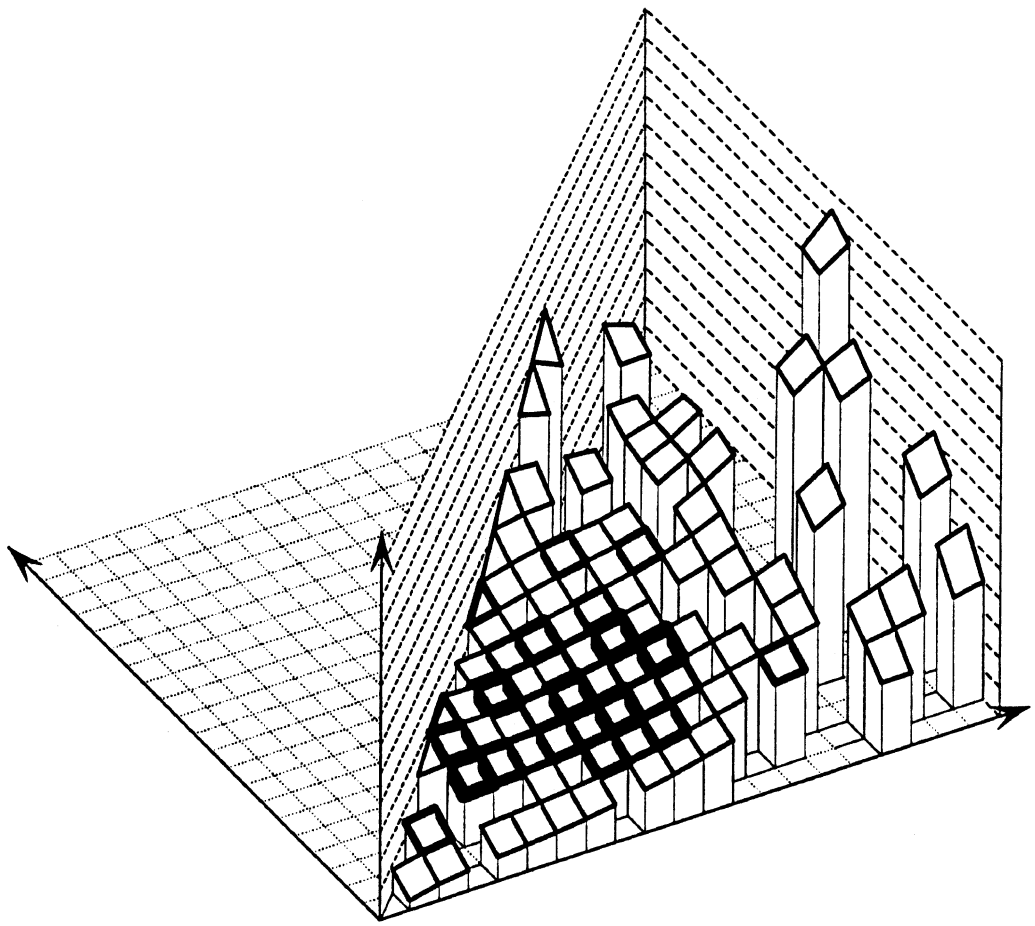


Figure 11. Signalized four-leg intersections in Washtenaw County, Michigan. Accident counts smoothed with a 6,000 x 6,000 window.

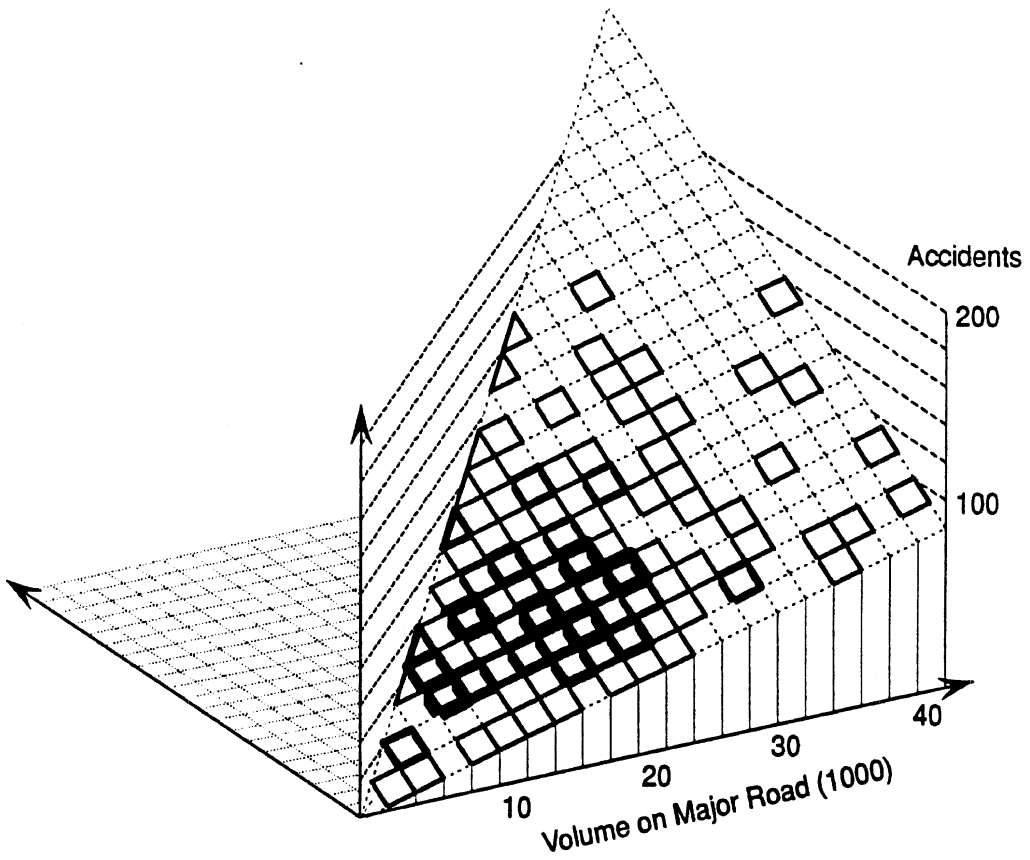


Figure 12. Signalized four-leg intersections in Washtenaw County, Michigan. Surface represents the analytical model 4.1.

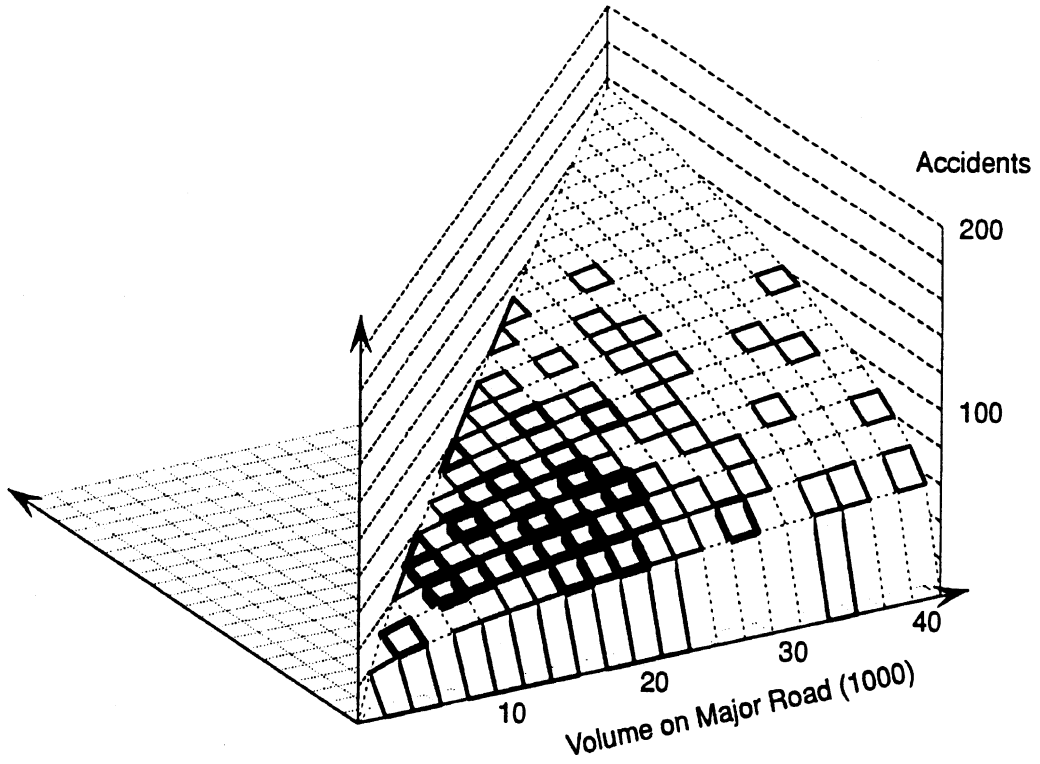


Figure 13. Signalized four-leg intersections in Washtenaw County, Michigan. Surface represents the analytical model 4.4.

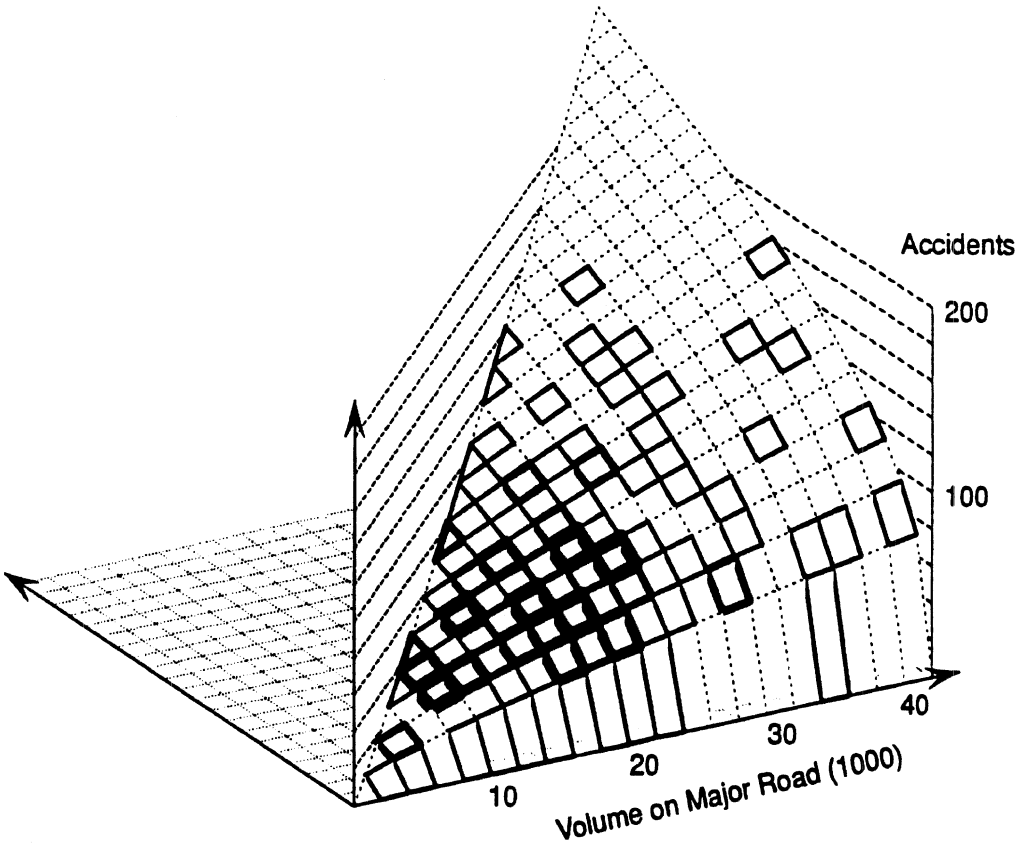


Figure 14. Signalized four-leg intersections in Washtenaw County, Michigan. Surface represents the analytical model 4-5.

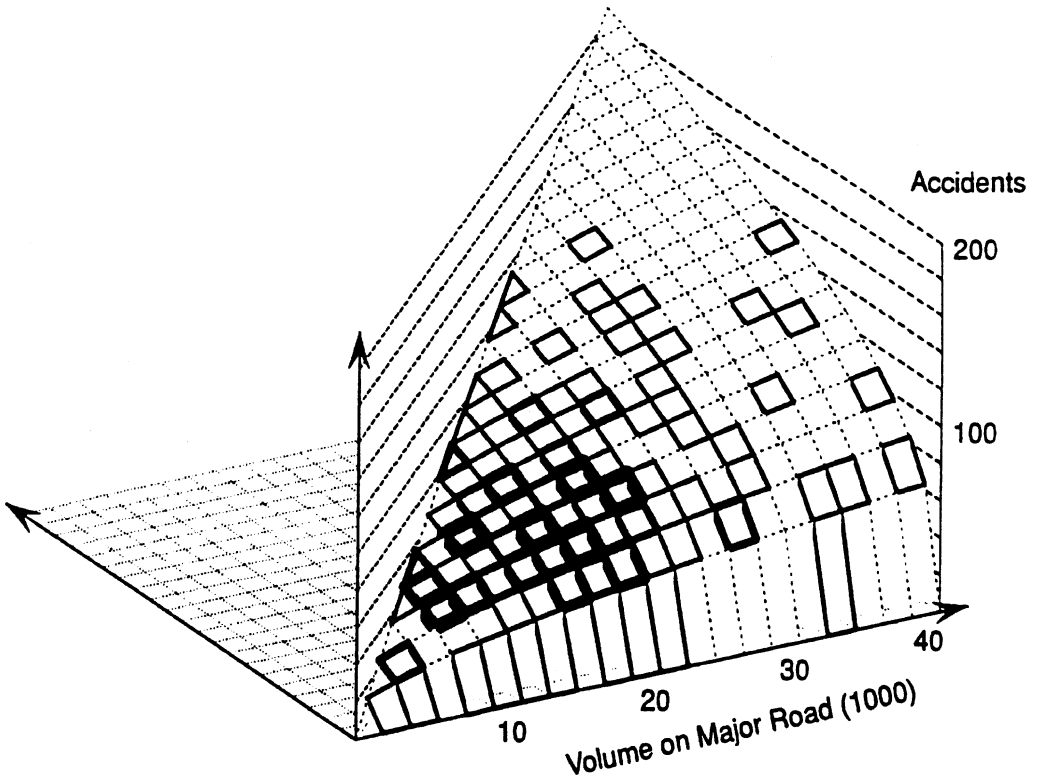


Figure 15. Signalized four-leg intersections in Washtenaw County, Michigan. Surface represents the analytical model 4-6.

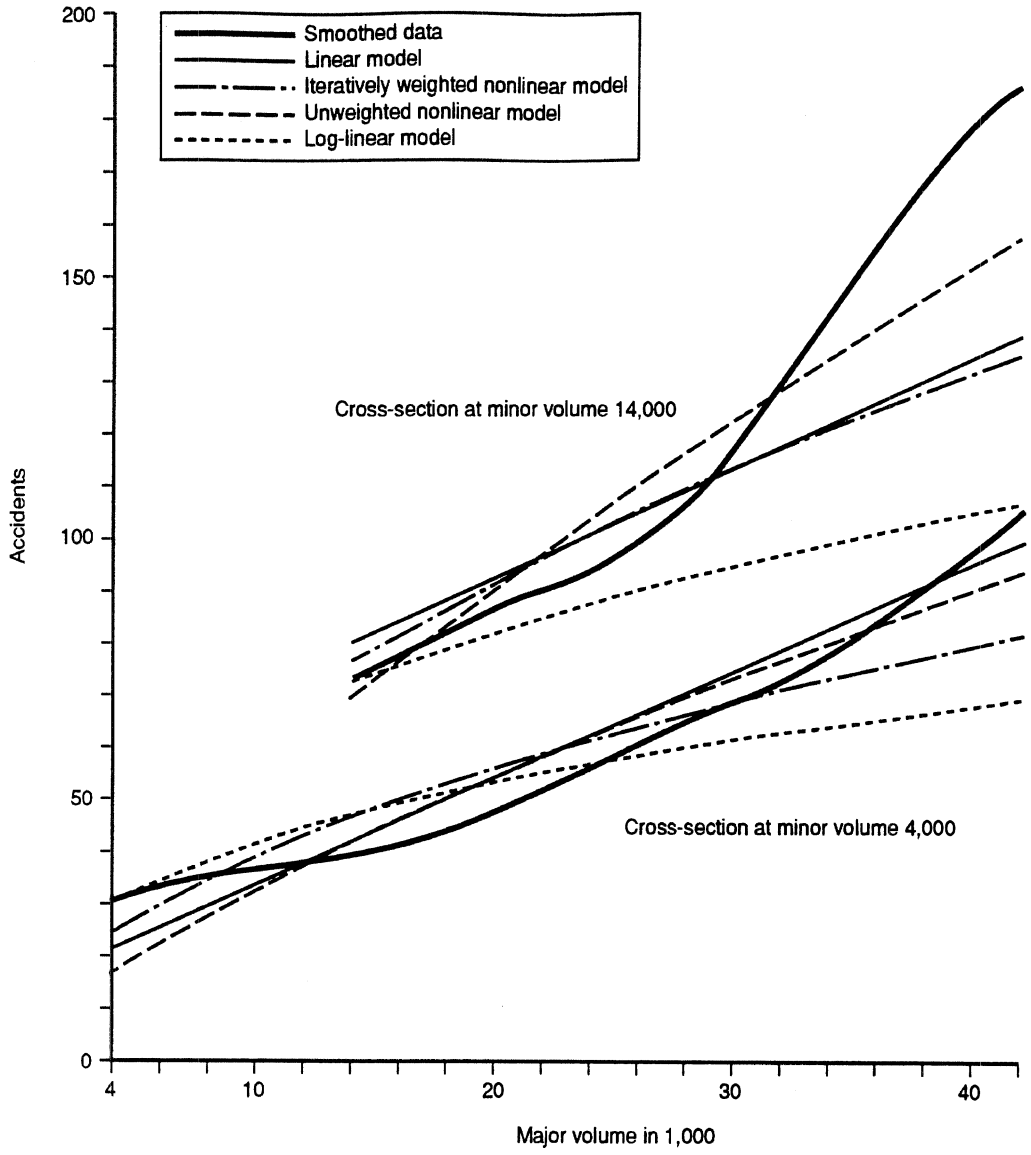


Figure 16. Signalized four-leg intersection in Washtenaw County, Michigan. Cross-sections through the surfaces in Figures 11, 12, 13, and 14 at minor volumes of 4,000 and 14,000.

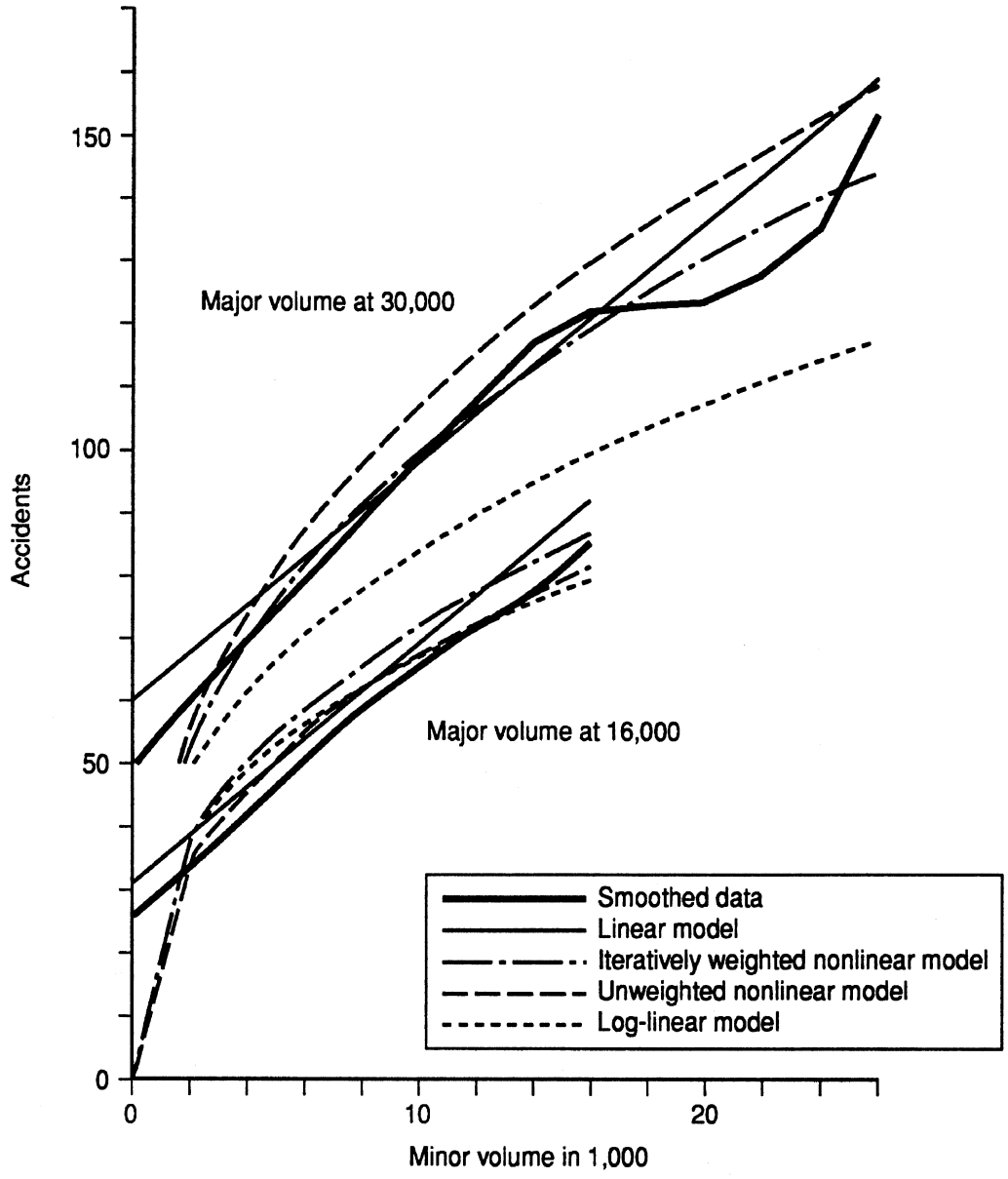


Figure 17. Cross-sections at major volume of 16,000.

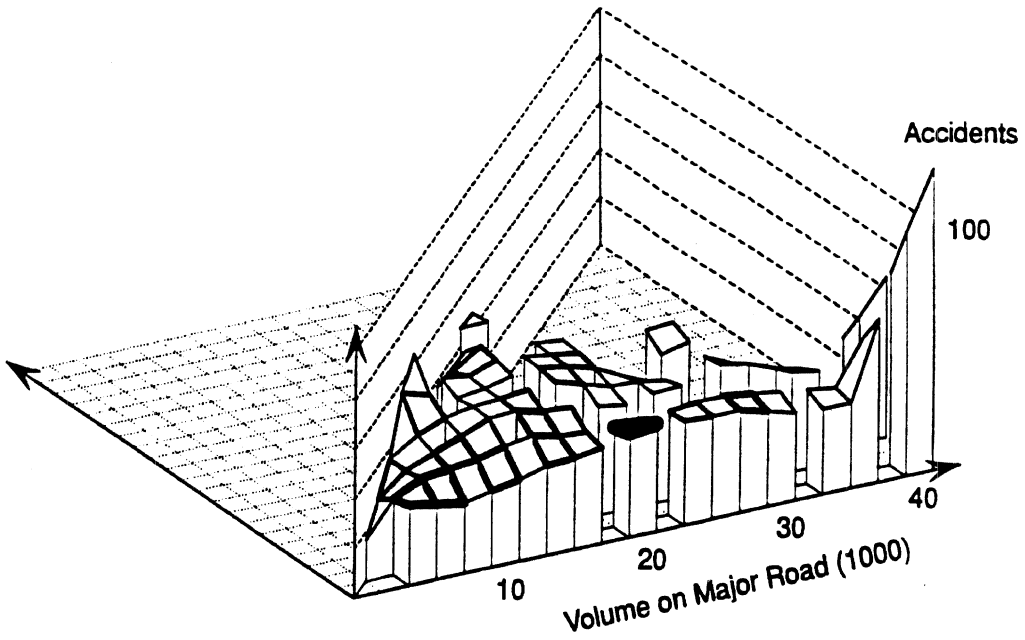


Figure 18. Stop-controlled four-leg intersections in Washtenaw County, Michigan. Accident count smoothed with a 3,000 x 3,000 window.

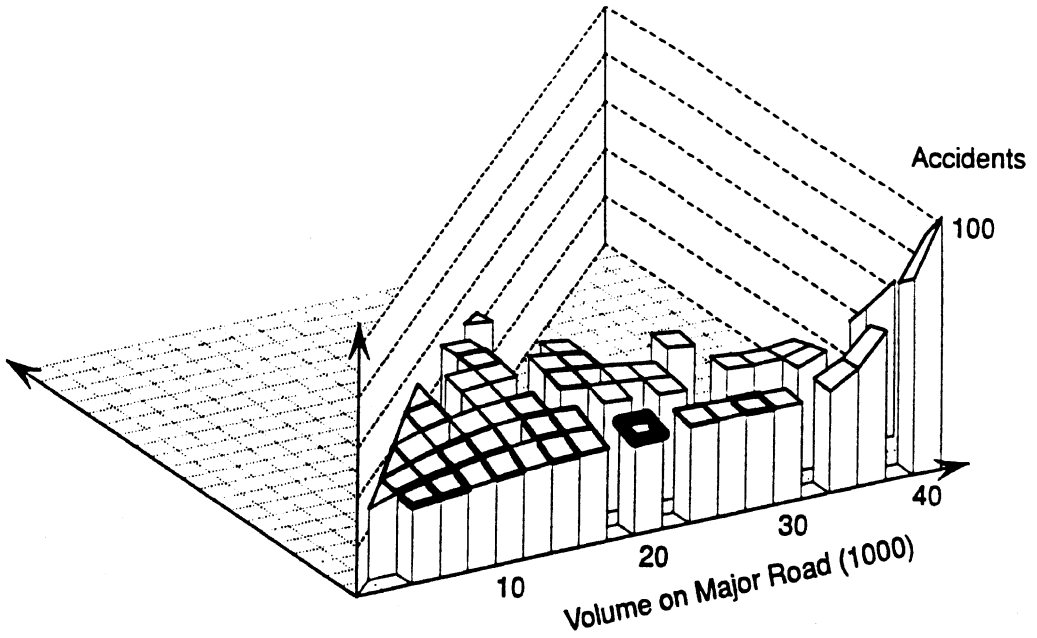
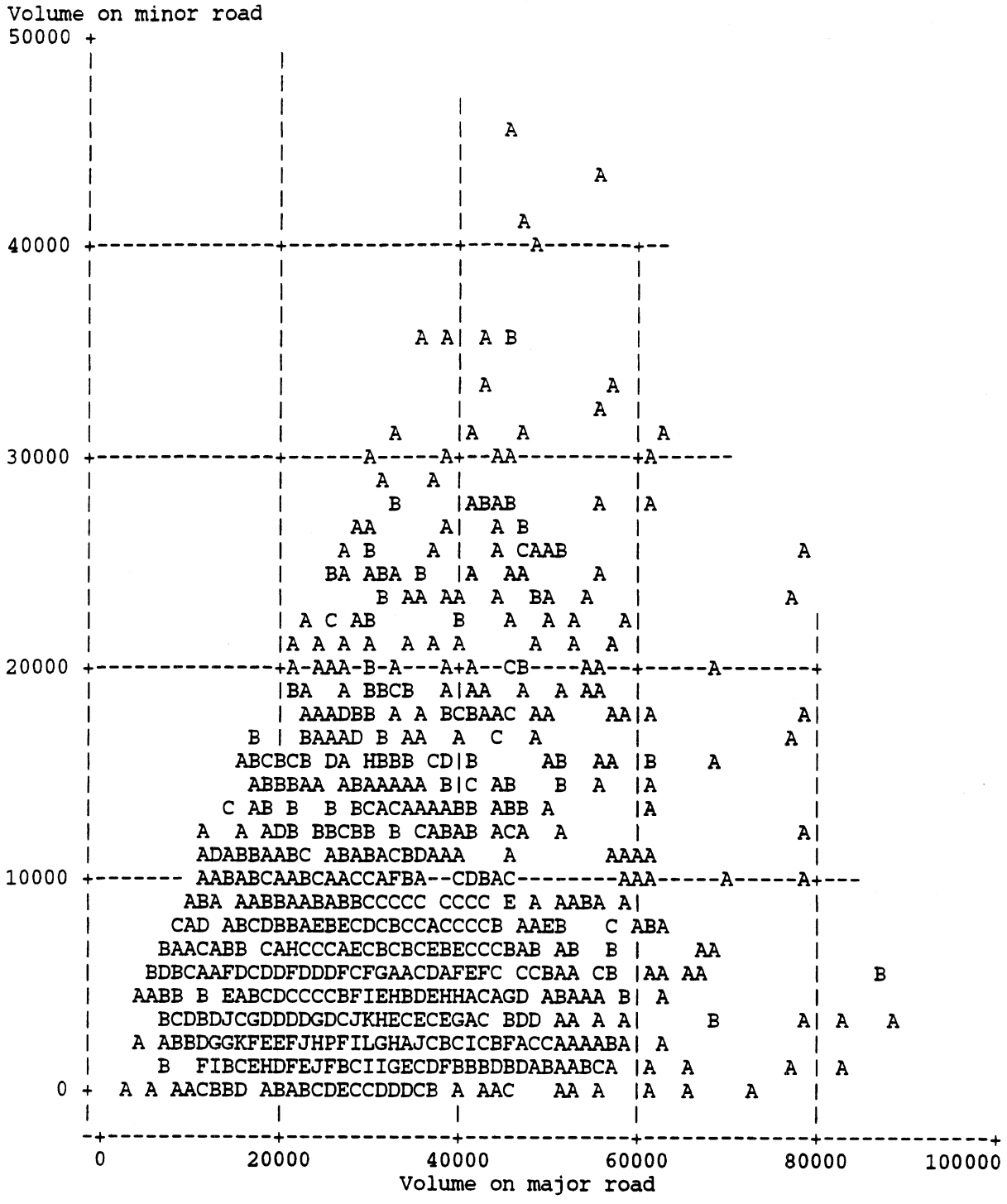


Figure 19. Stop-controlled four-leg intersections in Washtenaw County, Michigan. Accident count smoothed with a 6,000 x 6,000 window.



California, signalized urban intersections

Figure 20. Distribution of volumes at signalized urban intersections from California data file.

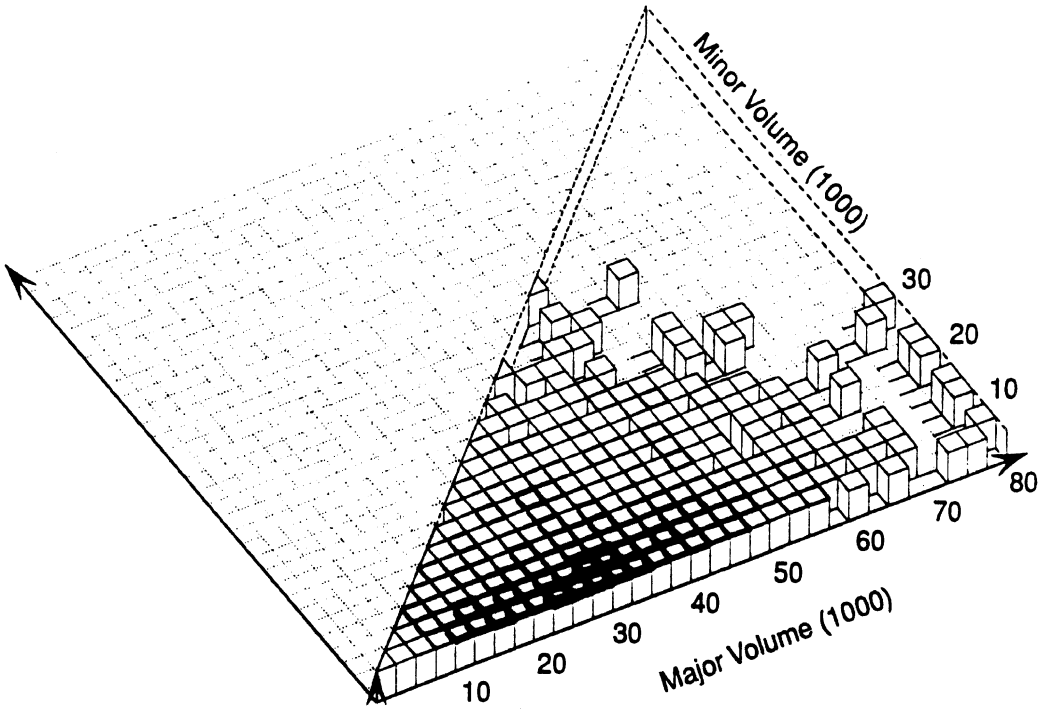


Figure 21. Distribution of four-leg signalized intersections in California by volume of the major and minor approaches. The width of the lines is proportional to the number of cases in each cell.

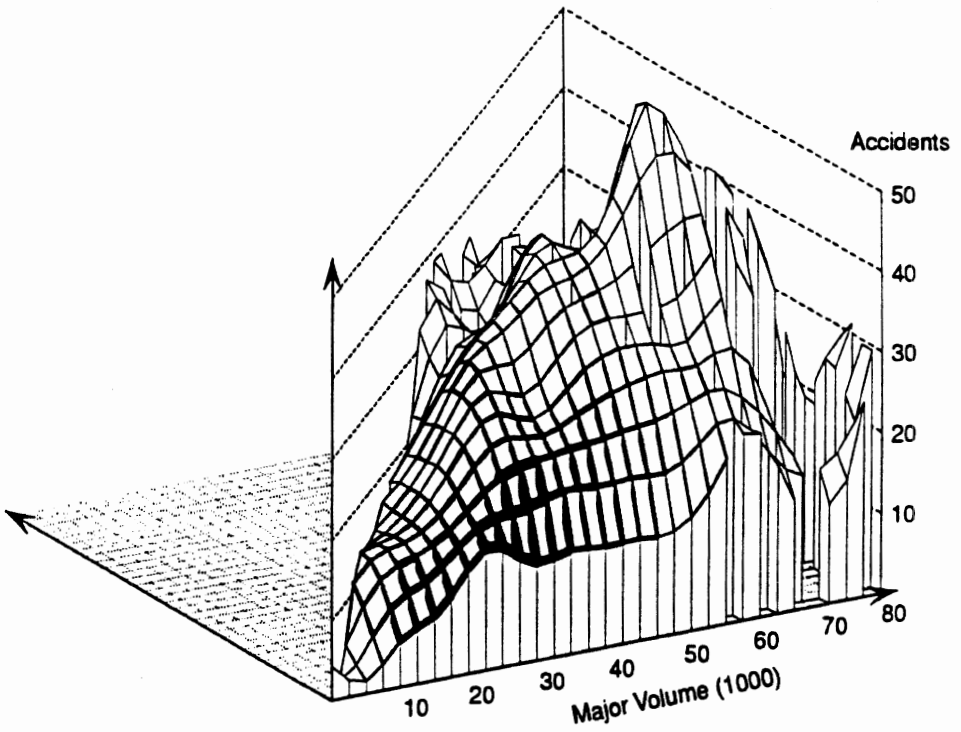


Figure 22. California four-leg signalized urban intersections. Total accident count smoothed with a 5,000 x 5,000 window.

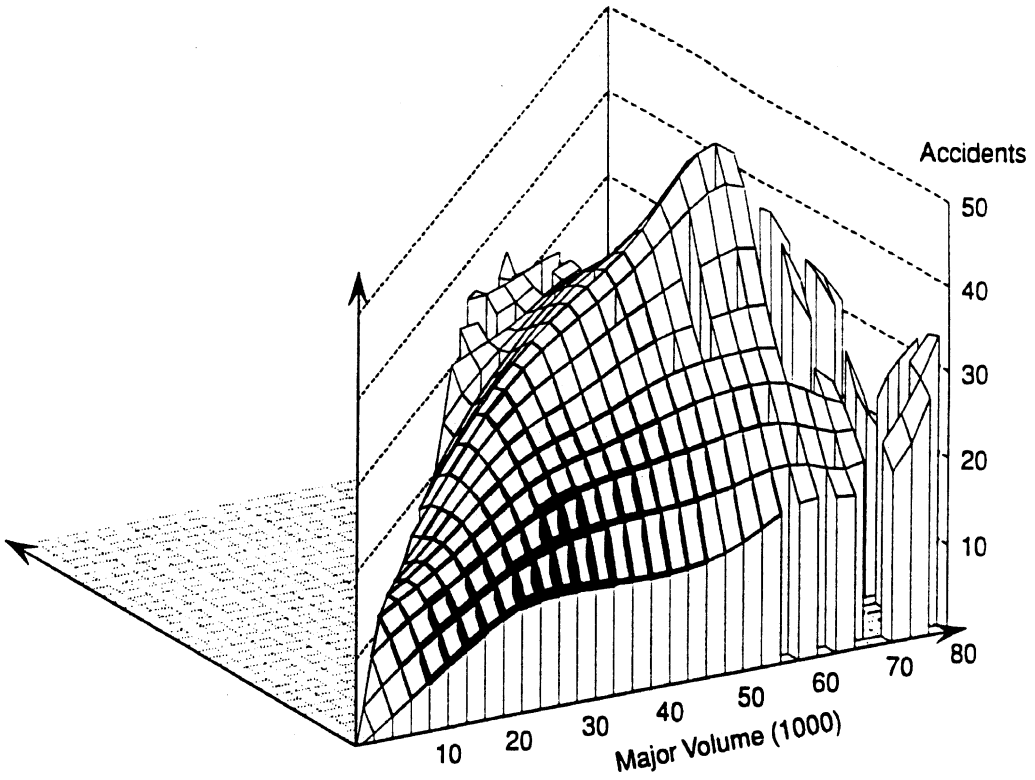


Figure 23. California four-leg signalized urban intersections. Total accident count smoothed with a 10,000 x 5,000 window.

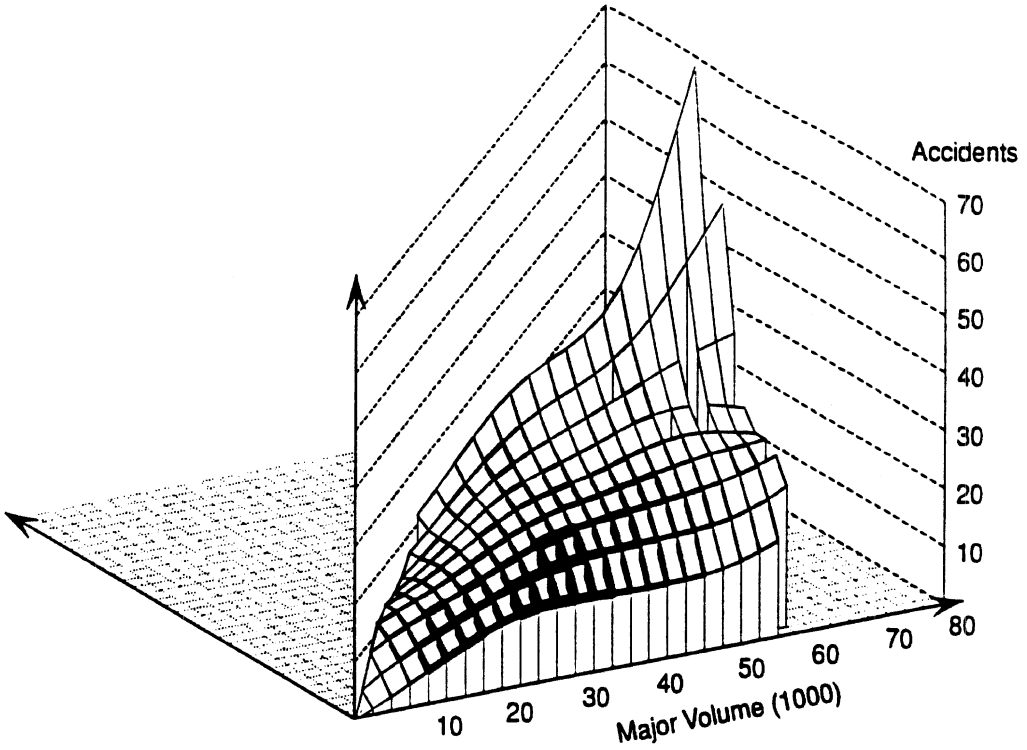


Figure 24. California four-leg signalized urban intersections. Major volume # 60,000, minor volume # 20,000. Total accidents, smoothed with a 10,000 x 5,000 window.

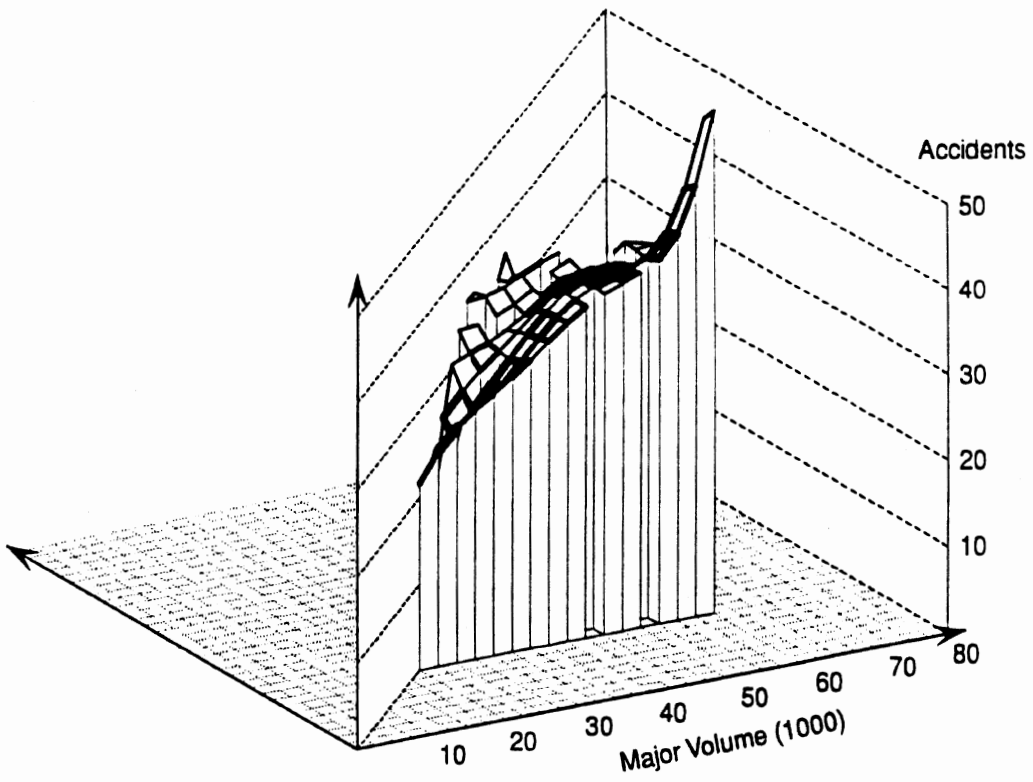


Figure 25. California four-leg signalized urban intersections. Total accidents for intersections with major volume # 60,000, minor volume > 20,000, smoothed with a 10,000 x 5,000 window.

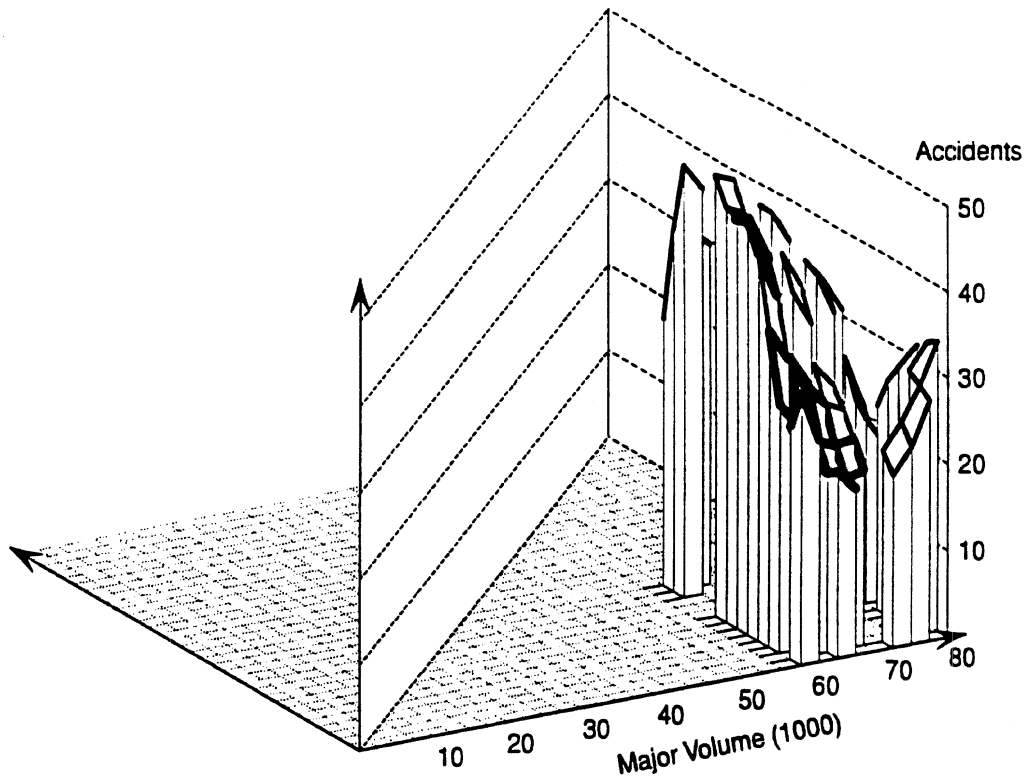


Figure 26. California four-leg signalized urban intersections. Total accidents for intersections with major volume > 60,000.

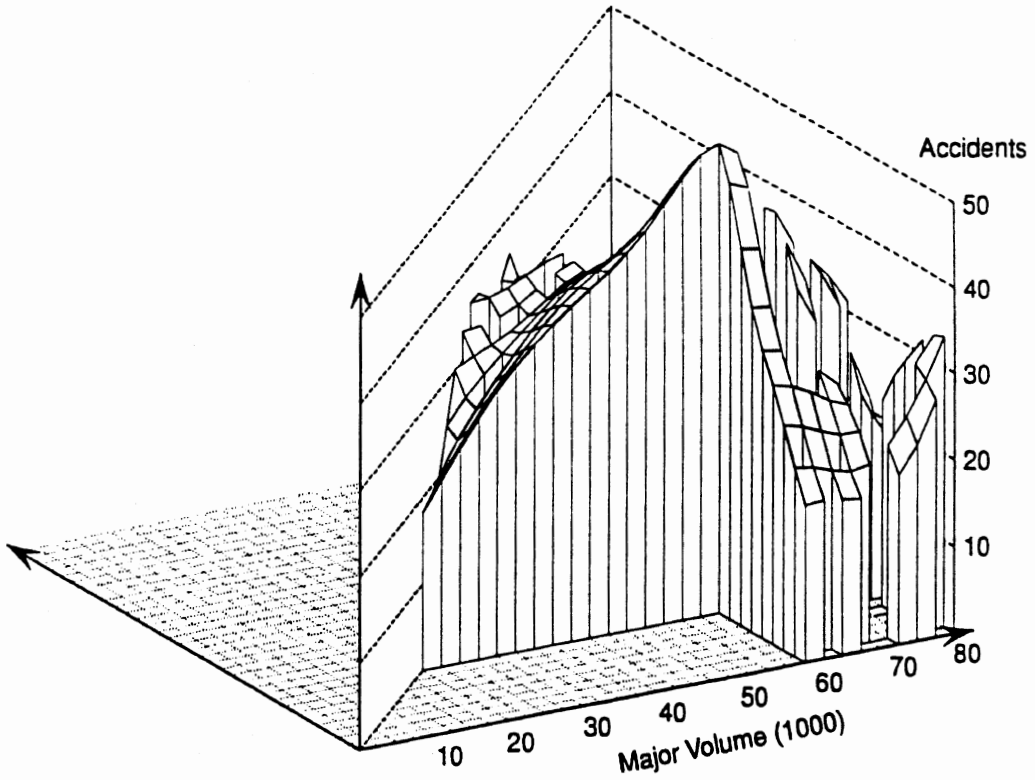


Figure 27. California four-leg signalized urban intersections with the same data and surface as in Figure 23, but with surface for major volume # 60,000 and minor volume # 20,000 not shown.

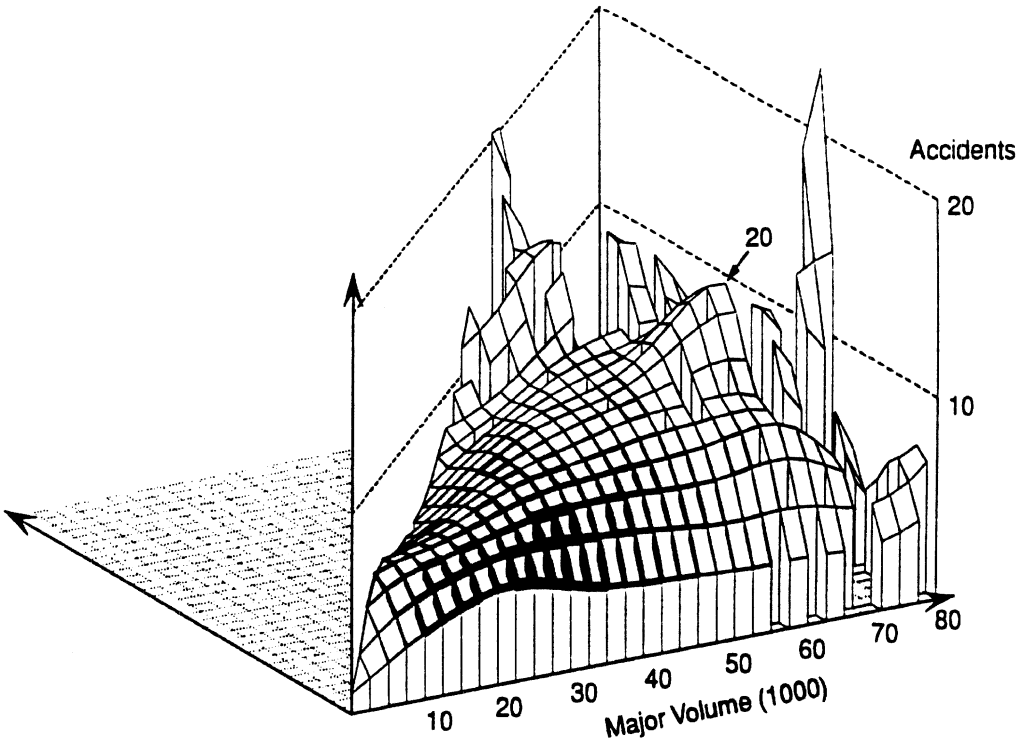


Figure 28. California four-leg signalized urban intersections. Total accidents within the intersection, smoothed with a 10,000 x 5,000 window.

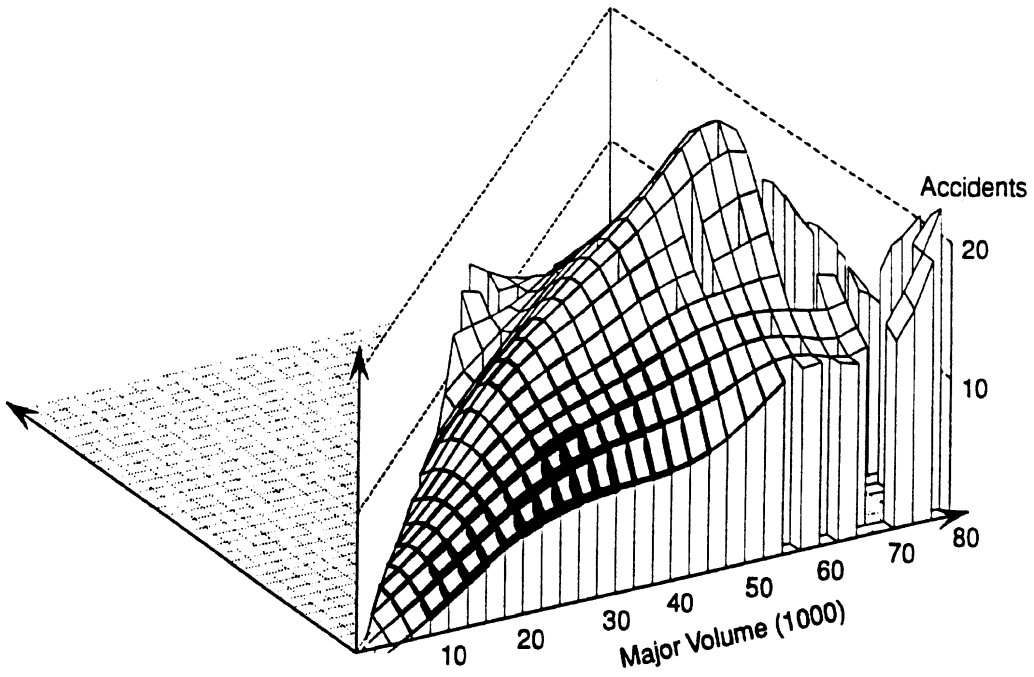


Figure 29. California four-leg signalized urban intersections. Total accidents on major approaches, smoothed with a 10,000 x 5,000 window.

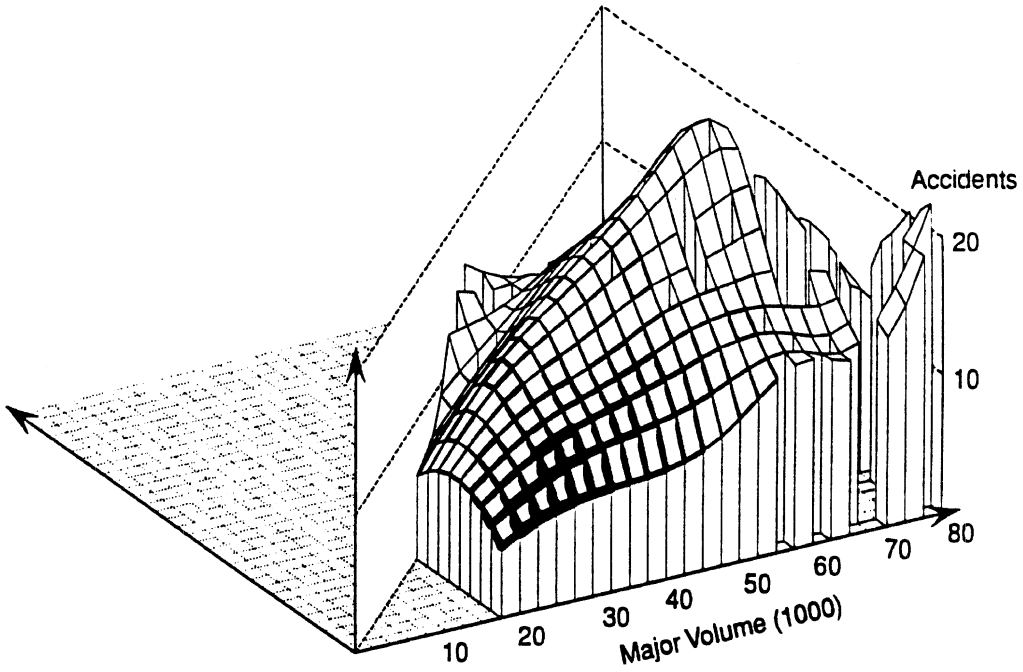


Figure 30. California four-leg signalized urban intersections. The same data and smoothing as in Figure 29, but with the surface not shown below minor volume of 20,000.

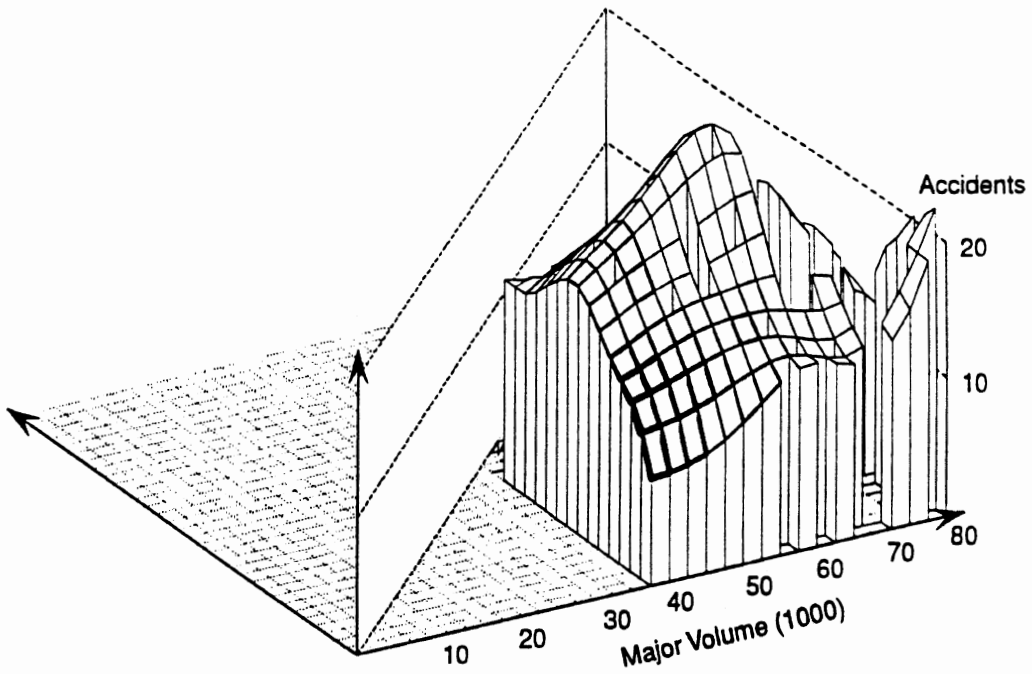


Figure 31. California four-leg signalized urban intersections. The same data and surface as in Figure 29, but with the surface below minor volumes of 40,000 not shown.

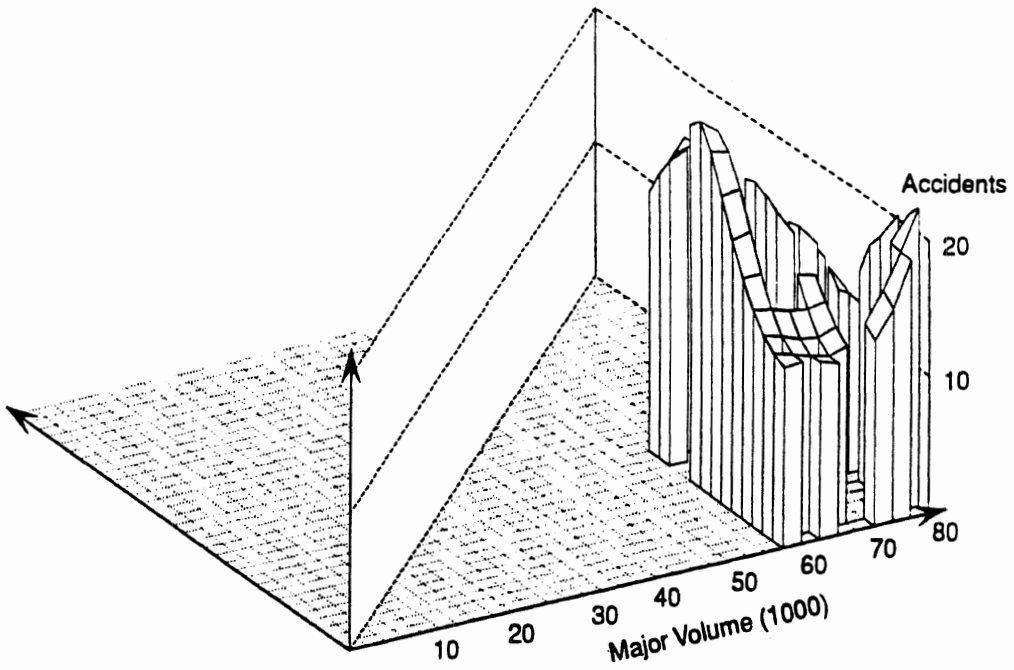


Figure 32. California four-leg signalized urban intersections. The same data and surface as in Figure 29, but with the surface below minor volume of 60,000 not shown.

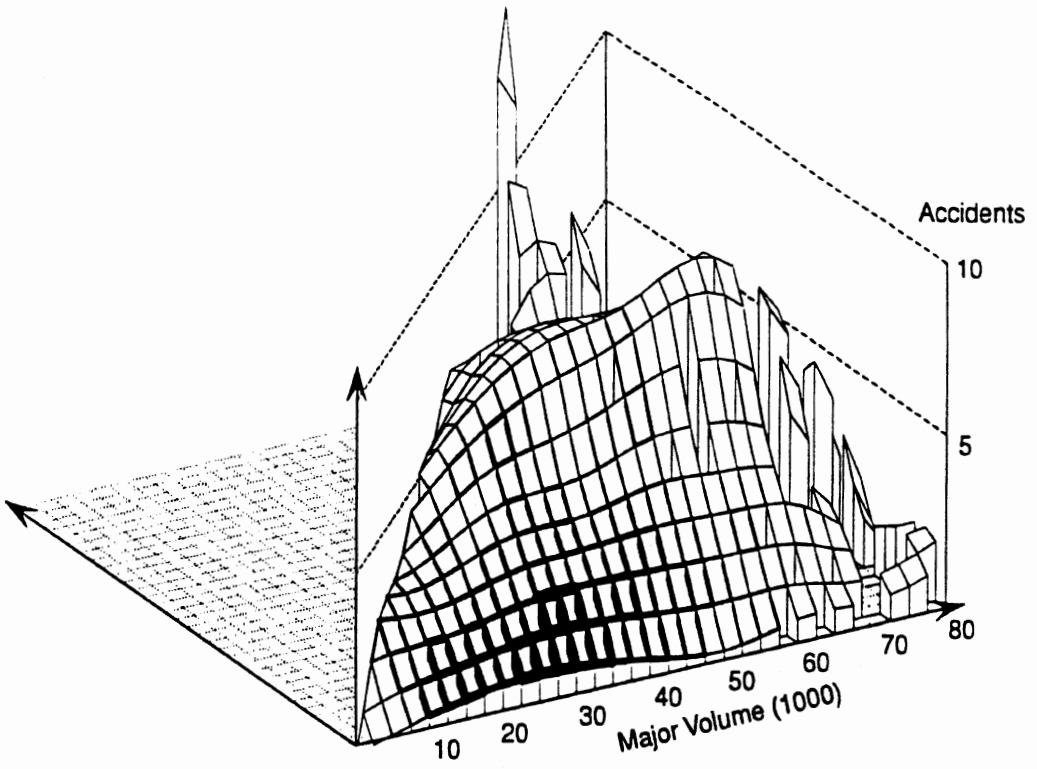


Figure 33. California four-leg signalized urban intersections. Total accidents on minor approaches, smoothed with a 10,000 x 5,000 window.

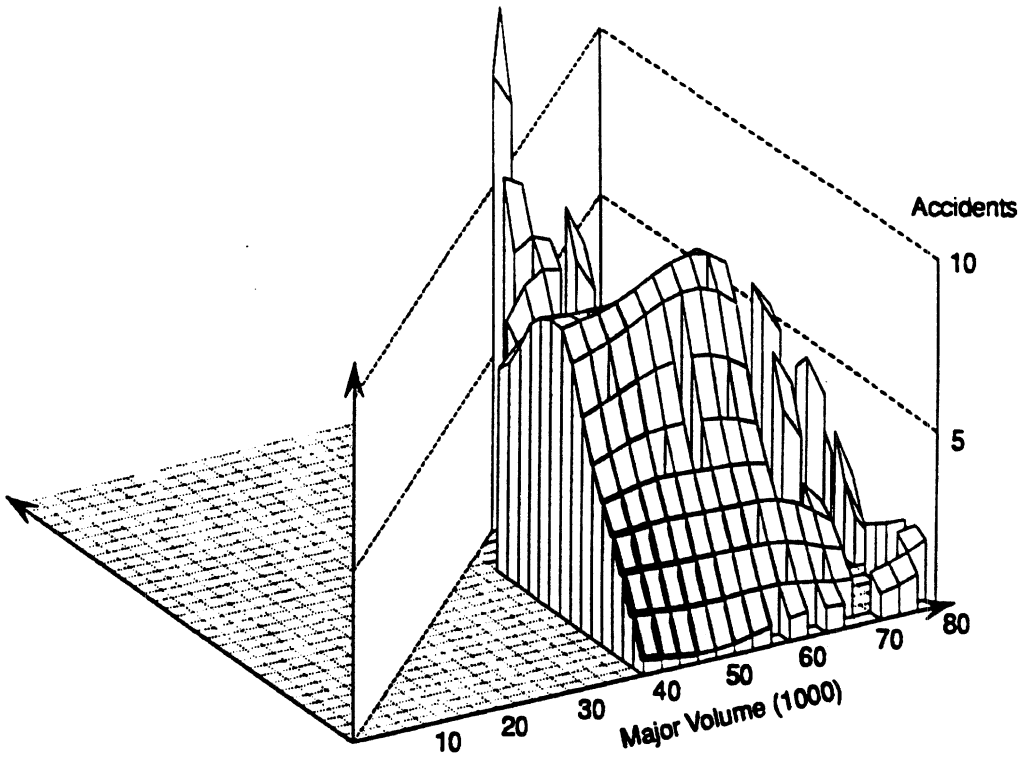


Figure 34. The same surface as in Figure 33, but not shown for major volume below 20,000.

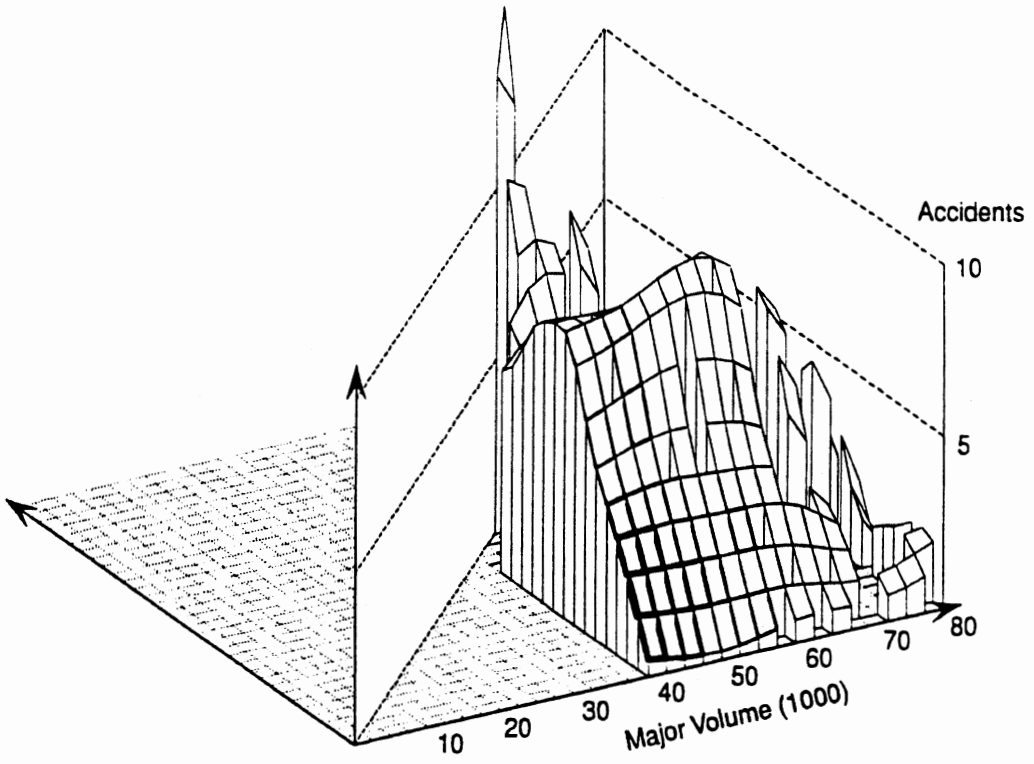


Figure 35. The same surface as in Figure 33, but not shown for major volumes below 40,000.

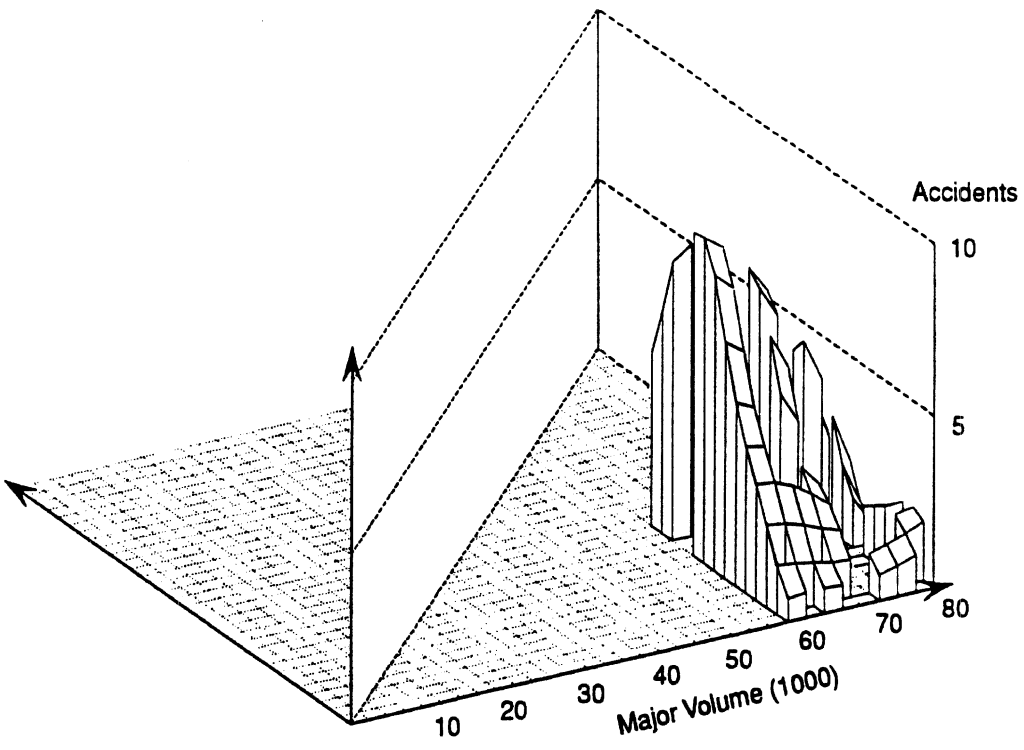


Figure 36. The same surface as in Figure 33, but with major volumes not shown below 60,000.

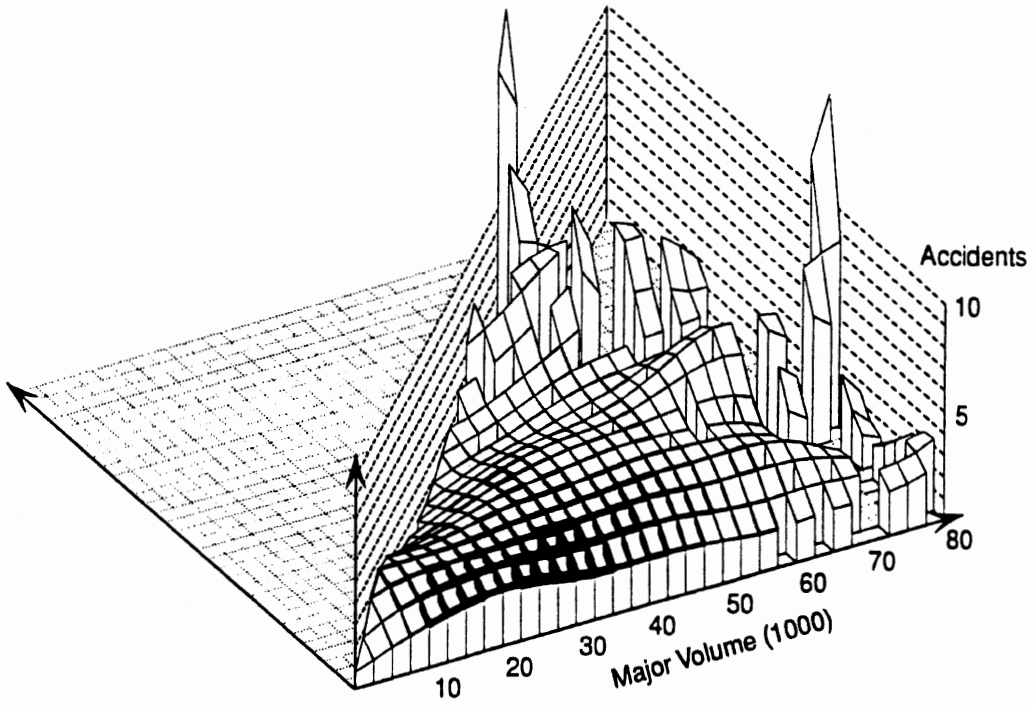


Figure 37. California four-leg signalized urban intersections. Left-turn accidents within the intersection, smoothed with a 10,000 x 5,000 window.

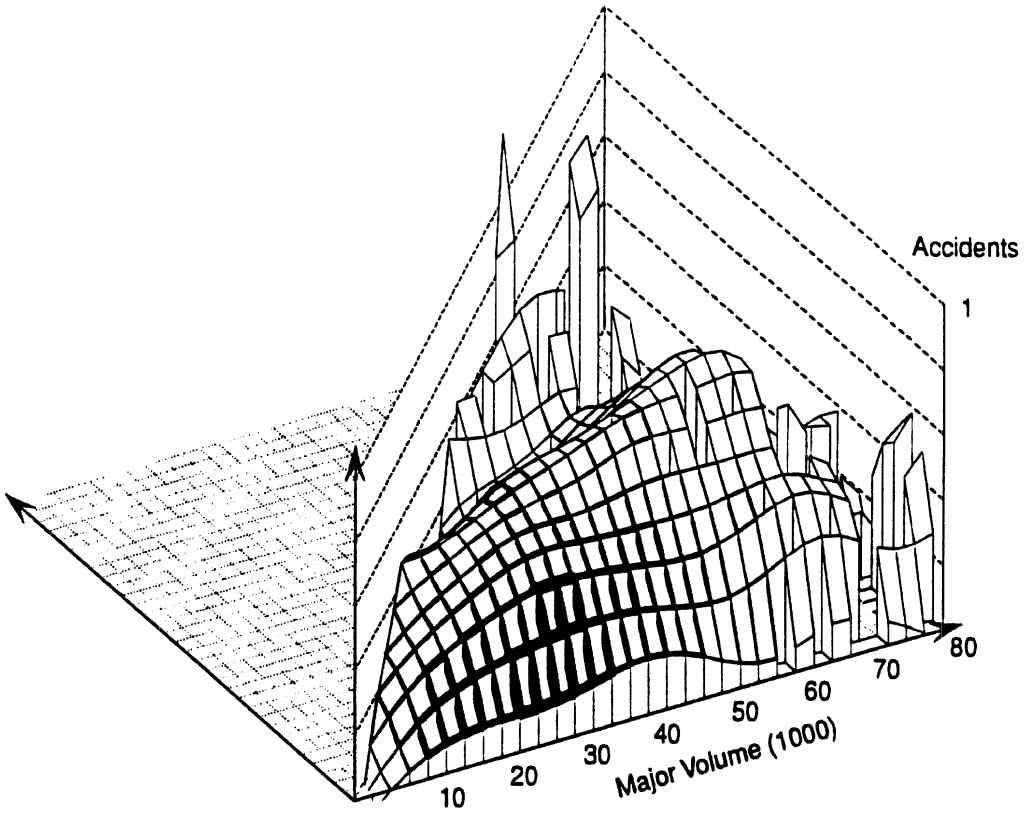


Figure 38. California four-leg signalized urban intersections. Right-turn accidents within the intersection, smoothed with a 10,000 x 5,000 window.

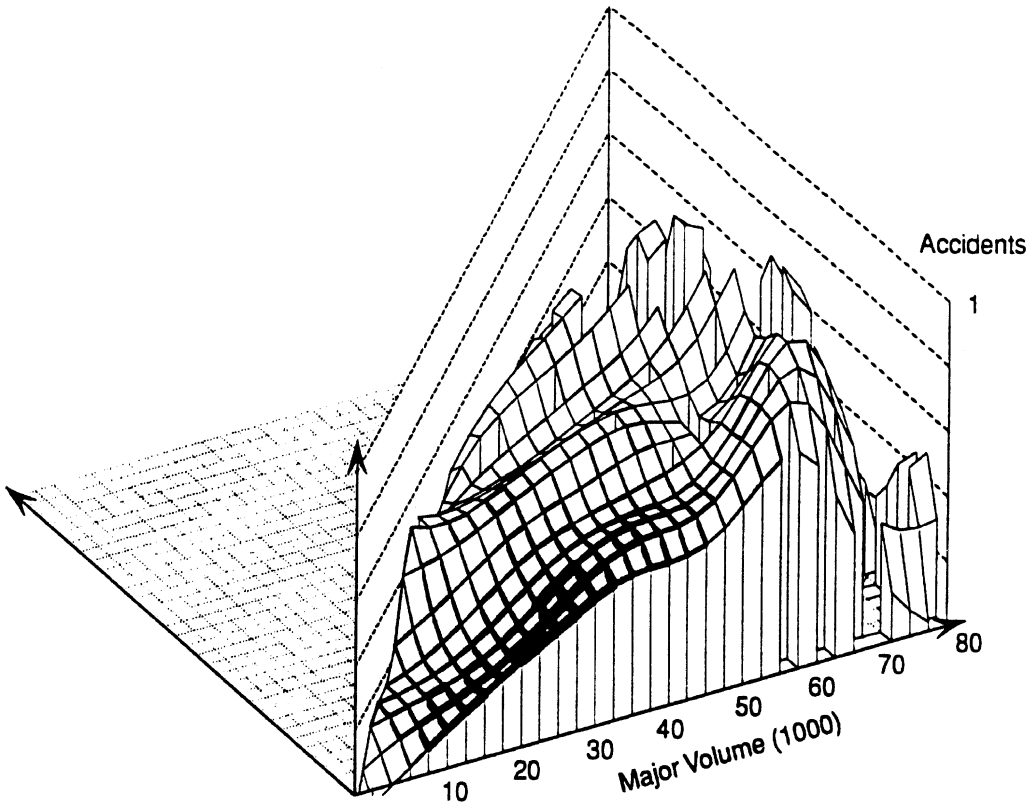


Figure 39. California four-leg signalized urban intersections. Rear-end collisions within the intersection, smoothed with a 10,000 x 5,000 window.

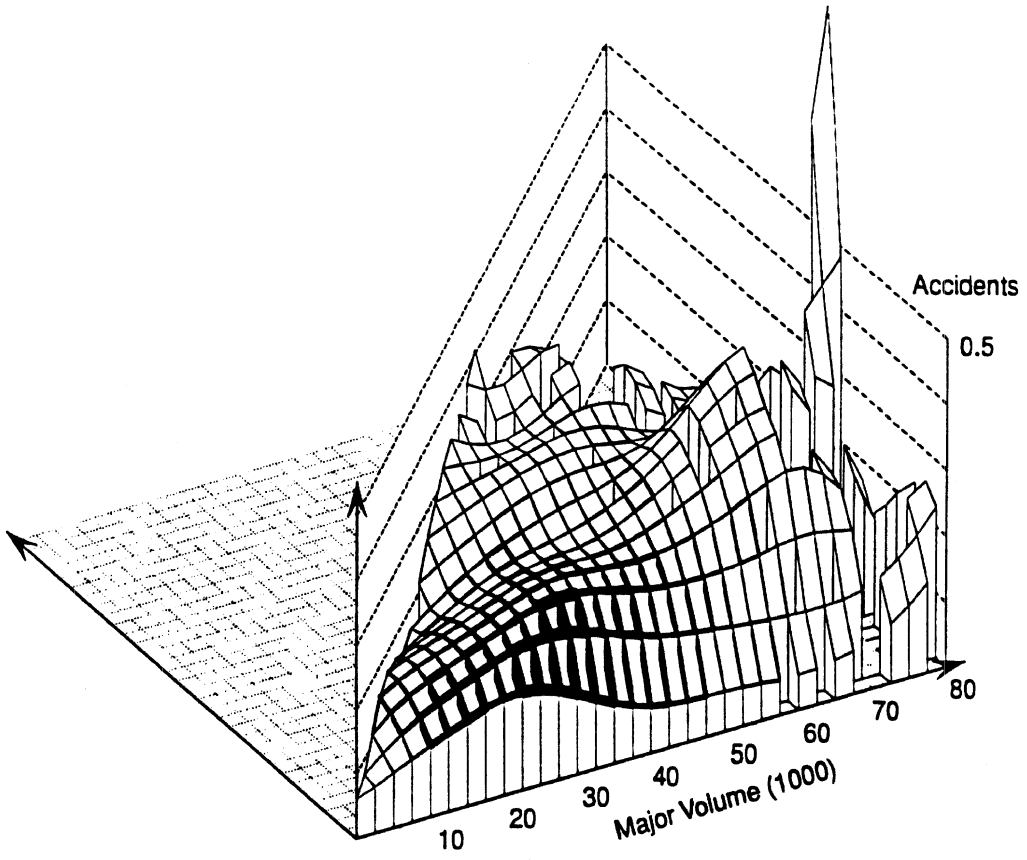


Figure 40. California four-leg signalized urban intersections. Angle collisions within the intersection, smoothed with a 10,000 x 5,000 window.

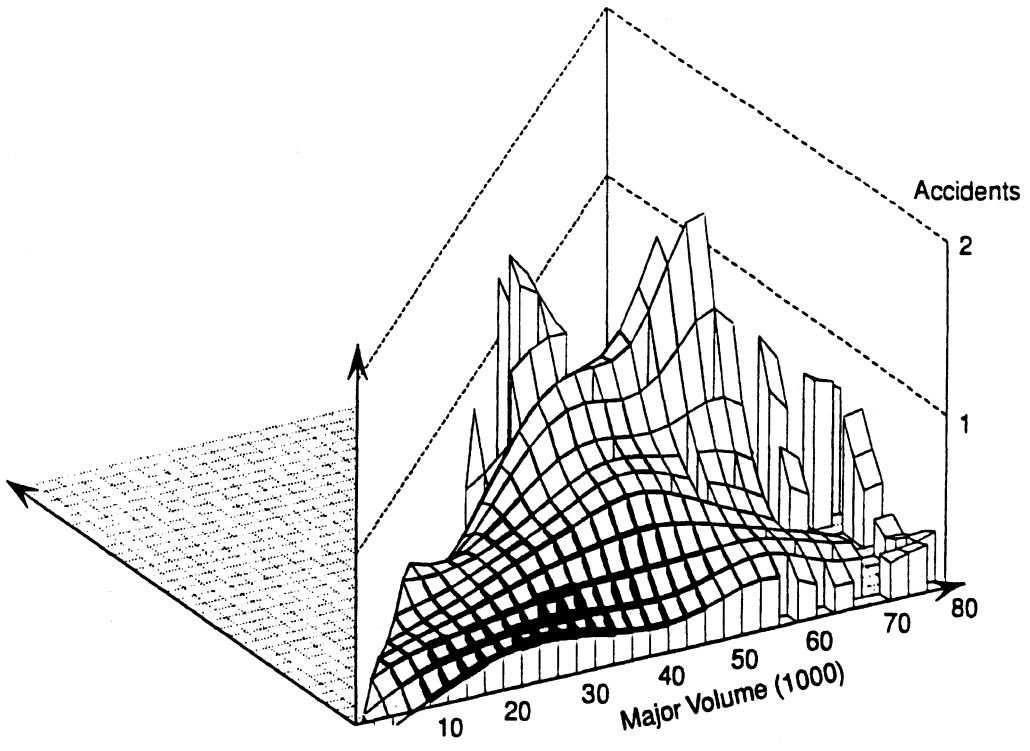


Figure 41. California four-leg signalized urban intersections. "Other" collisions within the intersection, smoothed with a 10,000 x 5,000 window.

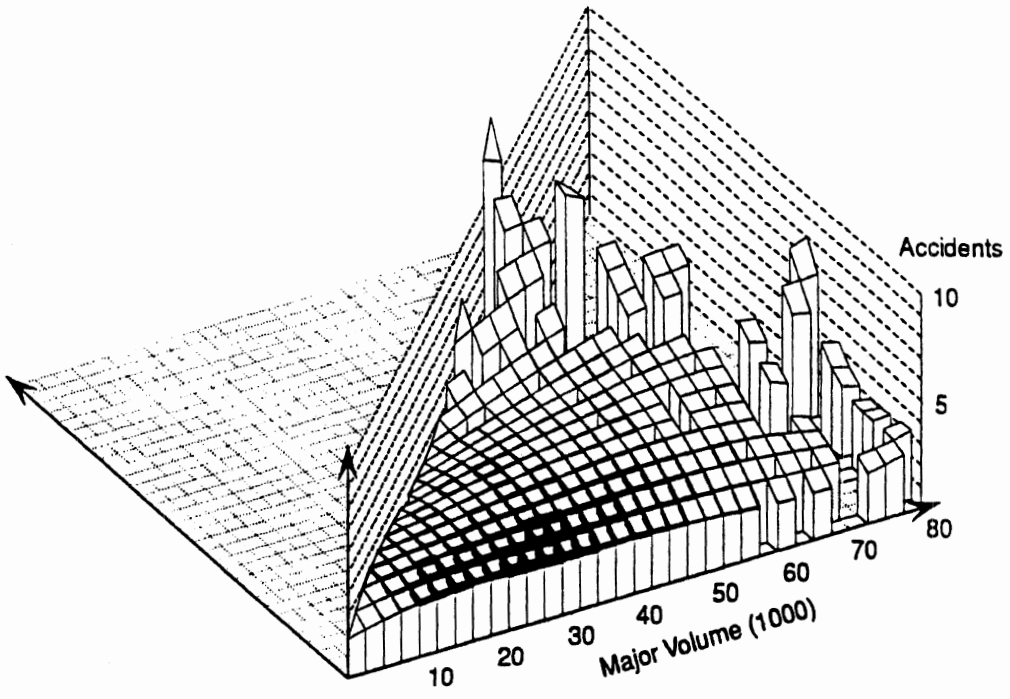


Figure 42. California four-leg signalized urban intersections. Left-turn collisions within intersection, smoothed with a 15,000 x 10,000 window. Based on the same data as Figure 37 but more smoothed.

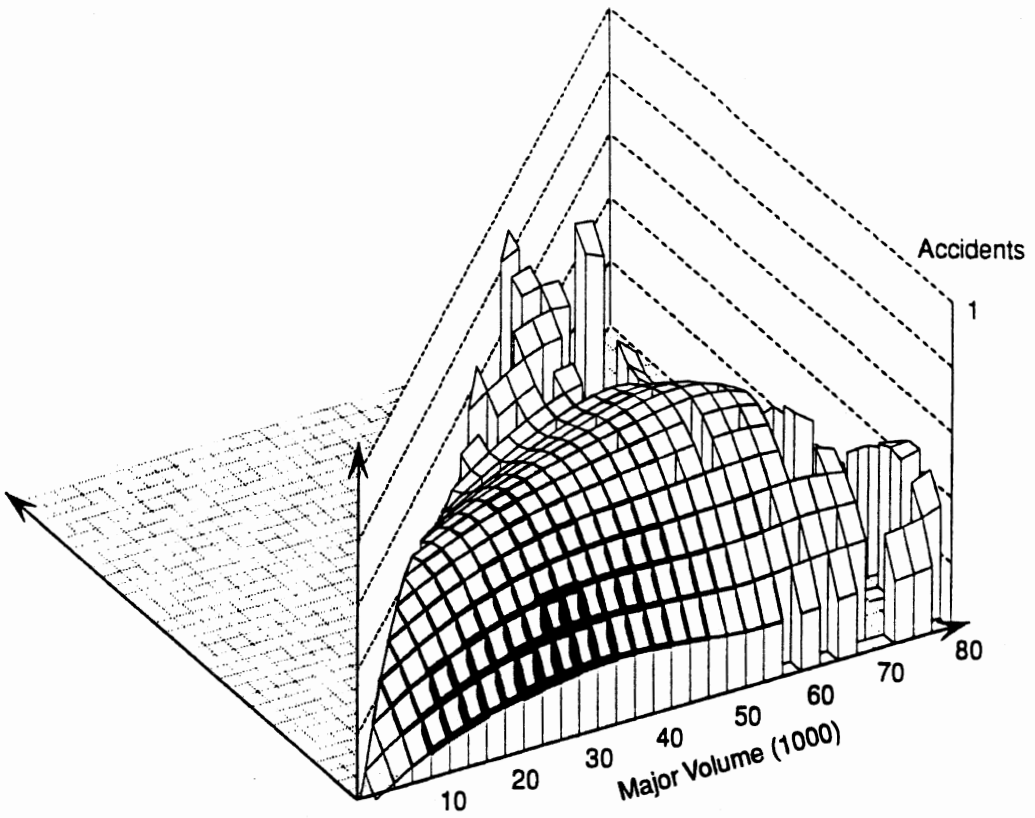


Figure 43. California four-leg signalized urban intersections. Right-turn collision within intersection, smoothed with a 15,000 x 10,000 window. Based on the same data as Figure 38 but more smoothed.

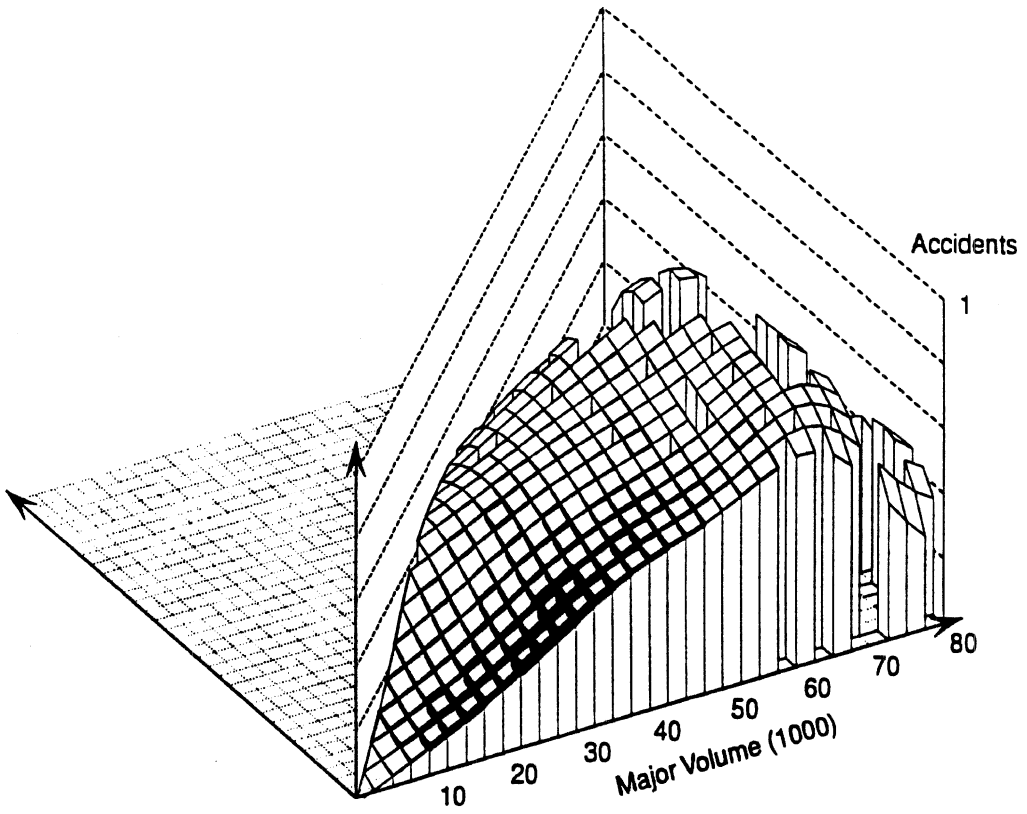


Figure 44. California four-leg signalized urban intersections. Rear-end collisions within intersection, smoothed with a 15,000 x 10,000 window. Based on the same data as Figure 39 but more smoothed.

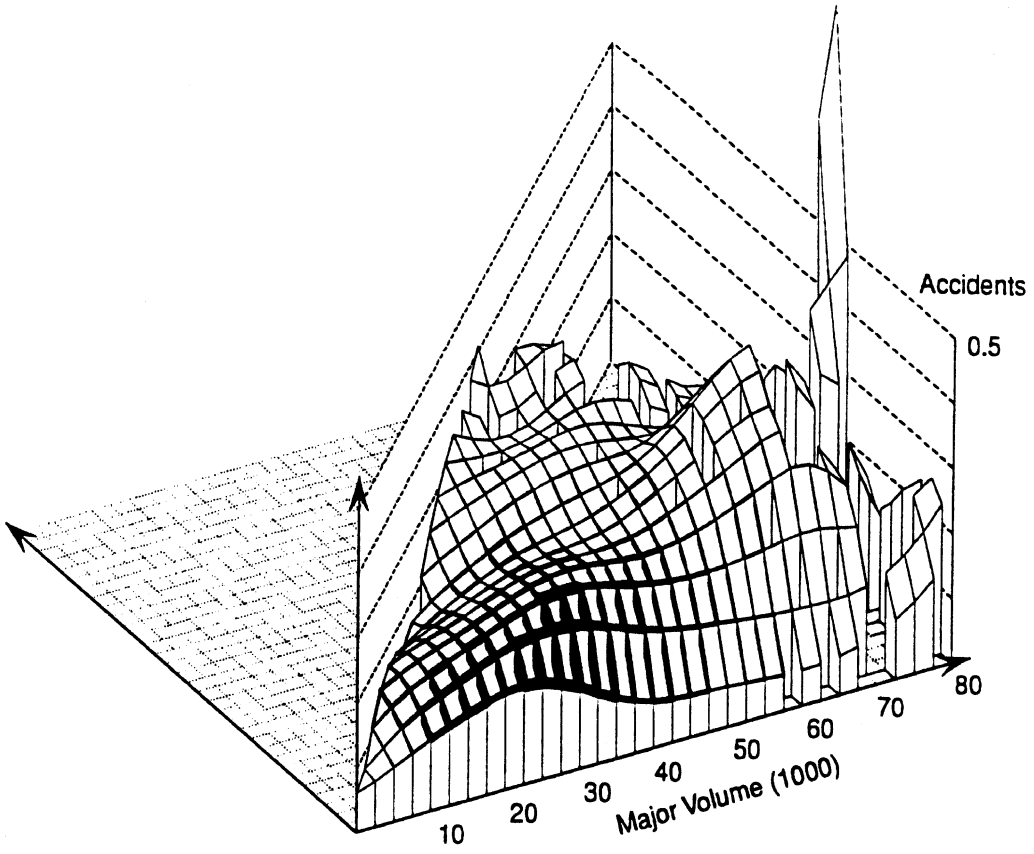


Figure 45. California four-leg signalized urban intersections. Angle - collision within intersection, smoothed with a 15,000 x 10,000 window. Based on the same data as Figure 40 but more smoothed.

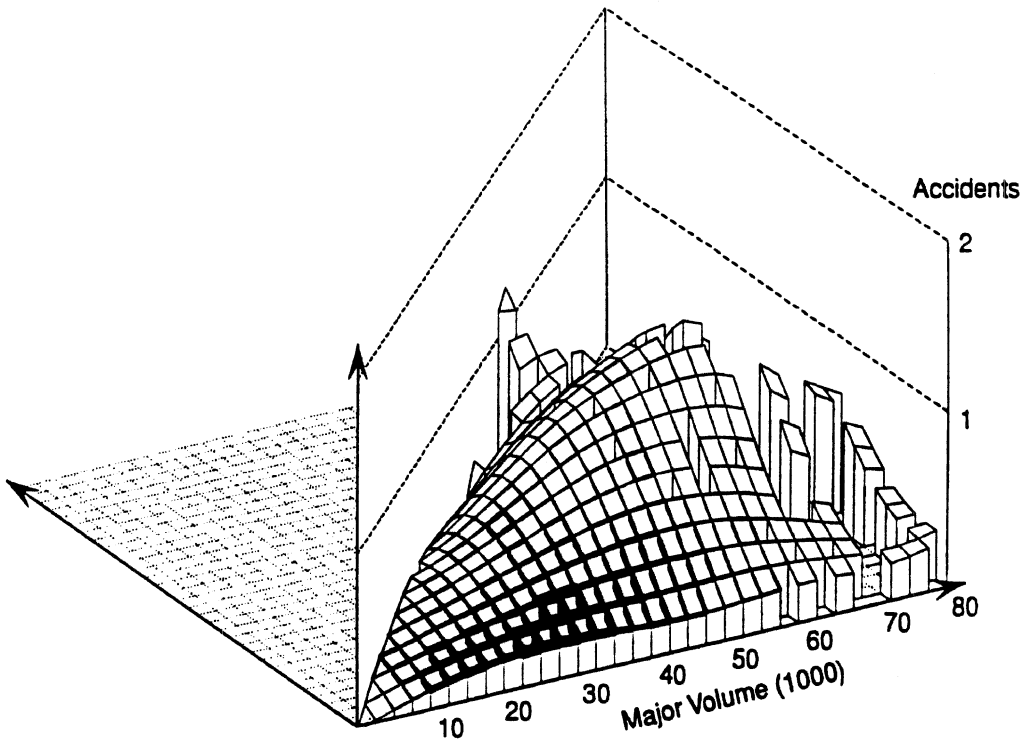


Figure 46. California four-leg signalized urban intersections. "Other" collisions within intersection, smoothed with a 15,000 x 10,000 window. Based on the same data as Figure 41 but more smoothed.

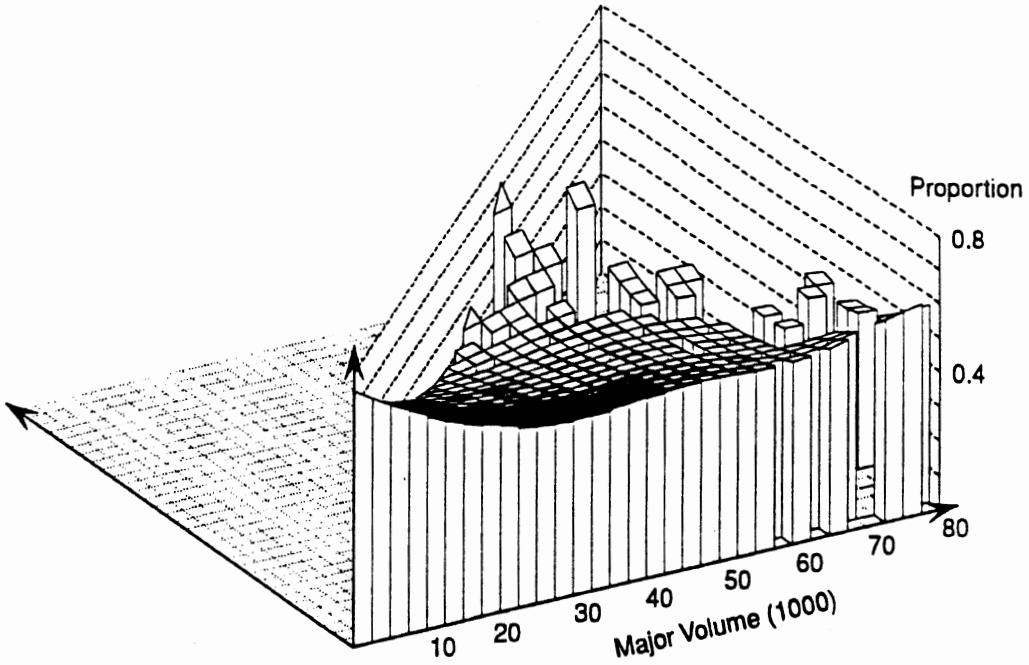


Figure 47. California four-leg signalized urban intersections. Proportion of left- and U-turn accidents within intersection, smoothed with a 15,000 x 10,000 window.

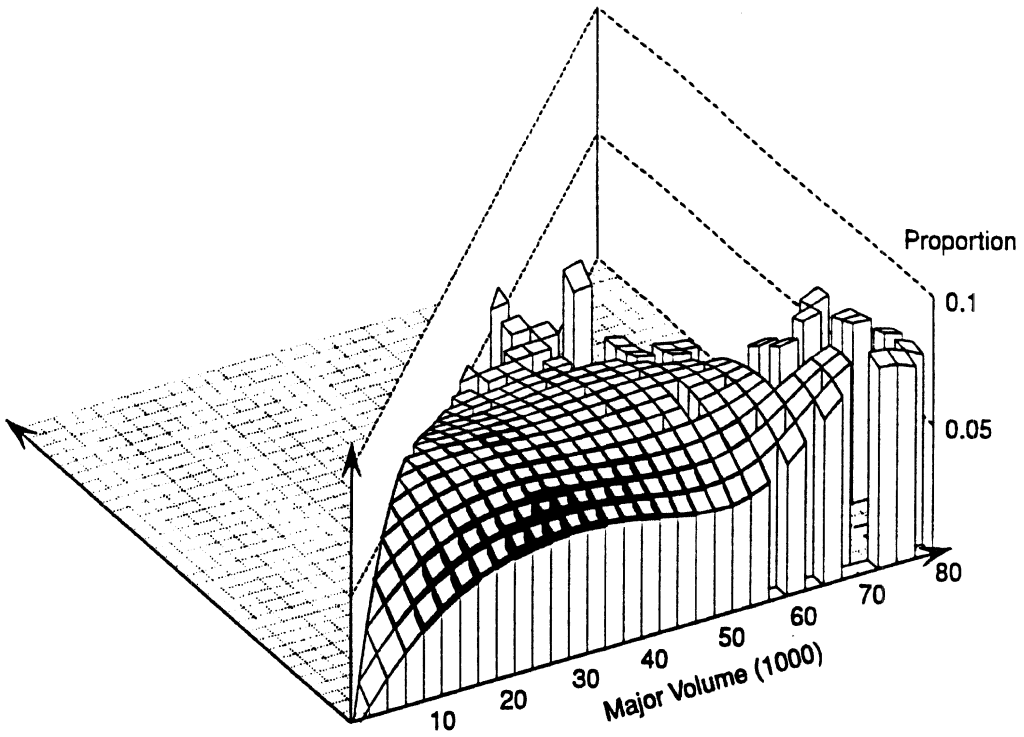


Figure 48. California four-leg signalized urban intersections. Proportion of right-turn accidents within intersections, smoothed with a 15,000 x 10,000 window.

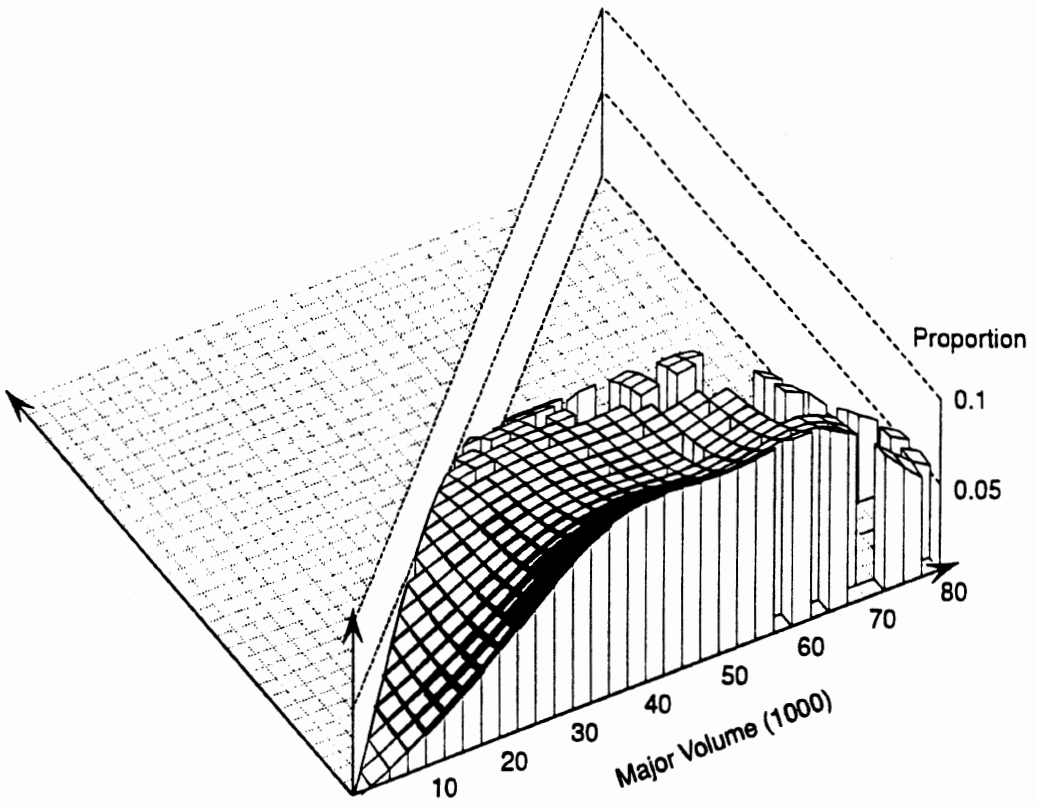


Figure 49. California four-leg signalized urban intersections. Proportion of rear-end accidents within intersections, smoothed with a 15,000 x 10,000 window.

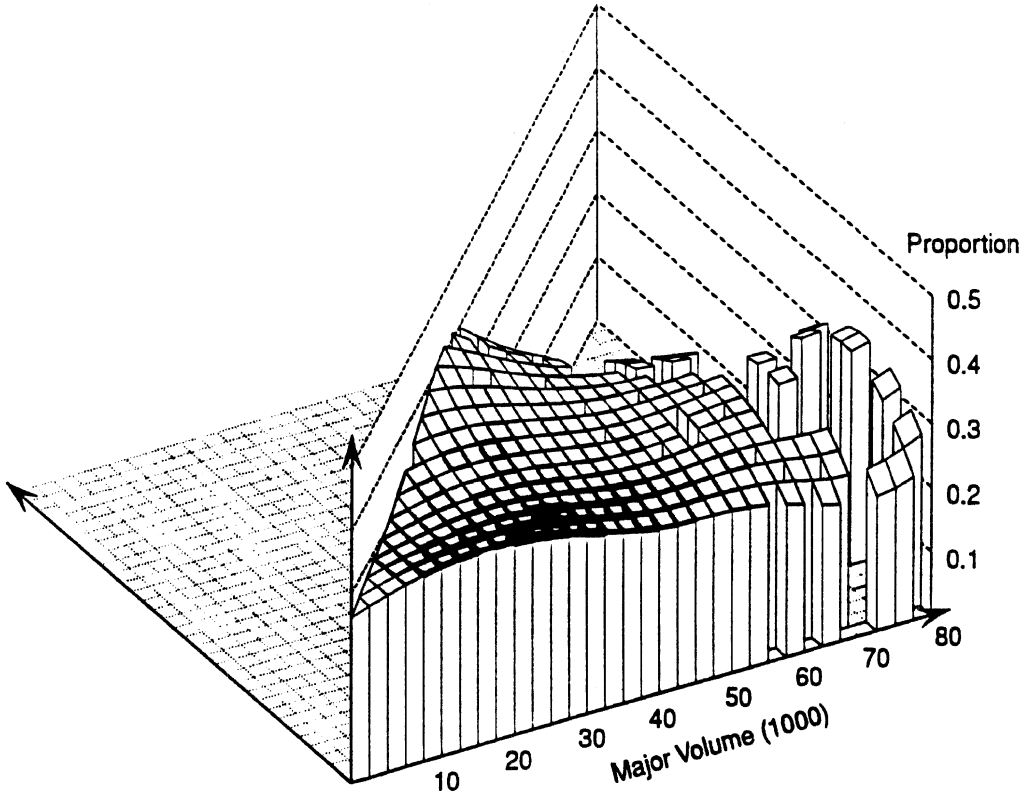


Figure 50. California four-leg signalized urban intersections. Proportion of angle accidents within intersection, smoothed with a 15,000 x 10,000 window.

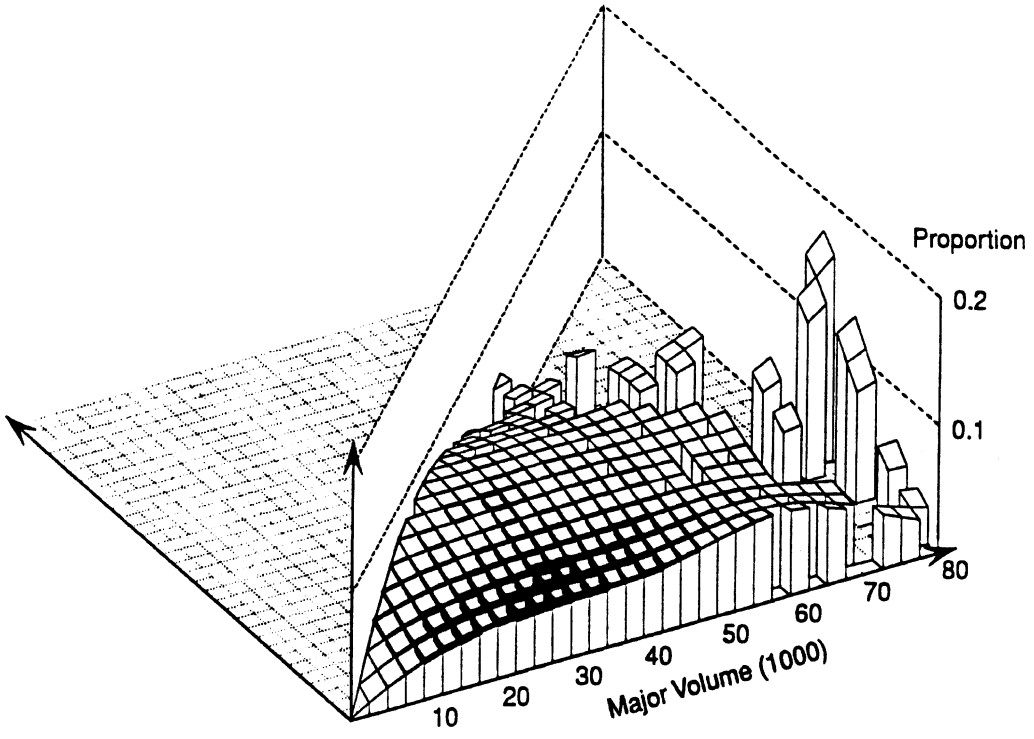


Figure 51. California four-leg signalized urban intersections. Proportion of "other" accidents within intersection, smoothed with a 15,000 x 10,000 window.

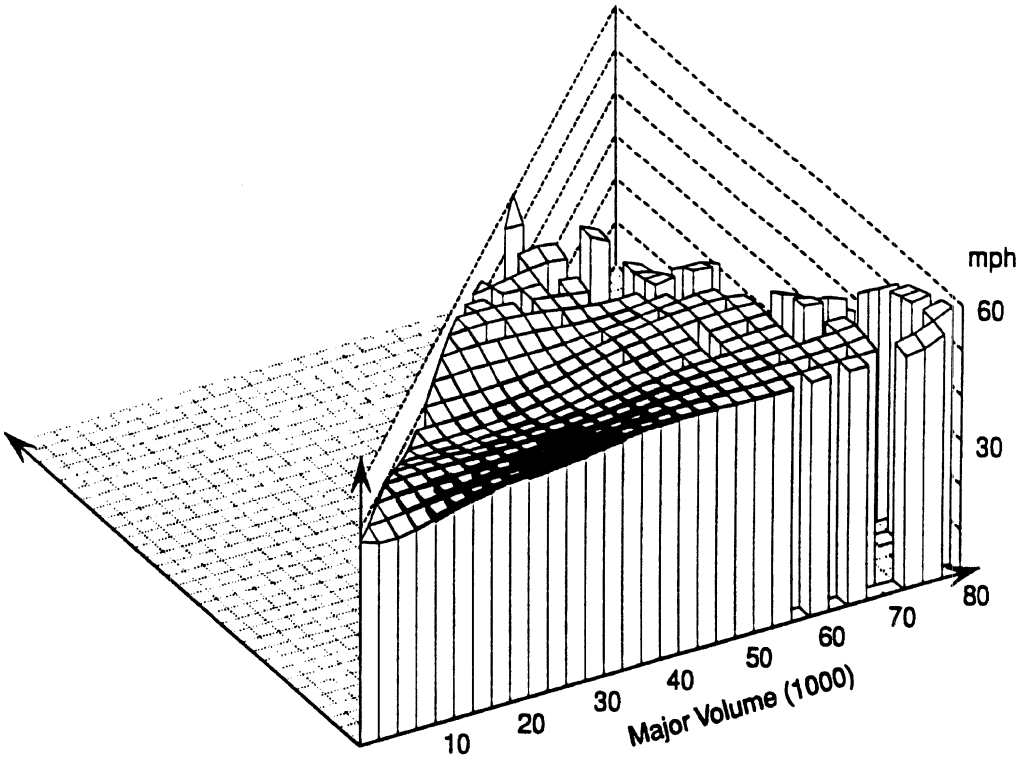


Figure 52. California four-leg signalized urban intersections. Design speed, smoothed with a 10,000 x 5,000 window.

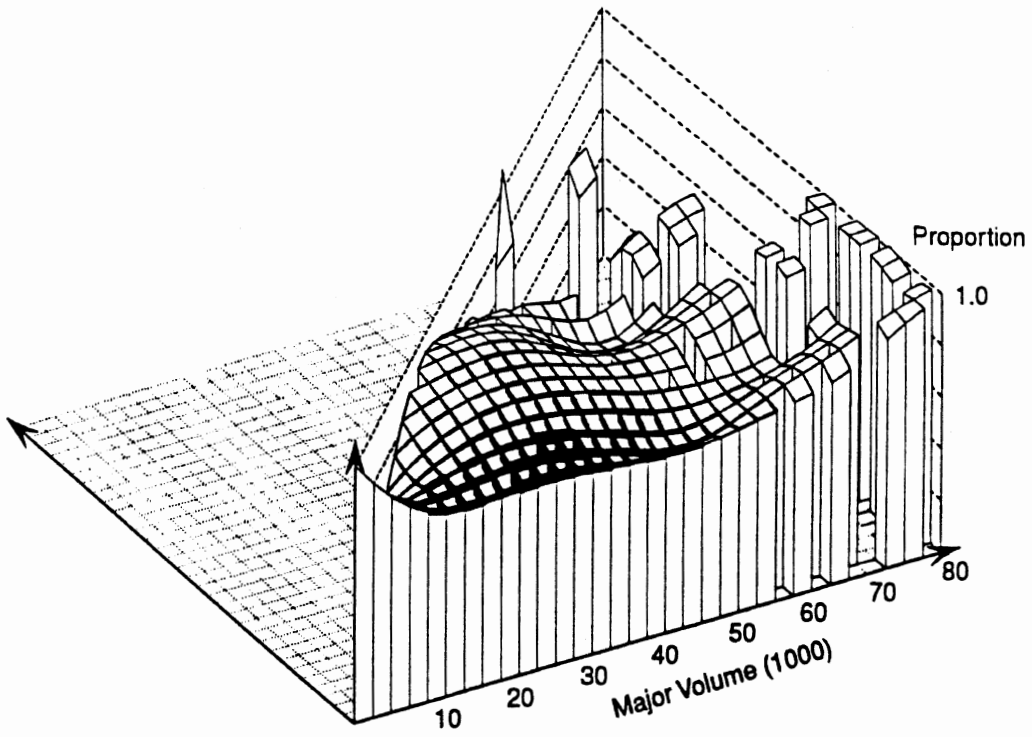


Figure 53. California four-leg signalized urban intersections. Proportion of intersections with multi-phase signals, smoothed with a 10,000 x 5,000 window.

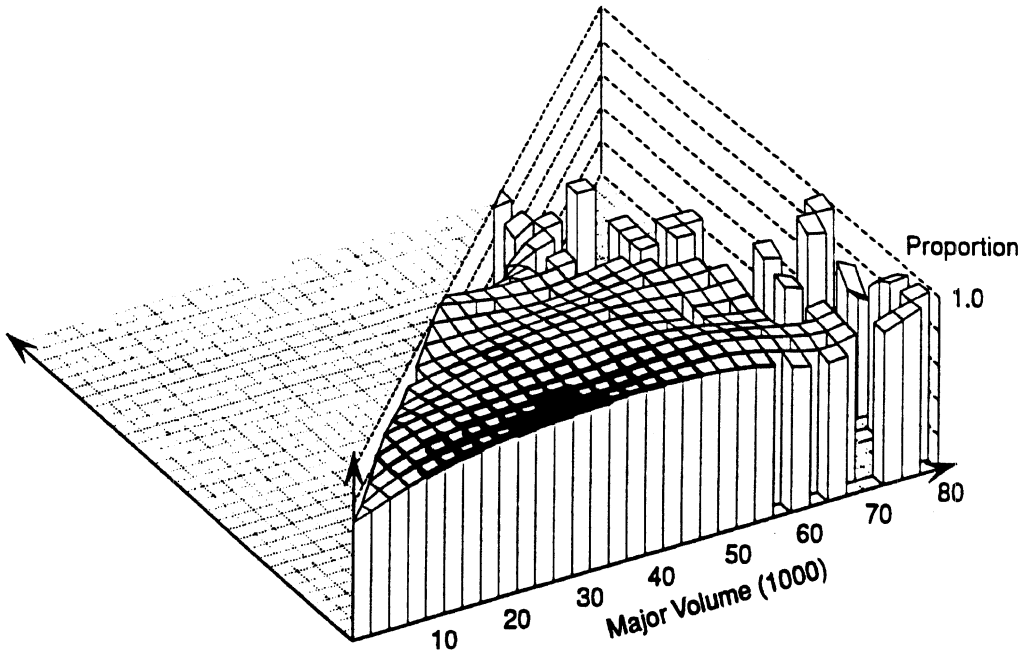


Figure 54. California four-leg signalized urban intersections. Proportion of intersections with left-turn channelization on the main road, smoothed with a 10,000 x 5,000 window.

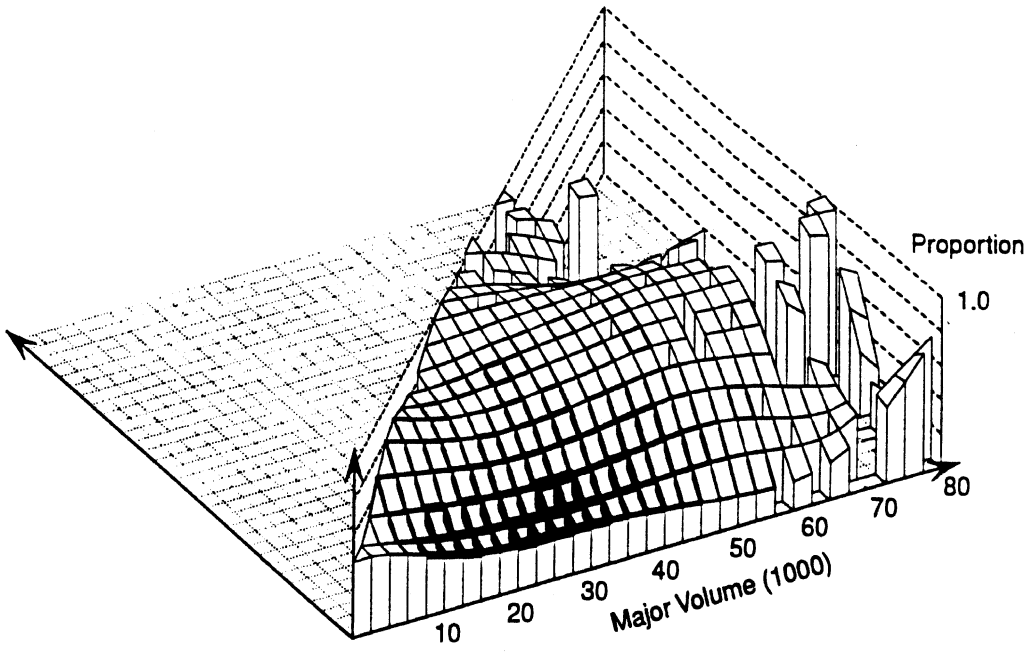


Figure 55. California four-leg signalized urban intersections. Proportion of intersections with left-turn channelization on the minor road, smoothed with a 10,000 x 5,000 window.

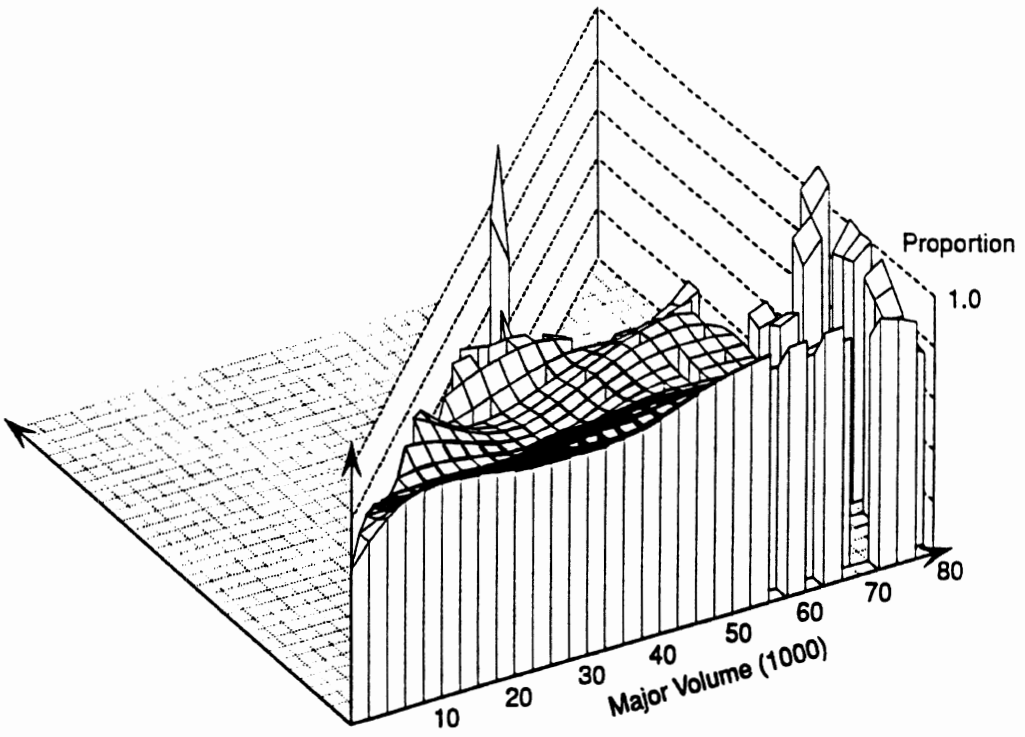


Figure 56. California four-leg signalized urban intersections. Proportion of intersections with free right turns on major road, smoothed with a 10,000 x 5,000 window.

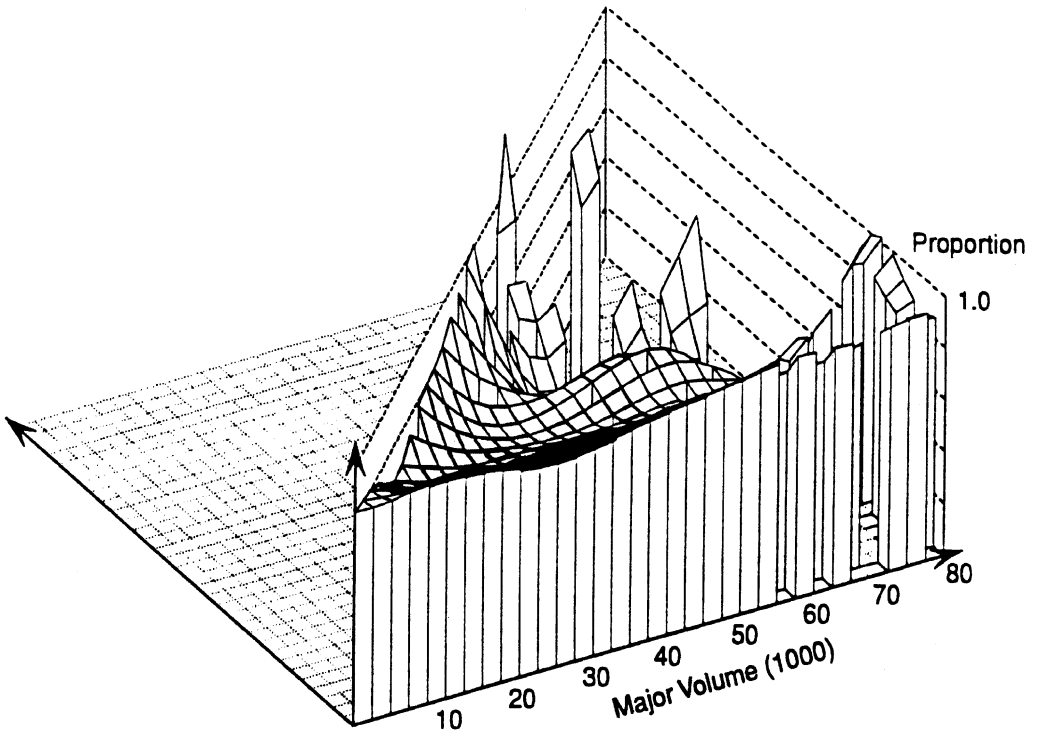


Figure 57. California four-leg signalized urban intersections. Proportion of intersections with free right turns on minor road, smoothed with a 10,000 x 5,000 window.

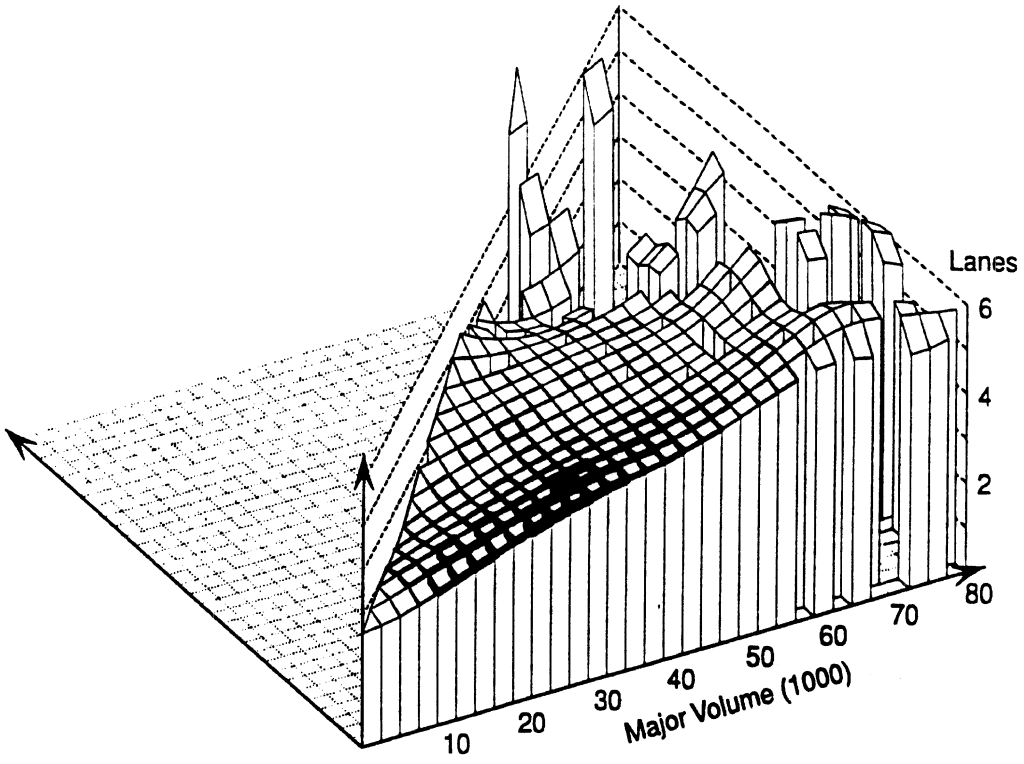


Figure 58. California four-leg signalized urban intersections. Number of lanes on major road, smoothed with a 10,000 x 5,000 window.

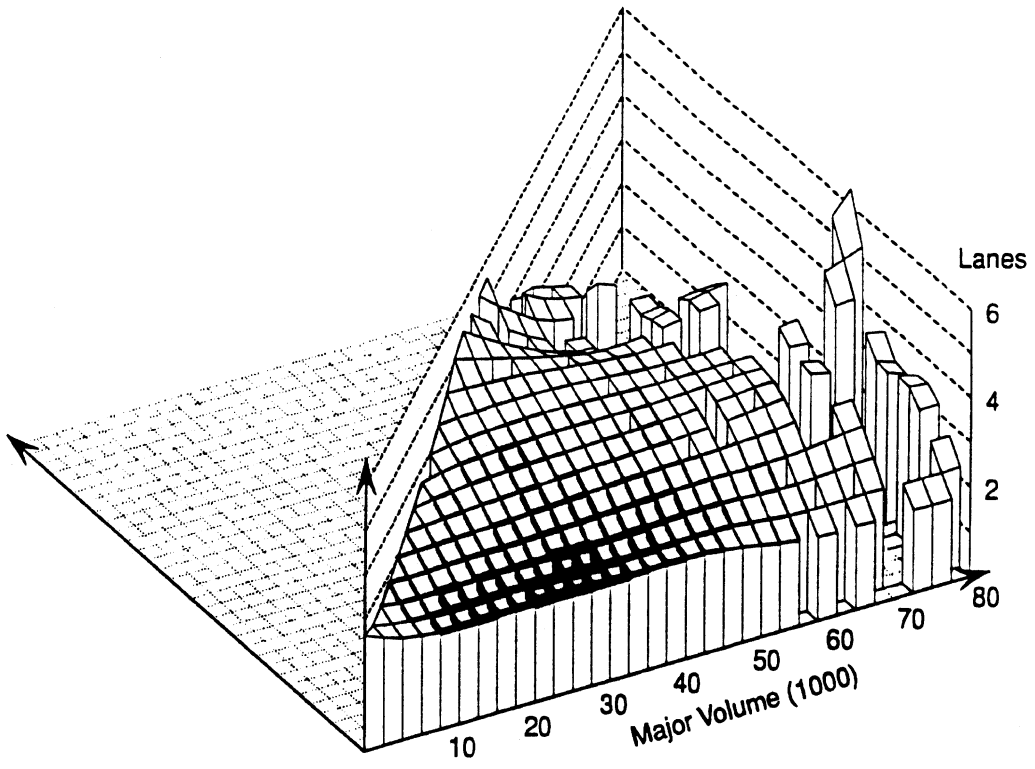


Figure 59. California four-leg signalized urban intersections. Number of lanes on minor road, smoothed with a 10,000 x 5,000 window.

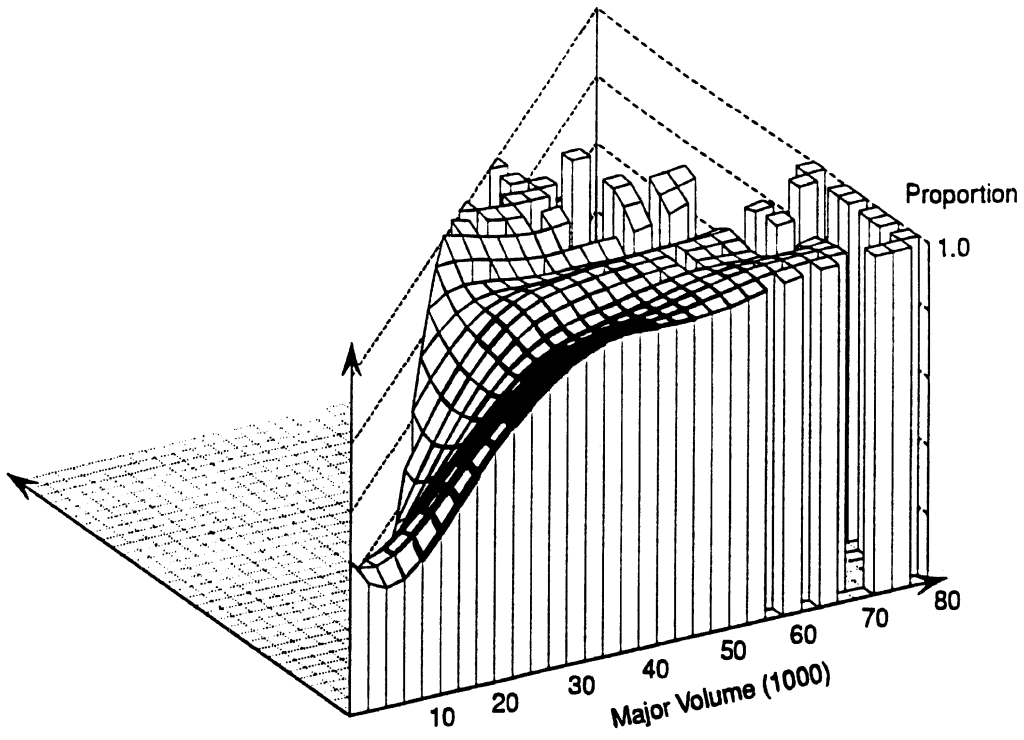


Figure 60. California four-leg signalized urban intersections with median on main road, smoothed with a 10,000 x 5,000 window.

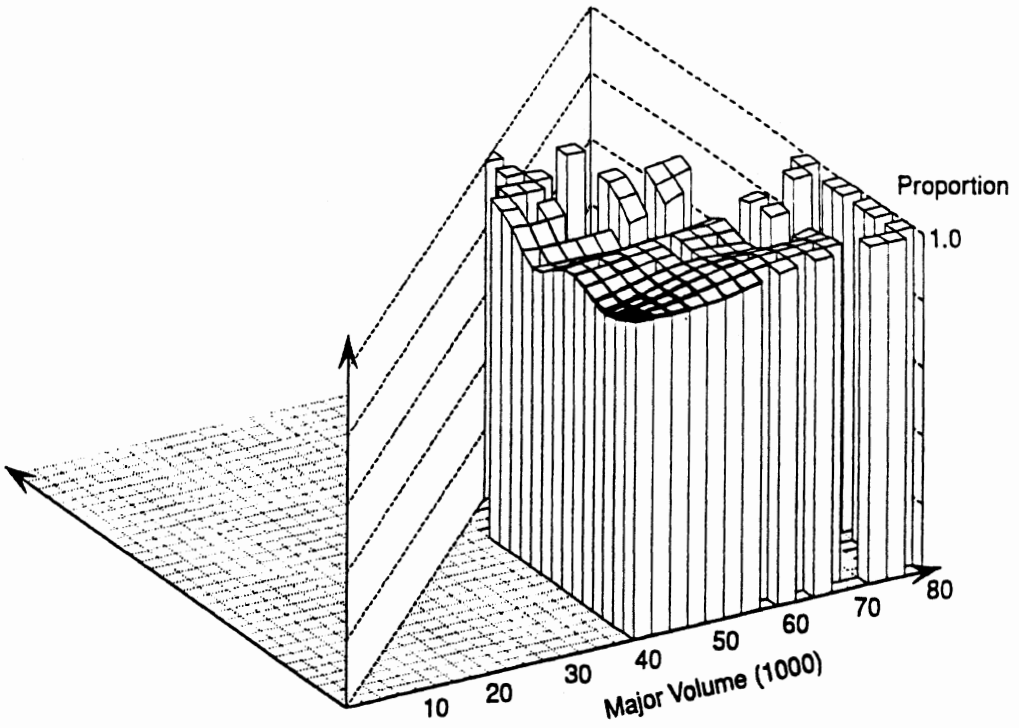


Figure 61. The same surface as in Figure 60, shown only for major volume above 40,000.

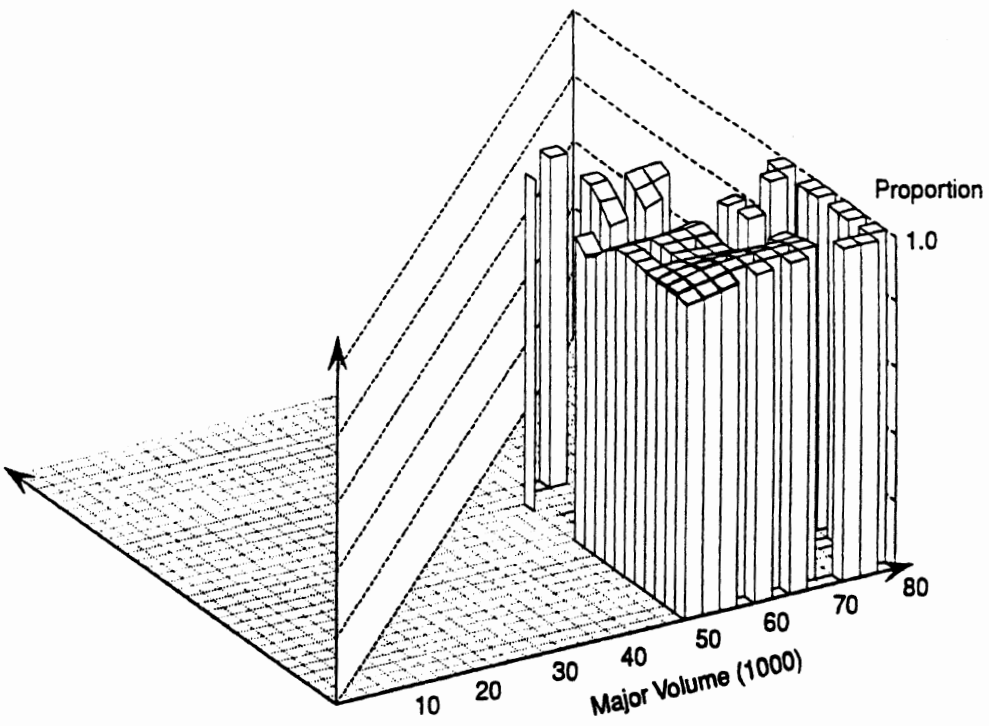


Figure 62. The same surface as Figure 60, shown only for major volume above 50,000.

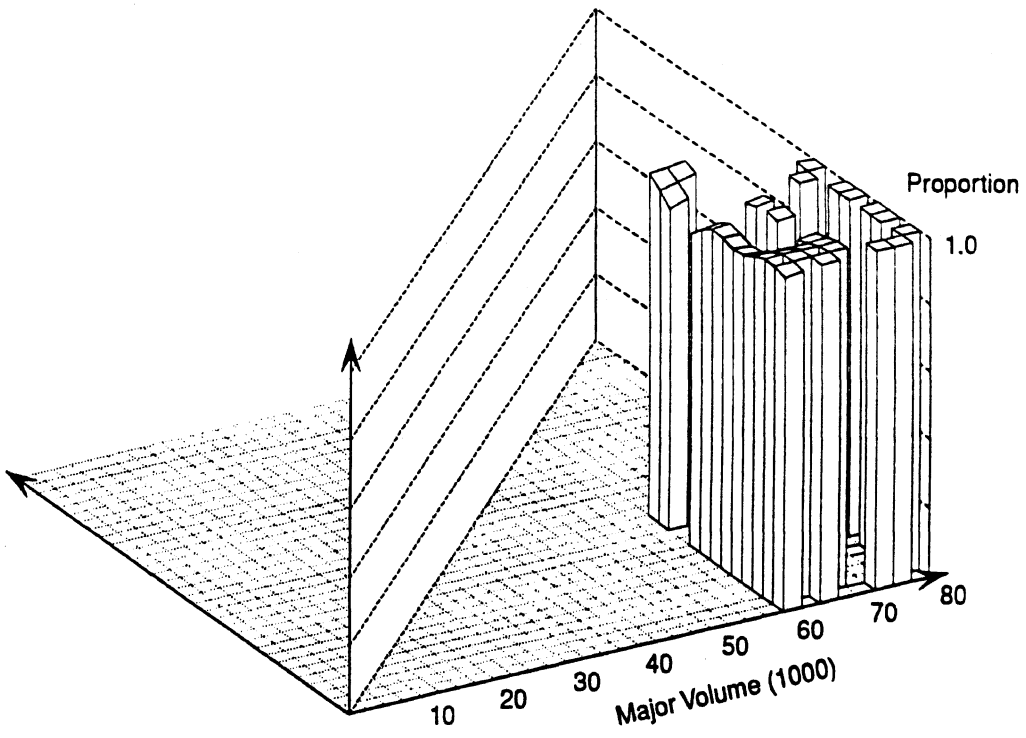


Figure 63. The same surface as Figure 60, shown only for major volume above 60,000.

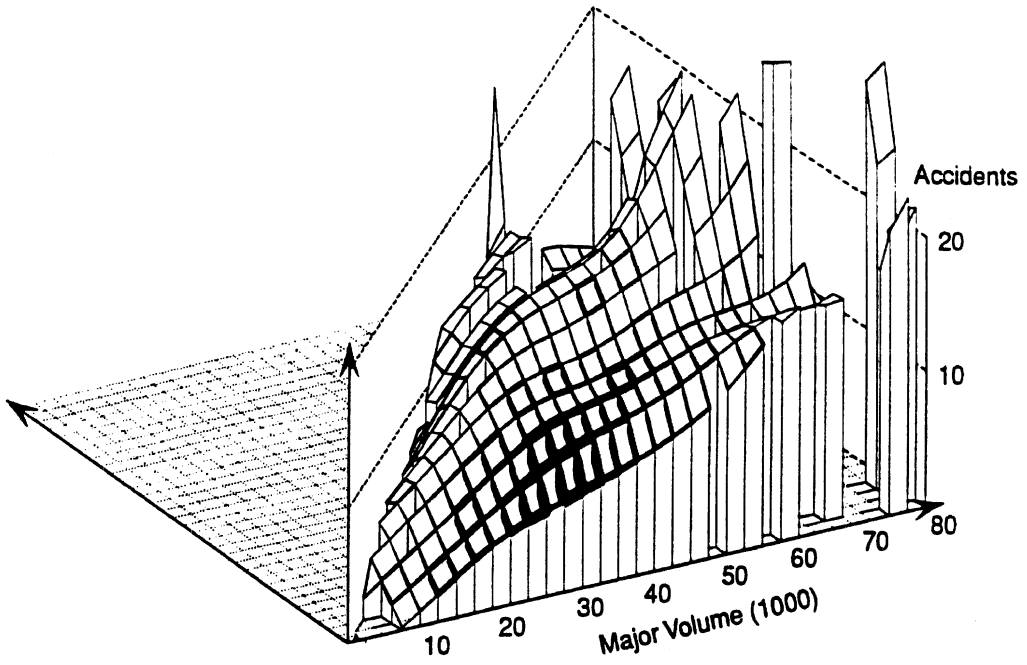


Figure 64. California four-leg signalized intersections with median on main road. Accidents on major approaches, smoothed with a 10,000 x 5,000 window.

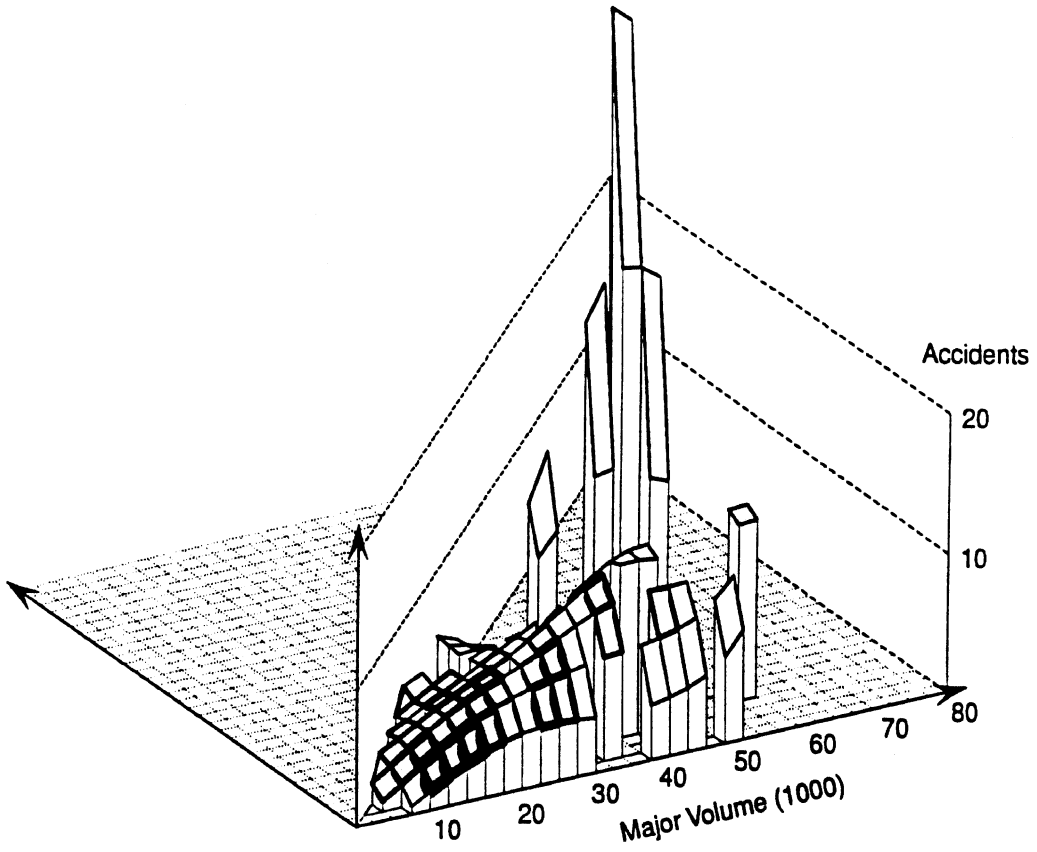


Figure 65. California four-leg signalized intersections with no median on main road. Accidents on major approaches, smoothed with a 10,000 x 5,000 window.

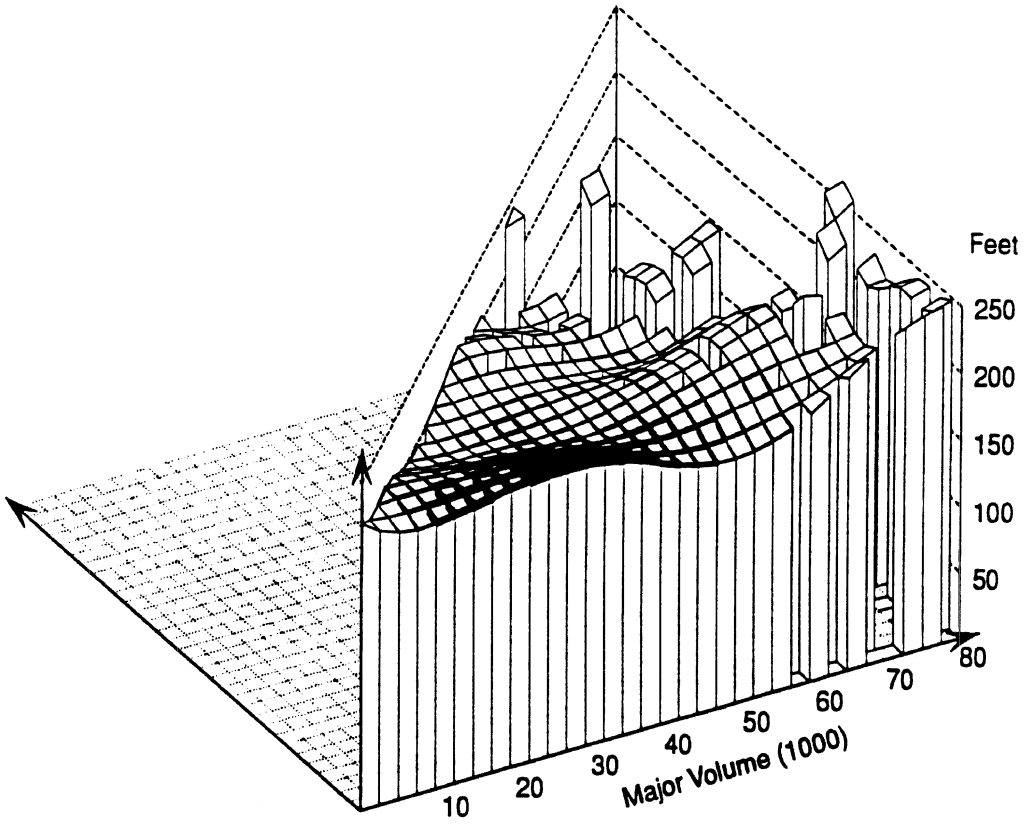


Figure 66. California four-leg intersection accidents. Length of influence zone on main (not major) road, smoothed with a 10,000 x 5,000 window.

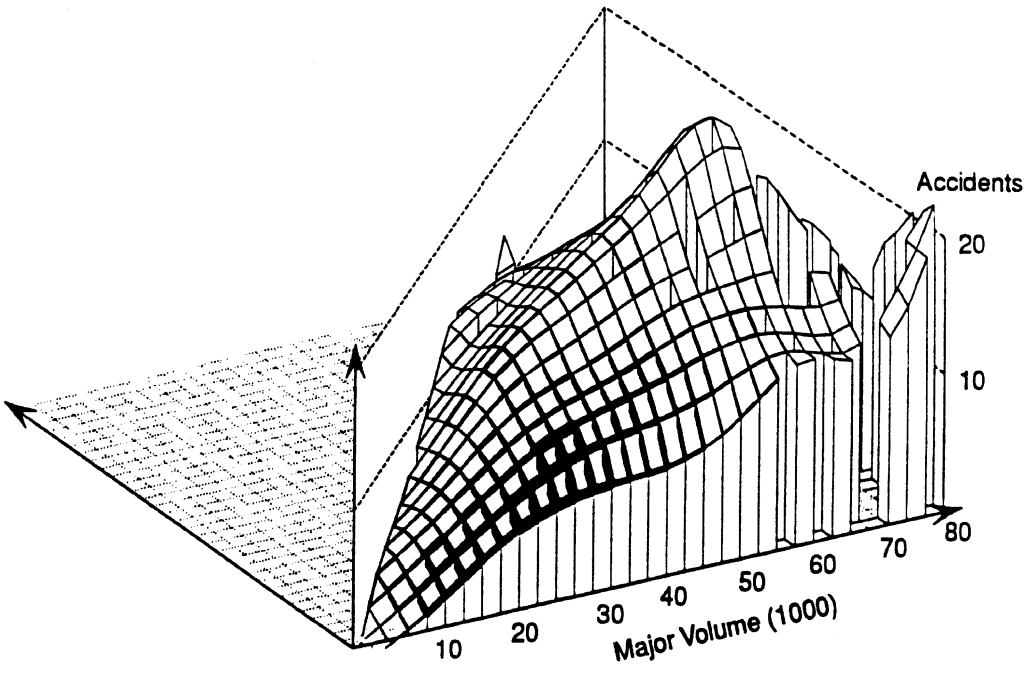


Figure 67. California four-leg signalized intersection. Number of collision accidents on main (not major) road, smoothed with a 10,000 x 5,000 window.

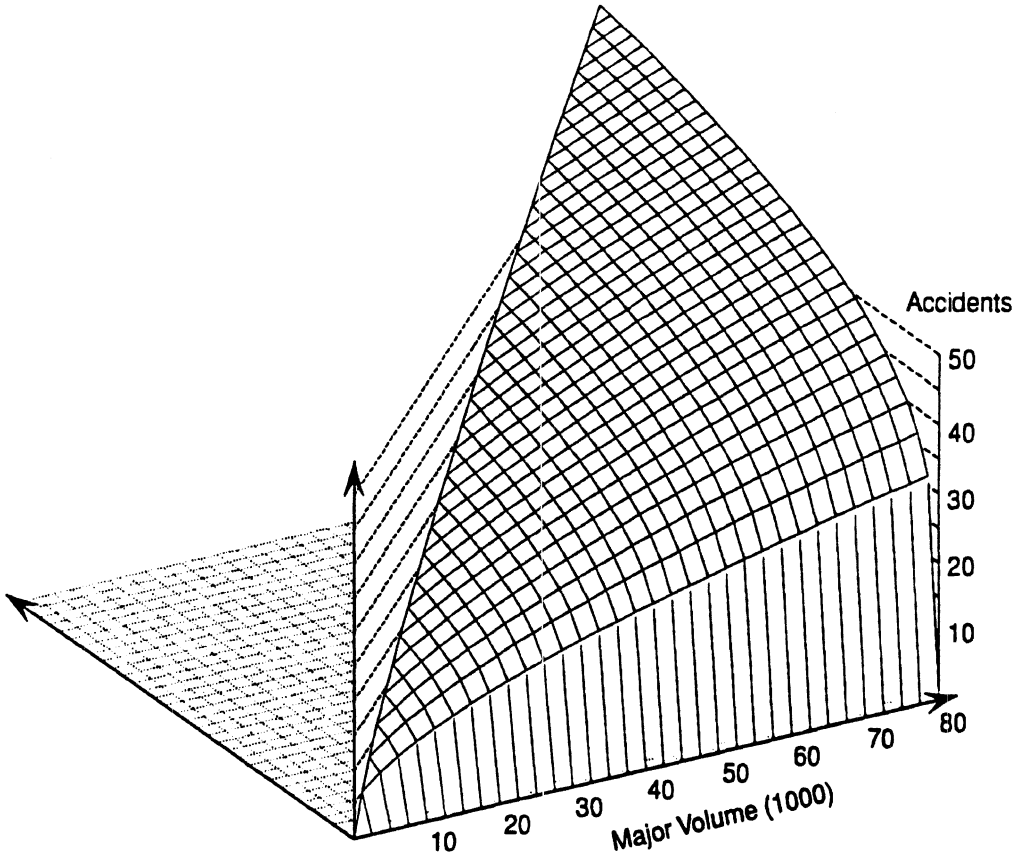


Figure 68. California four-leg signalized urban intersections. Model (5-2) fitted to total accidents.

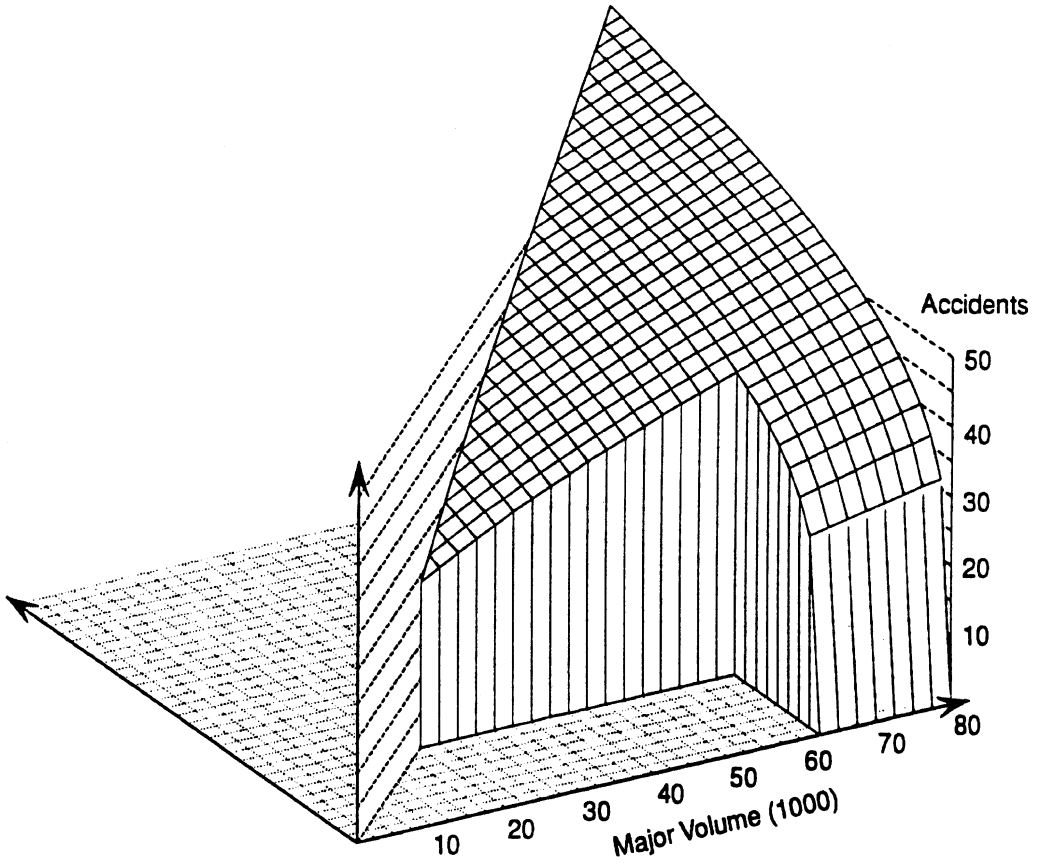
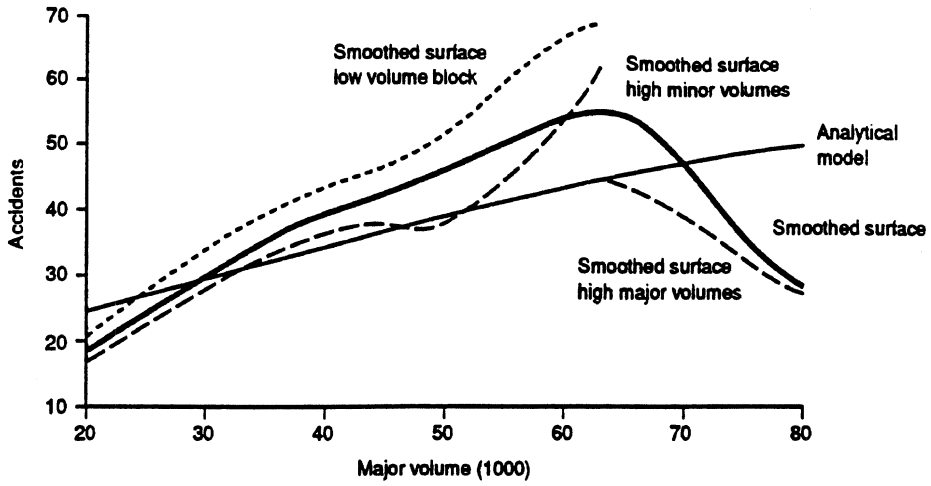


Figure 69. The same surface as in Figure 68, but cut out at major volume = 62,500, minor volume = 20,000.



(The heavy solid line refers to the surface in Figure 23, smoothed over all intersections. The dotted line, "low volume block," refers to the surface in Figure 24, smoothed over intersections with $x \leq 60{,}000$, $y \leq 20{,}000$. The broken line, "high volume cross traffic," refers to the surface shown in Figure 25, smoothed over intersections with $y > 20{,}000$ "High volume major road" refers to the surface shown in Figure 26, smoothed over intersections with $x > 60{,}000$. The light solid line refers to the analytical model shown in Figure 68.)

Figure 70. Cross-sections at $y = 20{,}000$ through the surfaces shown in Figures 23, 24, 25, 26, and 68.

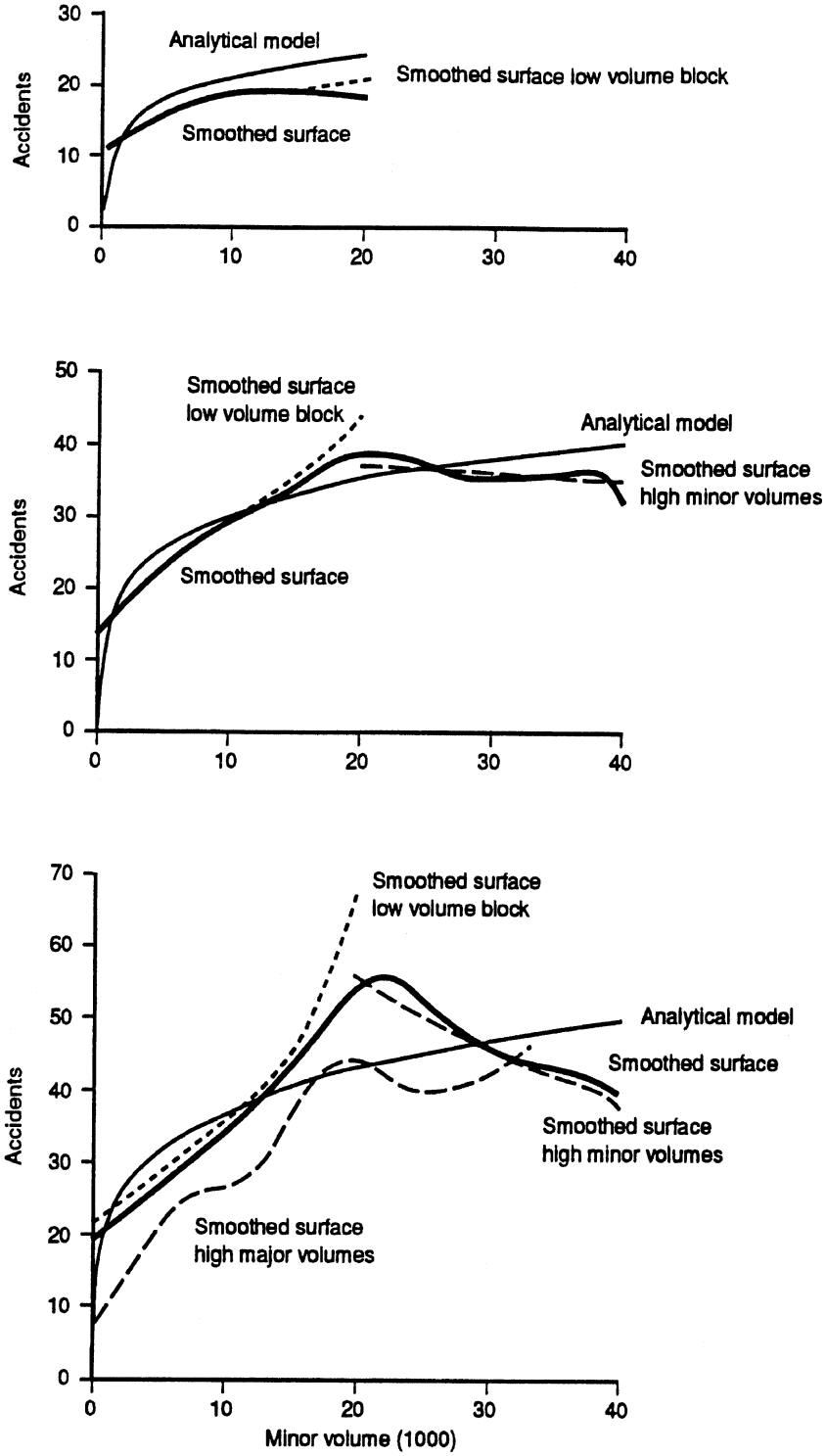


Figure 71. Cuts through the same surface as in Figure 70, but at (a) x = 20,000, (b) x = 40,000, (c) x = 60,000.

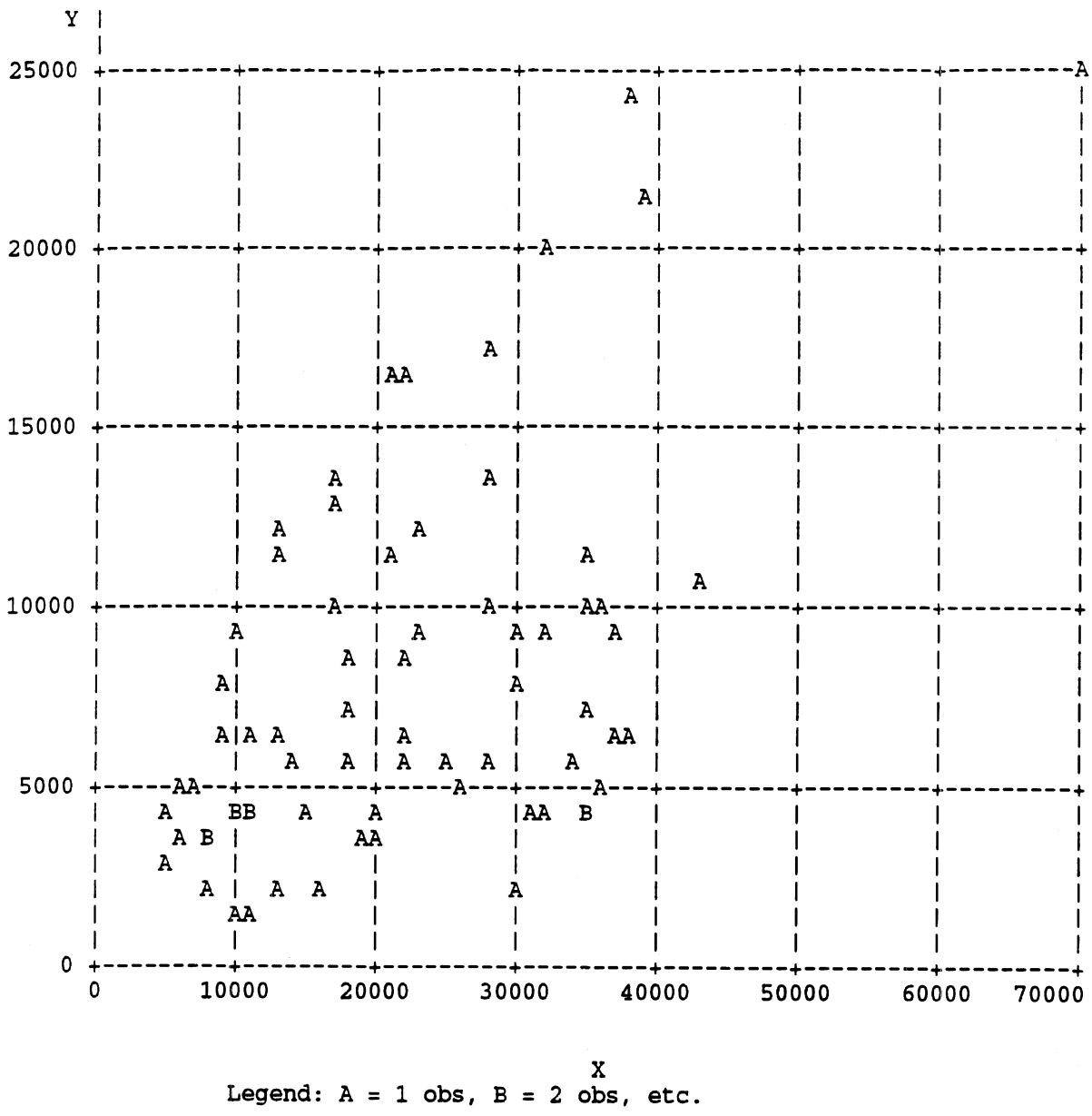


Figure 72. Distribution of traffic volumes for 71 signalized urban four-leg intersections from Minnesota data files.

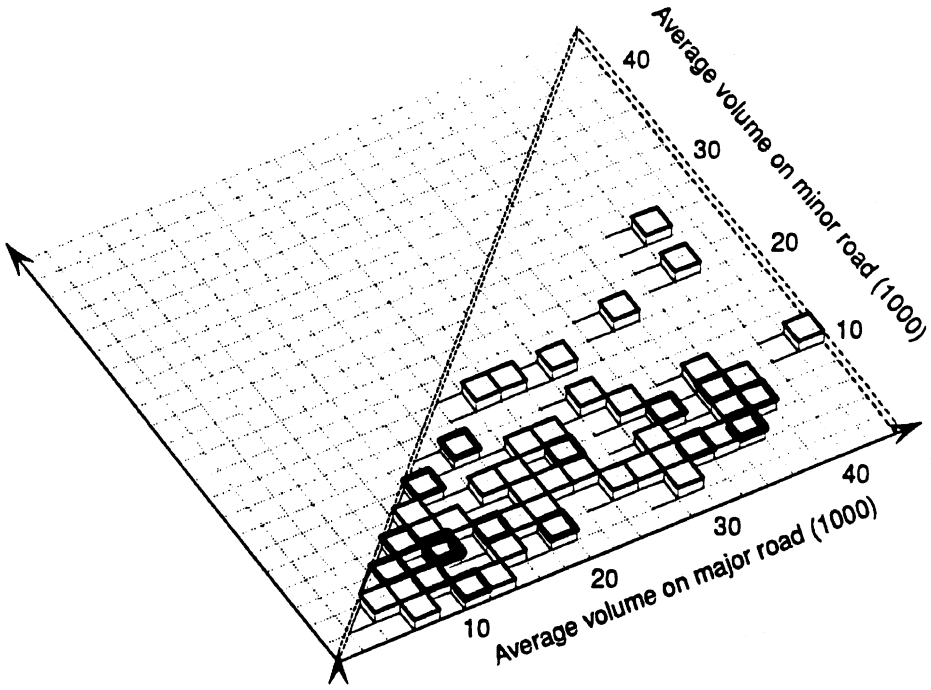


Figure 73. Distribution of approach volumes of four-leg signalized urban intersections in Minnesota. The width of the gridlines is proportional to the number of intersections in each cell.

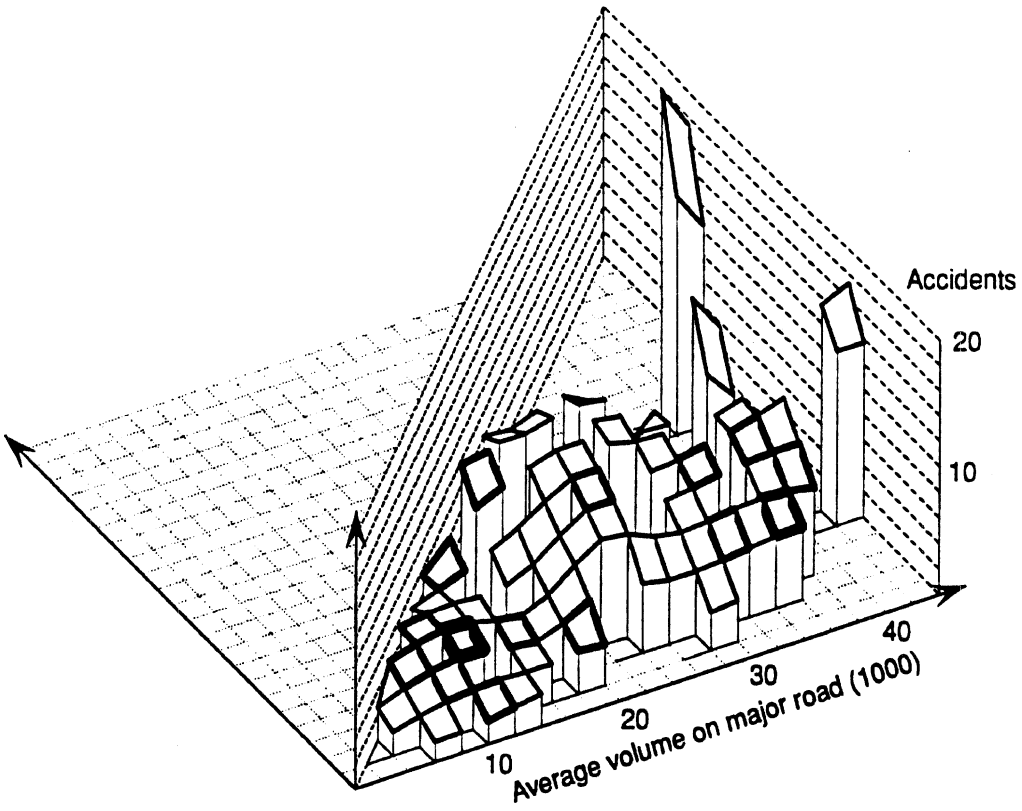


Figure 74. Signalized four-leg urban intersections in Minnesota. Counts of accidents within intersections smoothed with a 4,000 x 4,000 window.

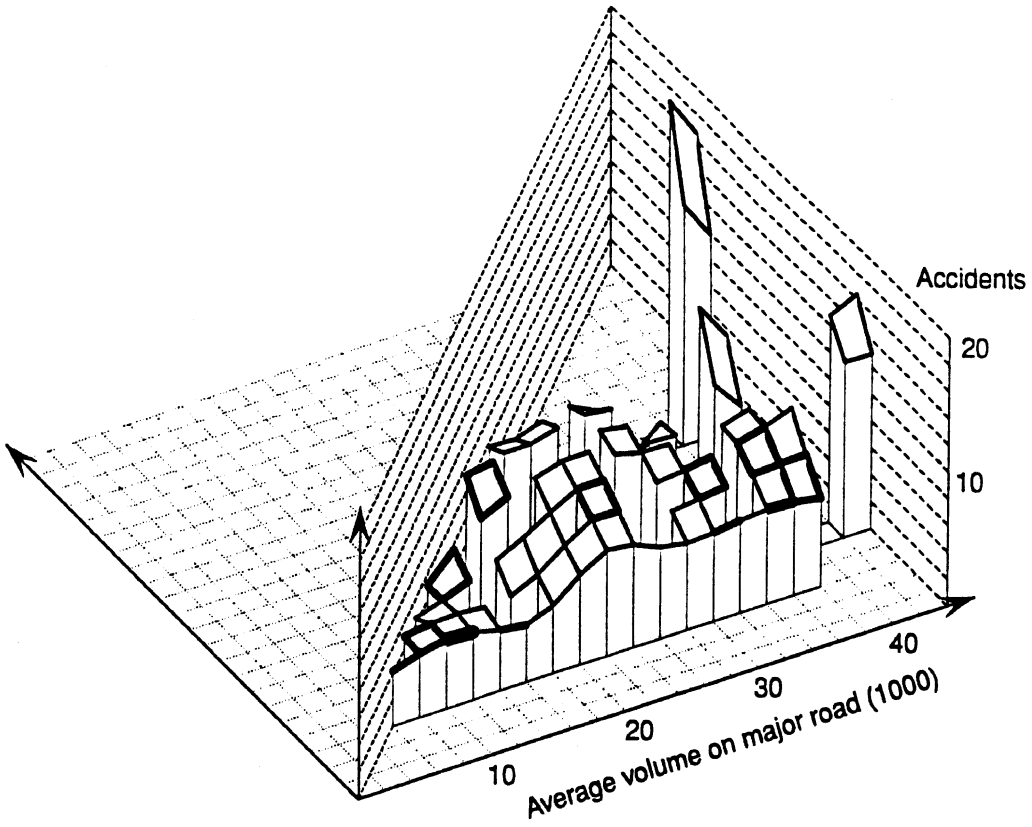


Figure 75. The same surface as in Figure 74, but cut-off at a minor volume of 6,000.

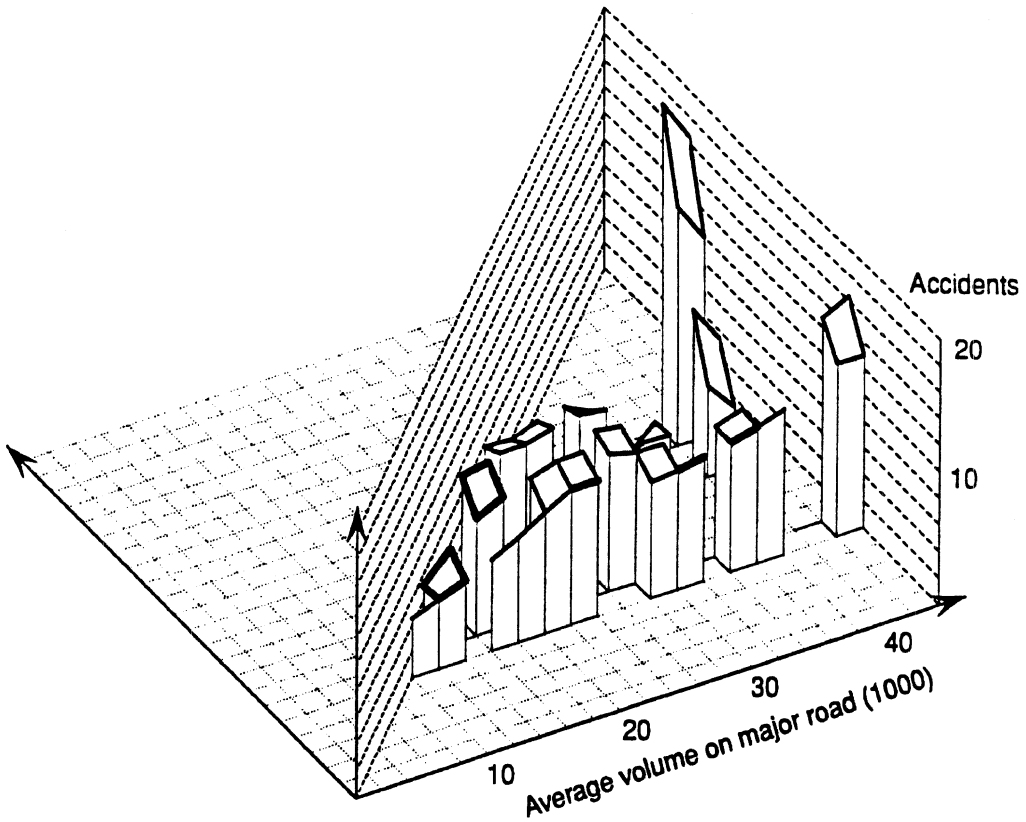


Figure 76. The same surface as in Figure 74, but cut-off at a minor volume of 10,000.

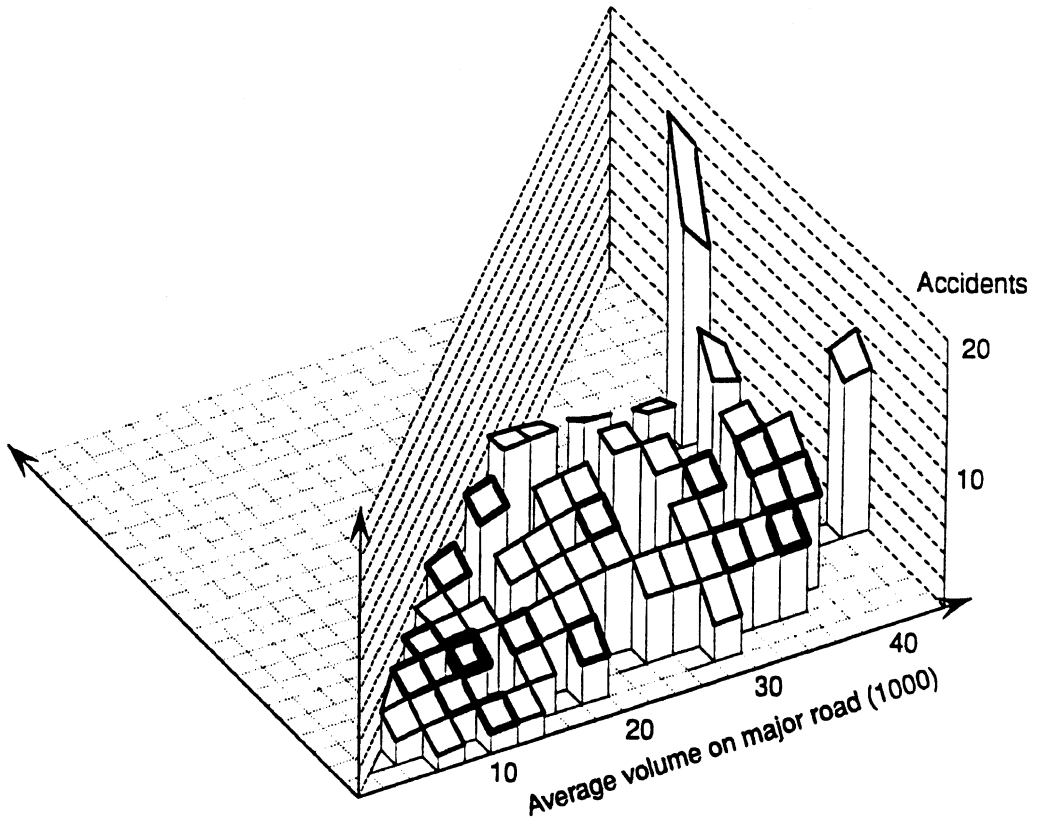


Figure 77. Signalized four-leg urban intersections in Minnesota. Accident counts within intersections smoothed with a 8,000 x 4,000 window.

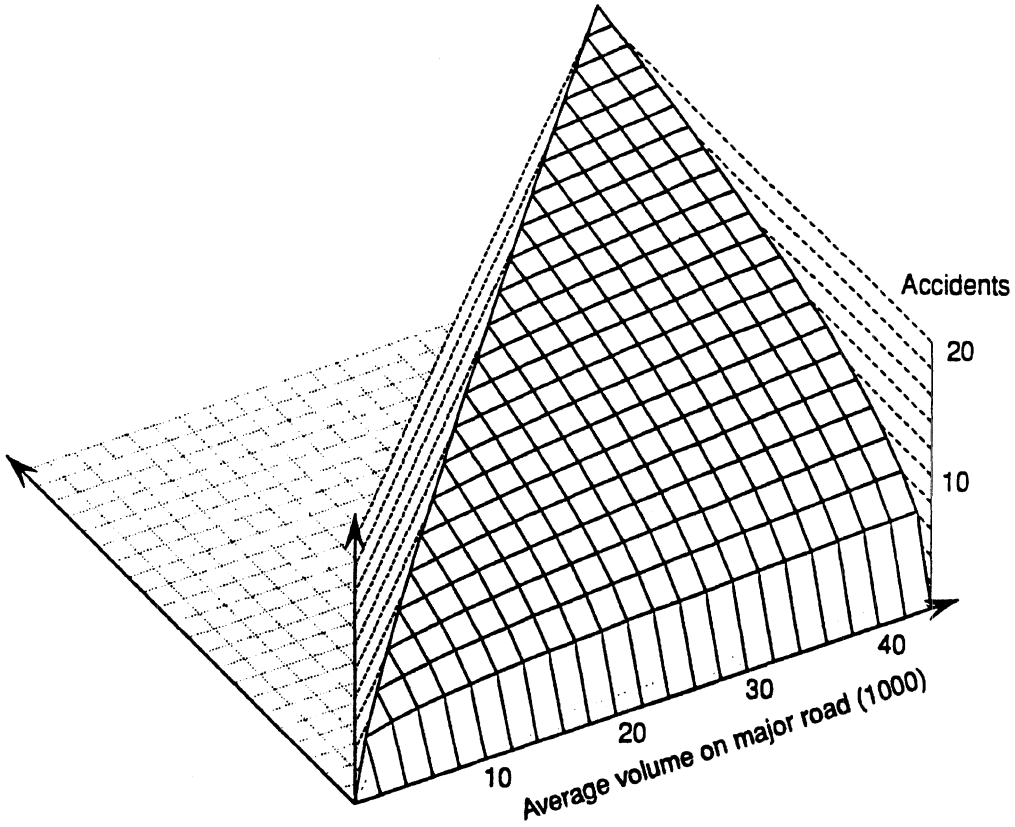


Figure 78. Signalized four-leg urban intersections in Minnesota. Surface represents model (6-1) for within intersection accident counts.

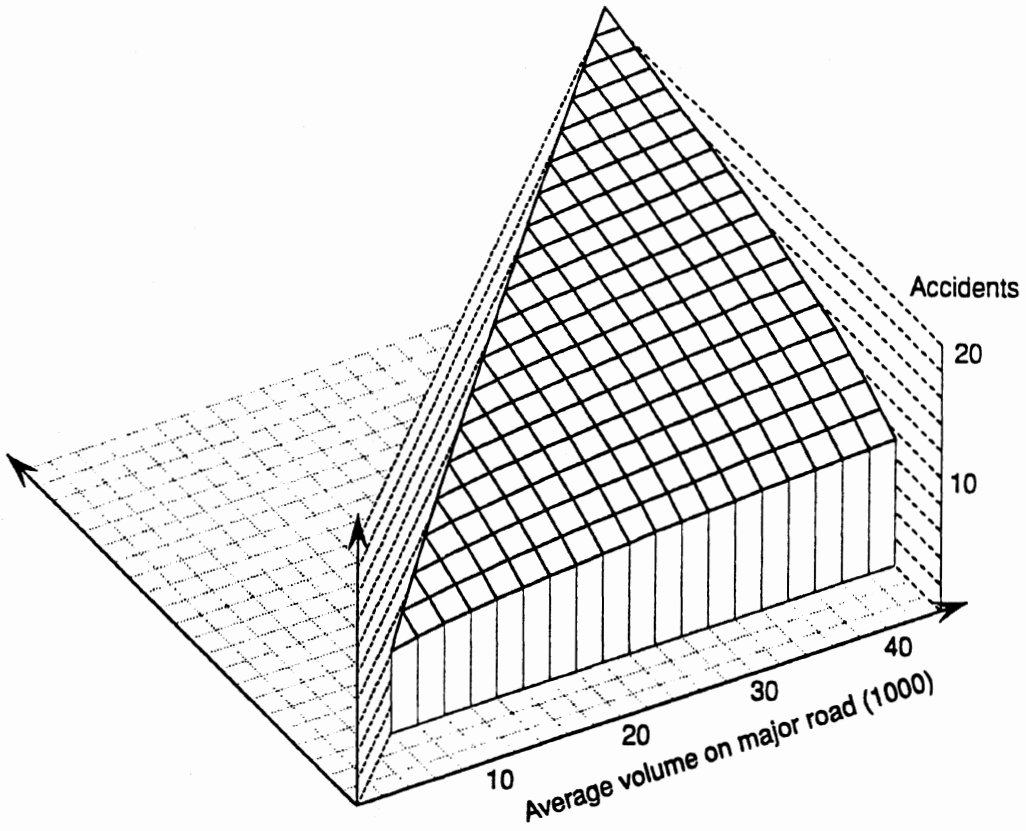


Figure 79. The same surface as in Figure 78, but cut-off at y = 6,000.

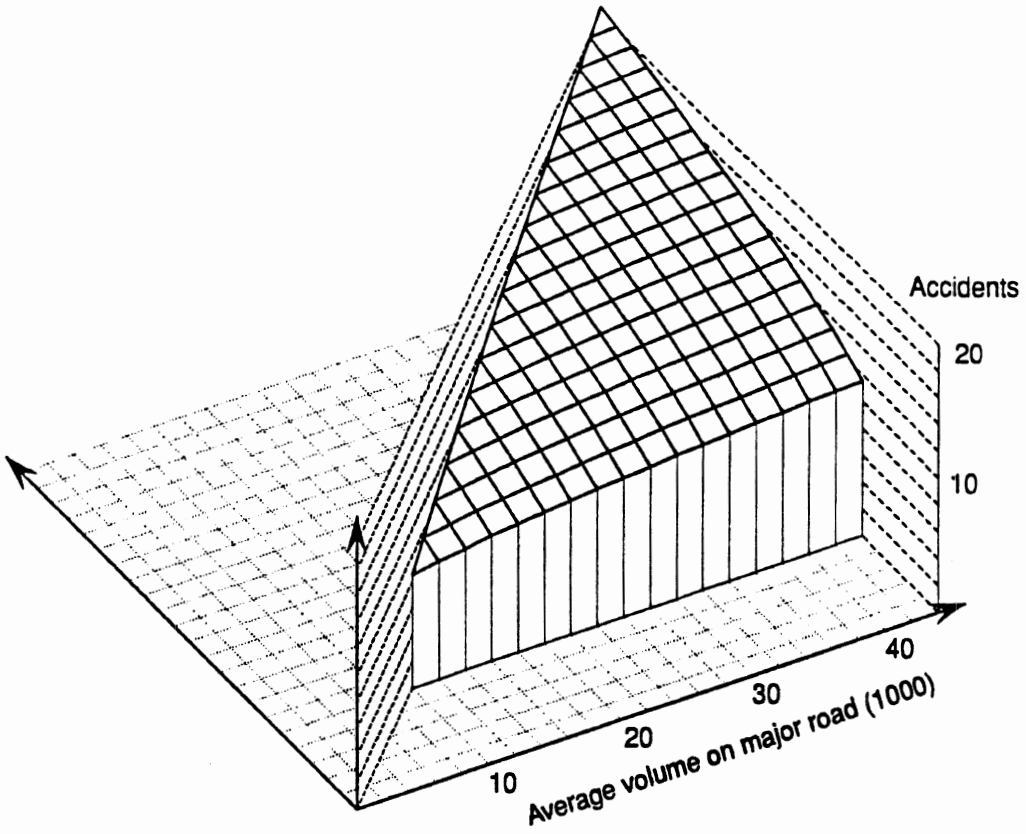


Figure 80. The same surface as in Figure 78, but cut-off at y = 10,000.

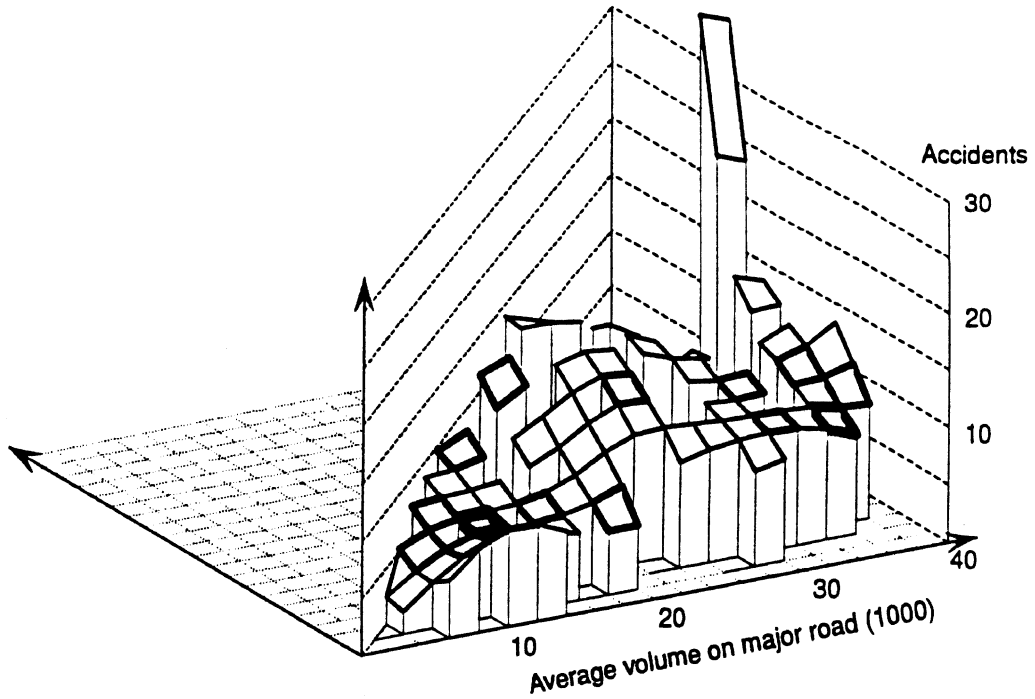


Figure 81. Signalized four-leg urban intersections in Minnesota. All accidents in the intersection and on the approaches within 60 meters, smoothed with a 5,000 x 5,000 window.

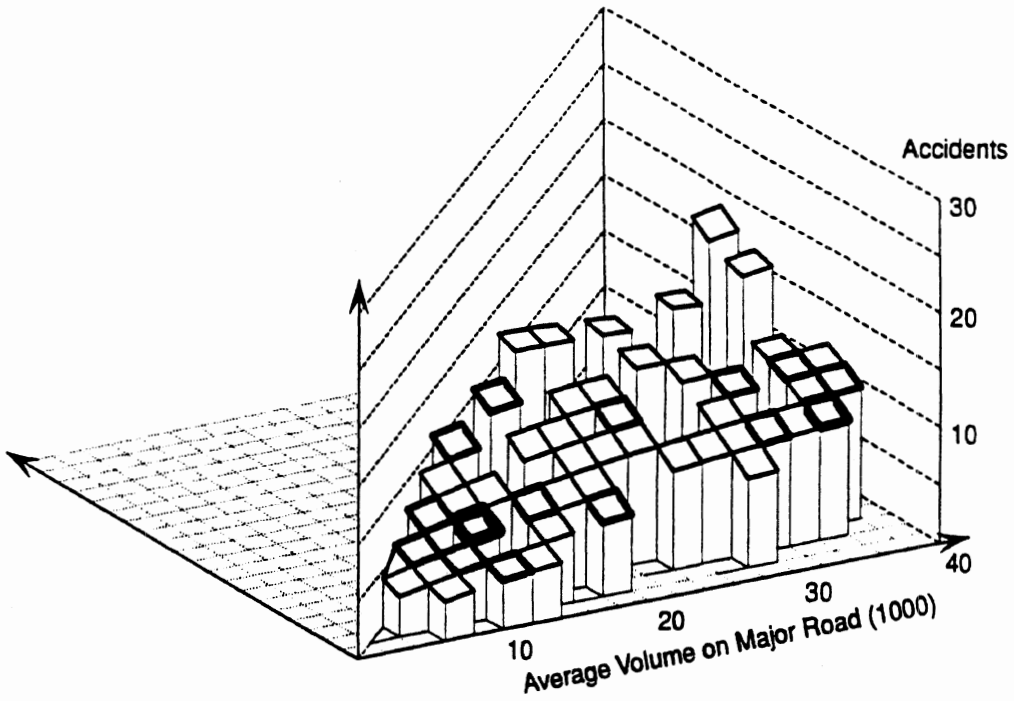


Figure 82. Signalized four-leg urban intersections in Minnesota. All accidents in the intersection and on the approaches within 60 meters, smoothed with a 15,000 x 10,000 window.

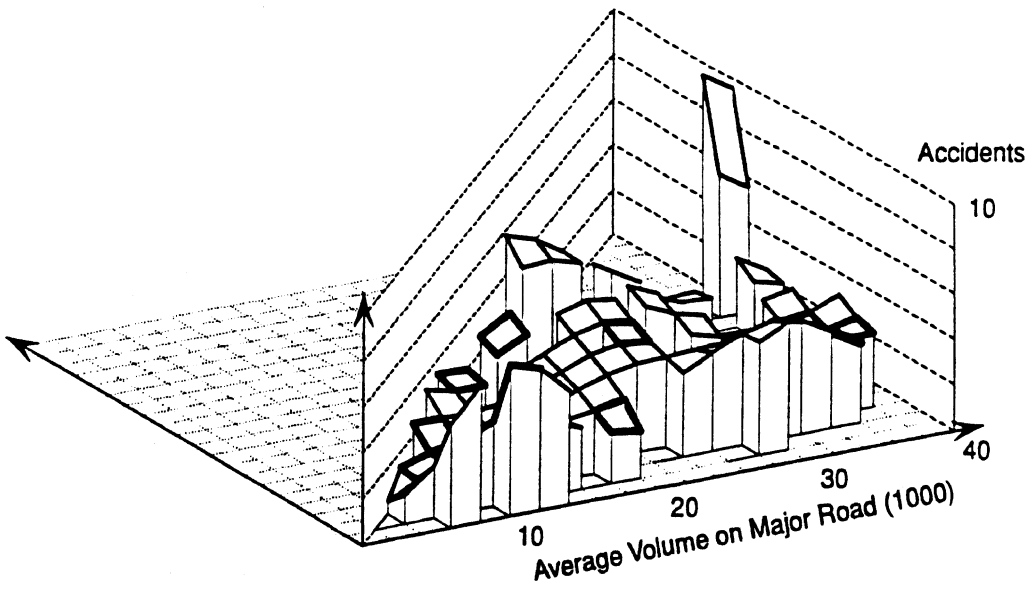


Figure 83. Signalized four-leg urban intersections in Minnesota. All accidents on the approaches outside the intersection within 60 meters, smoothed with a 5,000 x 5,000 window.

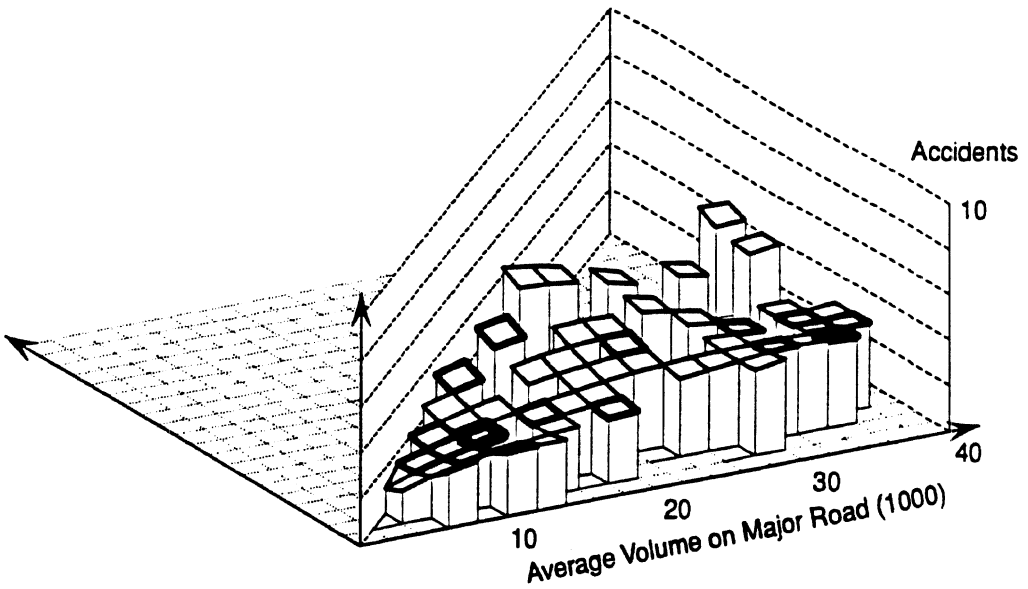


Figure 84. Signalized four-leg urban intersections in Minnesota. All accidents on the approaches outside the intersection within 60 meters, smoothed with a 10,000 x 10,000 window.

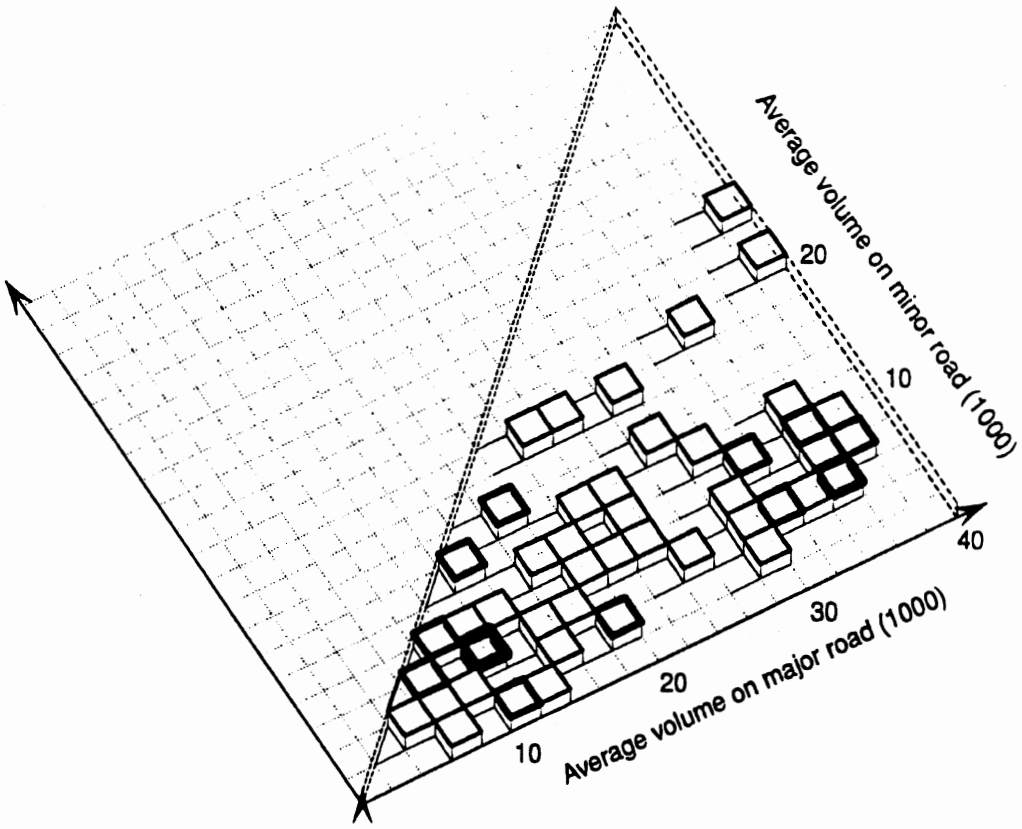


Figure 85. Signalized four-leg urban intersections in Minnesota. Distribution of intersections with typical intersection accidents by volumes of the two roads.

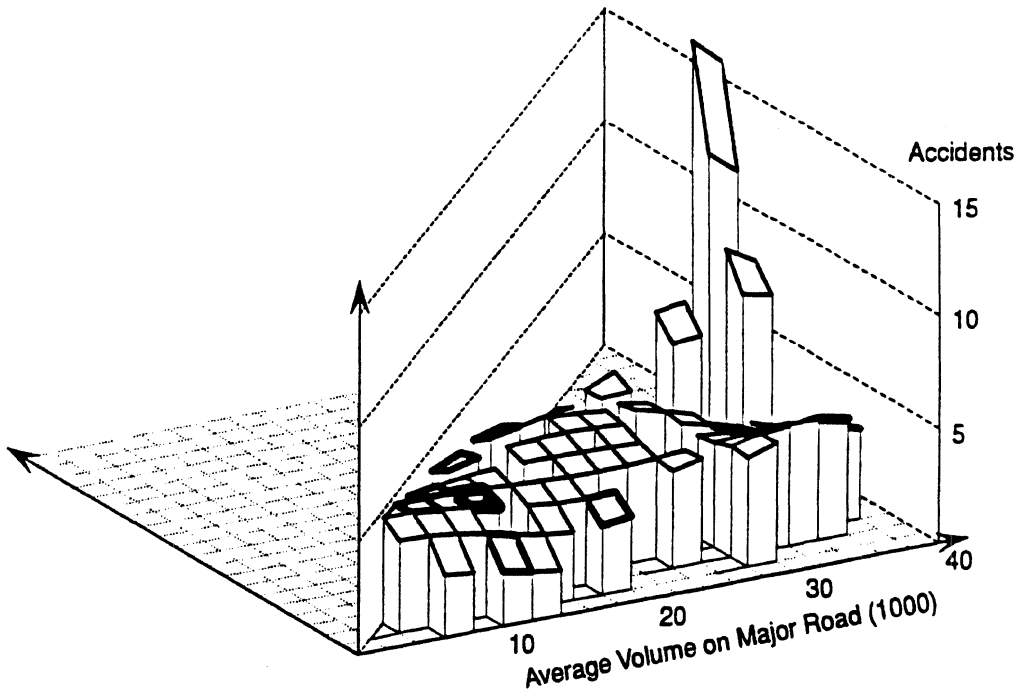


Figure 86. Signalized four-leg urban intersections in Minnesota. Typical intersection accident, smoothed with a 10,000 x 5,000 window.

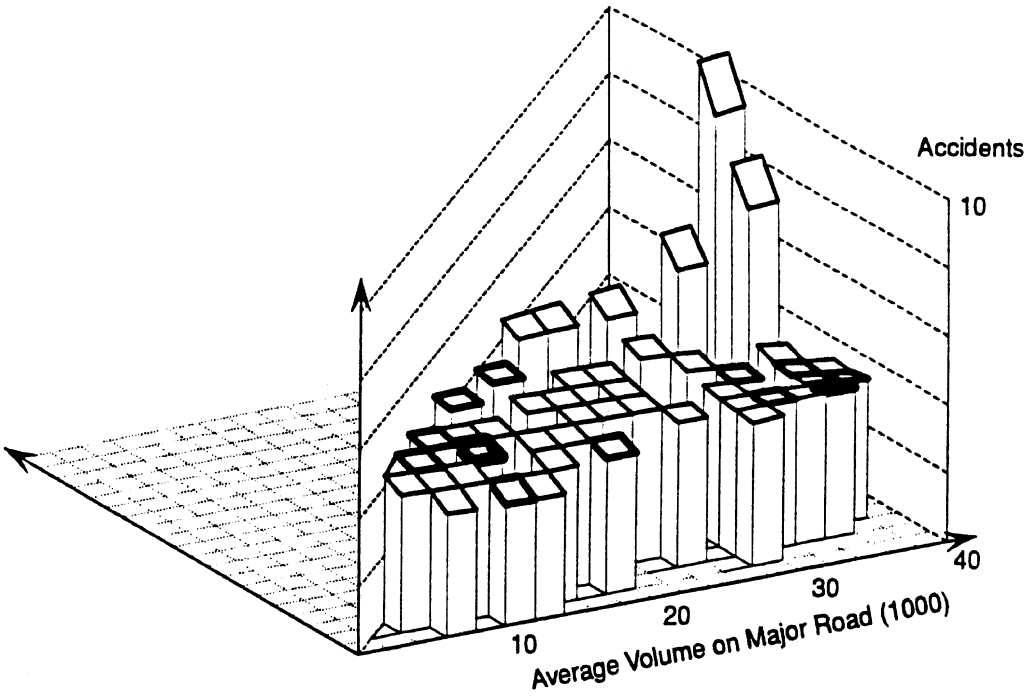


Figure 87. Signalized four-leg urban intersections in Minnesota. Typical intersection accident within the intersection, smoothed with a 20,000 x 10,000 window.

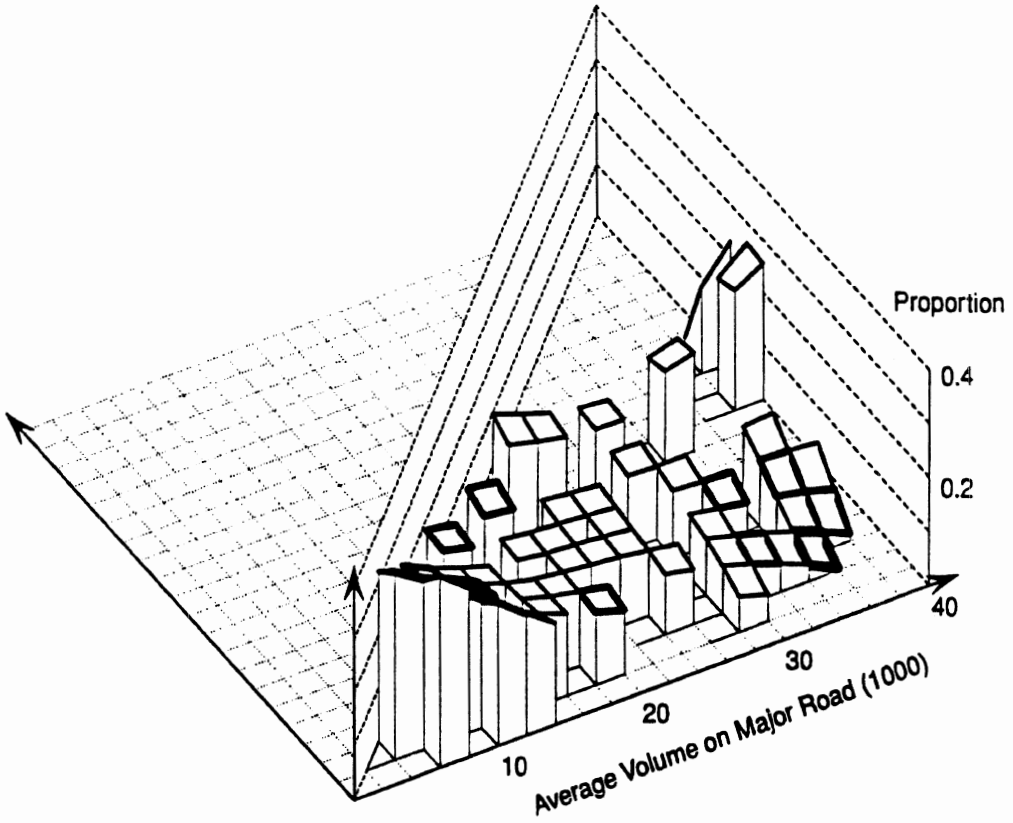


Figure 88. Signalized four-leg urban intersection in Minnesota. Left-turn accidents within the intersection as proportion of typical intersection accidents, smoothed with a 10,000 x 15,000 window.

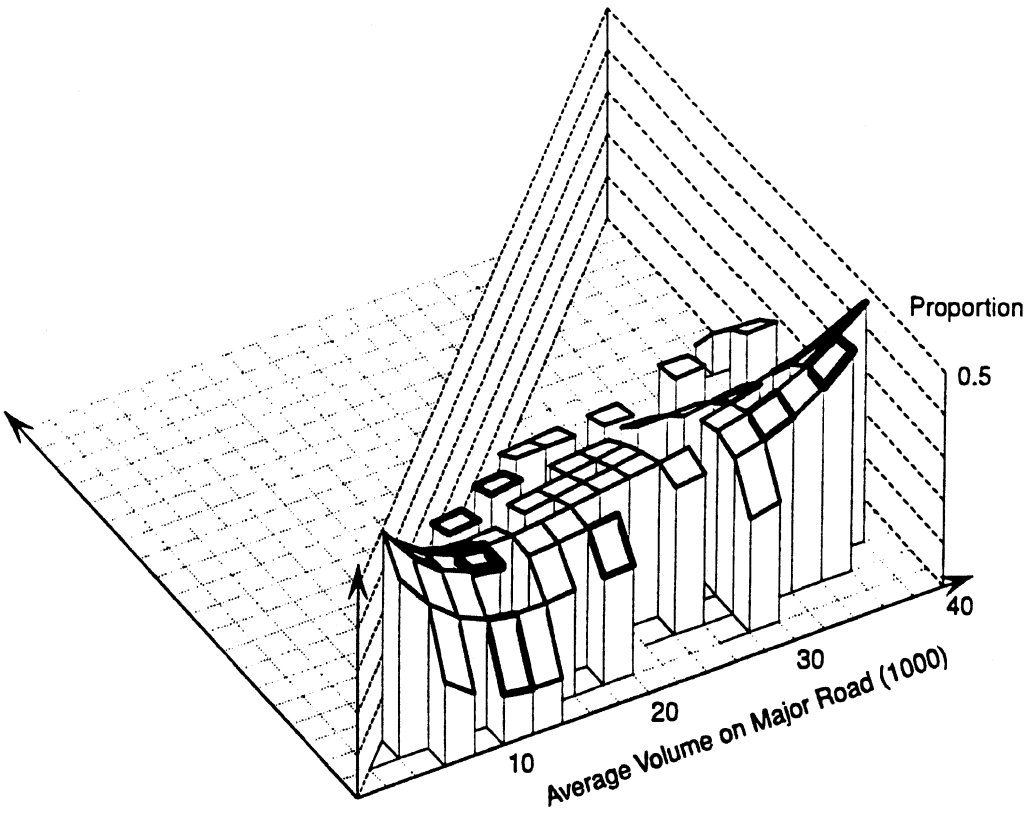


Figure 89. Signalized four-leg urban intersections in Minnesota. Angle collisions within the intersection as proportion of typical intersection accidents, smoothed with a 10,000 x 5,000 window.

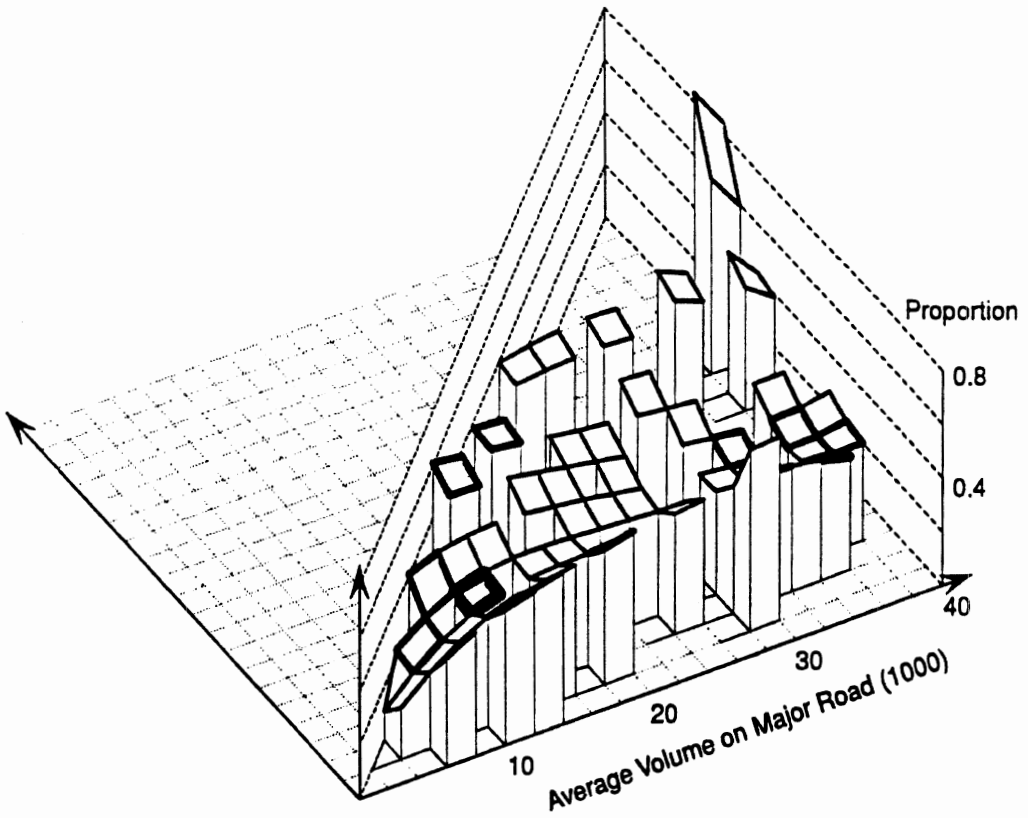


Figure 90. Signalized four-leg urban intersections in Minnesota. Rear-end collisions within the intersection as proportion of typical intersection accidents, smoothed with a 10,000 x 5,000 window.

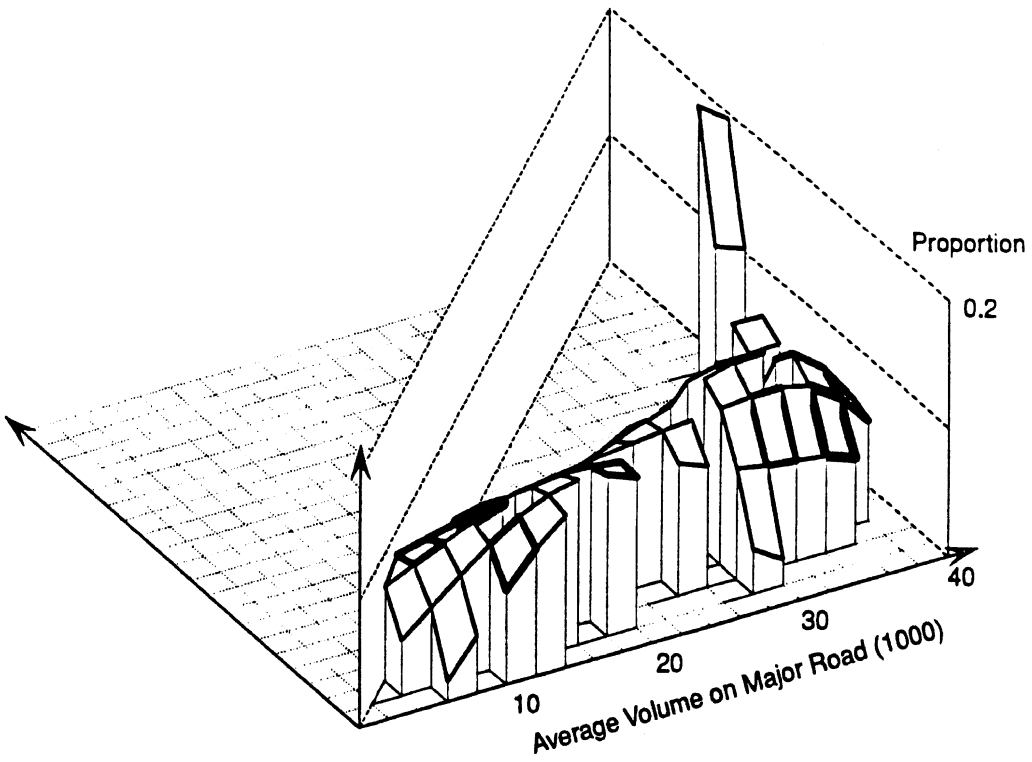
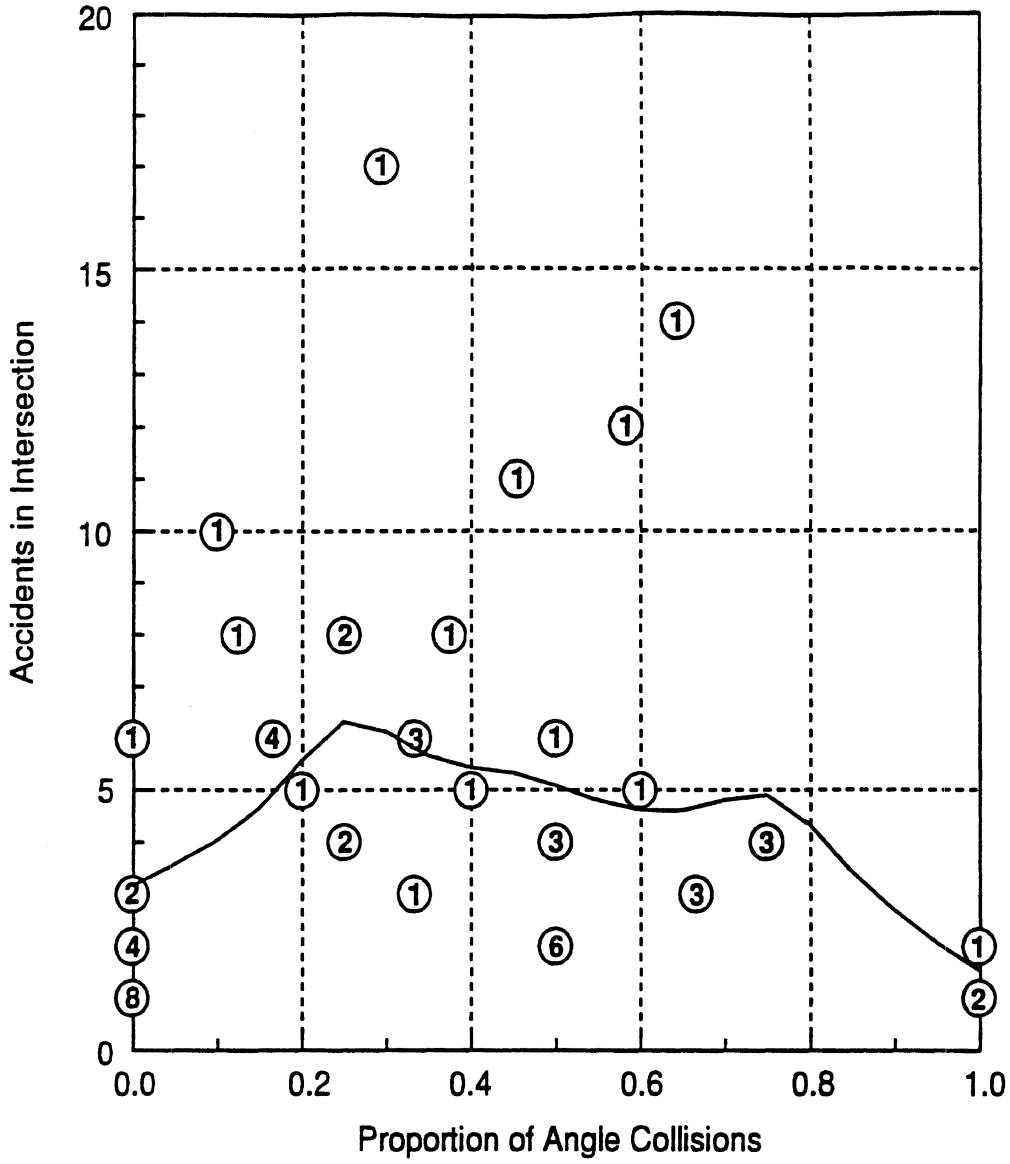
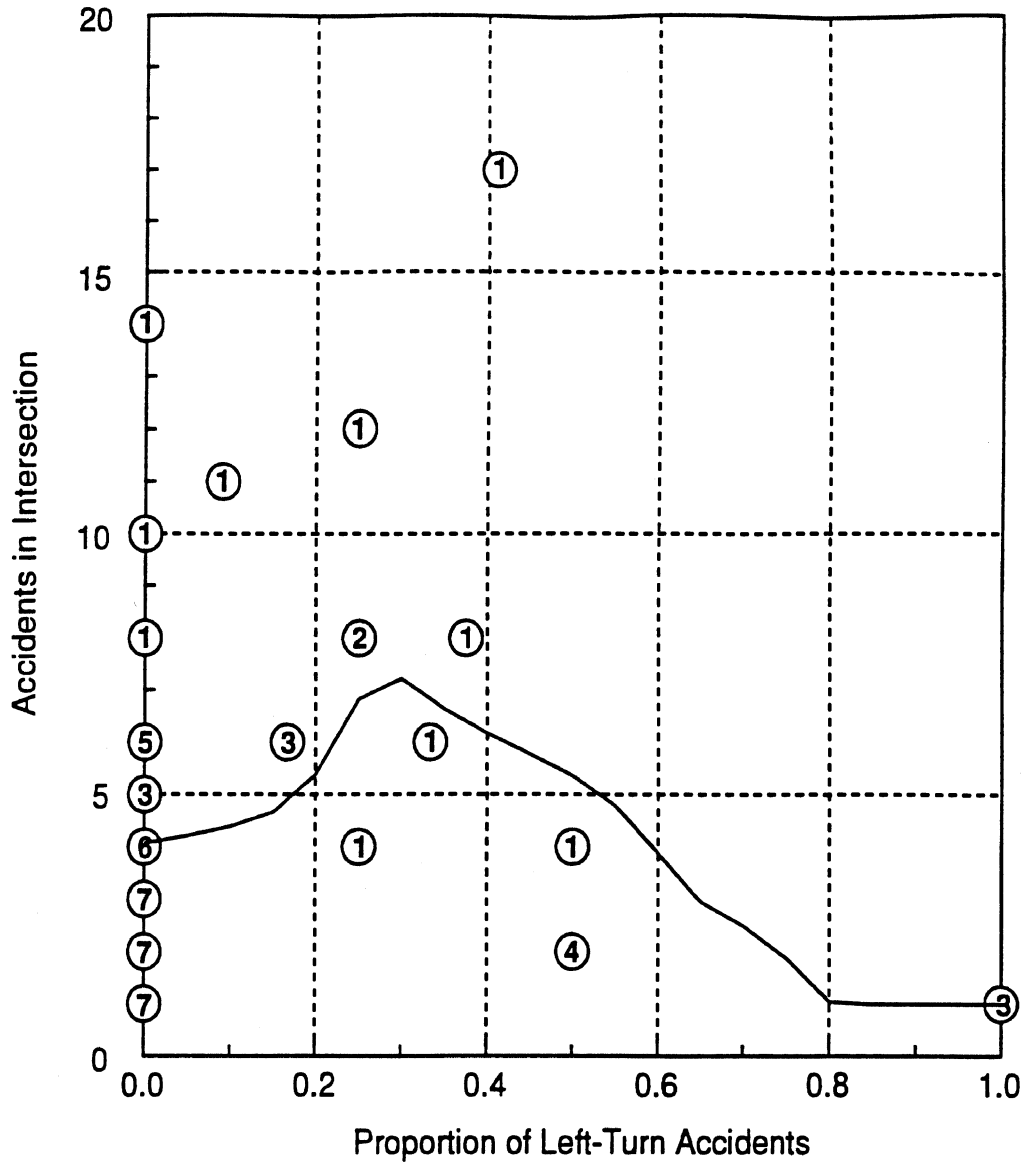


Figure 91. Signalized four-leg urban intersections in Minnesota. Other collisions within the intersection as proportion of typical intersection accidents, smoothed with a 10,000 x 6,000 window.



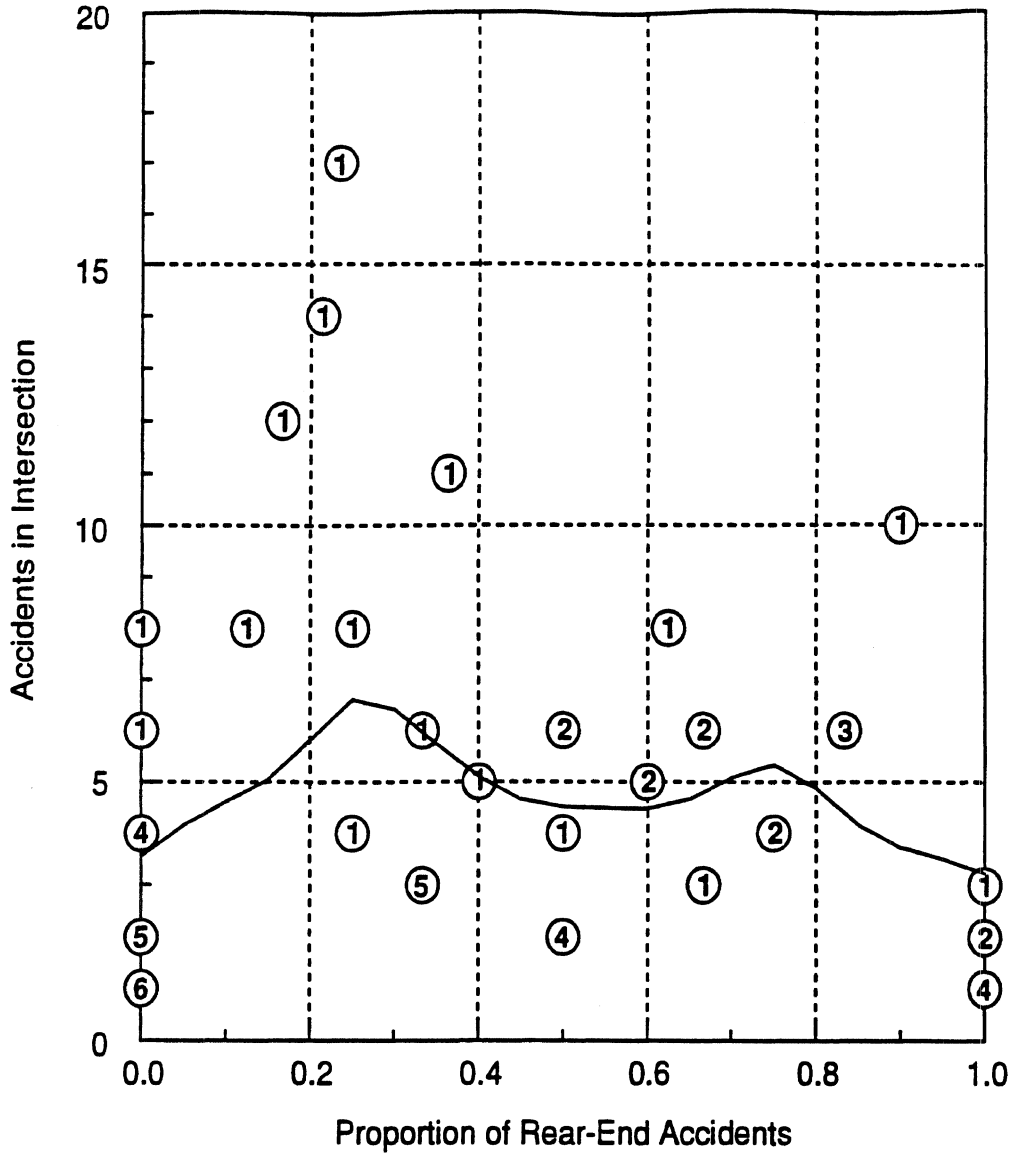
(The number of intersections are in the circles. The line shows smoothed values.)

Figure 92. Signalized four-leg intersections in Minnesota. Accidents in intersections versus proportion of angle collisions.



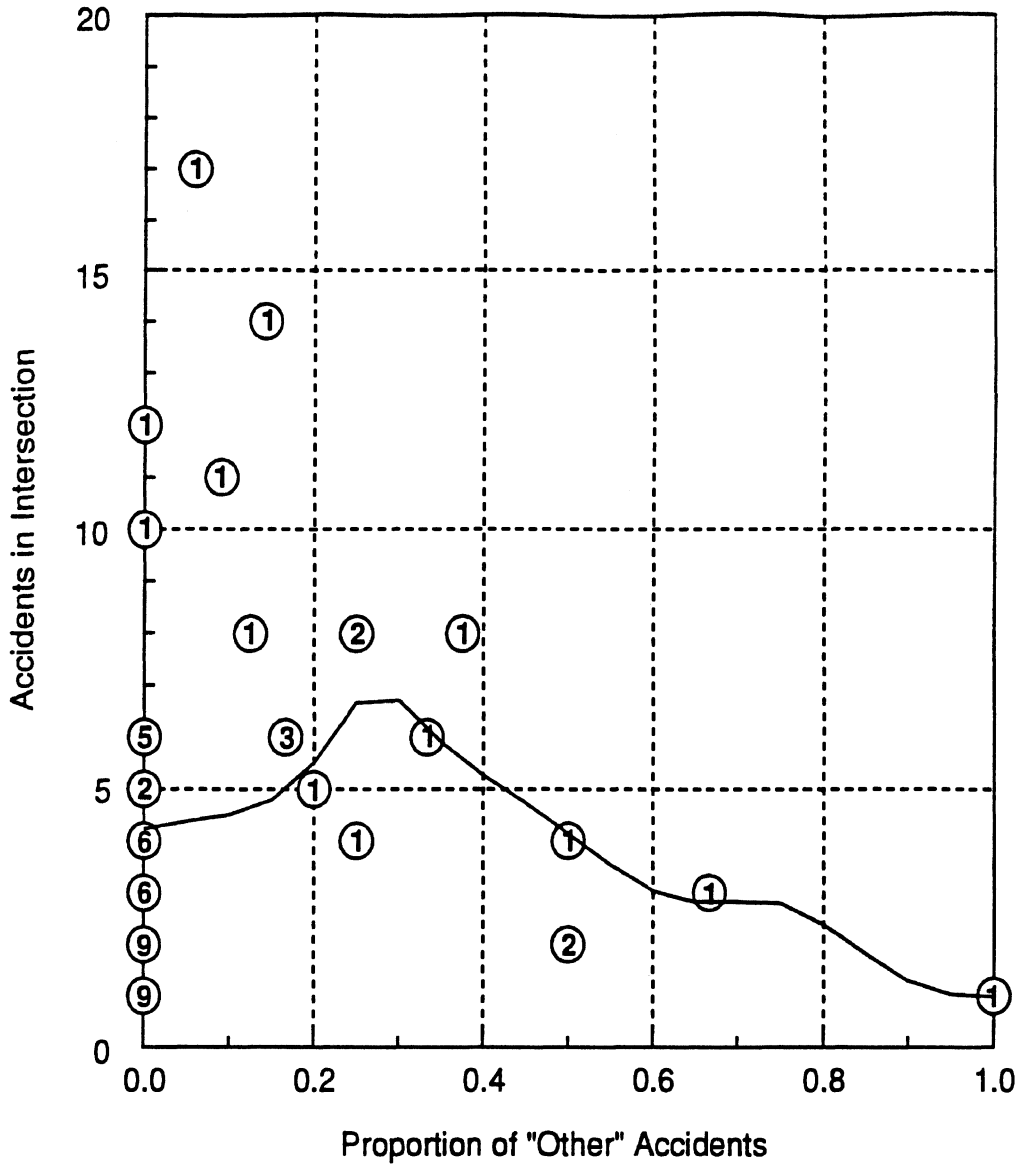
(The number of intersections are in the circles. The line shows smoothed values.)

Figure 93. Signalized four-leg intersections in Minnesota. Accidents in intersections versus proportion of left-turn accidents.



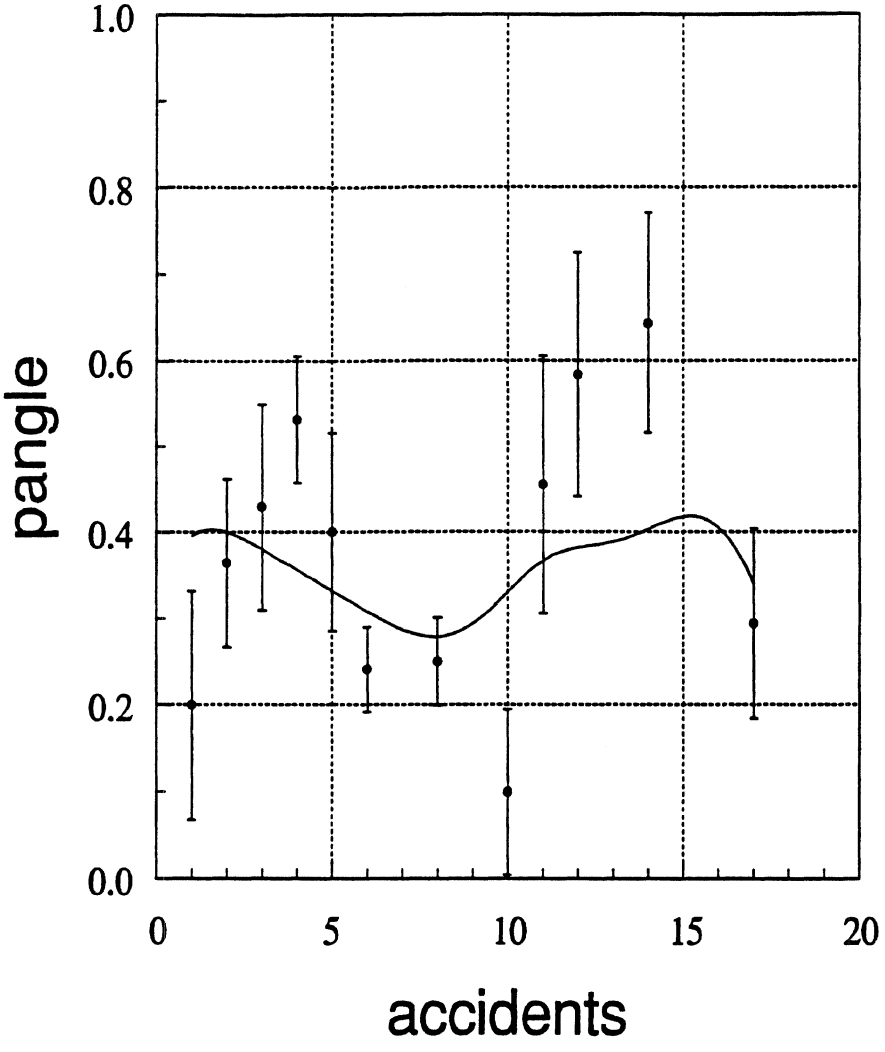
(The number of intersections are in the circles. The line shows smoothed values.)

Figure 94. Signalized four-leg intersections in Minnesota. Accidents in intersections versus proportion of rear-end accidents.



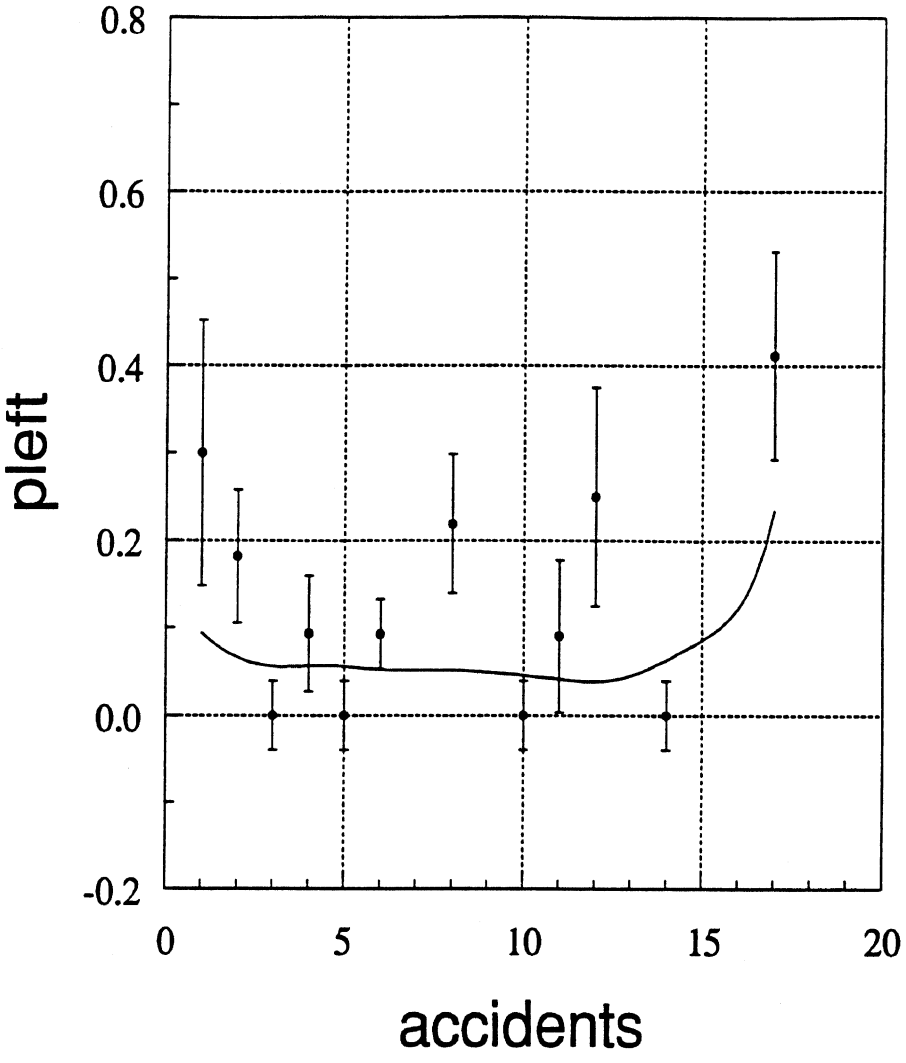
(The number of intersections are in the circles. The line shows smoothed values.)

Figure 95. Signalized four-leg intersections in Minnesota. Accidents in intersections versus proportion of "other" accidents.



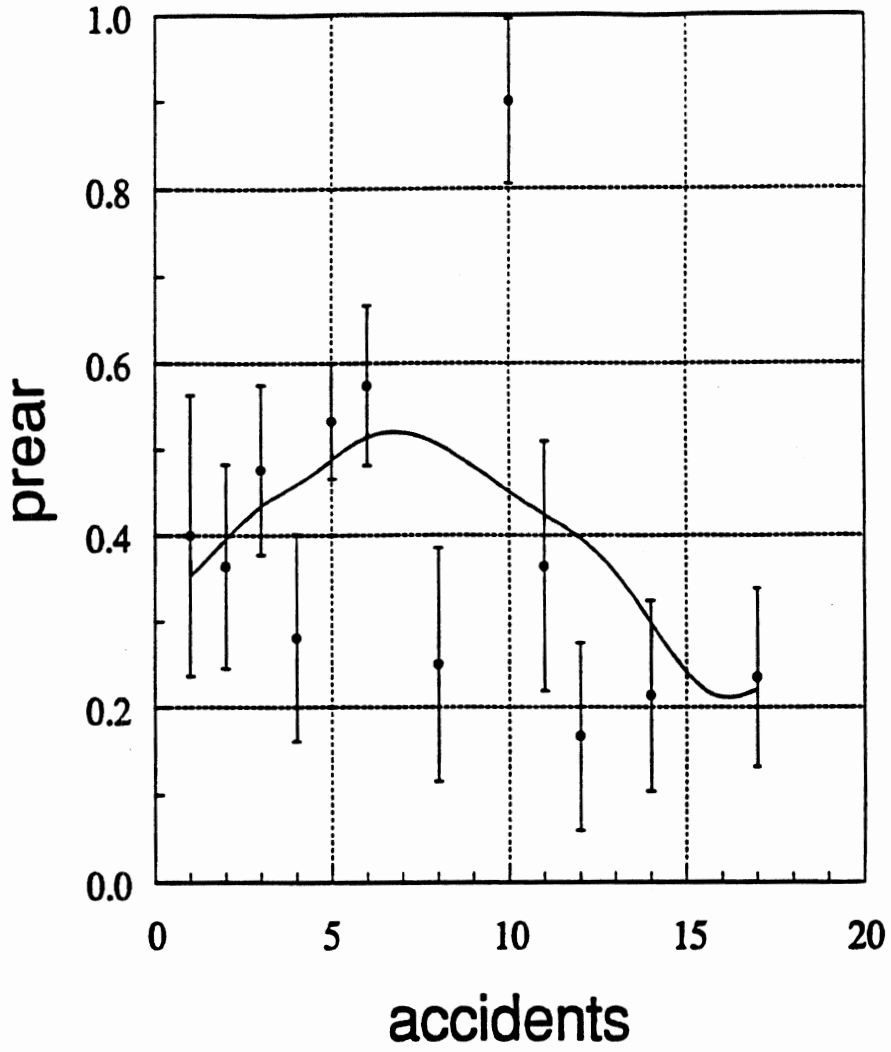
(The bars show +/- one standard deviation, the curve smoothed values.)

Figure 96. Signalized four-leg intersections in Minnesota. Proportion of angle collisions versus number of accidents in intersections.



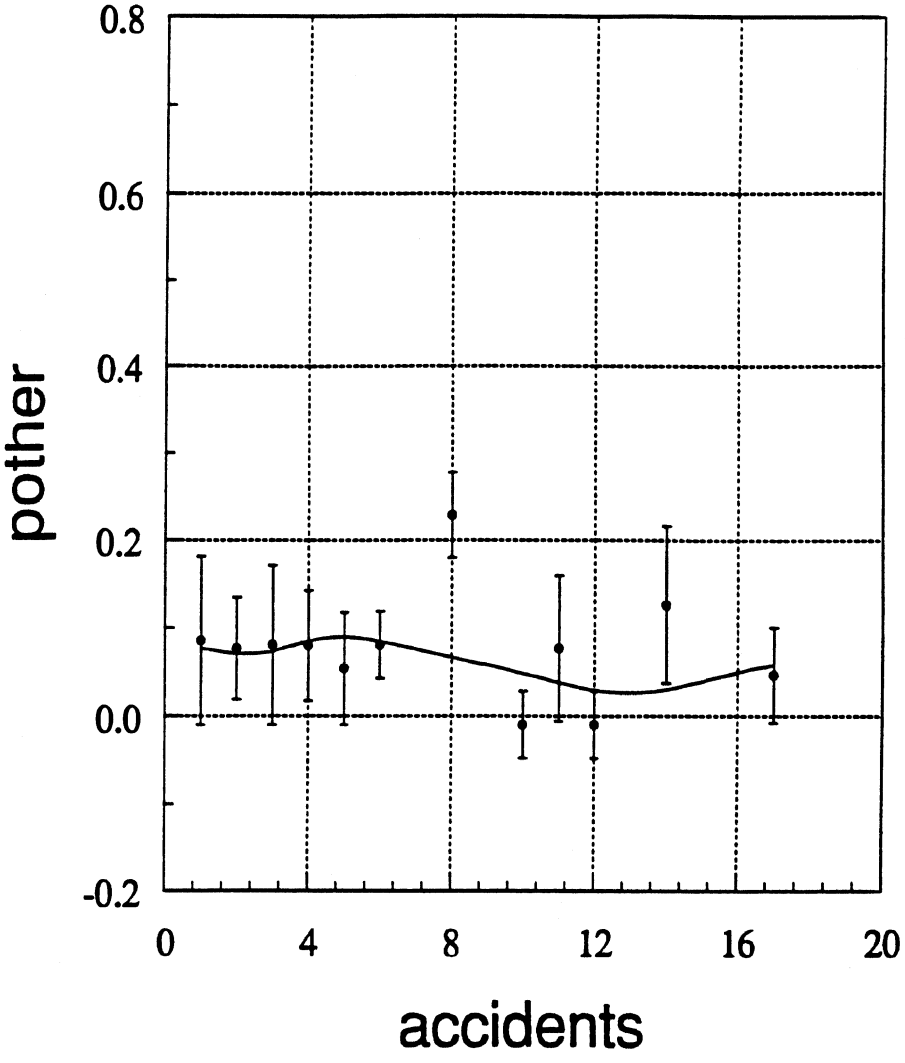
(The bars show +/- one standard deviation, the curve smoothed values.)

Figure 97. Signalized four-leg intersections in Minnesota. Proportion of left-turn collisions versus number of accidents in intersections.



(The bars show +/- one standard deviation, the curve smoothed values.)

Figure 98. Signalized four-leg intersections in Minnesota. Proportion of rear-end collisions versus number of accidents in intersections.



(The bars show +/- one standard deviation, the curve smoothed values.)

Figure 99. Signalized four-leg intersections in Minnesota. Proportion of "other" collisions versus number of accidents in intersections.