This is to certify that the

dissertation entitled

Analysis of Crash Rates on Horizontal Curves

presented by

Cyrus Safdari

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Civil Engineering

William C. Taylor
Major professor

Date 2 February, 1999

ANALYSIS OF CRASH RATES ON HORIZONTAL CURVES

By

Cyrus Safdari

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil and Environmental Engineering

1999

ABSTRACT

ANALYSIS OF CRASH RATES ON HORIZONTAL CURVES

By

Cyrus Safdari

In Michigan over 25% of fatal traffic crashes take place on non-freeway trunkline highways. Research has consistently demonstrated that crash rates on horizontal curves are significantly higher than that of the tangent sections on the same road, and most studies have found the degree of curvature to be the most significant single factor related to curve crashes. However, other roadway features, such as superelevation and skid resistance of the pavement surface, traffic control elements, driving environments and human factors, individually or in combination are major contributors as well.

The purpose of this study was to analyze horizontal curve crashes on two-lane trunkline roads in the State of Michigan and to devise procedures to identify road segment attributes that correspond to the crash rate on curves. A second goal was to identify curves that exhibited crash frequencies significantly higher than the mean for their group. Simple and multiple regression models were found to be poor predictors of crashes, explaining only a small percentage of the variation in the crash rate on curves.

Discriminant Analysis was used to determine variables that distinguish between high and low crash rate curves. The curve length, the presence of a turn or curve warning sign, the radius of the curve and the tangent crash rate are the discriminating variables identified. Using these variables 79.1% of the curves were correctly classified.

These variables were then used to identify those curves with a high crash rate that should (based on their characteristics) have a low crash rate. These curves are candidates for countermeasures implementation.

Cluster analysis was used to identify the variables with a strong association with the crash rate. The clustering of high, medium and low crash rate curves with other variables was clear, with cluster one having a crash rate of 3.08, cluster two a crash rate of 7.78 and cluster three a crash rate of 18.05.

The same variables identified in the discriminant analysis were important in the cluster analysis. The ADT, curve radius and length, and the presence of traffic control devices (arrow and chevron) are all important in defining the clusters.

ACKNOWLEDGMENTS

First and foremost I want to thank my academic advisor Dr. William Taylor, for his advice and guidance throughout this research and my Ph.D. work. My appreciation and gratitude are also extended to the members of my guidance committee, Dr. Nancy Scannell, Dr. James Stapleton, Dr. Richard Lyles and Dr. Thomas Maleck. I also would like to thank Dr. Bruce Piggozi for what I learned from him.

During my years as a student in MSU, I was fortunate to know many people which I am thankful to, Dr. John Hazard, Dr. Gail Blomquist, Dr. Charles Barr and Dr. Charles Cutts to name a few.

I also want to thank the Michigan Department of Transportation for the cooperation of the staff and for funding the project through the Michigan State University's Transportation Center of Excellence.

Special thanks are due to the support staff of the Department of Civil and Environmental Engineering, Linda Phillipich, Laura Taylor, Cheryl Mollitor, Linda Steinman and Mary Wiseman.

Finally, I wish to express my sincere gratitude to my family and friends, who shall remain unnamed, for their encouragement, support and making my efforts worthwhile.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION AND OBJECTIVES

In Michigan over 25 percent of fatal traffic crashes take place on non-freeway trunkline highways. These highways typically have all the elements associated with a high number of serious crashes. Lack of a separation buffer from the opposing traffic, combined with rather high speeds contributes to such crashes.

Research has consistently demonstrated that crash rates on horizontal curves are many times higher than the rate on the tangent sections on the same road, and most studies have found the degree of curvature to be the most significant single factor related to curve crashes. However, other roadway features, such as superelevation and skid resistance of the pavement surface, traffic control elements, driving environment and human factors, individually or in combination are major contributors as well.

Several models, most notably the Glennon Model and the Zegeer Model, have been developed to explain curve crashes. However, when applied to Michigan data, the results are not sufficiently reliable for establishing corrective or preventative programs.

The purpose of this study was to analyze horizontal curve crashes experienced on two-lane rural trunkline roads in the State of Michigan, and to devise procedures to identify curved road segment grouping attributes that are associated with the crash rate on these curves. A second goal was to identify curves that exhibited crash frequencies significantly higher than the mean for their group, or which potentially may exhibit such crash frequencies.

The specific objectives were to:

1) Identify the factors influential in horizontal curve crashes based on Michigan's crash data.

2) Prepare guidelines as to where and to what extent improvement of horizontal curves is warranted.

LITERATURE REVIEW

Modeling of Crashes on Horizontal Curves

Prior to 1985, modeling of crash frequencies or rates on horizontal curves was normally based on a single variable. For example, Jorgenson (1) in 1978 reported a linear relationship between crashes and the degree of curvature.

In 1985, Glennon et al. (2) published a report titled "Safety and Operational Consideration for Design of Rural Highway Curves." The research was performed to study the safety and operational characteristics of two-lane, rural highway curves. A series of independent research methodologies were employed, including; (a) multivariate crash analyses; (b) simulation of vehicle/driver operations using the Highway Vehicle Operation Simulation Model (HVOSM); (c) field studies of vehicle behavior of highway curves; and (d) analytical studies of specific problems involving highway curve operations.

The results of each of these approaches confirmed that, in general, as curve radius decreases, crash rate increases. However, radius of curve is not the only geometric element affecting safety. The crash and field studies showed that the design of highway curves must consider a series of trade-offs among the basic elements of a curve: radius, superelevation, and length.

The study also found that either very sharp or very long highway curves tend to produce more crashes. Larger angles (i.e., greater than 45 degree) require either sharp curvature, or a long curve length and should be avoided when possible.
Studies of crashes on highway curves showed single-vehicle run-off-road crashes to be of paramount concern. Roadside treatment countermeasures were found to offer the greatest potential for mitigating the frequency and severity of crashes on rural highway curves. Studies involving a single factor have generally reached the following conclusions:

a) As shoulder width increases, the probability that a highway curve will be a high crash location decreases.

b) Roadside character (roadside slope, clear zone width, and coverage of fixed-objects) is the most dominant contributor to the probability that a highway curve is a high-crash location.

c) As pavement skid resistance decreases, the probability that a highway curve will be a high-crash location increases.

d) Limited sight distance increases the probability that a curve will be a high crash location. Two special considerations of stopping sight distance are important:
(a) the increased friction demand of a vehicle that is both cornering and braking; and
(b) the loss of the eye height advantage for truck drivers on highway curves when the horizontal sight restriction is either a row of trees, a wall, or a vertical rock cut.

MODELING EFFORTS

Based on these analyses, Glennon developed and presented the following crash model in the Transportation Research Board's Special Report 214:

$A = ARs\,(L)(V) + 0.0336\,(D)(V) \qquad \text{for } L >= Lc$

where,

A = Total number of crashes on the roadway segment.

ARs = Crash rate on comparable straight roadway segments in crashes per million vehicle miles.

L = Length of roadway segment in miles

V = Traffic volume in millions of vehicles

D = Curvature in degrees

Lc = Length of curved component in miles

As noted in Special Report 214, the accuracy of this horizontal curve model "may be diminished for curves sharper than about 15 degrees, the approximate limit recorded in the data base from which the model was calibrated." This model does not include the following factors and curve design parameters: curve length, superelevation and superelevation run-off, spiral transitions, cross-slope break, roadside, geometric design consistency.

In 1986, Zegeer et al. (3) reported the results of their study "Safety Effects of Cross-section Design for Two-lane Roads, Volume I." In this study, they quantified the effects of lane width, shoulder width, and shoulder type on highway crash experience on extended sections of roadways based on an analysis of data for nearly 5,000 miles of two-lane highway from seven states. The following crash prediction model resulted from that study:

$$AO/M/Y = 0.0019\,(ADT)^{0.8824}\,(0.8786)^{W}\,(0.9192)^{PA}\,(0.9316)^{UP}\,(1.2356)^{H}\,(0.8822)^{TER1}\,(1.3221)^{TER2}$$

where:

AO/M/Y = related crashes (i.e., single-vehicle plus head-on plus opposite direction sideswipe plus same direction sideswipe crashes) per mile per year.

ADT = average daily traffic

W = lane width in feet.

PA = average paved shoulder width in feet.

UP = average unpaved shoulder width in feet.

H = roadside hazard rating, a subjective measure with values of 1 to 7 (least to most hazardous), based on a visual assessment.

TER1 = 1 if terrain is flat, otherwise 0.

TER2 = 1 if terrain is mountainous, otherwise 0.

This model is applicable only to:

- two-lane, two-way paved rural highways of state primary and secondary systems.
- lane widths of 8 to 12 feet.
- shoulder widths of 0 to 10 feet.
- ADT's less than 10,000 vpd.
- homogenous roadway sections.

The model does not include intersection related crashes or those crash types that are not expressly stated on the previous page. The model did not explain the variance in crash experience on horizontal curves. This model does not include the effects of horizontal or vertical alignment or the frequency of horizontal curves, or the frequency of sight-restricted crest vertical curves.

In 1991 Zegeer et al. (5) formulated the following model for predicting crashes on horizontal curves:

$$A=[1.552(L)(V)+0.014(D)(V)-0.012(S)(V)](0.978)^{(W-30)}$$

where:

A=number of total crashes on the curve in a 5-year period.

L=length of curve in miles (or fraction of a mile)

V=volume of vehicles in million vehicles in a 5-year period passing through the curve (both directions)

D=degree of curve

S=presence of spiral, S=0 if no transition spiral exists and S=1 if there is a transition spiral.

W=width of the roadway on the curve in feet.

The purpose of this study was to determine the horizontal curve features which affect safety and operations and to quantify the effects on crashes of various curve-related improvements. The primary data base developed and analyzed consisted of 10,900 horizontal curves in Washington State. Three existing federal data bases on curves were also analyzed. These data bases included the cross-section data base of nearly 5,000 miles of roadway from seven states, a surrogate data base of vehicle operations on 78 curves in New York, and 3,277 curve roadway segments from four other states.

Based on statistical analyses and model development, variables found to have a significant effect on crashes include degree of curve, roadway width, curve length, ADT, presence of a spiral, superelevation, and roadside condition.

In a comprehensive review of design features related to highway safety, McGee et al. (6) concluded that the Zegeer and Glennon models were the best models available for predicting crashes on horizontal curves. The authors of this report concluded that:

> "The Zegeer model relating crashes to horizontal alignment appears to represent the best available relationship to estimate the number of crashes on individual horizontal curves on two-lane rural roads, although it does have limitations. While the model explicitly considers curve length, degree of curvature, roadway width, and presence of a spiral transition, it does not explicitly consider roadside parameters or the effect of upstream or downstream alignment. The fact that it does not consider roadside or even some surrogate rating for roadside is a major limitation, especially since crash research has shown that roadside design is a determinant of horizontal curve safety.
>
> The model does not consider the effect of vertical alignment or the consistency with respect to the design of all curves within the highway section (e.g., geometric design consistency). The model also does not consider the frequency of horizontal curves greater than three degrees within the section, the frequency of sight-restricted vertical crest curves, or the percent grade. The average operating speeds or design speeds are also not considered explicitly. The model does not consider the influence of access points, driveways or intersections that may be in close proximity to the subject curve."

In 1992, Kach and Benac (7) tested both the Zegeer and Glennon models with Michigan Trunkline data, and found a poor fit between the predicted and actual crash frequency, as shown in Figures 1 through 3.

After reviewing the models developed by Glennon and Zegeer, the authors identified the following weaknesses of these models:

1. The model predicts total crashes instead of "curve related" crash types: fixed object, overturn, head-on and side swipe opposite direction crashes.
2. The models do not recognize an "influence zone" for curves.
3. The models do not adequately address the variability in crash experience for all the curves with a given length and degree of curvature.

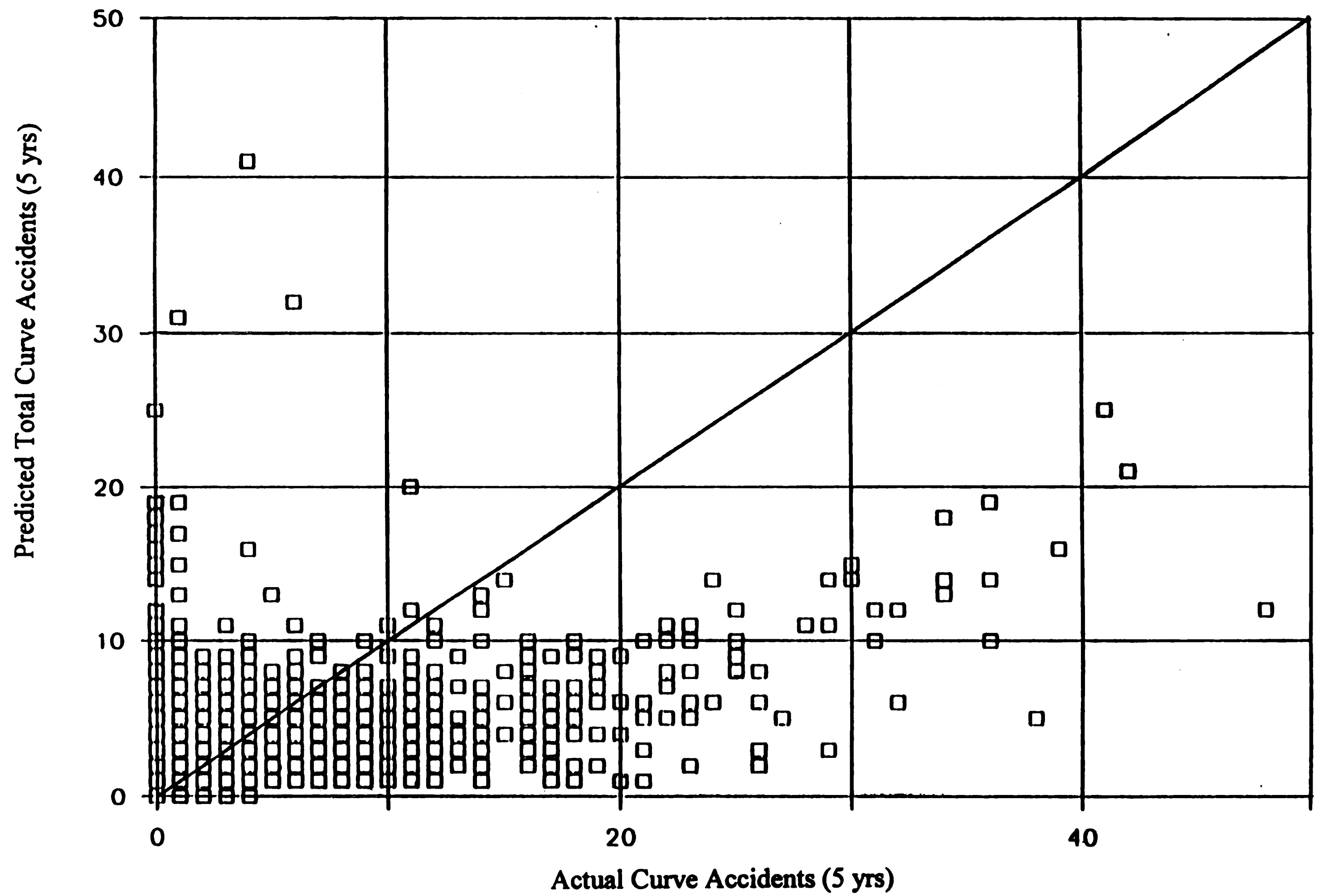


Figure 1 Actual vs Predicted Total Accidents (Glennon Model, Calculated ARs)

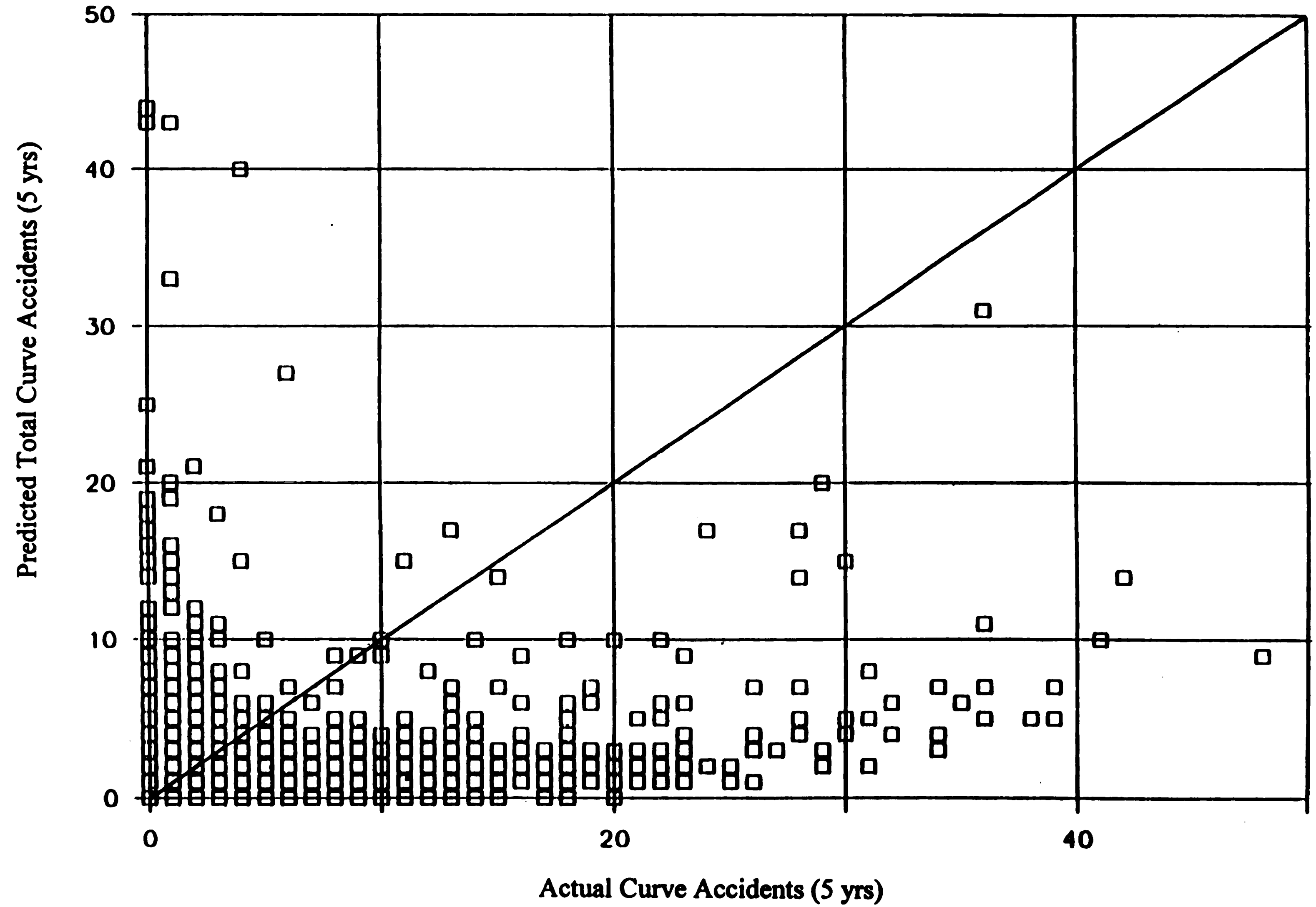


Figure 2 Actual vs Predicted Total Accidents (Glennon Model, 0.902 ARs)

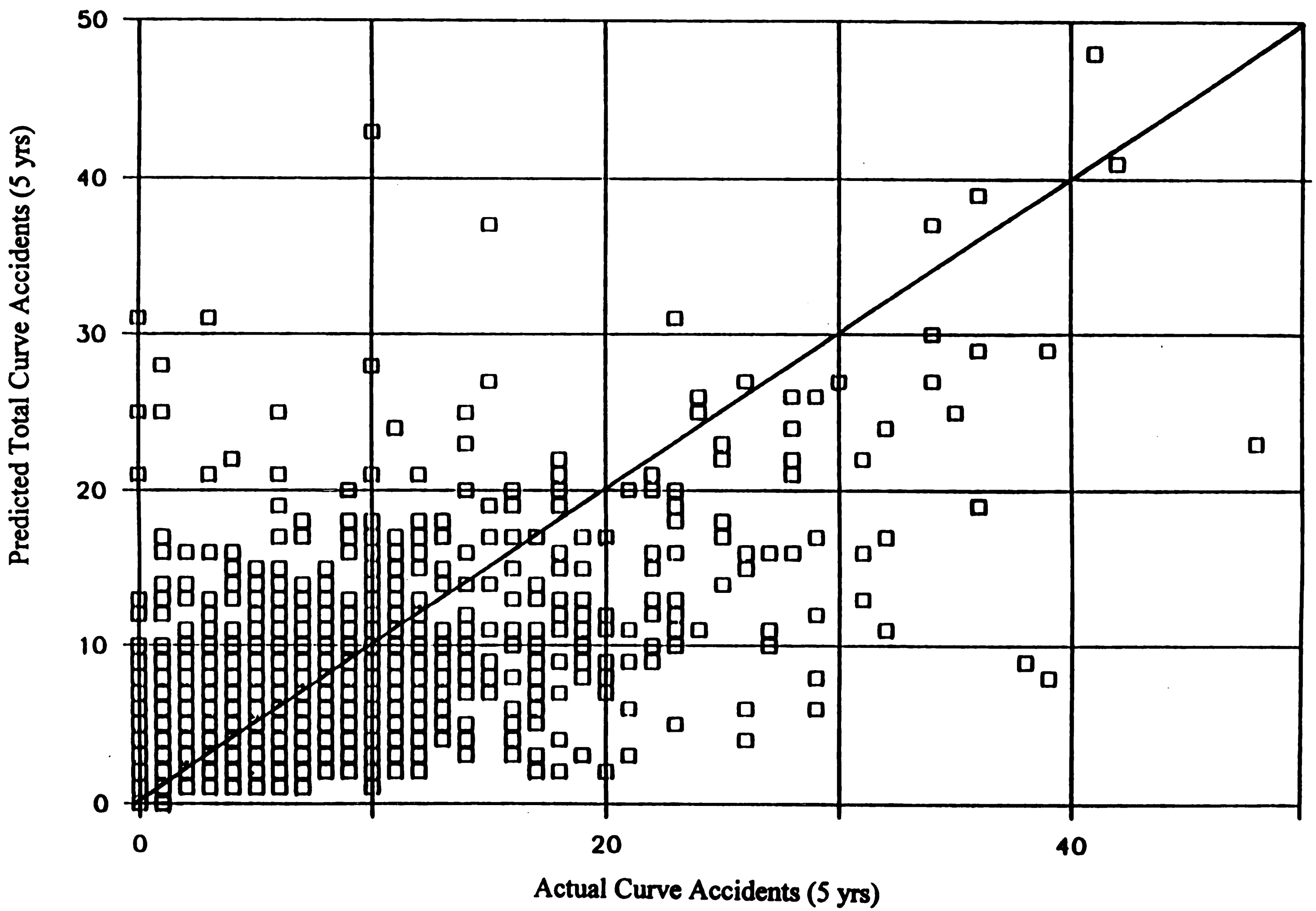


Figure 3 Actual vs Predicted Total Accidents (Zeger Model, lanewidth included)

In 1995, Fink and Krammes (8) reported on a study of the effect of the length of the tangent preceding a horizontal curve and the approach sight distance on crashes at horizontal curves. To add insight on the effects of these variables on safety and operations at horizontal curves, a base relationship between crash rates at horizontal curves and degree of curvature was established, and the effects of approach tangent length and approach sight distance on this relationship were examined.

The results confirm that degree of curvature is a good predictor of crash rates on horizontal curves. Although the effects of approach tangent length and sight distance were not as clear, the results suggest that the adverse safety effects of long approach tangent length and short approach sight distance become more pronounced on sharp curves.

Four other studies considered tangent length among a set of candidate predictors of crash rates at horizontal curves (10-13). Their findings with respect to tangent length were mixed. Datta et al. (10) found tangent length to be a significant predictor of outside-lane crash rates for one subset of 25 curve sites in Michigan. Terhune and Parker (11) evaluated tangent length (among other variables) using data bases of 78 curves in New York, 40 curves in Ohio, and 41 curves in Alabama, and concluded that tangent length was not significant. Matthews and Barnes (12) studied 4,666 curves on the rural two-lane

portion of State highways in New Zealand. They found a significant relationship that involved tangent length in combination with other variables and concluded that crash risk was particularly high on short radius curves at the end of long tangents, on steep down grades, and on relatively straight sections of roads.

Zegeer et al. (13) evaluated the significance of the minimum and maximum distance to the adjacent curve; although neither variable was significant, they observed, "there appears to be evidence that tangents above a certain length may result in some increase in crashes on the curve ahead."

Glennon et al. (14) concluded that approach sight distance was not a significant variable in a discriminate analysis of curve sites with high and low crash rates. Fambro et al. (15) concluded that available stopping sight distance is not a good indicator of crashes, with the exception that "when there are intersections within limited sight distance portions of crest vertical curves, there is a marked increase in crashes."

The report by Fink and Krammes (8) presented two models:

1) A regression model for predicting mean crashes per million vehicle kilometers versus mean degree of curvature:

$$\text{mean crash rate} = 0.05 + 0.23 \text{ mean degree of curvature}$$

The model has an r^2 value of 0.94. The r^2 is much higher than typically observed in crash analyses, because the unit of observation is a grouping of curve sites into nine degree-of-curvature categories which eliminates much of the variability among individual sites.

2) A regression model for predicting the crash rate based on the approach tangent length. Three categories of curves were defined representing those with the shortest 25 percent (<=107 m [350 ft]), middle 50 percent (107 m[350 ft] to 427 m [(1400 ft]), and longest 25 percent (>427 m [1400 ft]) of tangent lengths in the database. The regression models were as follows:

mean crash rate = 0.35 + 0.16 mean degree of curvature for the shortest 25%

mean crash rate = -0.30 +0.32 mean degree of curvature for the middle 50%

mean crash rate = 0.52 + 0.20 mean degree of curvature for the longest 25%

The results indicate that the slope and intercept for the middle 50 percent of tangent lengths are significantly different from the slope and intercept for the shortest and the longest 25 percent. (See Figure 4)

These models, like those of Zegeer and Glennon, fail to explain the variation in crash rate experienced at different curves with the same degree of curvature or the same approach tangent length.

While the models found in the literature may have some value when the design engineer is developing the alignment for a new road, none are suitable for identifying hazardous curves on an existing road system.

They also provide no assistance in determining countermeasures once a location is identified as being hazardous.

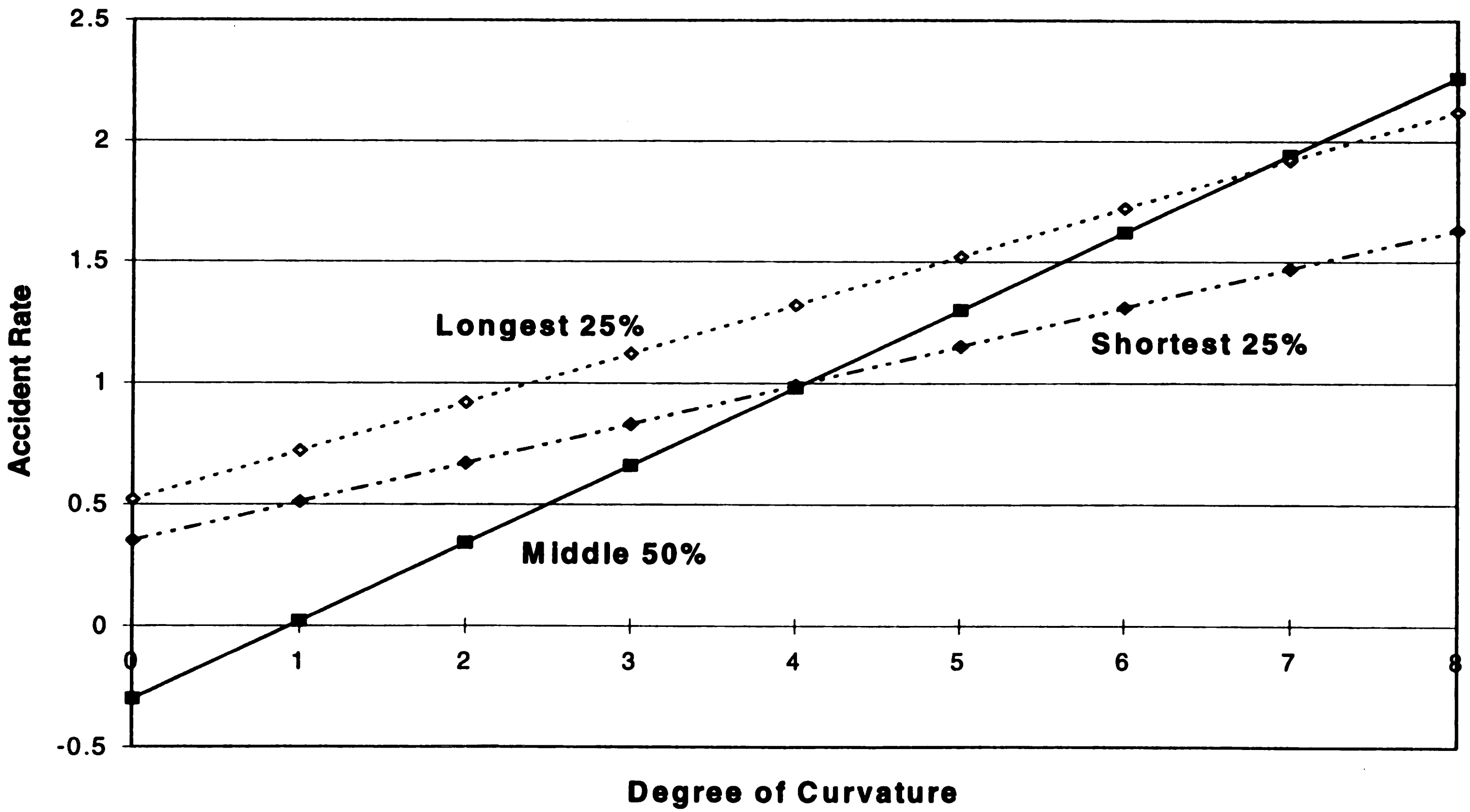


Figure 4 Accident rate versus degree of curvature

STUDY DESIGN

Data Preparation

To accomplish the objectives of this study, a multi-step approach was utilized. Step one was to acquire geometric data for all the rural, two-way, two-lane trunkline highways in Michigan from the Michigan Department of Transportation (MDOT). Based on the selection criteria listed in Table 1, the candidate curves were selected and the control section (reference system used by MDOT for trunklines) and the mile points of the beginning and ending of the curves were noted.

Table 1 Curve Selection Criteria

Rural two-lane, two-way trunkline highways
No taper, no extra lanes
No curb, no parking
No median, and preferably no intersections
At least 306 meters (0.19 mile, about 1000 feet) of tangent at each end of each curve
Preferably at least 611 meters (0.38 mile) of tangent between the two curves
Degree of curvature greater than one

The geometric selection criteria yielded a total of 285 roadway segments, each consisting of a curve and two tangents. Based on the photo log observations, 50 of the segments did not fit the specified criteria and were eliminated from the study. Fifteen more were eliminated based on the field observation. Examples of such cases are listed in Table 2. The final data set consisted of 220 valid roadway segments.

Table 2 Examples of the Disqualified Roadway Segments

(based on the photo log or field observations)

Control Section	Listed BMP*	Listed EMP*	Length km	Actual Length km	Comments
23111	3670	3710	0.06	0.21	Intersection Corner
32092	60	190	0.21	-	Intersection widening (M52/M36)
38073	9810	9920	0.18	0.26	Curve not found
38073	14350	14500	0.24	-	Curve not found
46011	5770	5900	0.21	0.10	Three Lanes (intersection with left turn lane)
46012	110	300	0.31	-	Three Lanes (intersection with left turn lane)
46051	380	490	0.18	0.27	Not found. Two curves near listed location
46074	20	130	0.14	0.18	Intersection (with median and right turn lanes)

* (coded in 0.001 mile with implied decimal point)

The next step consisted of obtaining additional data from the photo log. In addition to data acquisition, data verification was also performed and locations which, based on this review, did not meet the selection criteria were removed from the database. While this step was in progress, field data collection was performed to obtain the curve superelevation and pavement friction. Field data collection resulted in the elimination of some additional curves where it was determined that the lane width or shoulder width were modified on the curve or there was an intersection within the study limits. After this step 220 curves were left for the final analysis. For each of the 220 curves, all the crashes corresponding to the mile points from 306 meters (0.19 mile) before the start of the curve to 306 meters (0.19 mile) after the end of the curve were extracted from the MDOT crash files. This procedure was performed for each year of the six year period of 1989 to 1994, yielding 3107 total crashes (Table 3).

Table 3 Number of crashes in the database

	Total Crashes	**Related Crashes**
1994	519	207
1993	491	166
1992	503	176
1991	532	164
1990	503	145
1989	559	136
Total	**3,107**	**994**

Of the 3107 total crashes, 991 were in curves and 2116 in the tangents. The total number of "related" crashes was 994 of which 463 were in the curves and 531 in the tangents.

Thirteen of the 220 roadway segments did not have any crashes in the curve or the two tangent sections. For the "Related" crashes 178 roadway segments had crashes in either the curve or the tangent sections.

The crash report forms for all these crashes were obtained and processed to locate the individual crash as being on the curve or on the tangent. After this step, various analyses were performed, including comparison of the actual curve crashes and those predicted by the models.

The Data

Data for the project consists of the four following sets:

1) Geometric data provided by the MDOT

2) Six years of crash data for the years 1989 through 1994

3) Data obtained from the photo log for all 220 segments

4) Field data for 81 segments.

The geometric data consisted of 44 variables such as control section, beginning mile point, ending mile point, average lane width, total shoulder width (right and left), etc. The variables selected from this file for use in this study are shown in Table 4.

The crash data are from the Michigan State Police "State of Michigan Master Crash File." This file contains information on up to three vehicles involved in a crash, but the data for the second and third vehicles were not used in the study. The original source of the data is the "State of Michigan Traffic Crash Report" (Form UD-10). The data consisted of 120 variables such as district, control section, mile point of crash, highway area type, highway area code, etc. The data were for the crashes for both traffic directions combined. The variables selected from this file for use in this study are shown in Table 5.

The photo log data were used for variables such as the presence of traffic signs (arrow, chevron, etc.) and other variables such as approach distance at which the curve was first observed, etc. The data also included a subjective measure of the roadside clearance or hazard, on a scale of one to seven. One being "Clear" (least hazardous) and seven being "Not Clear" (most hazardous). The data acquisition for this variable was performed twice, once for each direction of the traffic flow.

Table 4 Geometric Data Variables Coded for the Study

VARIABLE DESCRIPTION	VARIABLE NAME
District	DNO
Control Section	CS
Beginning Mile Point of Study Segment	BMP
Ending Mile Point of Study Segment	EMP
Average Lane Width	ALW
Total Shoulder Width (Right)	TSWR
Paved Shoulder Width (Right)	PSWR
Total Shoulder Width (Left)	TSWL
Paved Shoulder Width (Left)	PSWL
No Passing Zone Code	NPZC
Posted Speed Limit	PSL
Degree of Curvature, Number of Degrees and Minutes	HCD, HCM
Degree of Curvature, Number of Minutes	HCM
Roadway segment File Record Number	SFRN
Intersection File Record Number	IFRN
Average Daily Traffic (Divided by 10)	ADT
Using BMP, EMP, HCD and HCM, four more variables were calculated as follows:	
Degree of Curvature in decimal degrees	HCDD
Curve Length in feet	HCLFT
Curve Radius in feet	HCRFT
Central Angle in decimal degrees	CANG

Table 5 Crash data used in the Study

District	Driver 1 Violation
Control Section	Contrib. Circumst., Vehicle 1
Crash Mile Point	Visual Obstruction, Vehicle 1
Highway Area Type	Direction of Travel, Vehicle 1
Highway Area Code	Alcohol/Drug use, Vehicle 1
Hour of Occurrence	Object Hit, Vehicle 1
Route Class	Situation, Vehicle 1
Weather Condition	Vehicle Size, Vehicle 1
Lighting	Impact Code, Vehicle 1
Road Surface Condition	Vehicle Condition, Vehicle 1
"A" Injuries	Trailer, Vehicle 1
"B" Injuries	Road Type, Vehicle 1
"C" Injuries	Number of Lanes
Road Alignment	Average Daily Traffic
Traffic Control	Number of Persons Killed
Crash Type	Number of Persons Injured
Distance From Crossroads	Number of Occupants
Direction From Crossroads	Crash Location
Intersecting Street name	Crash Route Number

Table 5 (Continued)

Driver 1 Intent	Original Prime Street Name
Number of persons injured	Operator Number, Vehicle 1
Vehicle 1 Type	Year of Crash
Vehicle 1 Make	Film Reel Number
Age of Driver 1	Film Frame Number
Residence of Driver 1	PR Number
Sex of Driver 1	PR Mile Point
Degree of injury to Driver 1	

The field data collection was performed to obtain only two variables, a measure of the superelevation of the road, and a measure of the skid resistance of the pavement surface. The superelevation was obtained by use of an ordinary 48 inch level. The difficulty with superelevation is the fact that unlike some other variables, an average value will not substitute for the lowest value and the highest value. If there is an optimal value, any deviation from it, positive or negative, could result in decreased safety. However, since there was no procedure available to record continuous values of superelevation, representative locations on the curve were selected and the average value for each lane was coded. Occasionally the superelevations were in the opposite direction, i.e., banking towards the outside of the curve. In these cases the superelevation is coded with a negative sign.

The friction factor was obtained and calculated by dragging a piece of tire filled with concrete to weigh 22.7 kilograms (50 lbs) (16). The horizontal force required to pull it over the pavement (divided by its weight), would have been the friction factor, had the tire been smooth. However, the reading corresponded to a value higher than the actual friction factor because the treads of the tire and the gravel particles on the road would "engage" and to some extent act like teeth gears. Occasionally the required horizontal force exceeded 22.7 kilograms, yielding friction factors higher than one. Since this variable was for comparison across the curves and not for the absolute values, the resulting values were used for the study. However, to avoid confusion it was referred to as the drag factor rather than the friction factor. The variables obtained from the photo log and field observations are listed in table 6.

Table 6 Variables obtained from the photo log and field observations

VARIABLE DESCRIPTION	VARIABLE NAME
Curve Sign	CURVES
Turn Sign	TURNS
Advisory Speed Sign	MPHS
Guard Rail	GRAIL
Chevron	CHEVRON
Arrow Sign	ARROW
Delineator	DLNTR
Edge Line	EDGLN
Mile point when Curve Observed	OBSDSTW
Roadside Clearance/hazard	CLRNCW
Superelevation	SPRELVN
Drag Factor	DRGFCTR

Variable Modification

Since the crash data were for both directions, variables with two values, one for each traffic direction, were reduced to a single value. This included all the photo log data, some geometric data and the two field data variables.

For the following variables, if for either direction of traffic the variable had a value of YES, the variable was coded as 1. If neither direction had a value of YES, it was coded as 0 (zero). Variables in this category were: Curve Sign, Turn Sign, Guard Rail, Chevron, Arrow Sign and Delineator.

The variable "Mile Point Where Curve Observed", was converted to the approach sight distance to the curve. A value was obtained for each direction of travel, and the lower of the two was used. Similarly for the variable "Roadside Clearance", there were two values corresponding to the two traffic directions and the higher of the two values was used.

Since the designation of right and left are associated with the direction of increasing mile point, and not the direction of the vehicle involved in the crash, the variables Total Shoulder Width Right and Left were combined into one value equal to the sum of the two. Similarly the Paved Shoulder Width Right and Left was replaced by the sum of the two values. The Shoulder or Curb Type Right and Shoulder or Curb Type Left, each with a value of 1 or 2 were collapsed into one value. If both values were the same that value was used. If one value was 1 and the other 2, a value of 2 was used.

The drag factor and superelevation also had two values, one for each side of the road. For the drag factor the lower of the two was used. For the superelevation, the lower of the two was used for one analysis, and then the analysis was repeated using the higher value.

Crash Types

For the analyses used in this project, several types of crashes were eliminated from the crash data. Only the "Curve Related" crashes consisting of the following types of crashes were used: Miscellaneous 1 Vehicle, Overturn, Fixed Object, Other Object, Head-on and Side-Swipe Opposite.

Selection of the "Related" crashes yielded 994 crashes corresponding to the 178 roadway segments which had "Related" crashes. Not all selected roadway segments had crashes in both the tangent and curve portion of the roadway segment. Table 3 listed both the number of total crashes and related crashes for each of the years included in the study.

Crash Location Mile Points

The location of each crash along its control section is indicated by a mile point. Based on the mile point of the crash location compared with the mile points of the two ends of a curve, one could presumably determine if the crash was on the curve or tangent. However, the location of a crash as recorded by the investigating police officer is not accurate. A plot of crashes showed that the crashes tend to accumulate at tenths or quarters of a mile from the nearest intersection. This level of accuracy was not adequate for this study.

To remedy this problem the UD-10s for all crashes were manually checked. If the police sketch showed the crash occured on a curve, it was assigned to the curve, even if based on the mile point it would fall on the tangent. The UD-10 forms also provide a check box for the road alignment and if the box for curve was checked, the crash was assigned to the curve. The reason being that it was unlikely that an investigator would draw a tangent section of a road showing curve, however they may draw the curve section as a tangent but check the curve box and use the code for curve. If the sketch depicted a tangent section and the curve check box was not marked, the crash was assigned to the tangent section of the study segment.

Special Data Considerations

Even though typically each roadway segment consists of two tangents of 306 meters each, and the curve itself, there were exceptions. In 14 cases the control section number changed within the 306 meters of tangent section of the roadway segment. In these cases the 306 meters of tangents existed for both ends of the curve, however, the mileage of tangents within the same control section was less than 306 meters. Pro-rated values were used to determine the tangent crashes for 612 meters for these cases. There were no cases of different control section numbers within a curve.

In another 8 cases even though there were 306 meters of tangents at each end of the curves, the distance between the end of one curve and start of another was less than 612 meters. In other words there was an overlap between the two tangents. In two cases there were crashes in the overlap section of the two tangents, and one of these cases contained "Related" crashes. The crashes corresponding to this overlapping section of tangents, (3 crashes), were included in the data for each of the two study segments involved.

ANALYSIS

Data Presentation

The crash data described in the preceding pages is presented in graphical form in Figures 5 through 10.

Figure 5 shows the number of crashes per 380 feet of curve, (Cper380) for the 178 roadway segments which had "related" crashes in their tangent sections or their curved section. The Cper380 values are sorted in ascending order including the roadway segments which did not have any crashes in their curved section.

Figure 6 shows the number of crashes per 380 feet of tangent, (Tper380) for the same 178 roadway segments, some with no crashes in their tangent sections. Similarly, the Tper380 values are sorted in ascending order.

Figure 7 shows the Tper380 values when sorted by ascending values of Cper380.

Figure 8 is the superimposed graph of Figure 5 and Figure 7.

Figure 9 shows the values of Cper380-Tper380, referred to as (CmnsT), when sorted in ascending order. Figure 10 shows the same values sorted by ascending values of Cper380.

From Figures 7, 8 and 10 it is clear that the crash rate on the tangent section approaching the curve is not sufficient to predict the curve crash rate. This is evidenced by the fact that the values of Tper380 and CmnsT do not display a consistent pattern when compared with the sorted values of Cper380.

Data Analysis

As a first step in the analysis, two sets of simple regressions, one for the curve crash rates (Cper380), and the other for the difference between the curve crash rates and tangent crash rates CmnsT, versus each of the independent variables. The results of these analyses for ADT, Tper380, HCLFT, HCRFT, CLRNCW and OBSDSTW are

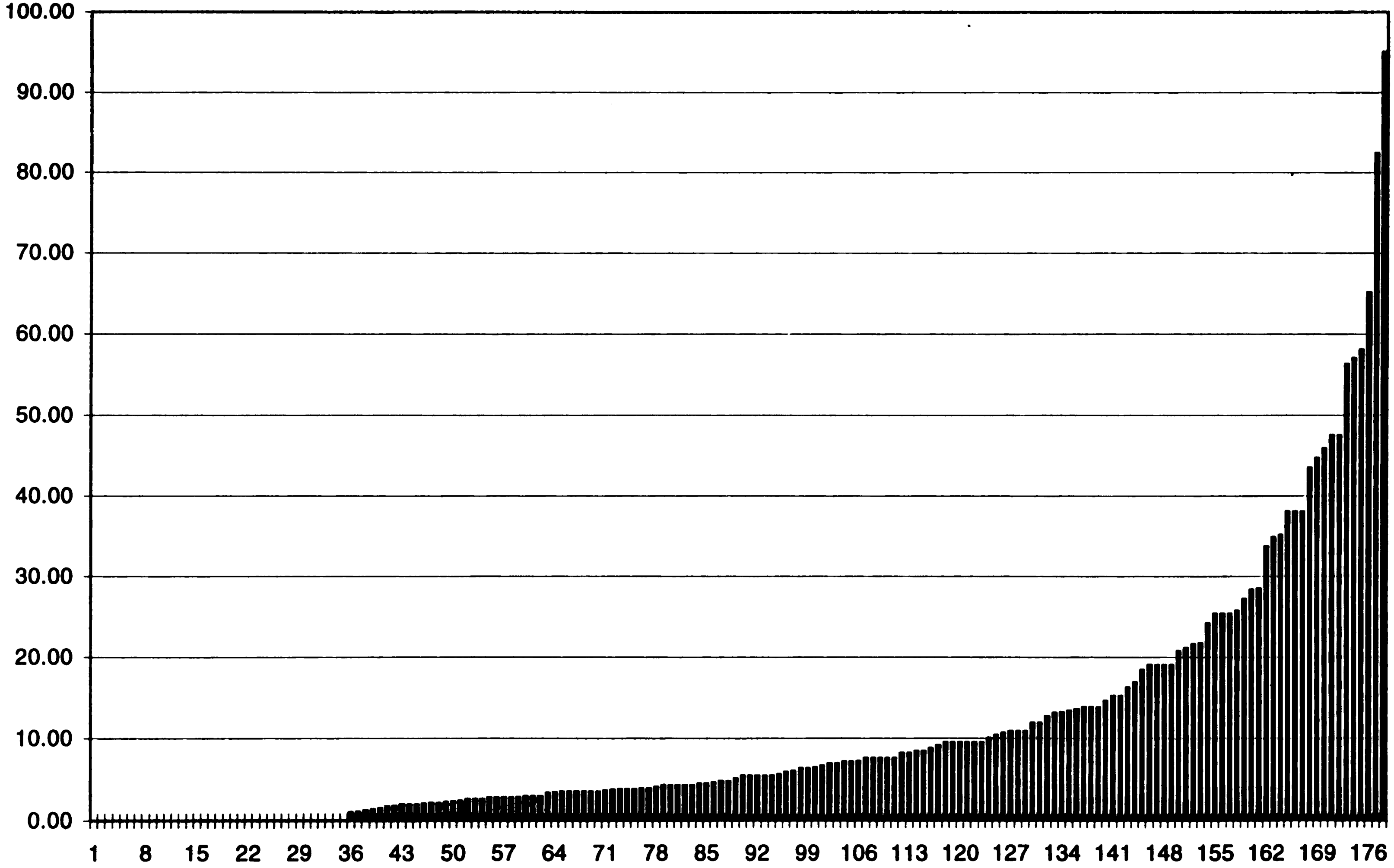


Figure 5 Curve crash rate (Cper380), arranged in ascending order

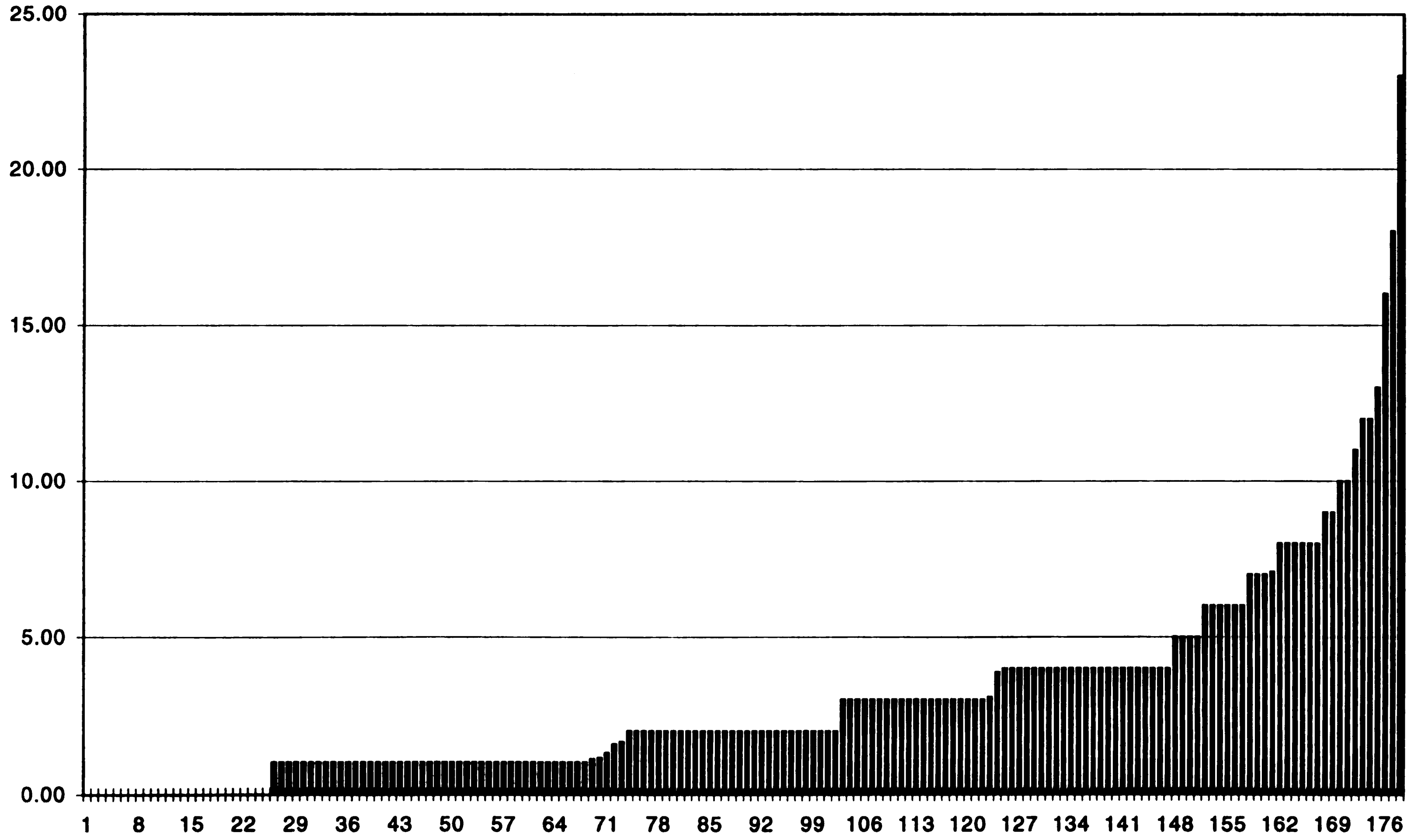


Figure 6 Tangent crash rate (Tper380), arranged in ascending order

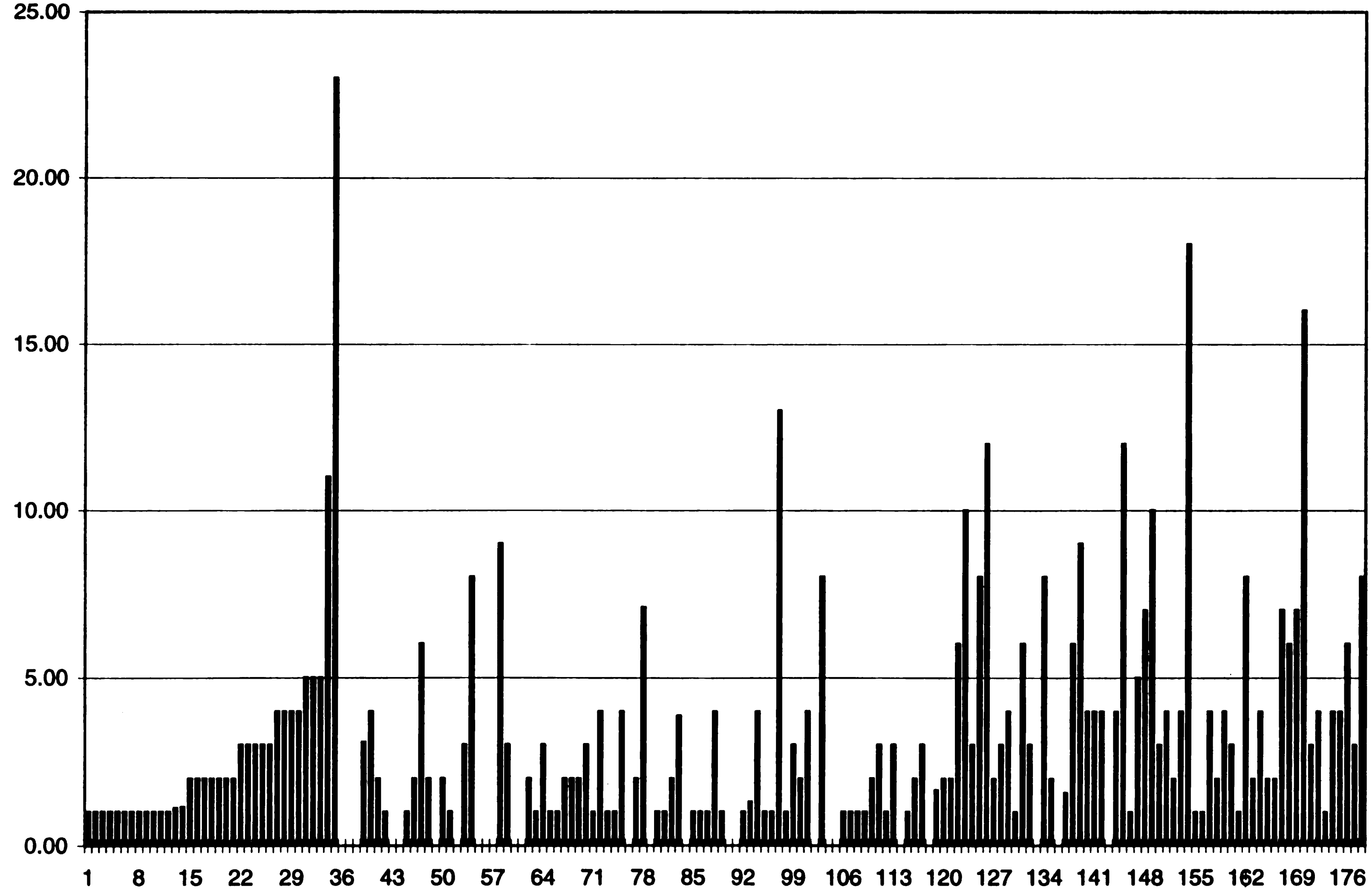


Figure 7 Tangent crash rate (Tper380), aaranged in ascending order of curve crash rate (Cper389)

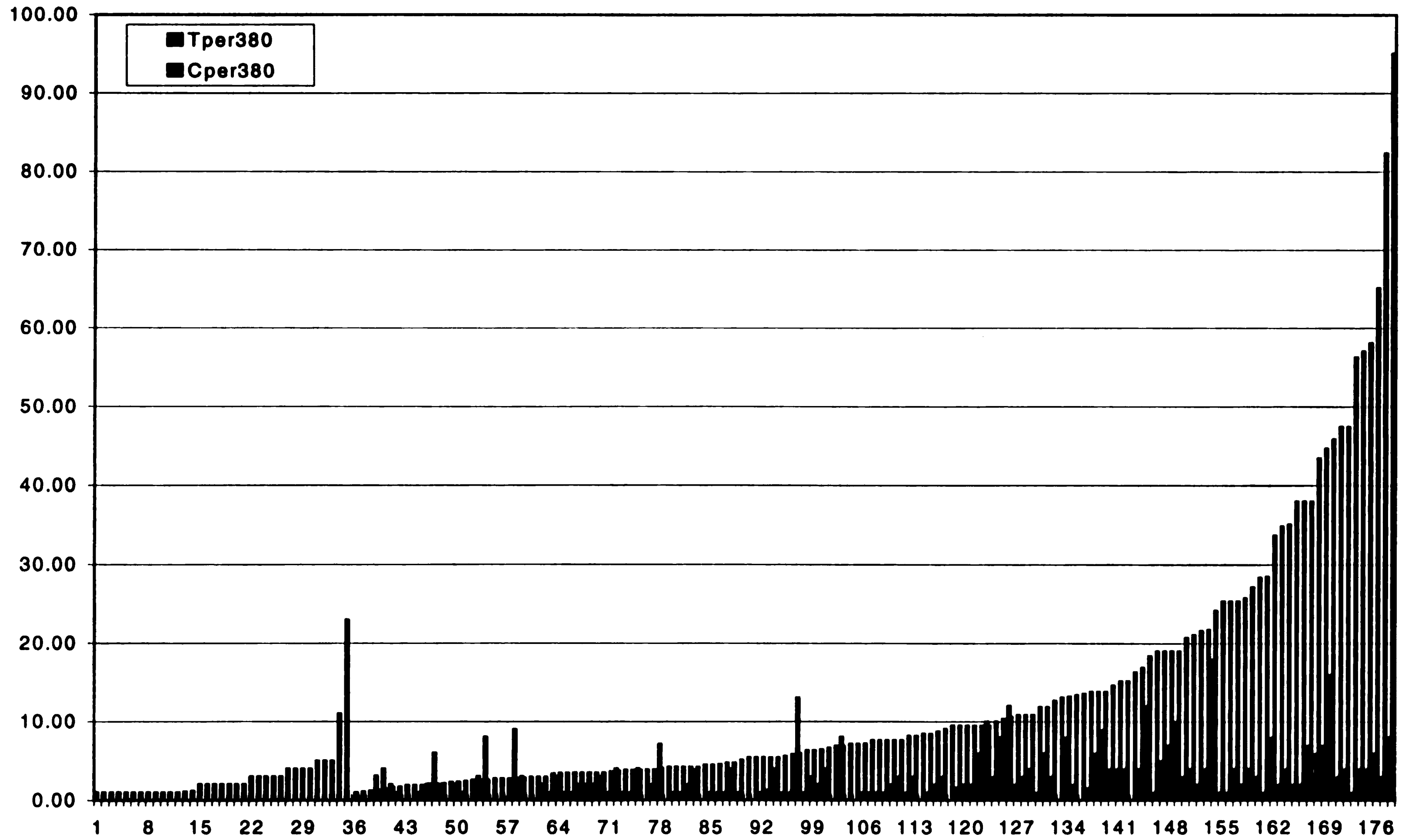


Figure 8 Curve crash rate (Cper380), and Tangent crash rate (Tper380),arranged in ascending order of Cper 380

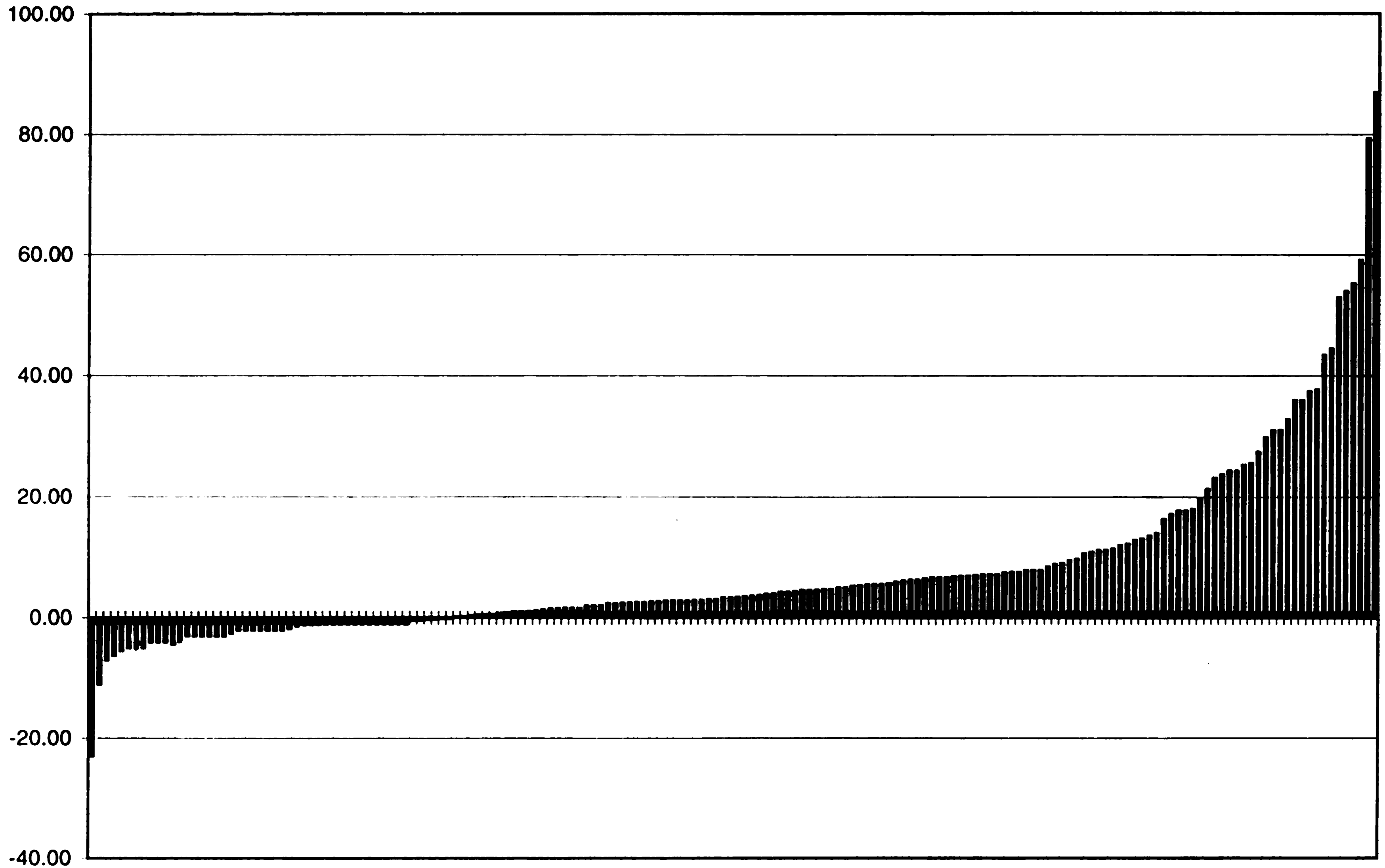


Figure 9 Curve crash rate minus tangent crash rate (CmnsT), arranged in ascending order

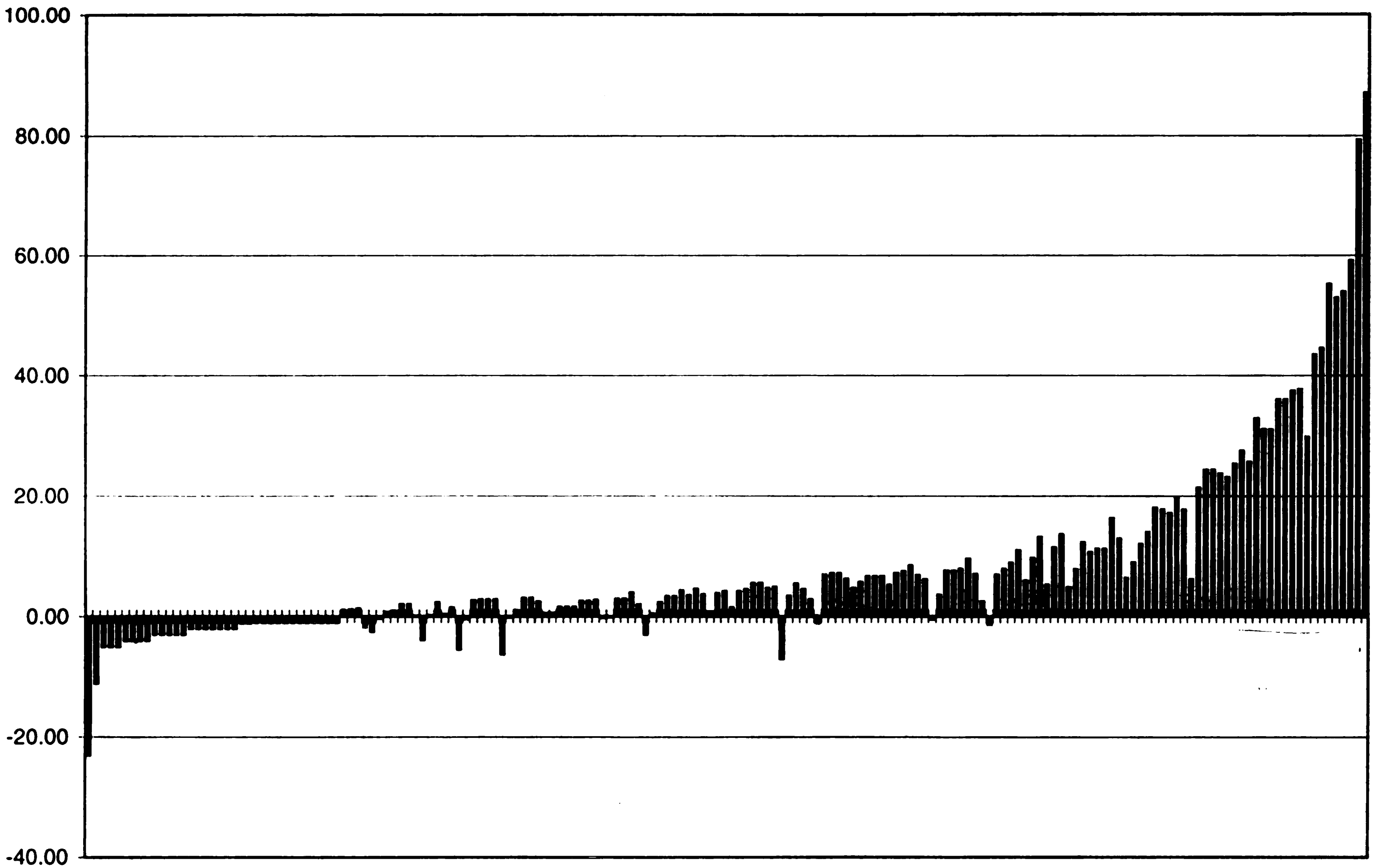


Figure 10 Curve crash rate minus Tangent crash rate (CmnsT), arranged in ascending order of curve crash rate (Cper380)

shown in Figures 11 through 22. The scatter plots, the regression lines, and the coefficients of regression all indicate that simple regression models are poor predictors of either the crash rate on curves or the difference in the crash rate between tangent sections and curved sections on the same segment of the road.

Table 7 Correlations of the linear regression models

Cper380 vs.	Corr.	Cper380 vs.	Corr.	CmnsT vs.	Corr.	CmnsT vs.	Corr.
ADT	.216	HCLFT	-.377	ADT	.062	HCLFT	-.366
ALW	-.077	HCRFT	-.408	ALW	-.074	HCRFT	-.427
ARROW	.063	MPHS	.326	ARROW	.087	MPHS	.302
CHEVRON	.280	NPZC	.187	CHEVRON	.268	NPZC	.195
CLRNCW	.049	OBSDSTW	-.130	CLRNCW	.018	OBSDSTW	-.100
CTSIGN	.175	PSL	-.073	CTSIGN	.163	PSL	-.030
DLNTR	.025	PSW	-.091	DLNTR	.060	PSW	-.108
EDGLN	.057	SCT	.060	EDGLN	.047	SCT	.012
GRAIL	.068	TSW	.051	GRAIL	.040	TSW	.019

(Table 23 shows the distribution of several of variables used in the study).
For the variable HCLFT (curve length), it appeared that there might be a nonlinear relationship. However the quadratic and cubic regression lines showed little improvement over the linear model, as shown in Figures 24 and 25.

The scatter plots indicate that there is no linear relationship between the curve crash rate and these variables that can be used to establish reliable crash reduction policies. Several nonlinear regression models were then constructed with the resulting R^2 shown in Tables 8 and 9.

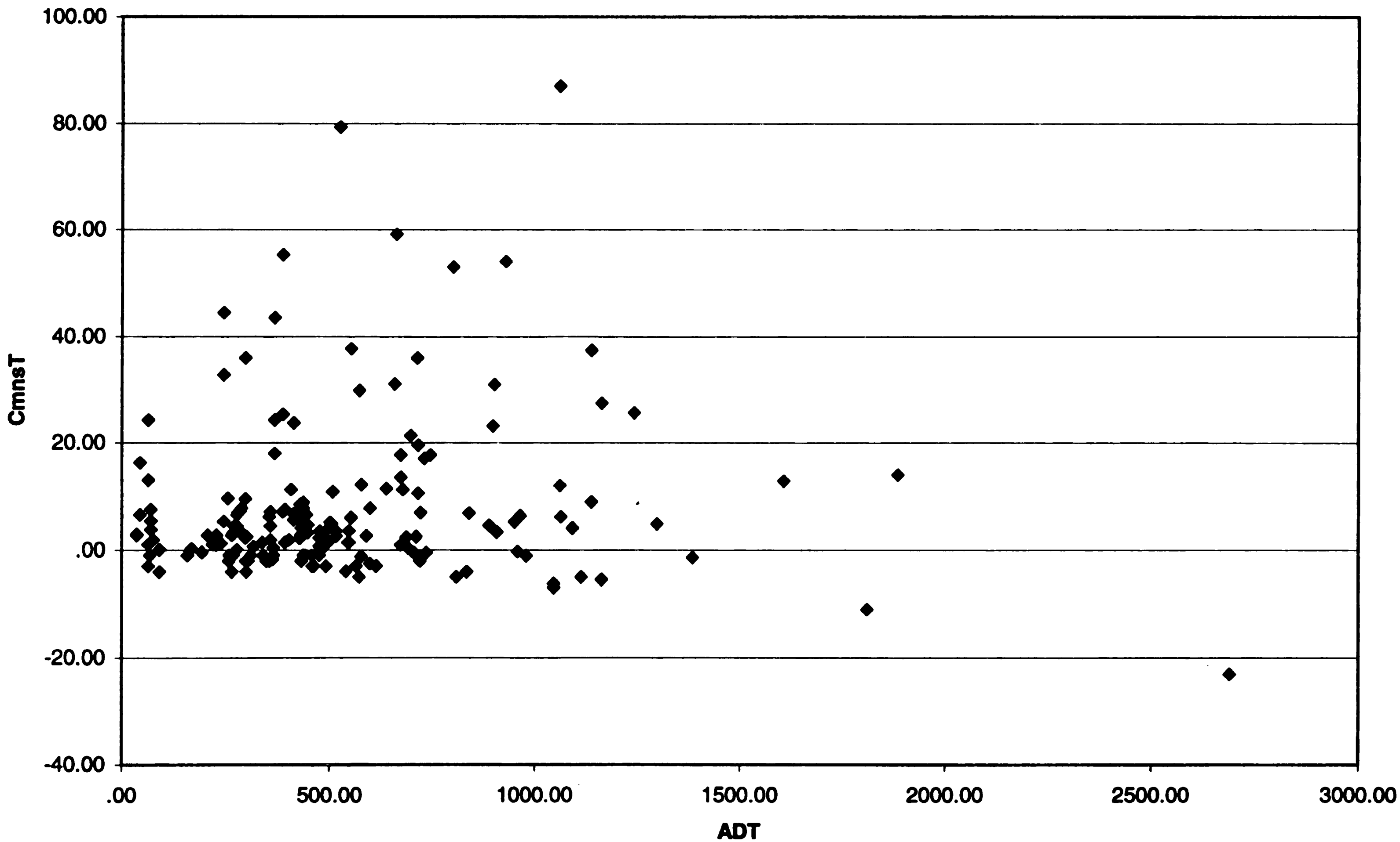


Figure 11 Curve crash rate minus tangent crash rate (CmnsT), for various values of average daily traffic (ADT)

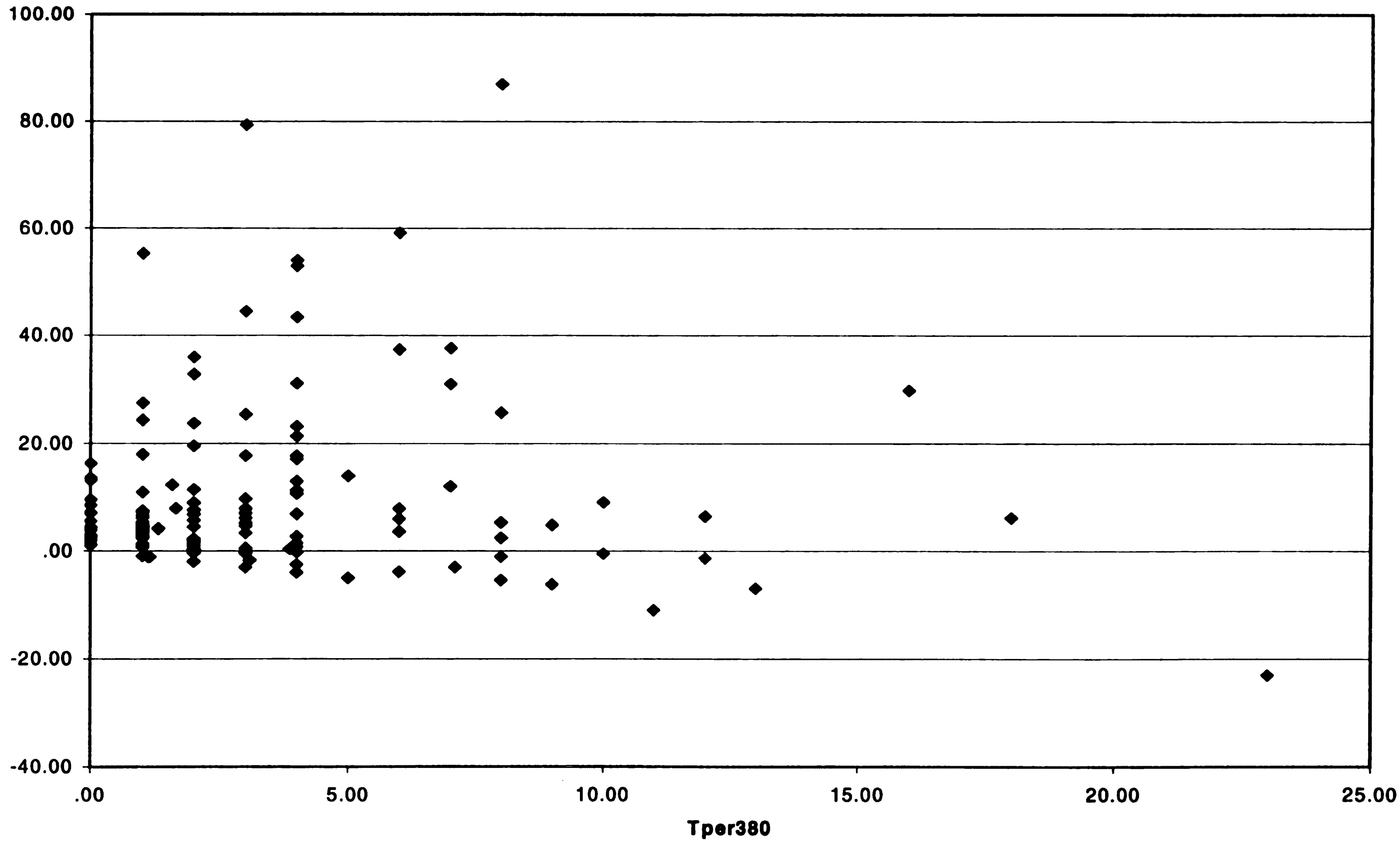


Figure 12 Curve crash rate minus tangent crash rate (CmnsT), for various values of tangent crash rate (Tper380)

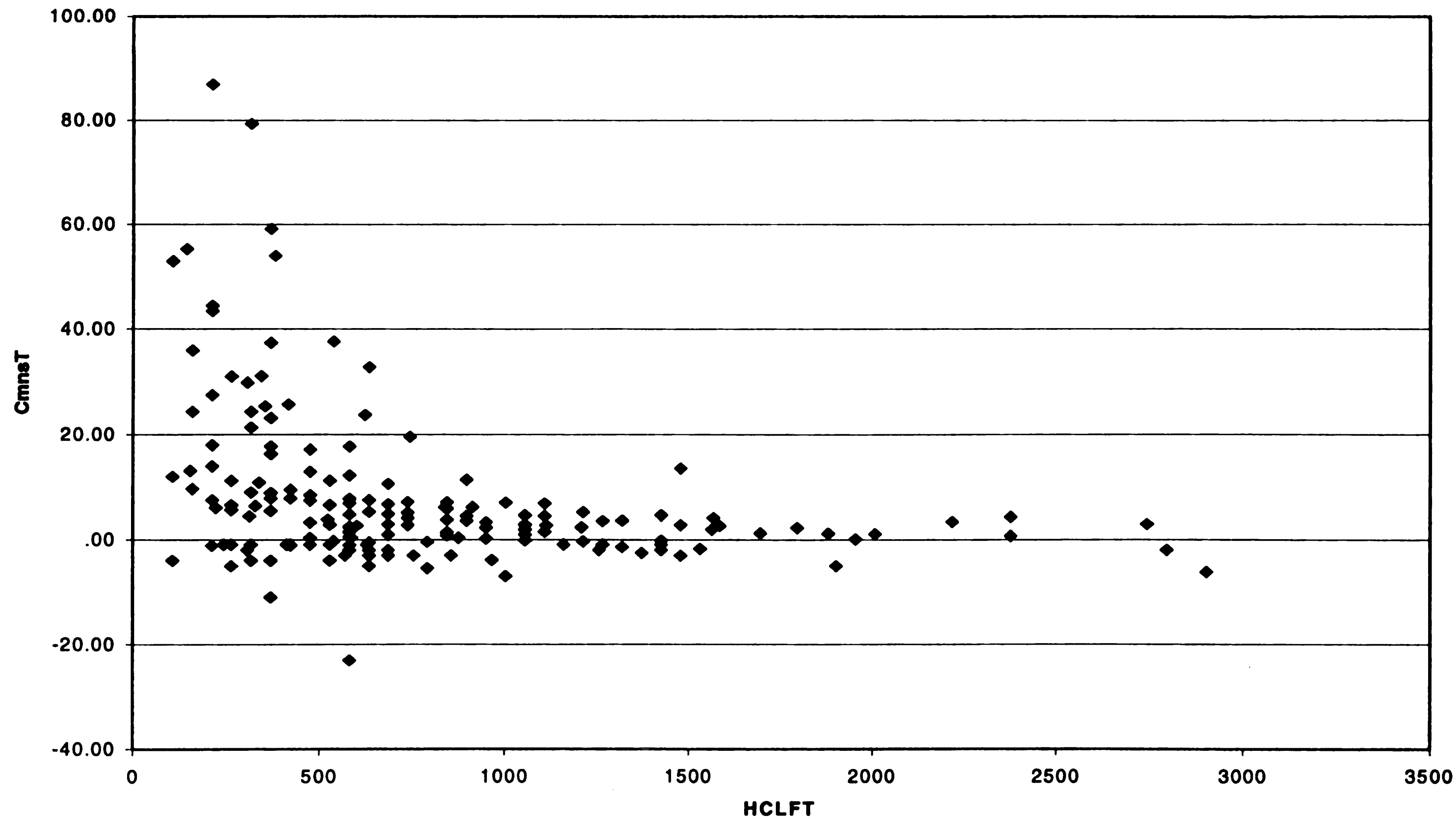


Figure 13 Curve crash rate minus tangent crash rate (CmnsT), for various values of curve length in feet (HCLFT)

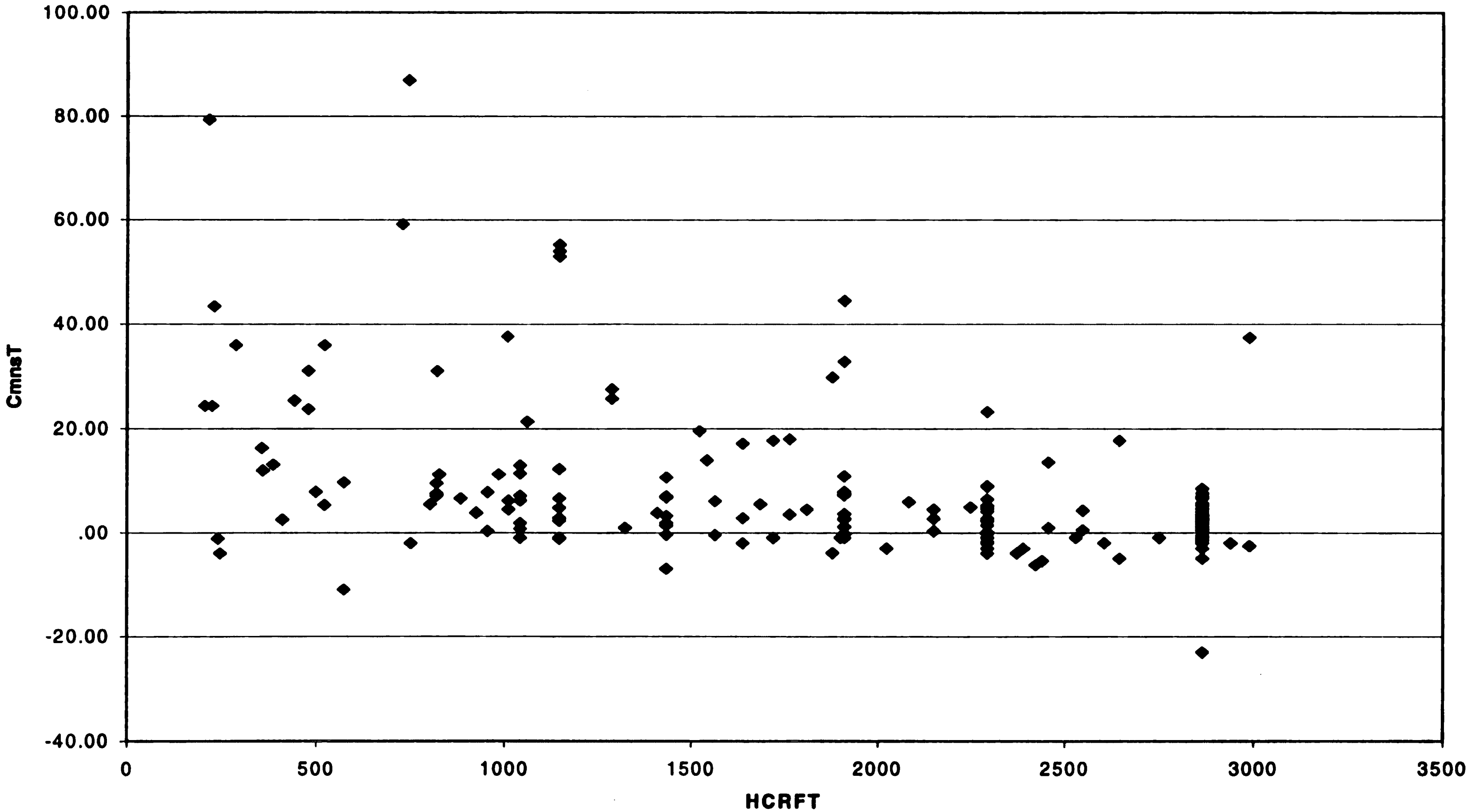


Figure 14 Curve crash rate minus tangent crash rate (CmnsT), for various values of curve radius in ft (HCRFT)

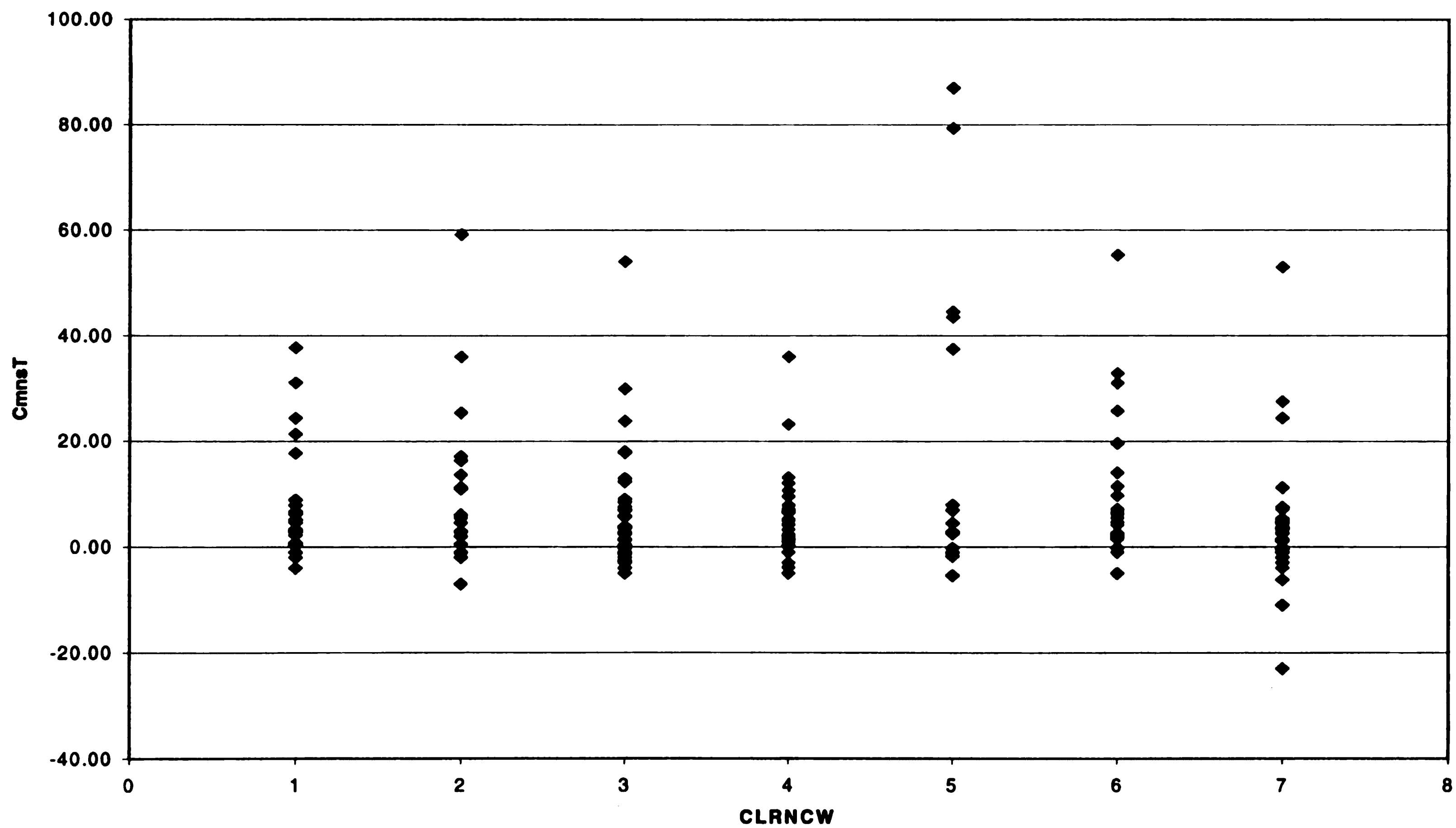


Figure 15 Curve crash rate minus tangent crash rate (CmnsT), for various values of roadside clearance

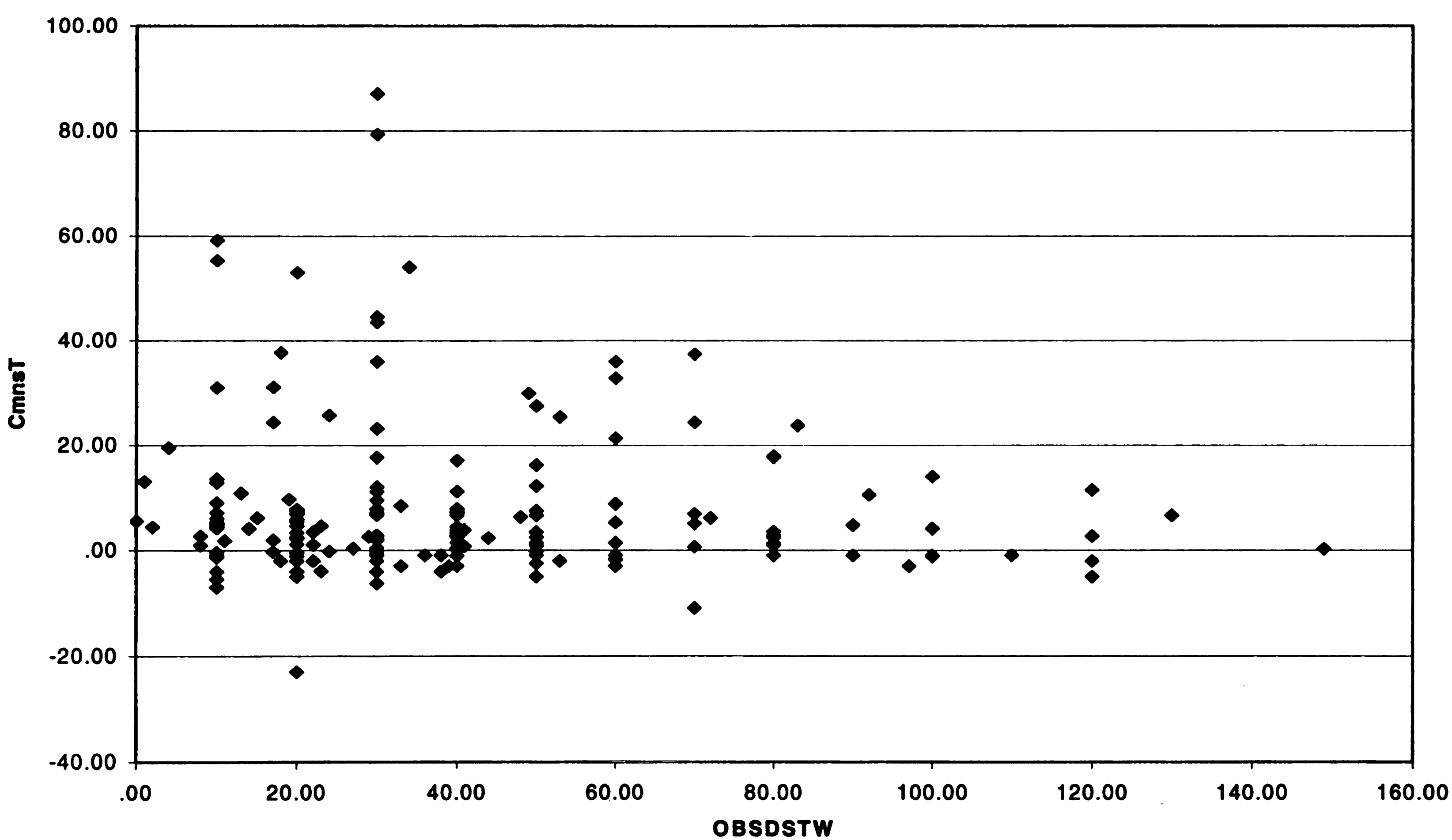


Figure 16 Curve crash rate minus tangent crash rate (CmnsT), for various values of sight distance to the beginning of curve (OBSDTW)

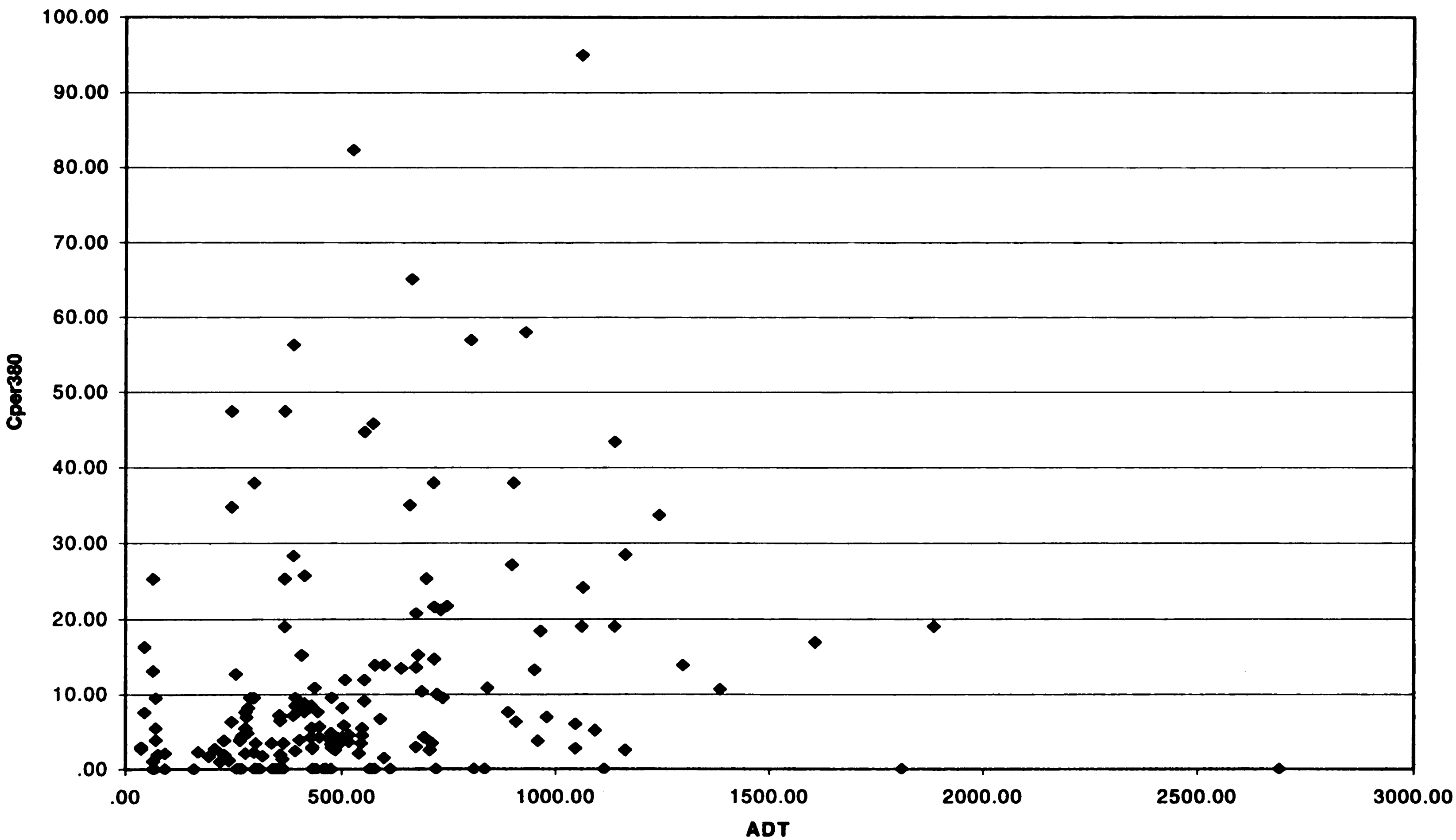


Figure 17 Curve crash rate (Cper389), for various values of average daily traffic (ADT)

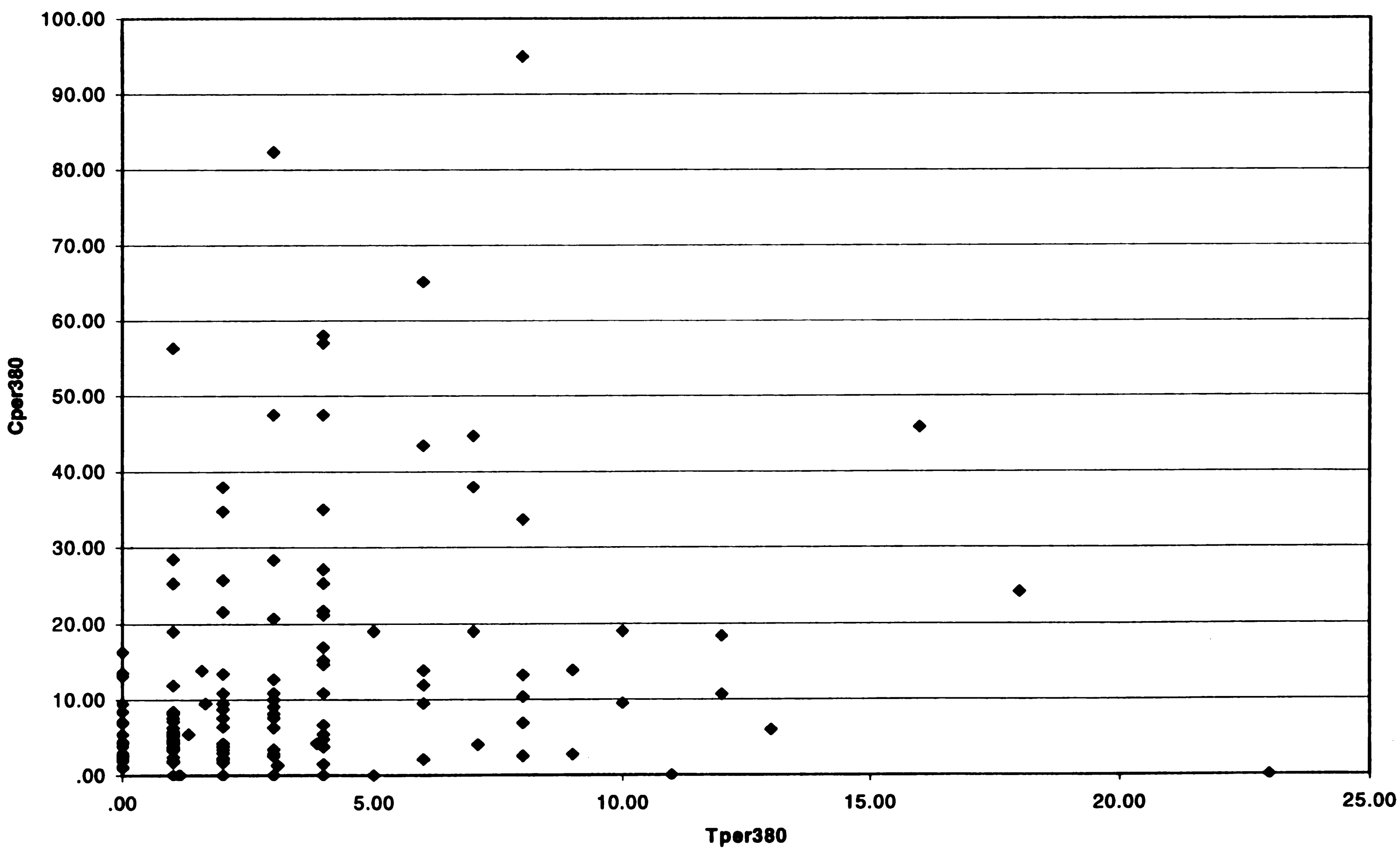


Figure 18 Curve crash rate (Cper380), for various values of tangent crash rate (Tper380)

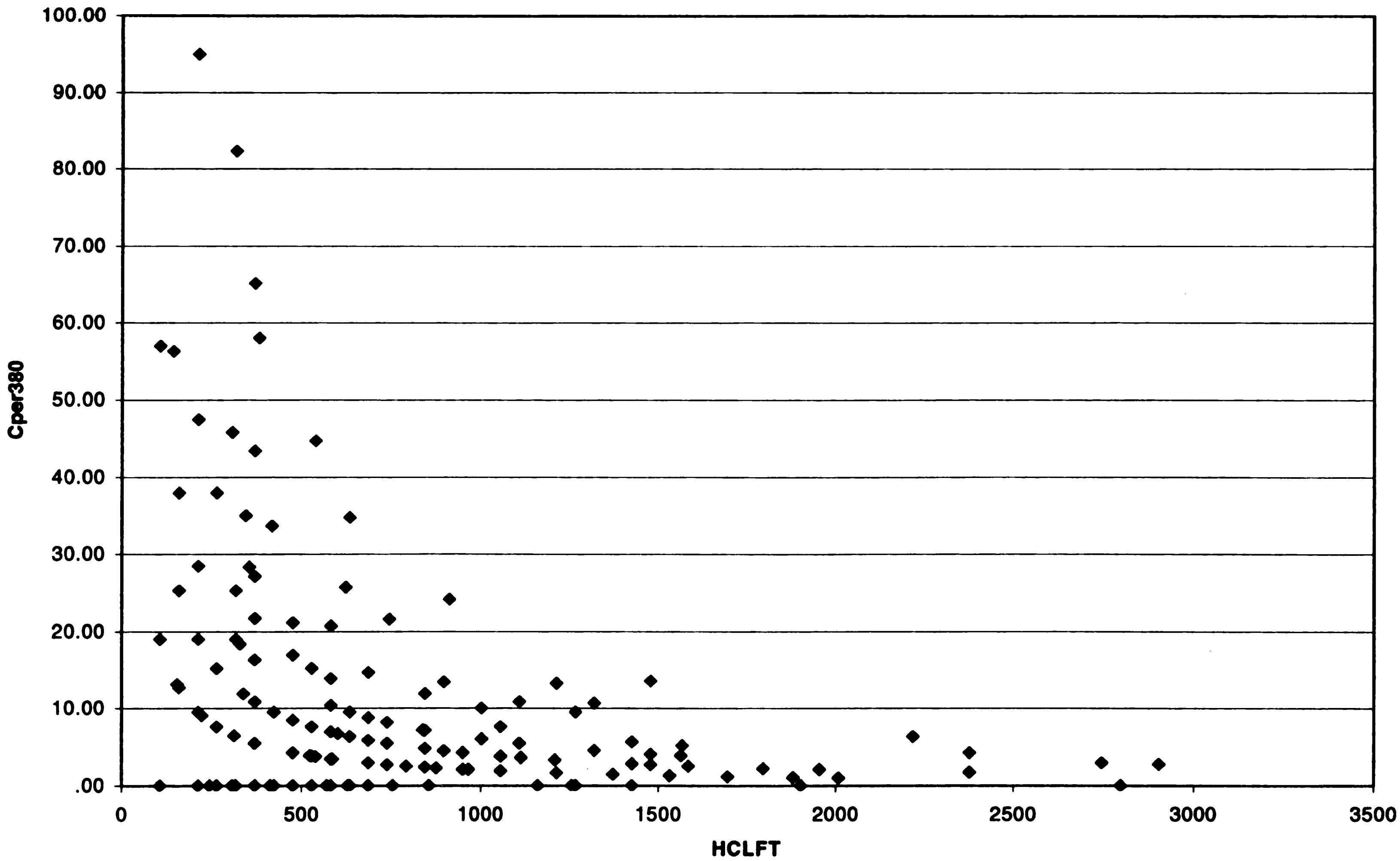


Figure 19 Curve crash rate (Cper380), for various values of curve length in feet (HCLFT)

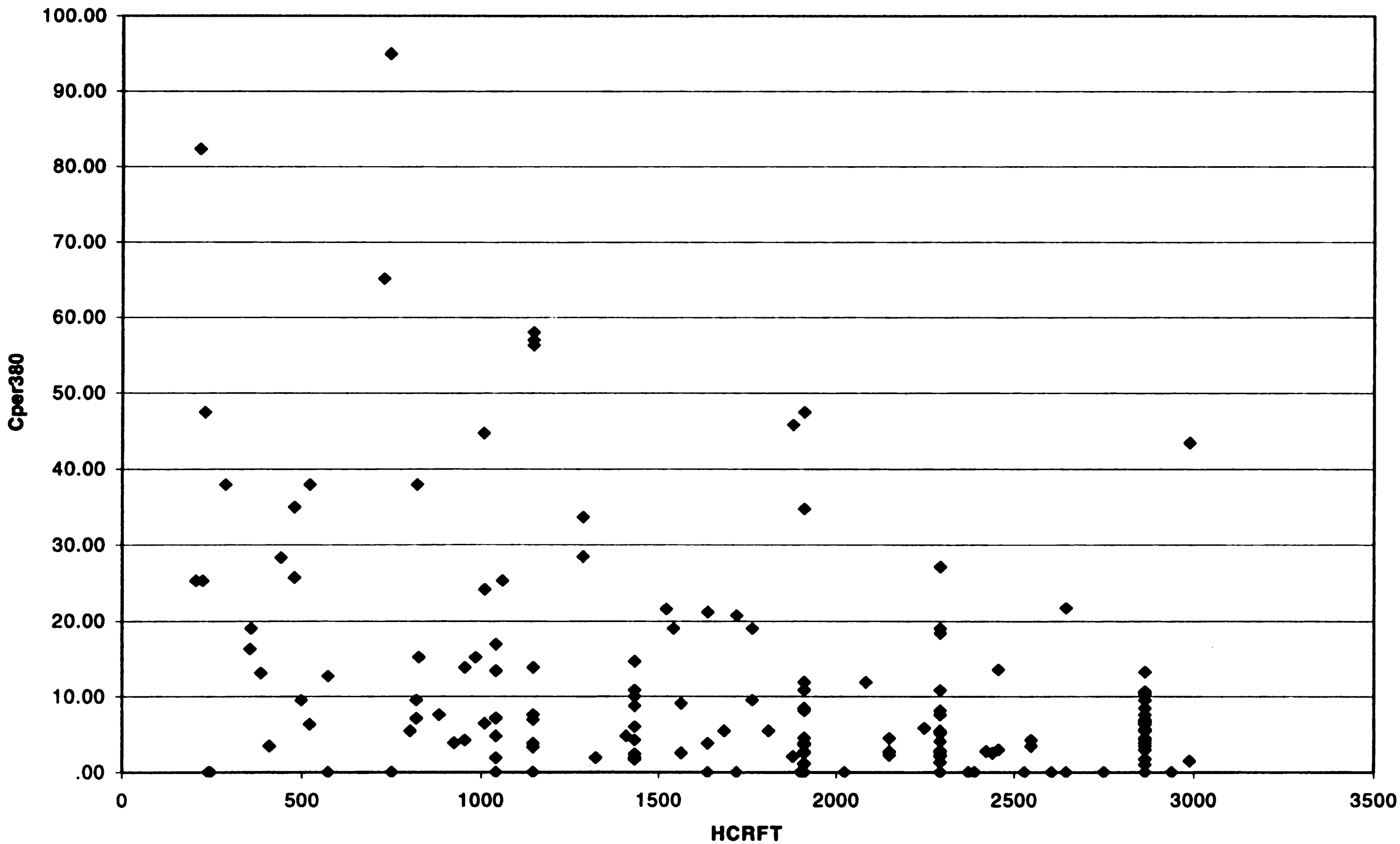


Figure 20 Curve crash rate (Cper380), for various values of curve radius in ft (HCRFT)

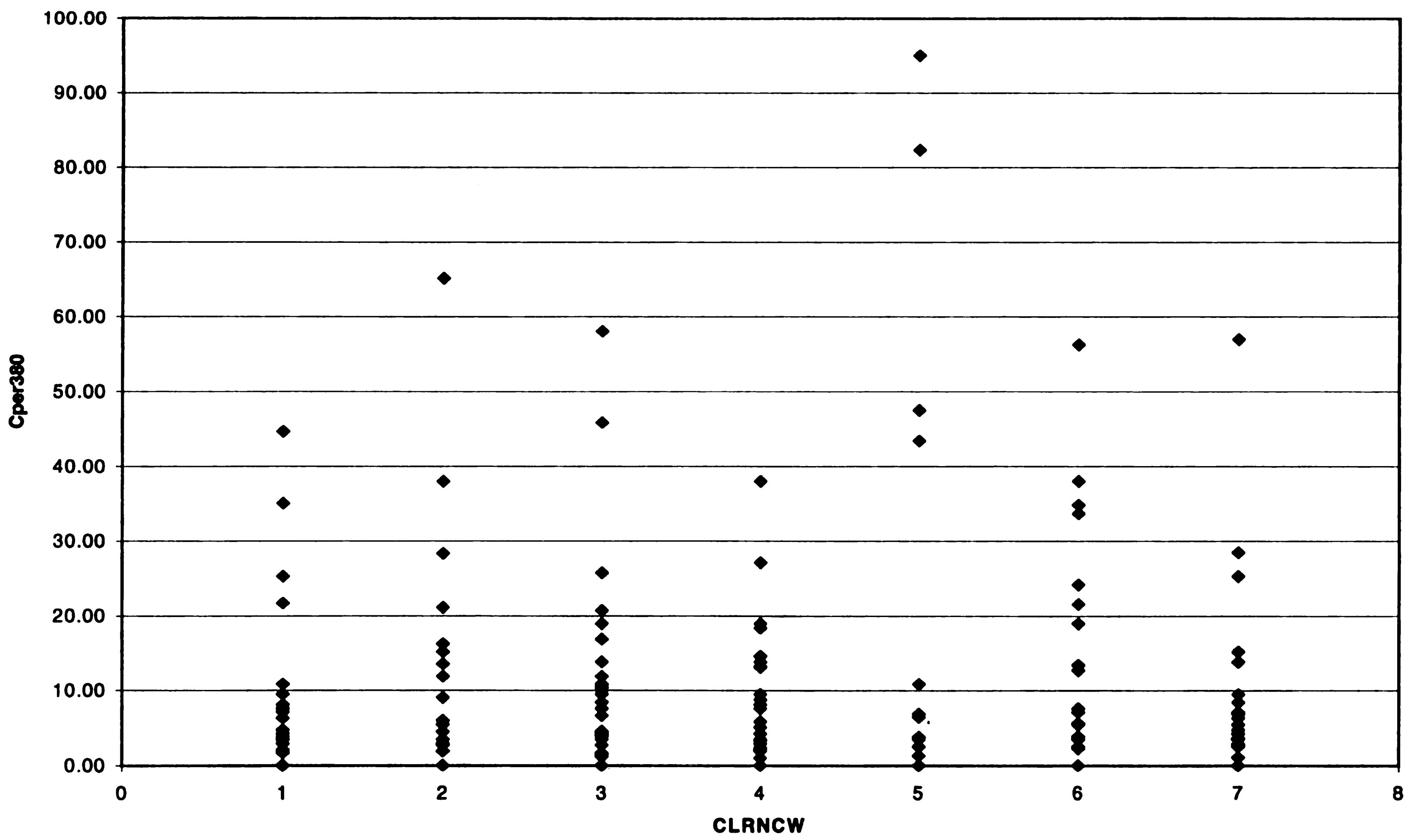


Figure 21 Curve crash rate (Cper380), for various values of roadside clearance (CLRNCW)

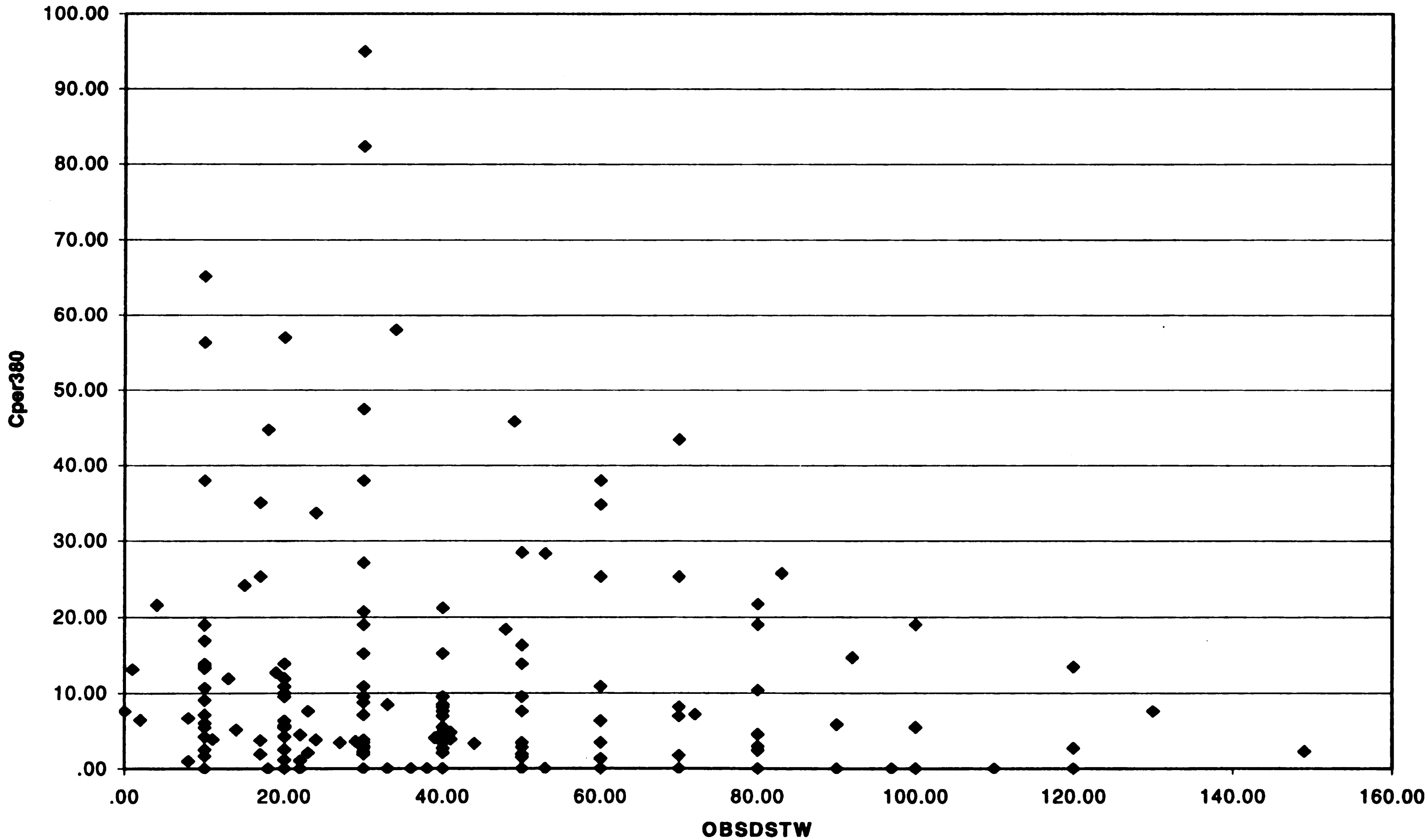
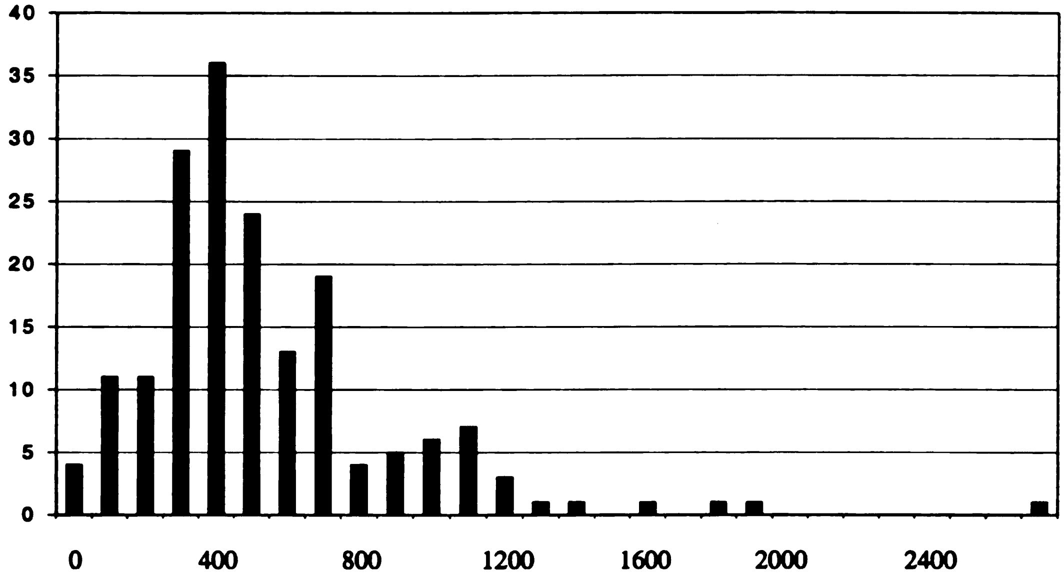
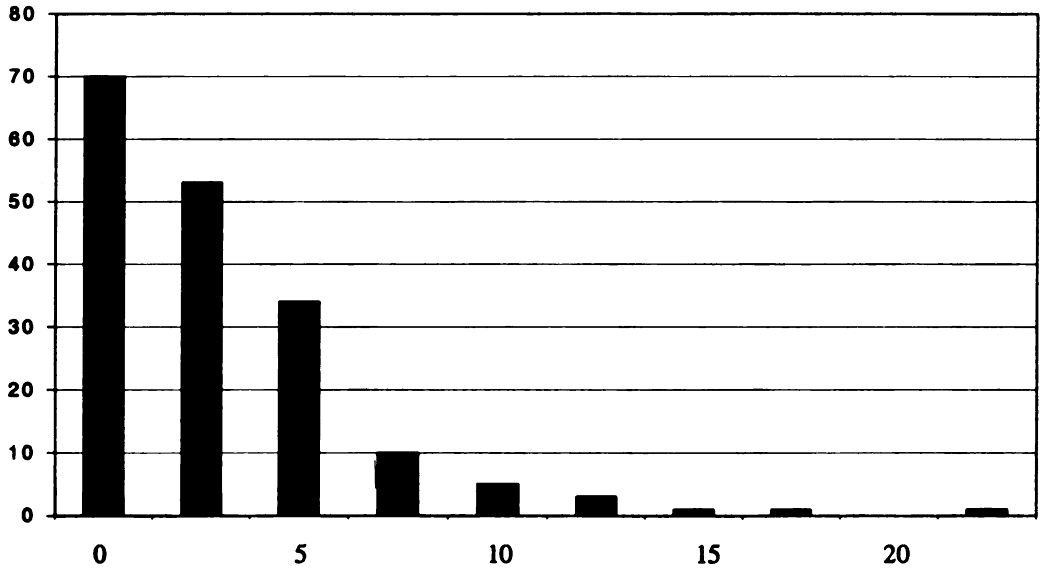


Figure 22 Curve crash rate (Cper380), for various values of sight distance to the beginning of curve (OBSDSTW)

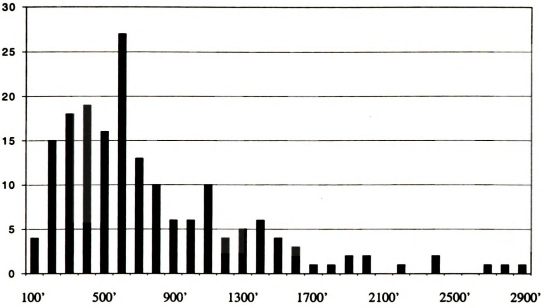


ADT Std. Dev. = 367 Mean = 532

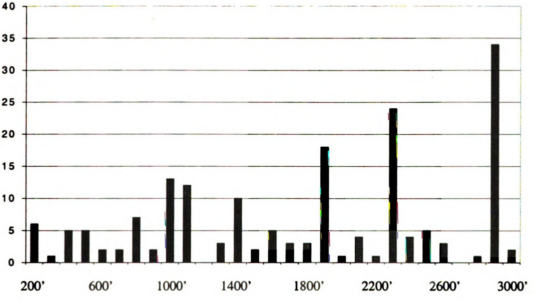


Tper380 Std. Dev. = 3.4 Mean = 3.0

Figure 23 Distribution of selected independent variables
(based on correlation with curve crash rate)

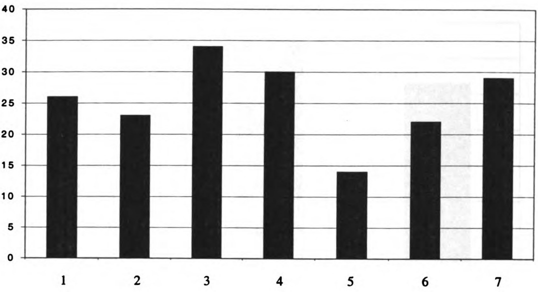


HCLFT Std. Dev. = 545 Mean = 769

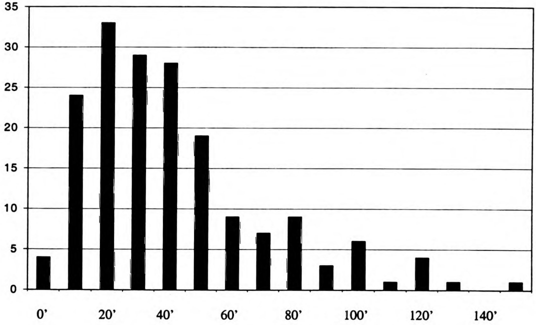


HCRFT Std. Dev. = 824 Mean = 1804

Figure 23 (continued)

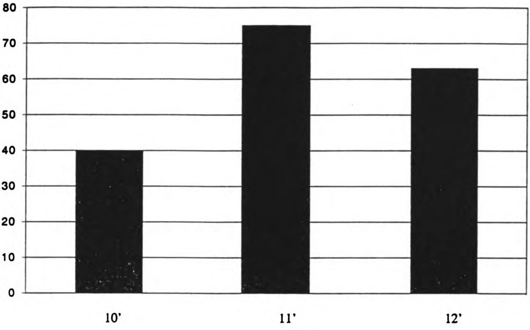


CLRNCW Std. Dev. = 2.0 Mean = 3.9



OBSDSTW Std. Dev. = 29 Mean = 41

Figure 23 (continued)



ALW Std. Dev. = .75 Mean = 11

Figure 23 (continued)

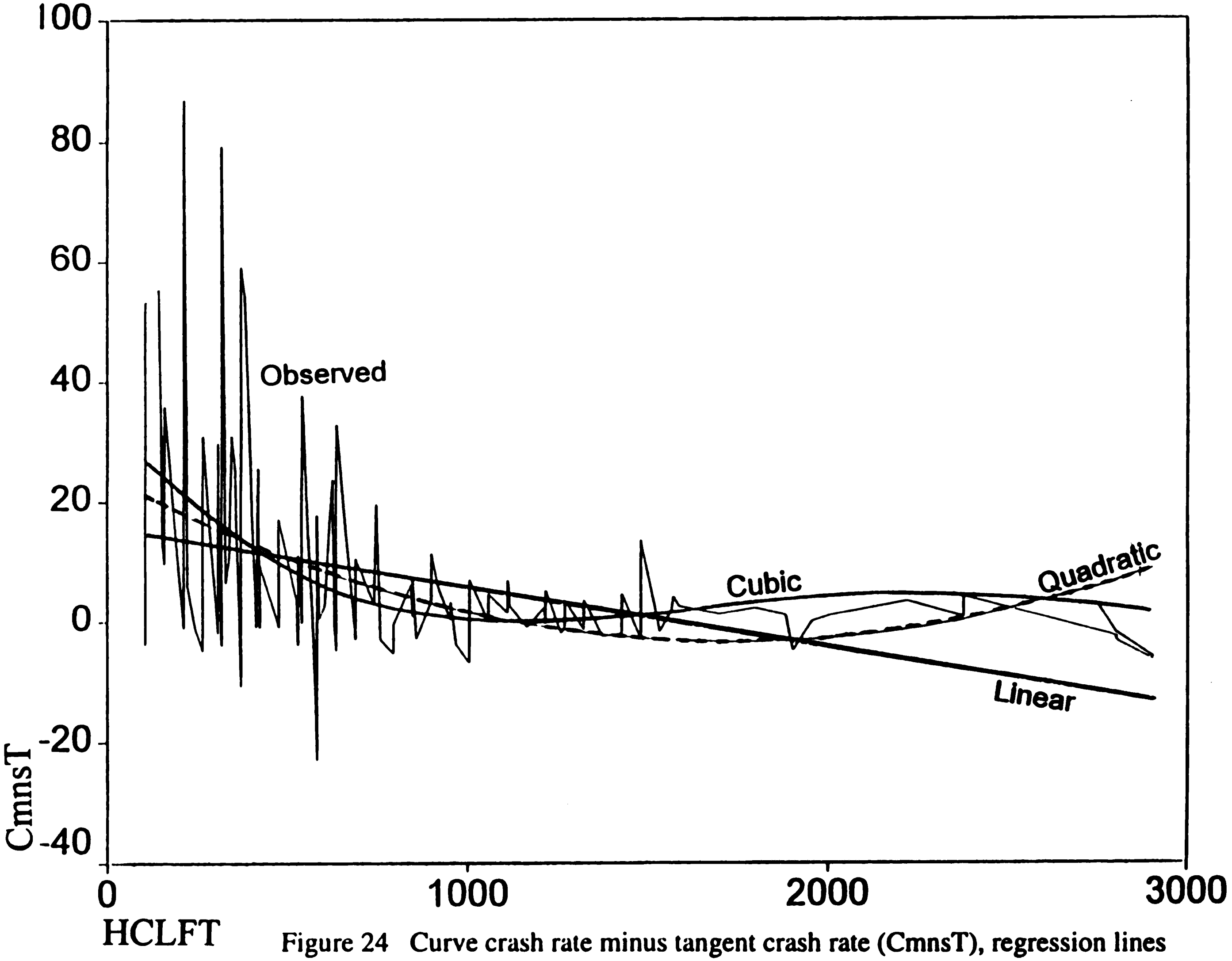


Figure 24 Curve crash rate minus tangent crash rate (CmnsT), regression lines for various values of curve length in feet (HCLFT)

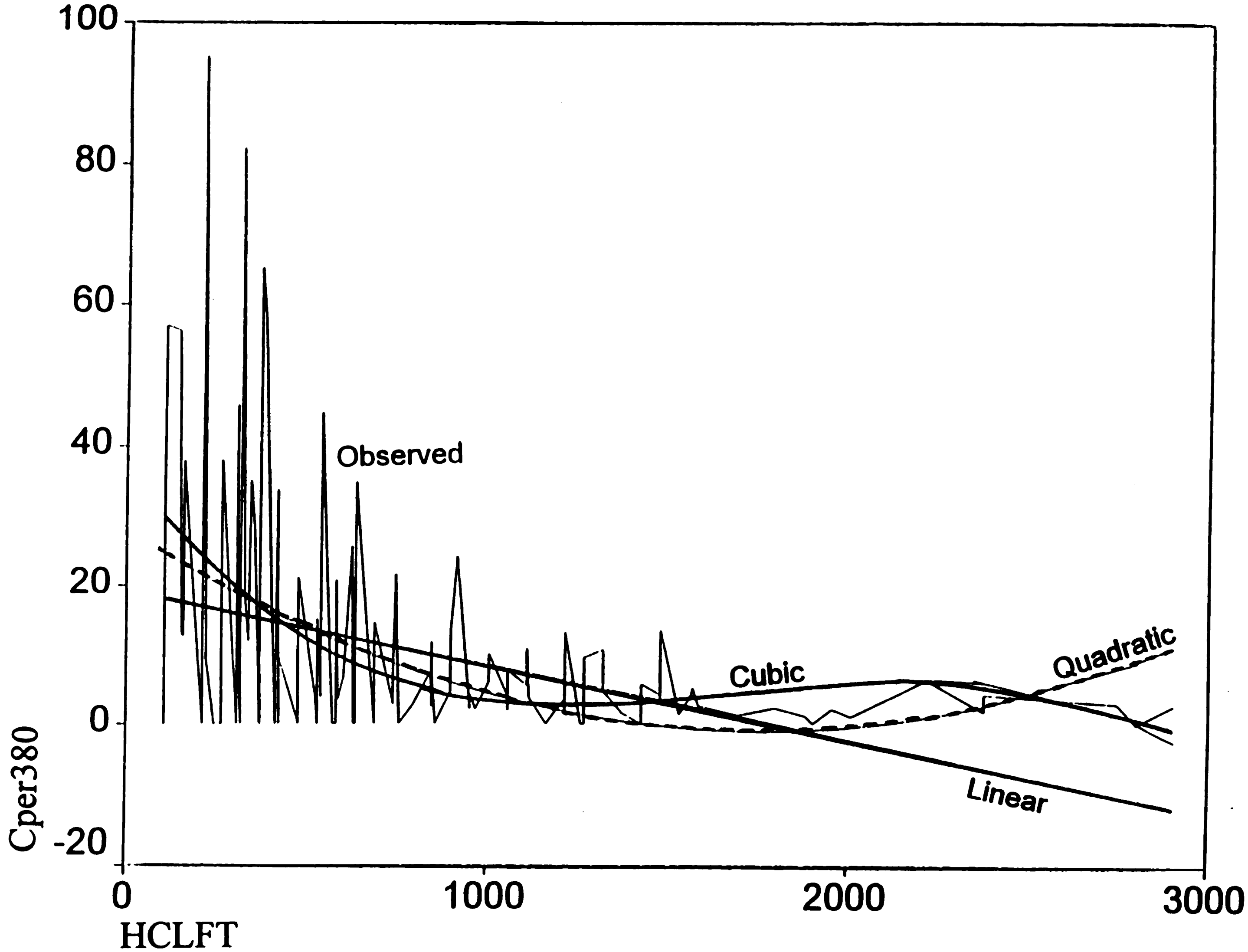


Figure 25 Curve crash rate (Cper380), regression lines for various values of curve length in ft (HCLFT)

Table 8 R^2 of non linear (Ln) models

143 cases with non-zero values for Cper380 (4 variables)			
Depend. Var.		**Independ. Var.**	**Rsq.**
LnCper380	1	LnHCRFT, LnHCLFT & ADT	0.614
143 cases with non-zero values for Cper380 (20 variables)			
Depend. Var.		**Independ. Var.**	**Rsq.**
LnCper380	1	HCLFT	0.385
	2	ADT	0.495
	3	HCRFT	0.546
	4	Tper380	0.566
92 cases with non-zero values for Cper380 and other 9 variables			
Depend. Var.		**Independ. Var.**	**Rsq.**
LnCper380	1	HCLFT	0.381
	2	HCRFT	0.463
	3	ADT	0.542
	4	Tper380	0.566
92 cases with non-zero values for Cper380 and other 9 variables			
Depend. Var.		**Independ. Var.**	**Rsq.**
LnCper380	1	HCLFT	0.381
	2	HCRFT	0.463
	3	ADT	0.542
	4	LnTper380	0.564
92 cases with non-zero values for Cper380 and other 9 variables			
Depend. Var.		**Independ. Var.**	**Rsq.**
LnCper380	1	LnHCLFT	0.486
	2	LnTper380	0.549
	3	LnHCRFT	0.597
	4	LnADT	0.618
92 cases with non-zero values for Cper380 and other 9 variables			
Depend. Var.		**Independ. Var.**	**Rsq.**
Cper380	1	LnHCLFT	0.390
	2	Tper380	0.452
	3	LnHCRFT	0.493
92 cases with non-zero values for Cper380 and other 9 variables			
Depend. Var.		**Independ. Var.**	**Rsq.**
Cper380	1	LnHCLFT	0.390
	2	LnTper380	0.455
	3	LnHCRFT	0.497

Table 9 R^2 of linear and several non linear models

Independent Variable: Cper380

Dep. Var.	Mth	Rsq.	Dep. Var.	Mth	Rsq.	Dep. Var.	Mth	Rsq.
ADT	LIN	0.100	ALW	LIN	0.000	HCLFT	LIN	0.381
ADT	LOG	0.095	ALW	LOG	0.000	HCLFT	LOG	0.377
ADT	INV	0.060	ALW	INV	0.001	HCLFT	INV	0.259
ADT	QUA	0.102	ALW	QUA	0.000	HCLFT	QUA	0.387
ADT	CUB	0.105	ALW	CUB	0.004	HCLFT	CUB	0.387
ADT	COM	0.097	ALW	COM	0.000	HCLFT	COM	0.486
ADT	POW	0.092	ALW	POW	0.000	HCLFT	POW	0.420
ADT	S	0.058	ALW	S	0.001	HCLFT	S	0.243
ADT	GRO	0.097	ALW	GRO	0.000	HCLFT	GRO	0.486
ADT	EXP	0.097	ALW	EXP	0.000	HCLFT	EXP	0.486
ADT	LGS	0.097	ALW	LGS	0.000	HCLFT	LGS	0.486
HCRFT	LIN	0.293	OBSDSTW	LIN	0.004	PSL	LIN	0.010
HCRFT	LOG	0.252	OBSDSTW	LOG	0.006	PSL	LOG	0.012
HCRFT	INV	0.139	OBSDSTW	INV	0.011	PSL	INV	0.011
HCRFT	QUA	0.294	OBSDSTW	QUA	0.004	PSL	QUA	0.010
HCRFT	CUB	0.295	OBSDSTW	CUB	0.033	PSL	CUB	0.022
HCRFT	COM	0.297	OBSDSTW	COM	0.006	PSL	COM	0.009
HCRFT	POW	0.241	OBSDSTW	POW	0.013	PSL	POW	0.011
HCRFT	S	0.122	OBSDSTW	S	0.020	PSL	S	0.010
HCRFT	GRO	0.297	OBSDSTW	GRO	0.006	PSL	GRO	0.009
HCRFT	EXP	0.297	OBSDSTW	EXP	0.006	PSL	EXP	0.009
HCRFT	LGS	0.297	OBSDSTW	LGS	0.006	PSL	LGS	0.009
PSW	LIN	0.001	Tper380	LIN	0.096	TSW	LIN	0.001
PSW	LOG	0.000	Tper380	LOG	0.070	TSW	LOG	0.000
PSW	INV	0.003	Tper380	INV	0.028	TSW	INV	0.002
PSW	QUA	0.017	Tper380	QUA	0.100	TSW	QUA	0.008
PSW	CUB	0.022	Tper380	CUB	0.105	TSW	CUB	0.014
PSW	COM	0.001	Tper380	COM	0.116	TSW	COM	0.000
PSW	POW	0.000	Tper380	POW	0.077	TSW	POW	0.000
PSW	S	0.003	Tper380	S	0.023	TSW	S	0.003
PSW	GRO	0.001	Tper380	GRO	0.116	TSW	GRO	0.000
PSW	EXP	0.001	Tper380	EXP	0.116	TSW	EXP	0.000
PSW	LGS	0.001	Tper380	LGS	0.116	TSW	LGS	0.000

For the categorical variables such as the No Passing Zone Code (NPZC) with values of 0 (zero), 1, 2 or 3, or CLRNCW, with values equal to integers 1 to 7, a separate set of regression analyses were performed. These analyses consisted of replacing, for example, the variable CLRNCW with 7 dichotomous variables, each corresponding to the value of one category of the CLRNCW variable.

For instance the dichotomous variable related to the CLRNCW equal to 5 , were given a value of 1 when the value of CLRNCW was 5, and a value of 0 (zero), when CLRNCW was 1, 2, 3, 4, 6 or 7.

None of the categorical variables displayed significant correlation with Cper380 or CmnsT.

In addition to analyzing all related crashes, crashes occurring under different road surface conditions, weather conditions and lighting conditions were also analyzed.

A sub-set of curves consisting of only those with the field data, (superelevation and drag factor), was analyzed separately. None of these stratification resulted in a significant improvement in the prediction capability of the regression models.

Based on the field measurements of superelevation a new variable was defined and computed . This variable called the design speed, was defined as the speed at which the lateral friction force between the tire and the road surface would equal a value of 0.19 times the normal force of the tire on the road surface. The speed was calculated from the equation: $R=V^2 / 15(e+f)$ where R is the curve radius in feet, V is the design speed in MPH, e is the superelevation and f is assigned a value of 0.19. For each curve two design speeds were computed, the design speed based on the lower value of the superelevation of the two sides of the road was named "DsgnSpdL" and the one for the higher value was named "DsgnSpdH."

The difference between this design speed and the posted advisory speed was calculated and named "DiffSpdL" and "DiffSpdH" corresponding to the lower and higher values of the superelevation as described above. Where an advisory speed was not posted, 55 MPH was used as the posted speed limit.

Figures 26 and 27 show regression plots for Cper380 versus DsgnSpdL and DiffSpdL.

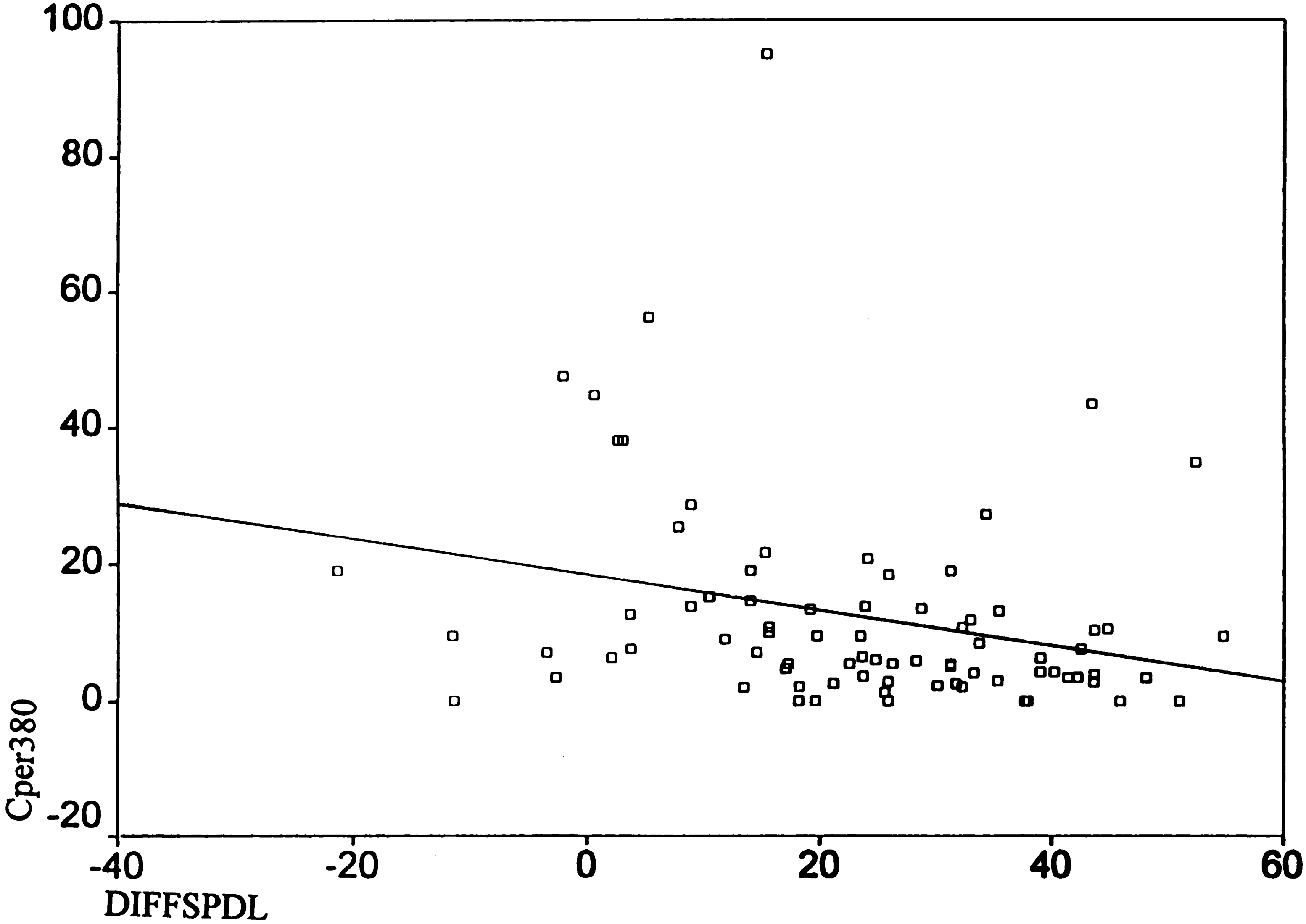


Figure 26 Curve crash rate (Cper380), regression line for various values of speed difference (DIFFSPDL)

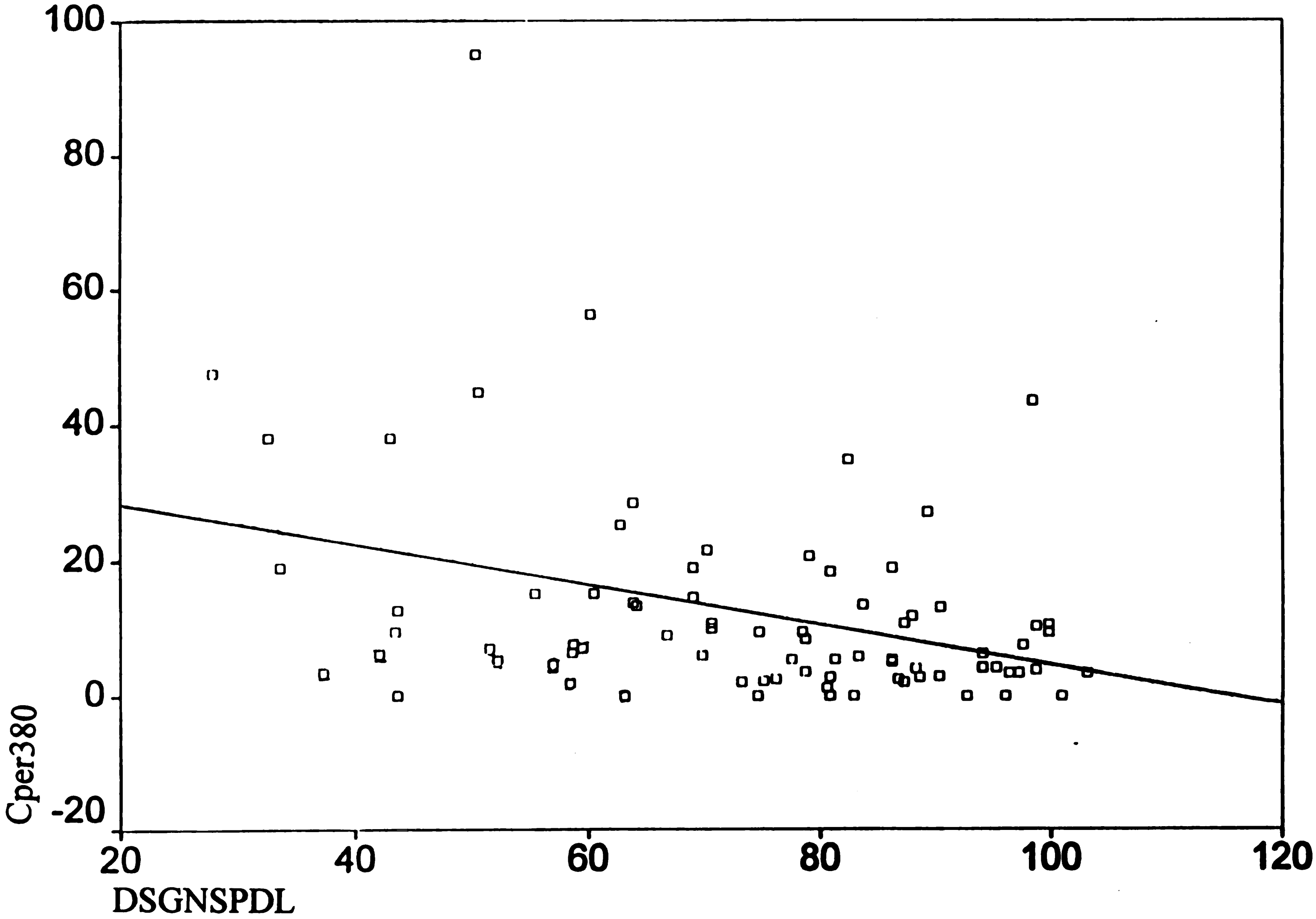


Figure 27 Curve crash rate (Cper380), regression line for various values of design speed difference (DSGNSPDL)

Test of Existing Models

The next step in the analysis was a comparison of the curve crash rate observed in the field versus the values predicted by the Glennon and Zegeer models identified in the literature review. In the Glennon model, the Tper380 was used as a surrogate for Ars, the crash rate on comparable straight roadway segments.
For the Zegeer model the predicted values were obtained for both the with spiral, ZegeerS, and without spiral, ZegeerM, assumptions.

The plots of the predicted values of curve crashes versus actual values of curve crashes, (Cacc), are shown in Figures 28-32. This analysis considered only "related" curve crashes with the model adjusted for the length of the individual curves, not for the 612 meters. While both the Zegeer model and the Glennon model appear to show the correct trend, neither model explains the variation in "related" crash rates observed in the Michigan data. Thus it does not appear that these models are beneficial in identifying curves that should be reviewed for possible safety improvements.

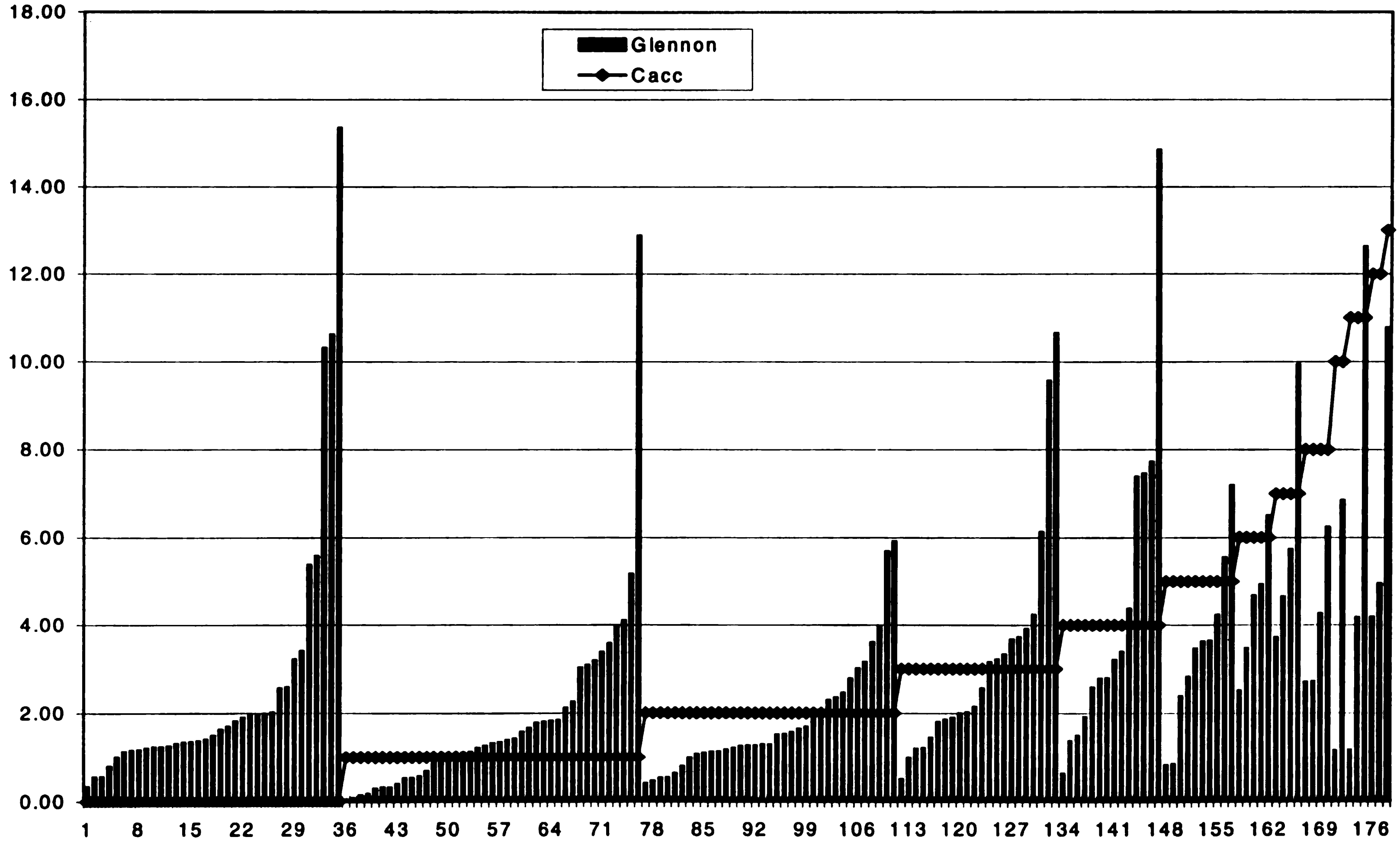


Figure 28 Comparison of the predicted number of curve crashes using Glennon's model (Glennon), and the actual number of curve crashes (Cacc)

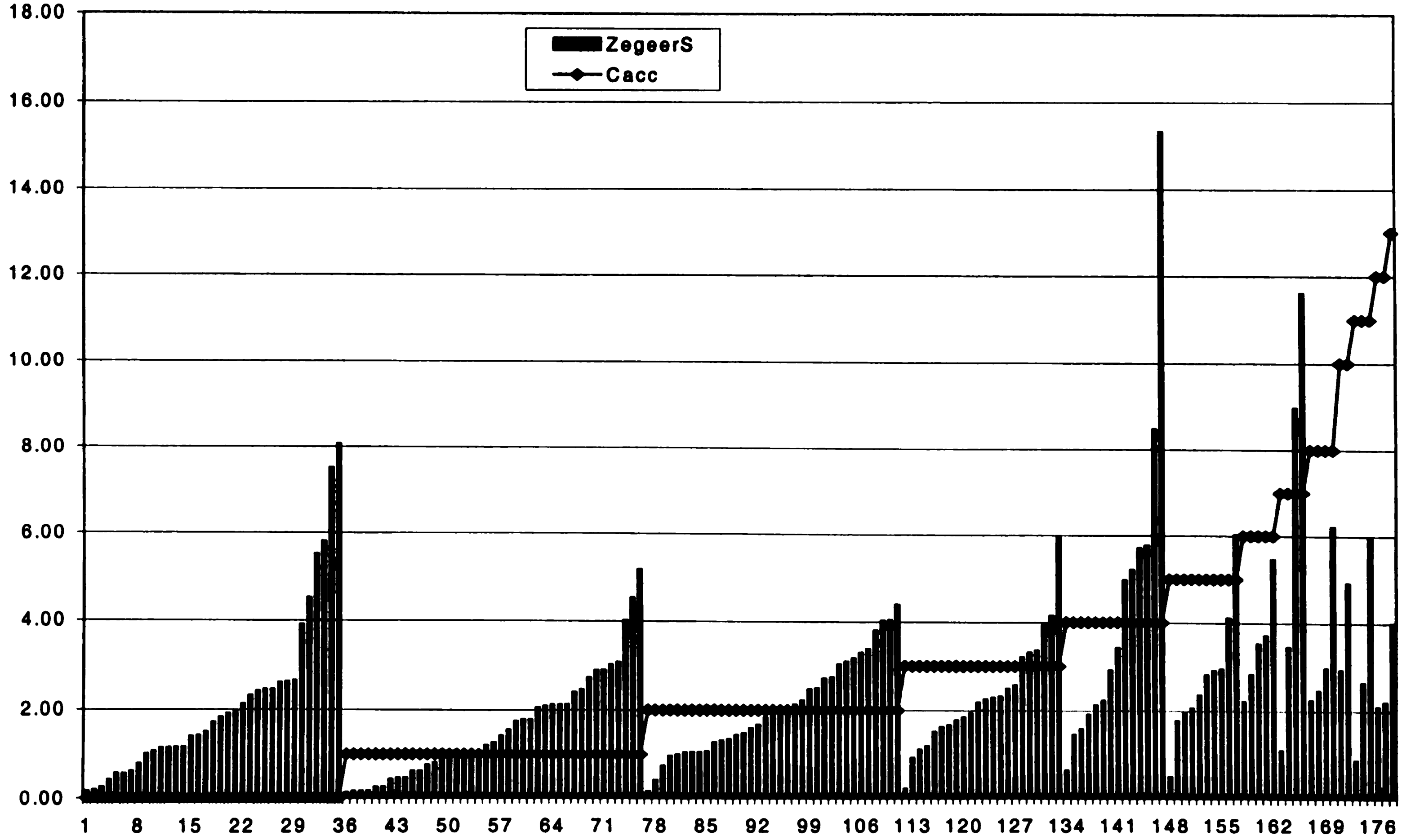


Figure 29 Comparison of the predicted number of curve crashes using Zegeer's model with spiral (ZegeerS), and the actual number of curve crashes (Cacc)

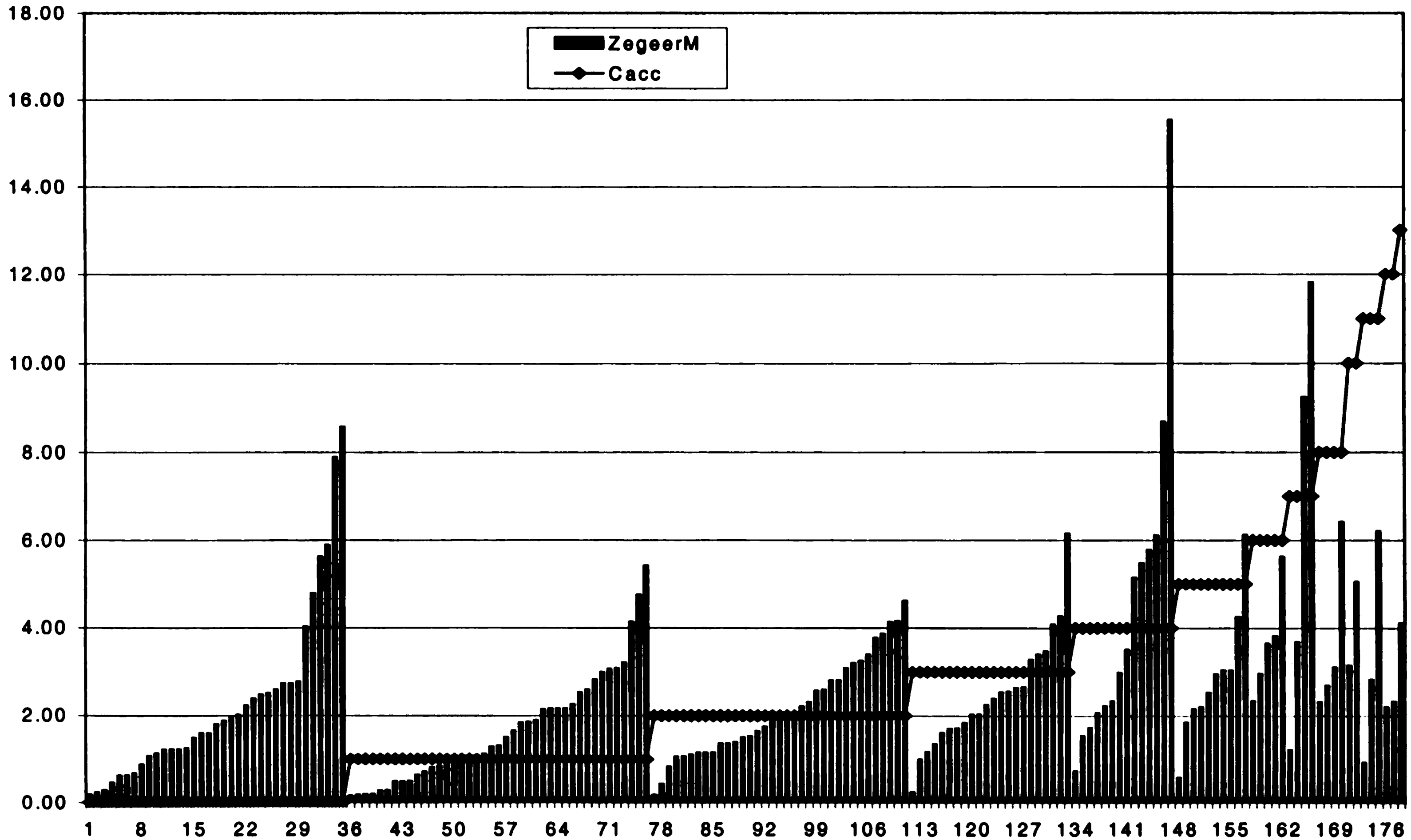


Figure 30 Comparison of the predicted number of curve crashes using Zegeer's model without spiral (ZegeerM), compared with the actual number of curve crashes (Cacc)

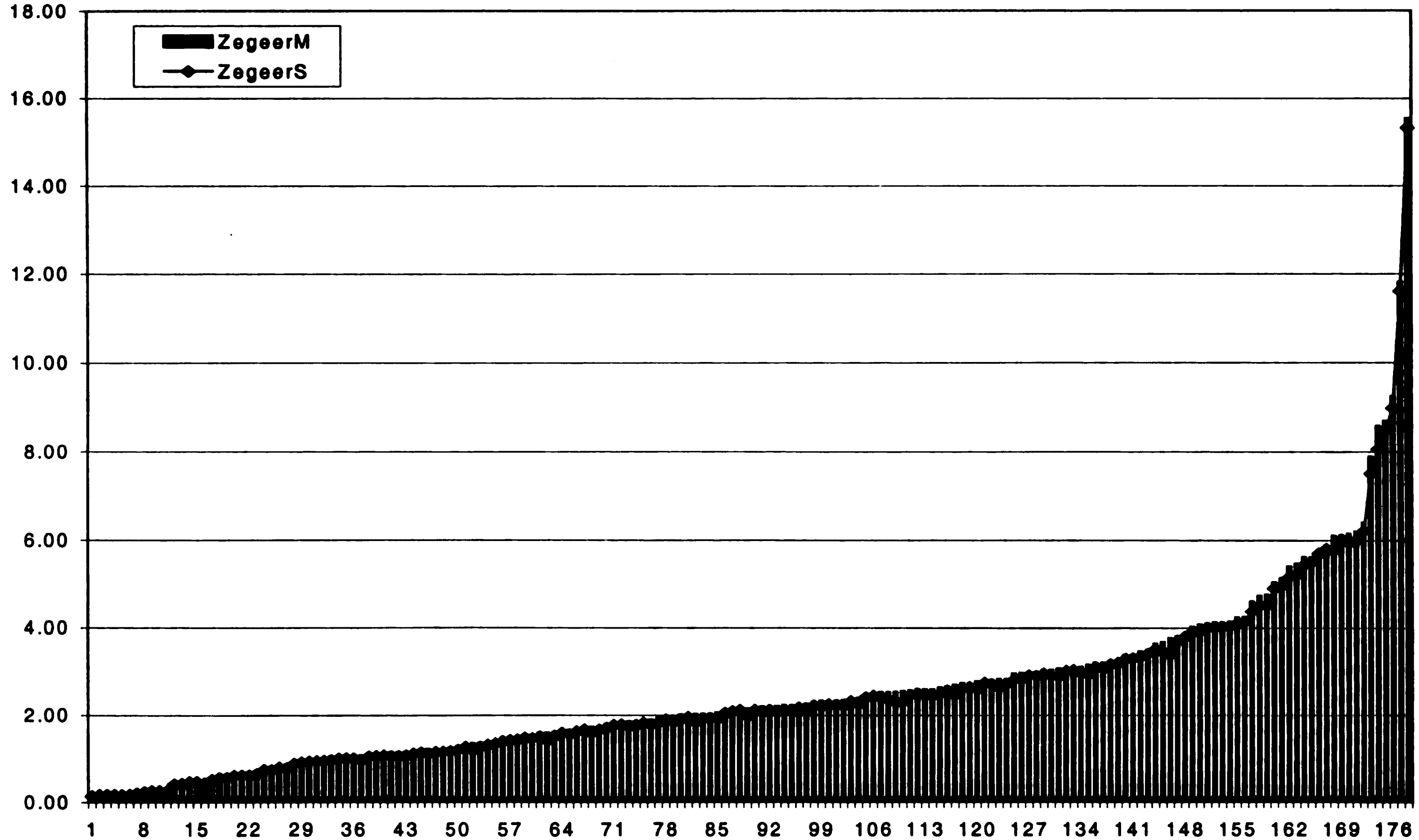


Figure 31 Comparison of the predicted number of curve crashes using Zegeer's model without spiral (ZegeerM), and that of the model with spiral (ZegeerS)

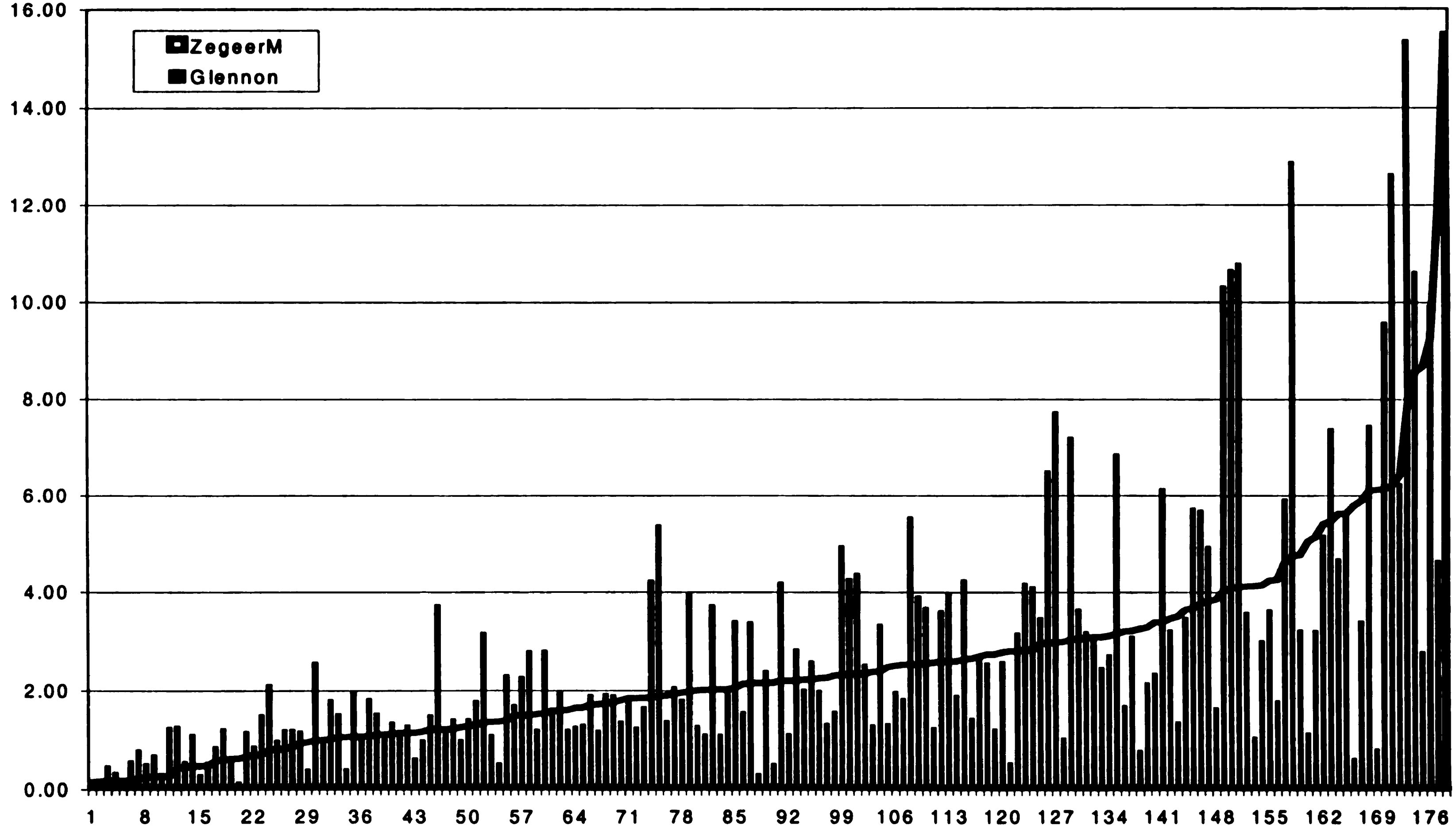


Figure 32 Predicted curve crashes using the Glennon's model (Glennon), arranged in ascending order of predicted curve crashes by Zegeer's model without spiral (ZegeerM)

MULTIVARIATE ANALYSIS

Having determined that the variation in crash frequency found on Michigan curves can not be satisfactorily explained by models based on simple linear or non-linear regression, various multivariate analysis techniques were performed.

Multivariate analysis almost always requires finding minimum or maximum value of a compound of some variables (usually a linear compound).

Factor analysis, for example, maximizes the variance of a linear compound of observed scores of sets of variables, while multiple regression deals with finding a linear compound of some variables such that this compound has the maximum correlation with a particular variable.

In some cases the maximum or minimum value needs to be found using additional constraints. An example would be finding a linear compound from one set of variables and another linear compound from another set of variables, such that the correlation between the two linear compounds is maximized.

Multiple Regression Analysis and Results

Many different multiple regression models were analyzed but with unsatisfactory results. Table 10 shows the results of one such model obtained by using stepwise regression analysis where each of the variables identified in the preceding discussion was available to be entered in the equation. In this model, the combination of variables HCRFT, Tper380, HCLFT and MPHS best explain the "related" curve crashes. These linear multiple regression models obtained by stepwise regression

Table 10 Results of the multiple linear regression analysis for curve crash rate (Cper380)

Regression Equation:

$$Cper380 = 7.35\ MPHS - .00632\ HCLFT + .936\ Tper380 - .00473\ HCLFT + 20.034$$

Model	R	R Square
1	.408	.166
2	.484	.234
3	.519	.269
4	.547	.299

Coefficients[a]

Model 4	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95% Confidence Interval for B Lower Bound	95% Confidence Interval for B Upper Bound
(Constant)	20.034	2.802		7.150	.000	14.504	25.564
HCRFT	-4.73E-03	.001	-.252	-3.328	.001	-.008	-.002
TPER380	.936	.297	.205	3.147	.002	.349	1.523
HCLFT	-6.32E-03	.002	-.223	-3.099	.002	-.010	-.002
MPHS	7.350	2.691	.187	2.731	.007	2.039	12.661

Dependent Variable: CPER380

(forward and backward) produced low coefficients of regression, which is consistent with previous research results.

Discriminant Analysis

Discriminant analysis is a multivariate technique used to distinguish between two or more groups of cases and for studying the overlap between groups, or divergence of one group from the others. Statistically, the objective is to define discriminating functions by weighting and linearly combining the variables such that the groups become associated with variables as distinctly as possible.

The variables with a high contribution toward explaining membership in each group, generally not all the original variables, are considered the predictor variables or the discriminating variables. It is then possible to predict group membership by their association with these discriminating variables.

The discriminant functions can be thought of as the axis of a geometric space in which each group centroid is a point. The weighting coefficients then can be interpreted as the contribution of a variable along the respective dimension of such space.

The only difference between discriminant analysis and multiple correlationanalysis is that in multiple correlation analysis a continuous criterion variable isgiven, whereas in discriminant analysis the criterion is dichotomic if we have two subgroups. In fact, we might distinguish between the two subgroups by introducing a dummy variable that has, say, a value of one for the first subgroup, and of zero for the second. Discriminant analysis can then be reduced to multiple correlation analysis by using this dummy variable as the criterion variable, and by calculating correlations between the observed variables and the dummy (17).

Discriminant analysis was used to determine the variables which distinguish between high and low crash rate curves. The analysis was conducted with the definition of high and low crash rates based on Cper380 and then again with some of the curves removed from the sample as explained in the next section.

Analysis and Results

All of the variables included in the database were used to conduct the first discriminant analysis. For this study, the analysis was used to define membership in one of two groups, either a high crash group or a low crash group.

A value of Cper380 equals 5 resulted in approximately half of the curves belonging to the high crash group and the other half belonging to the low crash group. Thus, this value of Cper380 was selected as the defining value between high and low crash rates.

The results of this analysis are shown in Table 11. The curve length and the curve radius were the two most important discriminating variables followed by ADT. Using only these three variables 71.9% of all cases were correctly classified. There were 26 curves that had a Cper380 value of greater than 5, that were classified in the low category and 24 curves that were misclassified in the other direction.

Similar results were obtained when multiple correlation analysis and a dummy variable was used, as described above. The dummy variable was given a value of one where Cper380 was 5 or greater and zero where it was less than 5.

Since our primary interest is determining whether it is possible to distinguish between high crash locations and low crash locations (rather than some intermediate group), the data set was reduced to eliminate the curves with a value of Cper380 approximately equal to five. A new variable called Modified Cper380 (ModCper)

Table 11 Results of the discriminant analysis for curve crash rate (Cper380)

Variables in the Analysis

Step		Tolerance	Sig. of F to Remove	Wilks= Lambda
3	HCLFT	.866	.001	.827
	HCRFT	.848	.002	.822
	ADT	.978	.002	.821

Classification Results

GRPCLT5			Predicted Group Membership 1.00	Predicted Group Membership 2.00	Total
Original	Count	1.00	64	24	88
		2.00	26	64	90
	%	1.00	72.7	27.3	100.0
		2.00	28.9	71.1	100.0

71.9% of original grouped cases correctly classified.

was defined. This variable is the same as Cper380 but 15 curves with a Cper380 value near the average for all curves were excluded from the analysis.

Table 12 shows the results of the analysis using the modified Cper380 as the grouping variable. Group 2 was defined as curves with ModCper > 7 and group 1 was defined as curves with ModCper < 5.

The independent variables curve sign and turn sign were replaced by a single variable called CTsign since these signs perform the same function and are mutually exclusive. If either sign were present, CTsign was assigned the value of 1 otherwise 0 (zero).

The curve length, the presence of a turn or curve warning sign, the radius of the curve and Tper380 are the discriminating variables identified in this analysis. Using these variables 79.1% of the curves were correctly classified. As expected, removing the cases near the average improved the predictive capability of the model. With this modification, only 16 curves were misplaced as low and 18 curves were misplaced as high.

For the next analysis the difference between the curve crash rate (Cper380) and the tangent crash rate (Tper380) is used as a grouping measure. This variable, (CmnsT), was also modified to more clearly distinguish the curves with high crash rates relative to their tangent crash rates. The cases with a curve crash rate nearly equal to the tangent crashes were eliminated. A total of 43 curves with CmnsT=-1.36 to CmnsT=1.90 were eliminated from the analysis.

As shown in Table 13, the variables curve radius, curve length and the presence of a warning sign are the three most important discriminating variables. For this analysis, 75.6% of the curves were correctly classified using these three variables. Using this model, 90.7% of the high crash rate curves were correctly identified, with only 10 curves being misclassified in this direction. The problem with this

Table 12 Results of the discriminant analysis for modified curve crash rate (ModCper380)

Variables in the Analysis

Step		Tolerance	Sig. of F to Remove	Wilks= Lambda
4	HCLFT	.882	.000	.725
	CTSIGN	.990	.000	.718
	HCRFT	.866	.001	.706
	TPER380	.971	.004	.697

Classification Results

GRPNT5T7			Predicted Group Membership 1.00	2.00	Total
Original	Count	1.00	70	18	88
		2.00	16	59	75
	%	1.00	79.5	20.5	100.0
		2.00	21.3	78.7	100.0

71.9% of original grouped cases correctly classified.

Table 13 Results of the discriminant analysis for modified curve minus tanget crash rate (ModC-T)

Variables in the Analysis

Step		Tolerance	Sig. of F to Remove	Wilks= Lambda
1	HCRFT	1.000	.000	
2	HCRFT	.987	.003	.917
	CTSIGN	.987	.004	.912

Classification Results

			Predicted Group Membership		
		GRPNT5T	1.00	2.00	Total
Original	Count	1.00	5	23	88
		2.00	10	97	75
	%	1.00	17.9	82.1	100.0
		2.00	9.3	90.7	100.0

75.6% of original grouped cases correctly classified.

model is that too many low crash rate curves, (23) were placed in the high crash category.

Despite this problem, Discriminant analysis produced satisfactory results and provided the information useful in meeting the objectives of this study. Specifically the results can be used to identify those characteristics of low crash rate curves which distinguish them from high crash rate curves. However, having reached the above results and conclusions does not mean other mathematical models need not to be tried.

We see that the researcher has to make decision in two stages, so to speak. The first decision stage is to choose a model, i.e., to specify the mathematical assumptions that he takes to be valid. The second stage is to determine whether the data are consistent with a solution of the model.
But if a solution can be accepted, it does not follow that other models would not lead to equally good solutions. .... researchers sometimes tend to reject one model because another is found to work (17).

The choice of mathematical model for the remaining part of this study, arbitrarily and due to practical considerations was set to be limited to the standard statistical procedures among which Factor analysis and Cluster analysis were considered for experimentation.

Cluster Analysis

Cluster Analysis is a systematic technique to look for regularities in a data set. Once the regularities are depicted, this procedure groups the data based on these regularities and their interpretations. Unlike discriminate analysis, which requires prior knowledge of the group membership for the data cases, cluster analysis does not require such knowledge.

Cluster analysis uses the concept of "distance" and "similarity" in generating new clusters or collapsing them into a lesser number of clusters. There are many methods of calculating "distance" and the analyst must use interpretative judgment and inspection in addition to the quantitative analysis.

Cluster analysis was used to identify the variables with a strong association with the crash rate. While any number of clusters can be created, three clusters were used in this study. While it was not predetermined that the crash rate would be included in each cluster, one cluster included curves that have a low crash rate, a second cluster was formed around curves with an intermediate crash rate, and the third around high crash rate curves. This was a useful outcome when interpreting the results.

Analysis and Results

Cluster analysis results proved to be useful for the objectives of this study. Table 14 shows the output for a three cluster case in which Modified Cper380, as discussed previously, was used to identify the curves to be included in the analysis.

The clustering of high, medium and low crash rate curves with other variables is clear, with cluster one having a crash rate of 3.08, cluster two a crash rate of 7.78 while the third cluster has a crash rate of 18.05. Variables such as lane width (ALW), that show little variance between the three clusters indicate that either this variable is unimportant in predicting the curve crashes, or that there is little variance in the variable across all curves. For this variable the latter is true. Other variables, such as curve length and curve radius, show large variations between at least two of the three clusters. This is an indication of an important variable in the prediction model. The important variables are shown in Table 15.

The same variables identified in the discriminant analysis were important in the cluster analysis. The ADT, curve radius, curve length, and the presence of traffic

Table 14
The numerical values of all variables in defining the clusters grouped by the modified curve crash rate (ModCper)

Final Cluster Centers

	Cluster		
	1	2	3
ADT	472.72	536.05	549.14
ALW	11.31	11.19	11.06
ARROW	.21	.09	.29
CHEVRON	.03	.03	.13
CLRNCW	3.69	3.66	4.09
CTSIGN	.34	.44	.56
DLNTR	.31	.19	.27
EDGLN	1.00	.98	1.00
GRAIL	.21	.13	.23
HCLFT	1704	590	520
HCRFT	2471	2383	963
MODCPER	3.08	7.78	18.05
MPHS	.10	.09	.30
NPZC	.90	1.06	1.96
OBSDSTW	45.24	44.27	38.37
PSL	54.66	54.53	53.29
PSW	10.79	6.56	7.03
SCT	1.66	1.53	1.60
TPER380	2.52	3.44	2.98
TSW	19.45	18.72	18.56

Number of Cases in each Cluster

Cluster	1	29.000
	2	64.000
	3	70.000
Valid		163.000
Missing		15.000

Table 15
The numerical values of the important variables in defining the clusters grouped by the modified curve crash rate (ModCper)

Final Cluster Centers

	Cluster		
	1	2	3
ADT	472.72	536.05	549.14
ALW			
ARROW	.21	.09	.29
CHEVRON	.03	.03	.13
CLRNCW			
CTSIGN			
DLNTR			
EDGLN			
GRAIL			
HCLFT	1704	590	520
HCRFT	2471	2383	963
MODCPER	3.08	7.78	18.05
MPHS			
NPZC			
OBSDSTW			
PSL			
PSW			
SCT			
TPER380			
TSW			

Number of Cases in each Cluster

Cluster	1	29.000
	2	64.000
	3	70.000
Valid		163.000
Missing		15.000

control devices (arrow and chevron)are all important in defining the clusters. Interestingly, the high crash rate curves are associated with the highest probability of having chevrons and target arrows deployed. However, this is explained by the fact that this cluster contains the short radius curves, where these devices tend to be deployed.

An analysis using Cper380 instead of ModCper shows similar results (Table 16). Most notably, the clustering of high crash rates with short curves and low radii while the low crash rate curves are clustered with long curves with large radii. This finding is consistent with prior research. Using this measure of the crash rate, ADT was replaced by the presence of an advisory speed plate and the paved shoulder width as explanitory variables. Perhaps the most interesting cluster is the third one, which clusters moderately high crash rate curves with curves of large radius but short length. These tend to not have traffic control devices deployed because of their large radius and subsequently their high design speed.

Tables 17 and 18 show two more cluster analysis results. These results are also consistent with the previous findings. In Table 17 the difference between the curve crash rate and the tangent crash rate, (CmnsT), is used as the crash rate variable, while in Table 18 the variable, ModCmnsT, as described before, is used.

It was hypothesized that the variation in crash rates within each cluster could be explained better by regression analysis than was possible for all curves combined To test this hypothesis, simple and multiple regression analyses were conducted on each of the three clusters obtained from the cluster analysis. However, the correlation coefficients within each cluster were still very low. (As examples Figures 33 through 38 show the regression plots of Cper380 with HCRFT and HCLFT for each of the referenced three clusters).

Table 16
The numerical values of the important variables in defining the clusters grouped by the curve crash rate (Cper380)

Final Cluster Centers

	Cluster		
	1	2	3
ADT			
ALW			
ARROW	.19	.30	.10
CHEVRON	.03	.14	.03
CLRNCW			
CPER380	3.33	17.10	7.62
CURVES			
DLNTR			
EDGLN			
GRAIL			
HCLFT	1707	522	608
HCRFT	2490	974	2392
MPHS	.09	.32	.10
OBSDST			
PSL			
PSW	11.09	7.26	6.53
SCT			
TPER380			
TSW			
TURNS			

Number of Cases in each Cluster

Cluster	1	32.000
	2	76.000
	3	70.000
Valid		178.000
Missing		.000

Table 17
The numerical values of the important variables in defining the clusters grouped by the curve minus tangent crash rate (C-T)

Final Cluster Centers

	Cluster		
	1	2	3
ADT			
ALW			
ARROW	.10	.30	.19
CHEVRON	.03	.14	.03
CLRNCW			
CMNST	4.32	14.05	.88
CTSIGN			
DLNTR			
EDGLN			
GRAIL			
HCLFT	608	522	1707
HCRFT	2392	974	2490
MPHS	.10	.32	.09
OBSDSTW			
PSL			
PSW	6.53	7.26	11.09
SCT			
TSW			

Table 18
The numerical values of the important variables in defining the clusters grouped by the modified curve minus tangent crash rate (ModC-T)

Final Cluster Centers

	Cluster		
	1	2	3
ADT			
ALW			
ARROW	.11	.33	.25
CHEVRON	.04	.18	
CLRNCW			
CTSIGN			
DLNTR			
EDGLN			
GRAIL	.07	.25	.20
HCLFT	607	471	1757
HCRFT	2351	902	2481
MODCMNST	5.59	17.67	1.40
MPHS	.09	.37	.15
PSL			
PSW	6.78	7.15	12.90
SCT			
TSW			
OBSDSTW			

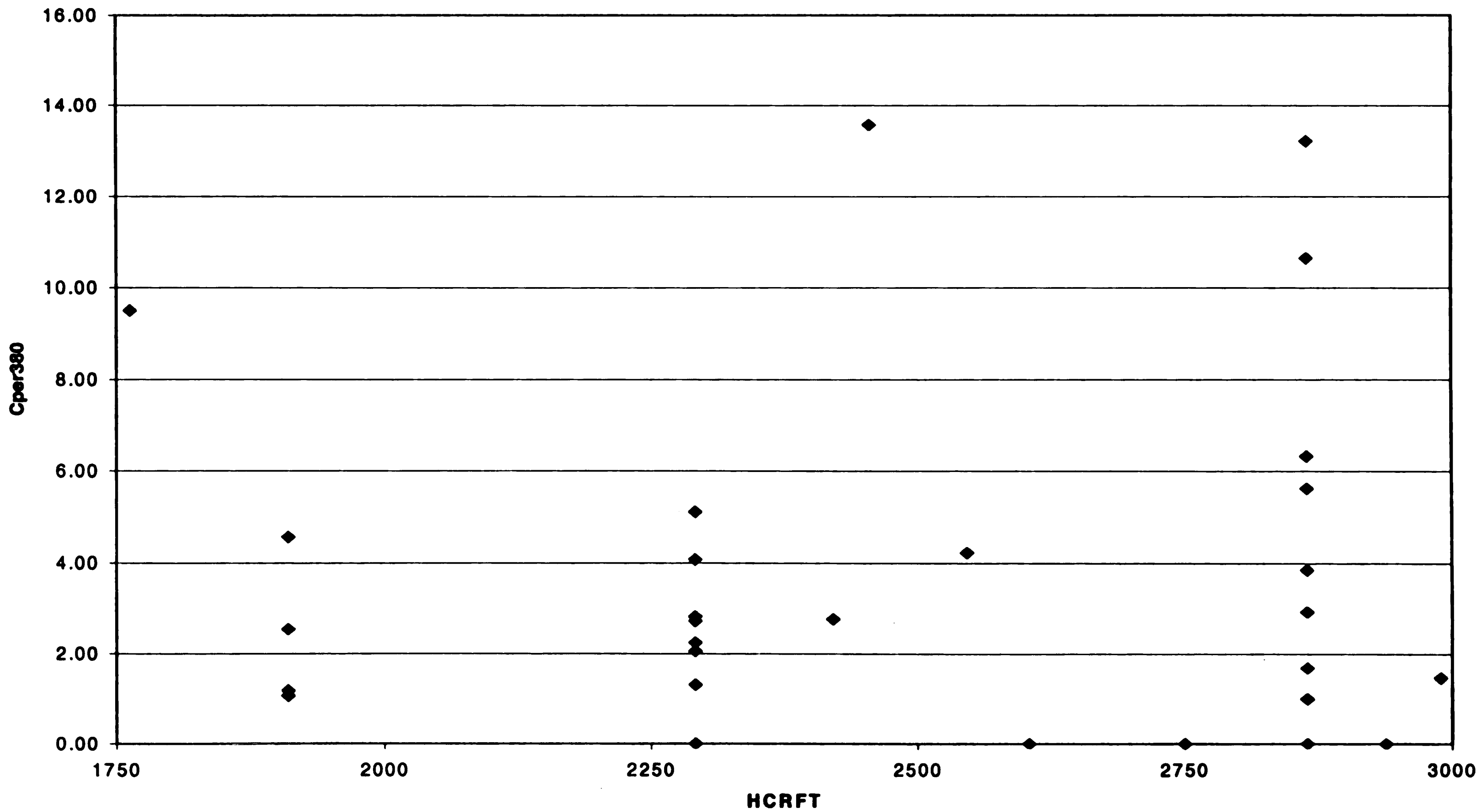


Figure 33 Curve crash rate (Cper380), for various values of curve radius in feet, cluster 1

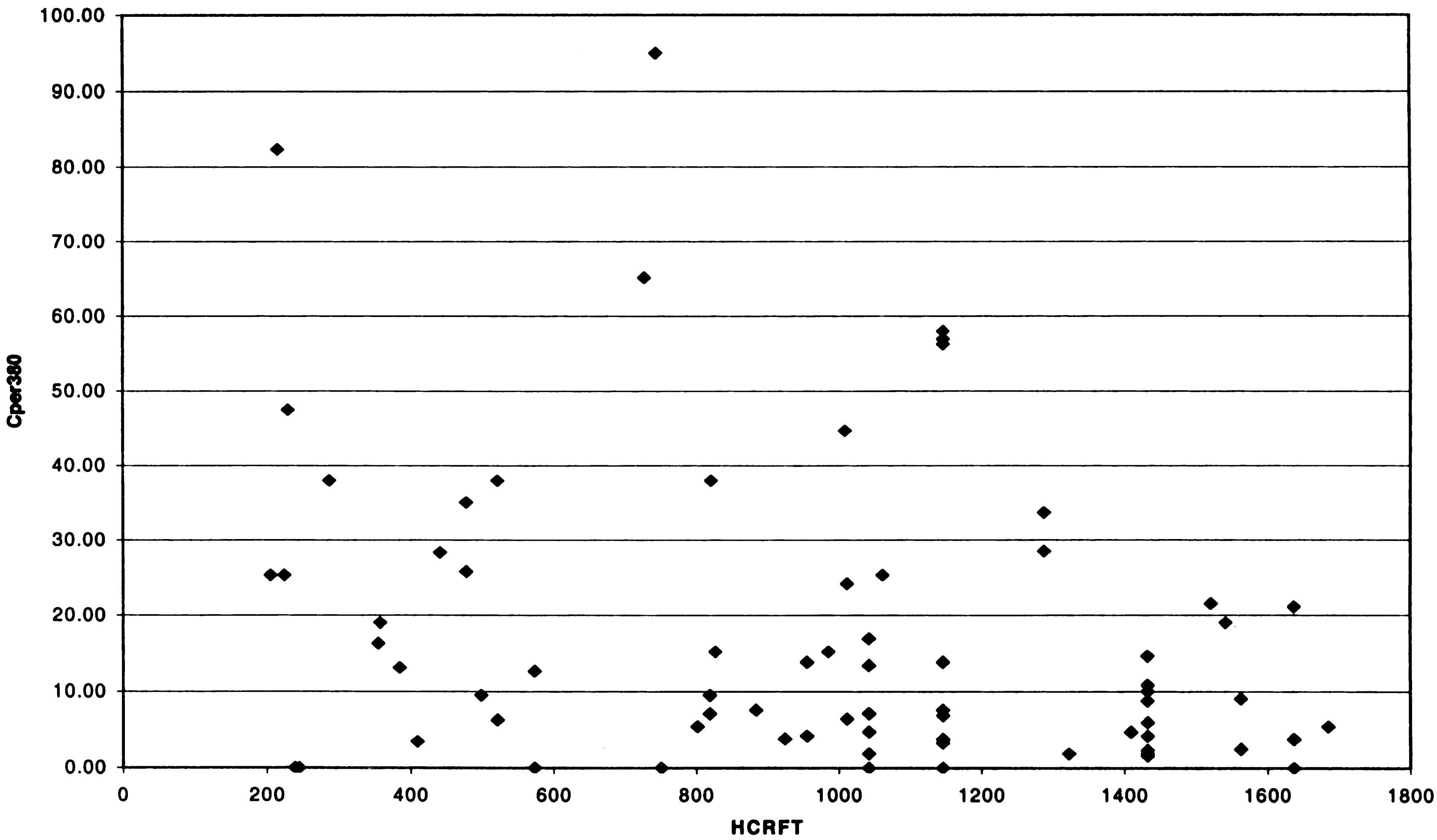


Figure 34 Curve crash rate (Cper380), for various values of curve radius in feet, cluster 2

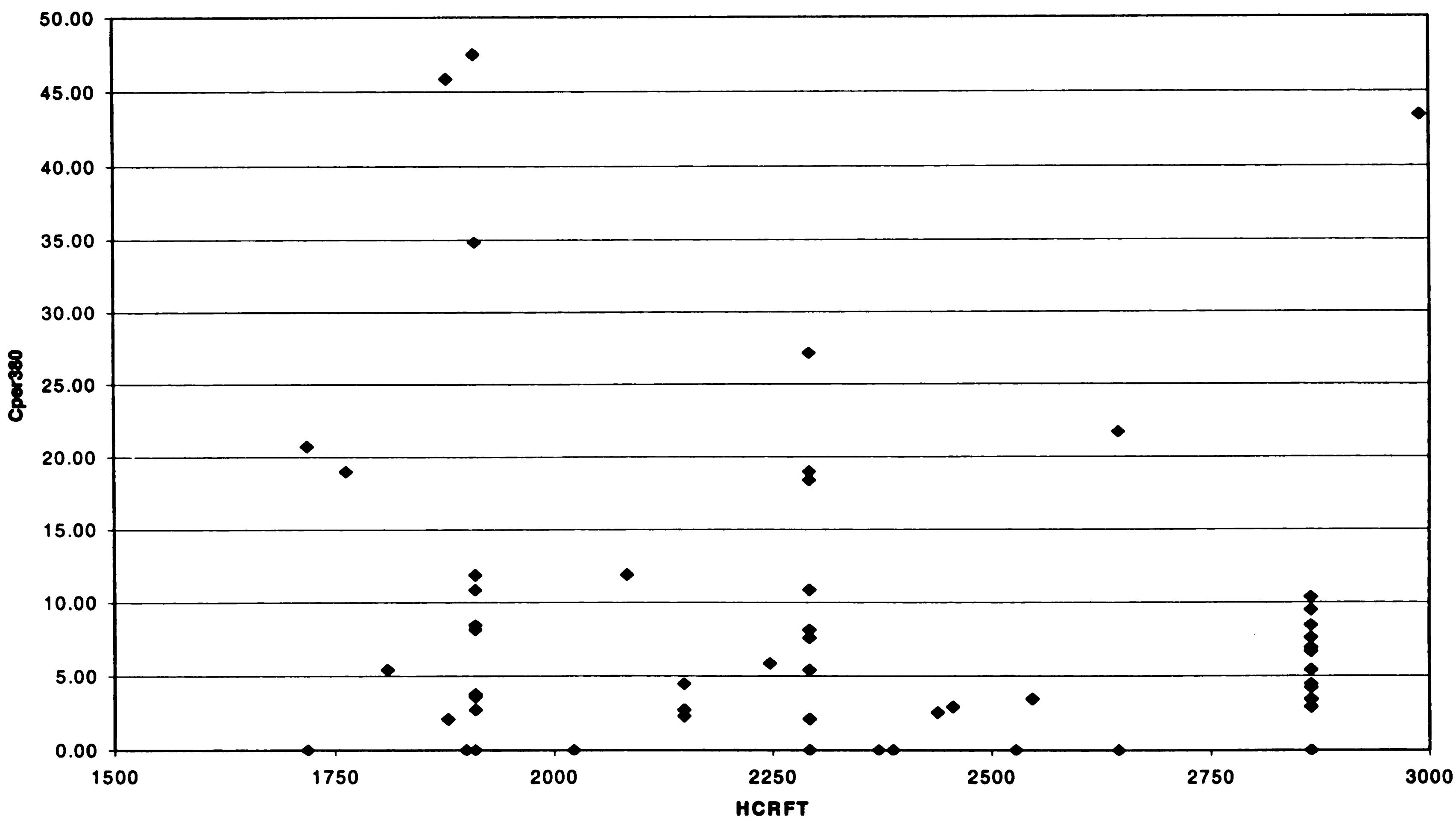


Figure 35 Curve crash rate (Cper380), for various values of curve radius in feet, cluster 3

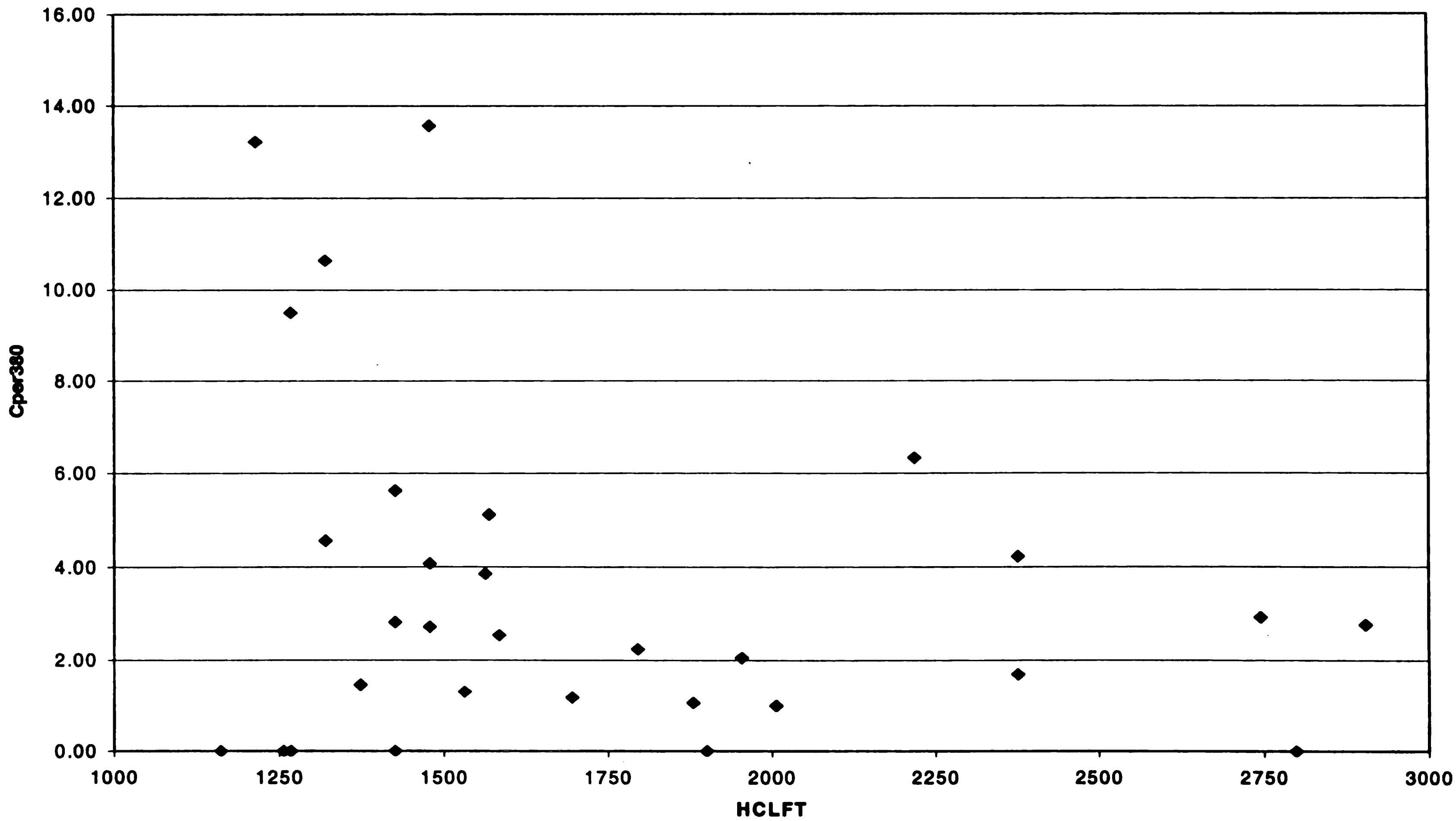


Figure 36 Curve crash rate (Cper380), for various values of curve length in feet, cluster 1

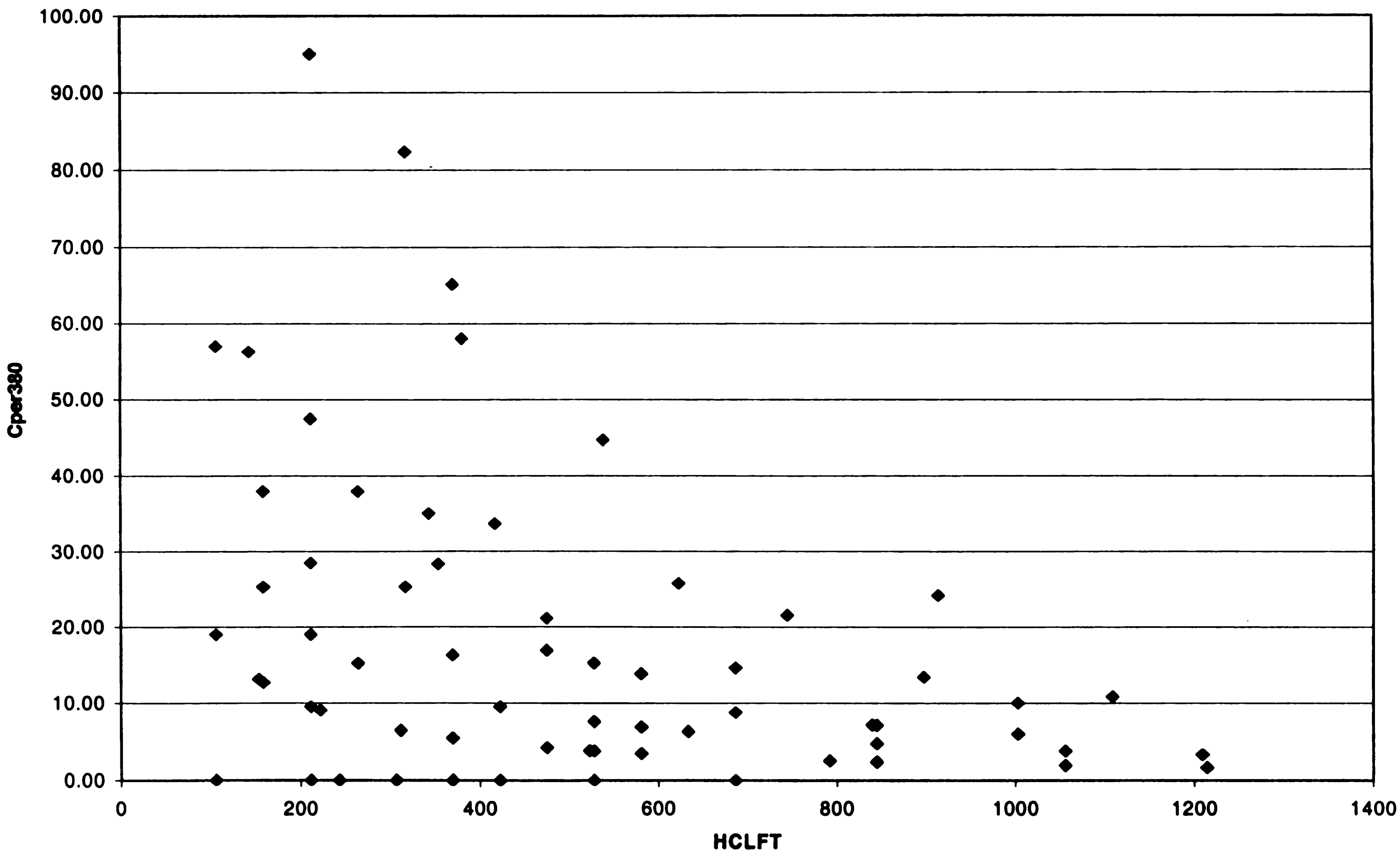


Figure 37 Curve crash rate (Cper380),for various values of curve length in feet, cluster 2

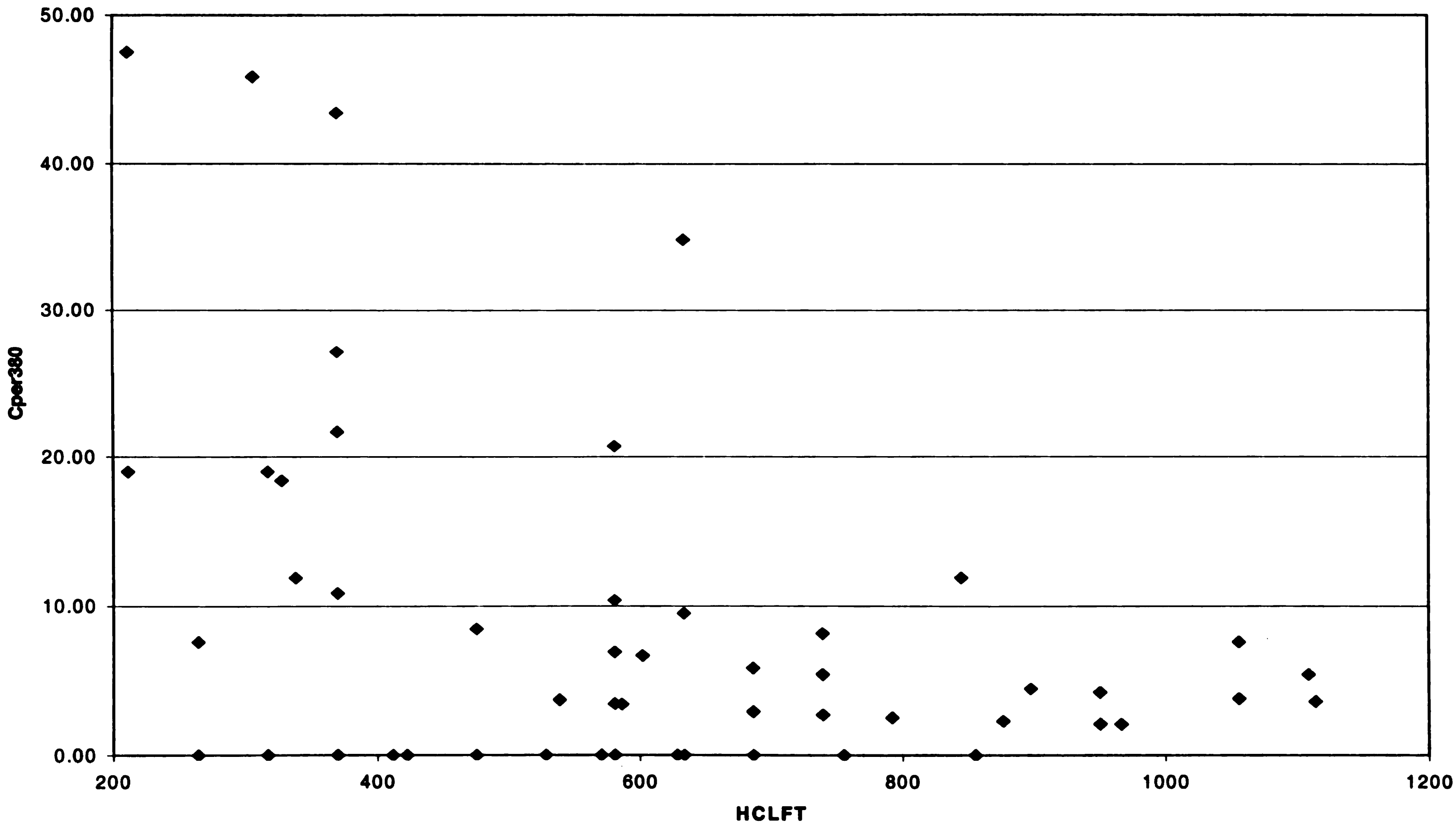


Figure 38 Curve crash rate (Cper380), foe various values of curve length in feet, cluster 3

Factor Analysis and Results

Factor analysis is a technique used to reduce many variables into a smaller set of factors. Each factor describes a "concept." Ideally the concept will be readily understood by individuals and there may even be an existing name for the concept. If not, the analyst can often understand the concept and give it an appropriate name.

Factor analysis starts with a set of variables, or better stated, the scores related to a set of variables. Next, a set of new variables is constructed based on the interrelations exhibited in the data. The first factor is defined as the best linear combination of variables explaining the variance in the data as a whole. The other factors are similarly defined as the best linear combination of variables which explains the variance remaining in the data as a whole. The first factor is thus more important than the second one and so on. The first few factors usually explain most of the variance in the data.

Factor analysis was conducted for many cases of differing variables, factoring criteria, rotation method and number of extracted factors. However, the use of this technique did not identify any relationships among the variables and crash rates that was not also identified by using discriminant analysis.

Table 19 shows the results of one factor analysis with the first three factors extracted. The variables that contribute the most to the factor score coefficients for these three factors are shown in Table 20. Only one of the three factors includes the crash rate (Cper380).

Factor 1 includes Cper380 and the presence of certain traffic control devices (chevron and advisory speed panels), curve length, radius, and roadside clearance

Table 19
Factor score coefficient matrix for all the variables

Factor Score Coefficient Matrix

	Factor		
	1	2	3
ADT	.014	.507	.277
ALW	-.032	.066	-.089
ARROW	.033	-.109	.053
CHEVRON	.128	.032	.009
CLRNCW	-.176	-.289	.611
CPER380	.321	.018	-.019
CURVE	.018	-.004	.049
DLNTR	.012	-.022	.006
EDGLN	.013	.018	.012
GRAIL	-.012	-.013	.101
HCLFT	-.124	.041	.024
HCRFT	-.347	.105	-.038
MPHS	.218	.039	.071
OBSDST	-.004	.006	.006
PSL	.002	-.033	-.018
PSW	-.029	.097	-.031
SCT	.013	.181	-.024
TPER380	.035	.142	.117
TSW	-.004	.124	-.044
TURNS	.103	-.062	-.063

Factor Score Covariance Matrix

Factor	1	2	3
1	.736	1.850E-03	4.219E-02
2	1.850E-03	.766	4.699E-02
3	4.219E-02	4.699E-02	.768

Table 20
Factor score coefficient matrix of relatively high values

Factor Score Coefficient Matrix

	Factor		
	1	2	3
ADT		.507	.277
ALW			
ARROW			
CHEVRON	.128		
CLRNCW	-.176	-.289	.611
CPER380	.321		
CURVE			
DLNTR			
EDGLN			
GRAIL			
HCLFT	-.124		
HCRFT	-.347		
MPHS	.218		
OBSDST			
PSL			
PSW			
SCT			
TPER380			
TSW			
TURNS			

(inversely). All of these variables, with the exception of the roadside clearance variable were also included in the discriminant analysis and cluster analysis results.

Factor 2 describes curves with high ADT and a safe roadside, while Factor 3 describes curves with more hazardous roadside conditions and a lower ADT. This can be interpreted to indicate that the high volume State Trunkline roads have a safer roadside than do those trunkline highways with lower volumes. However, nothing is revealed about the difference in crash rates between these two combinations of variables, nor does this information help to identify specific curves that contribute disproportionally to the crash rate.

Analyses Including Field Data

Since the superelevation of a curve and the coefficient of friction are both factors considered in selecting an advisory speed for curves, a separate set of analysis was performed using only the subset of curves for which these data were collected. A total of 81 roadway segments containing 531 crashes, (279 in tangents and 252 in curves), were included in these analyses. Seventy one of the 81 roadway segments had crashes on their curved section. The values of superelevation and drag factor for the 81 roadway segments are shown in Figures 39-41.

Analyses similar to those performed previously for all the roadway segments, were conducted for only the roadway segments with the field data. The analyses were conducted with the addition of the two field data variables, superelevation and drag factor, for each direction of traffic individually and combined. The analysis was done twice, once for the higher value of the superelevation for the two directions, SPRELVN and again for their lower value, SELELO.

Figures 42 and 43 show graphs sorted by ascending value of Cper380 for those curves with the field data. Figures 44-49 show the simple linear regression results of Cper380 and CmnsT with these variables.

Neither the drag factor nor the superelevation, individually or in combination, were significant in explaining the curve crash rate or improving the results of the discriminant analysis or cluster analysis. In fact, neither of these two variables entered the results of these tests as predicter variables.

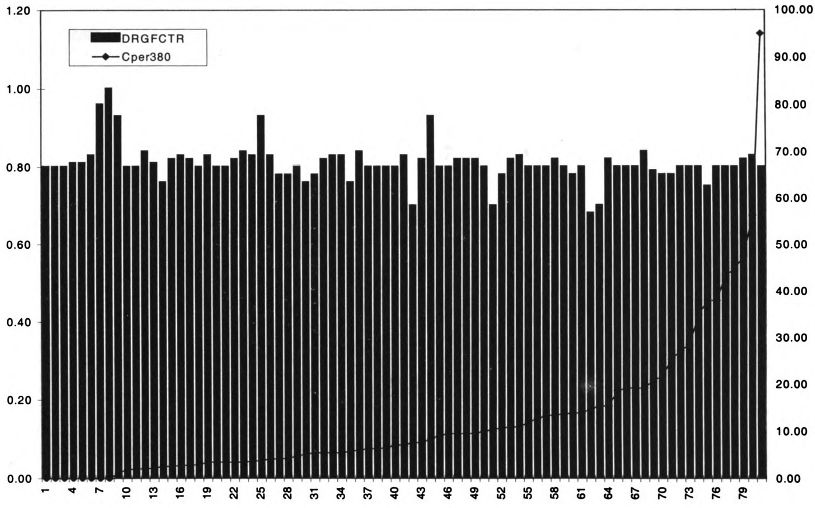


Figure 39 Drag factor (DRGFCTR), arranged in ascending order of curve crash rate (Cper380)

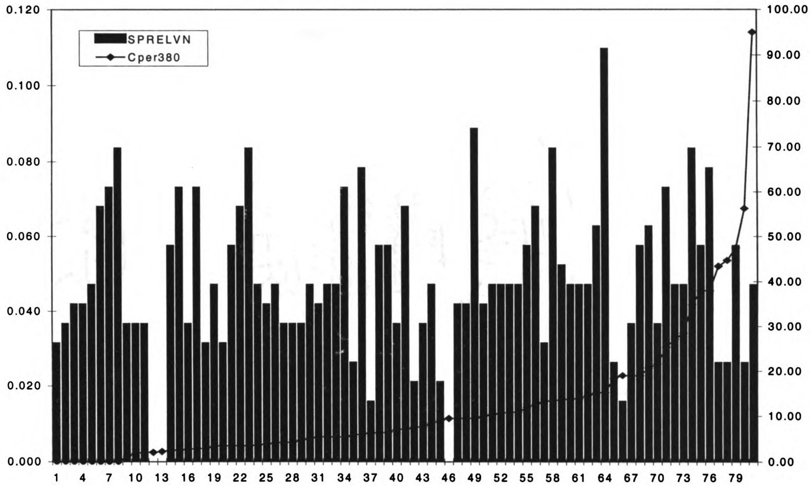


Figure 40 Superelevation high values (SPRELVN), arranged in ascending order of curve crash rate (Cper380)

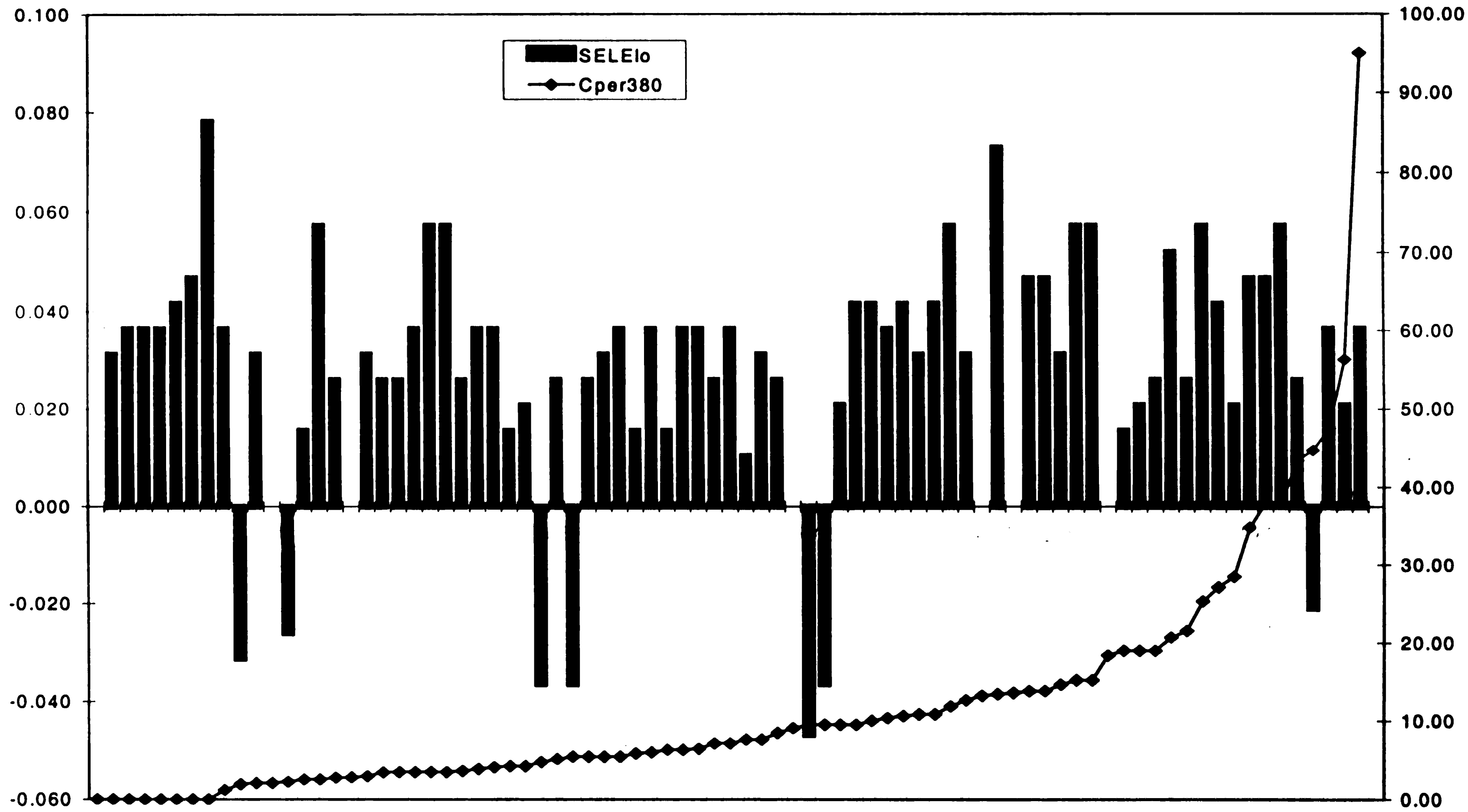


Figure 41 Superelevation low values, (SELElo), arranged in ascending order of curve crash rate (Cper380)

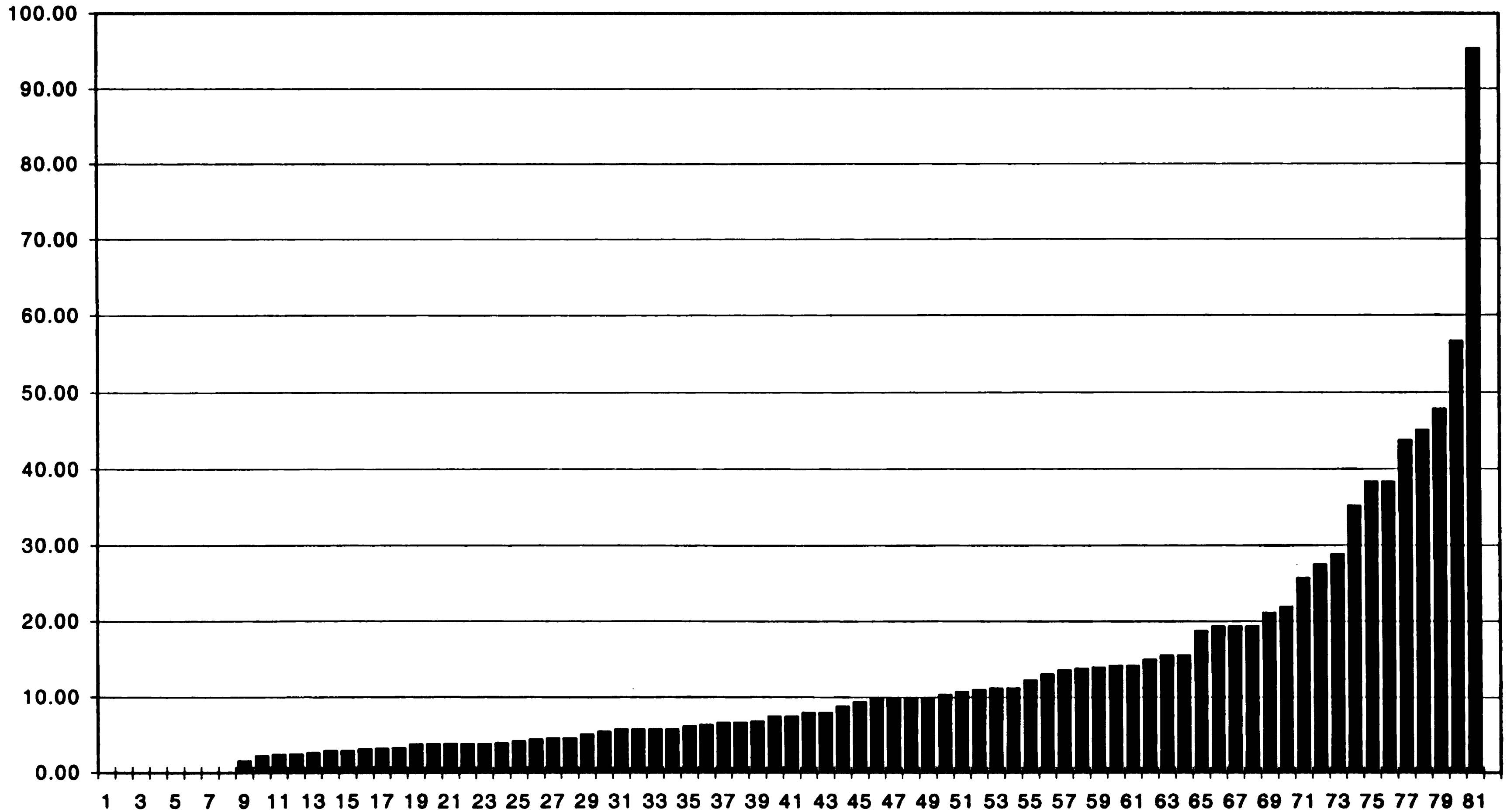


Figure 42 Curve crash rate (Cper380), arranged in ascending order

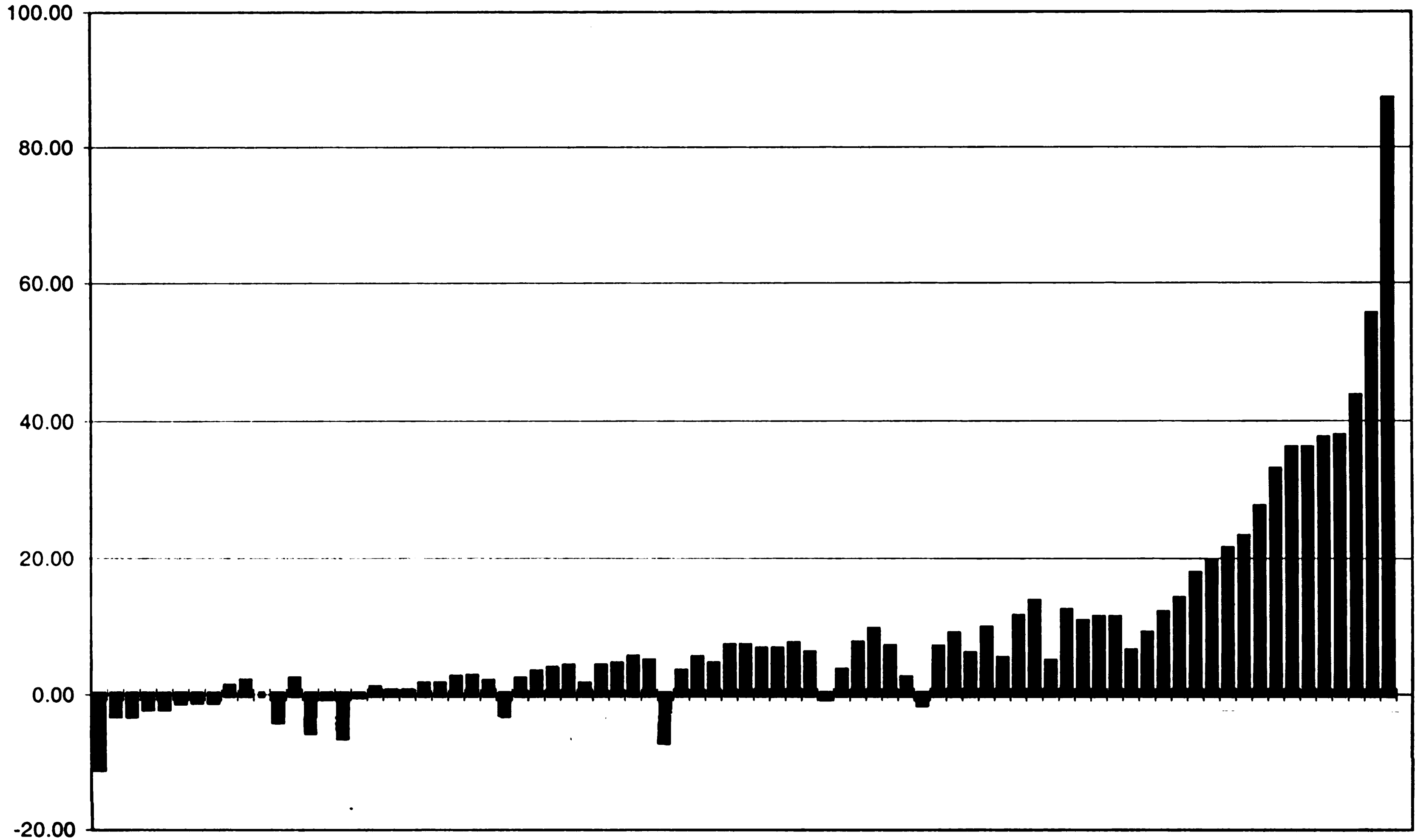


Figure 43 Curve crash rate minus tangent crash rate (CmnsT), arranged in ascending order of curve crash rate (Cper380)

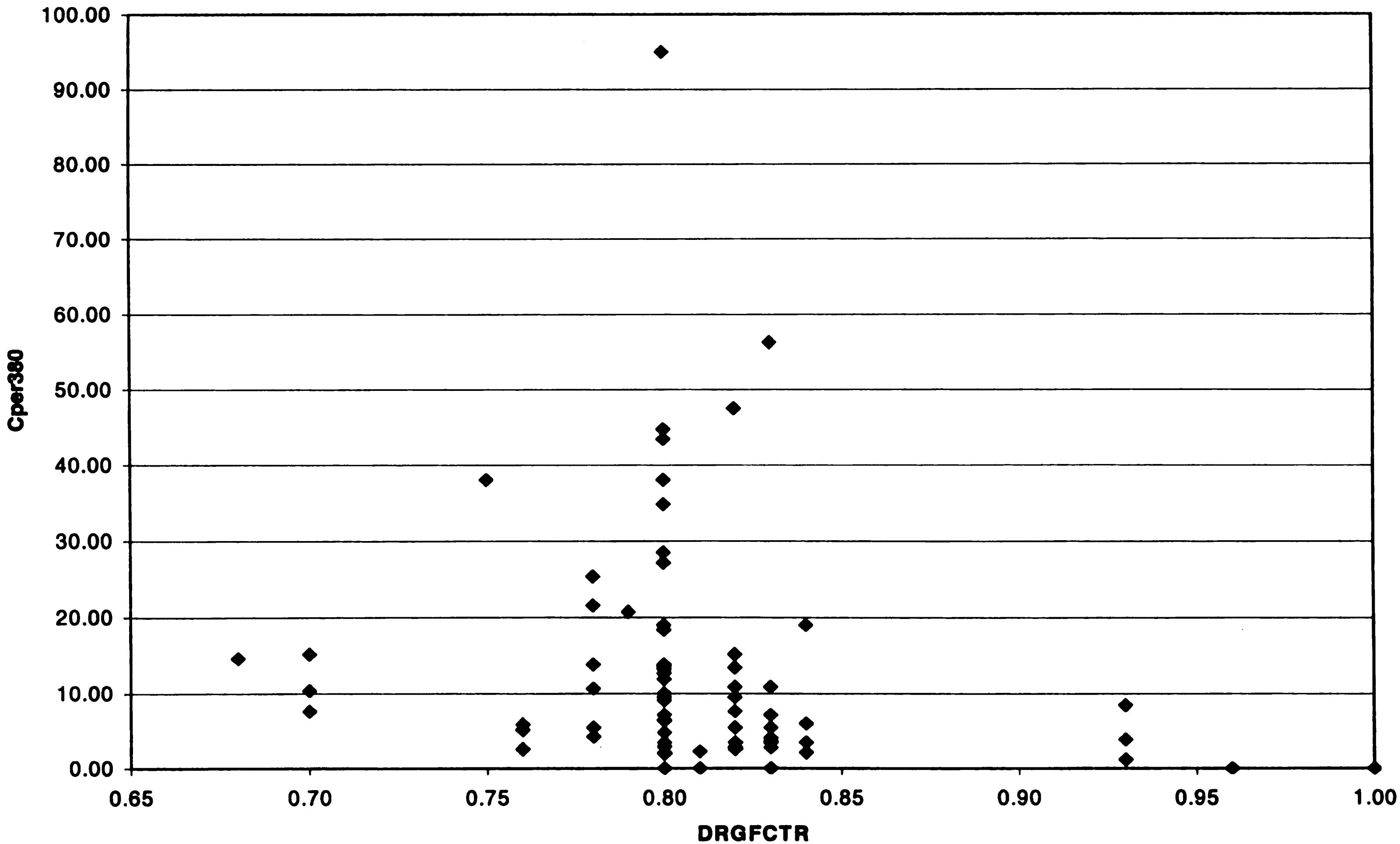


Figure 44 Curve crash rate (Cper380), for various values of drag factor (DRGFCTR)

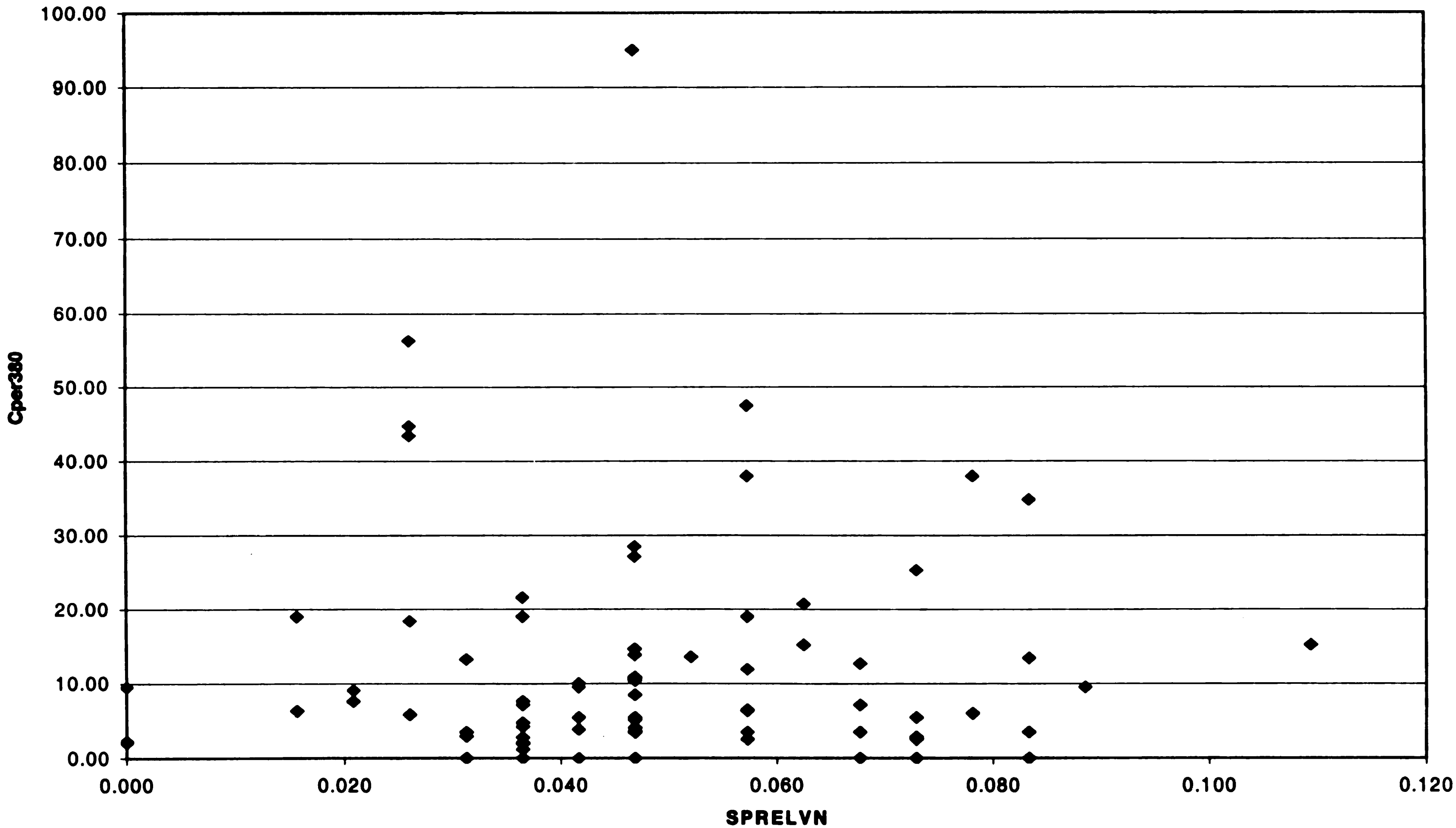


Figure 45 Curve crash rate (Cper380), for various values of superelevation high values (SPRELEV)

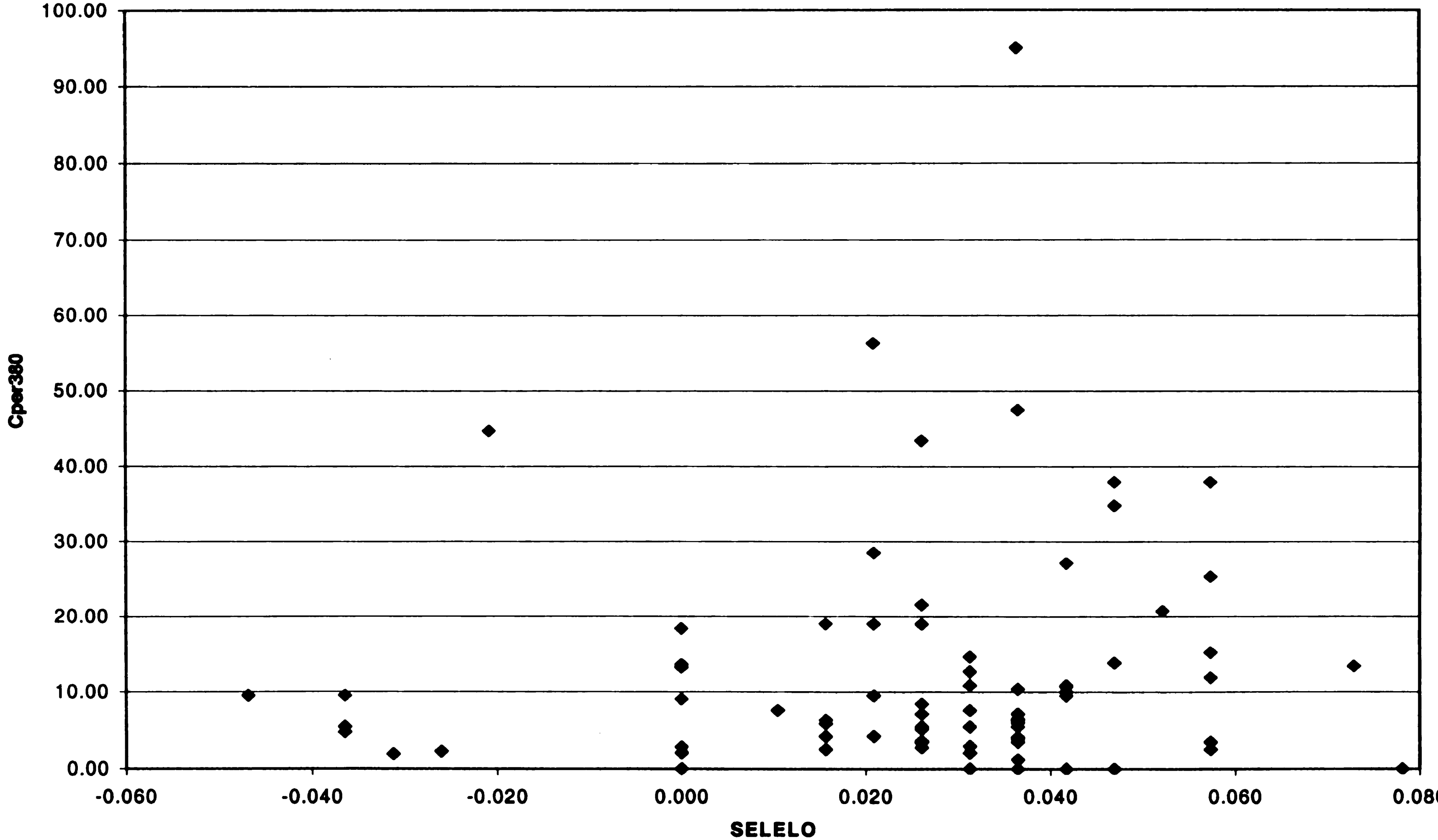


Figure 46 Curve crash rate (Cper380), for various values of superelevation low values (SELElo)

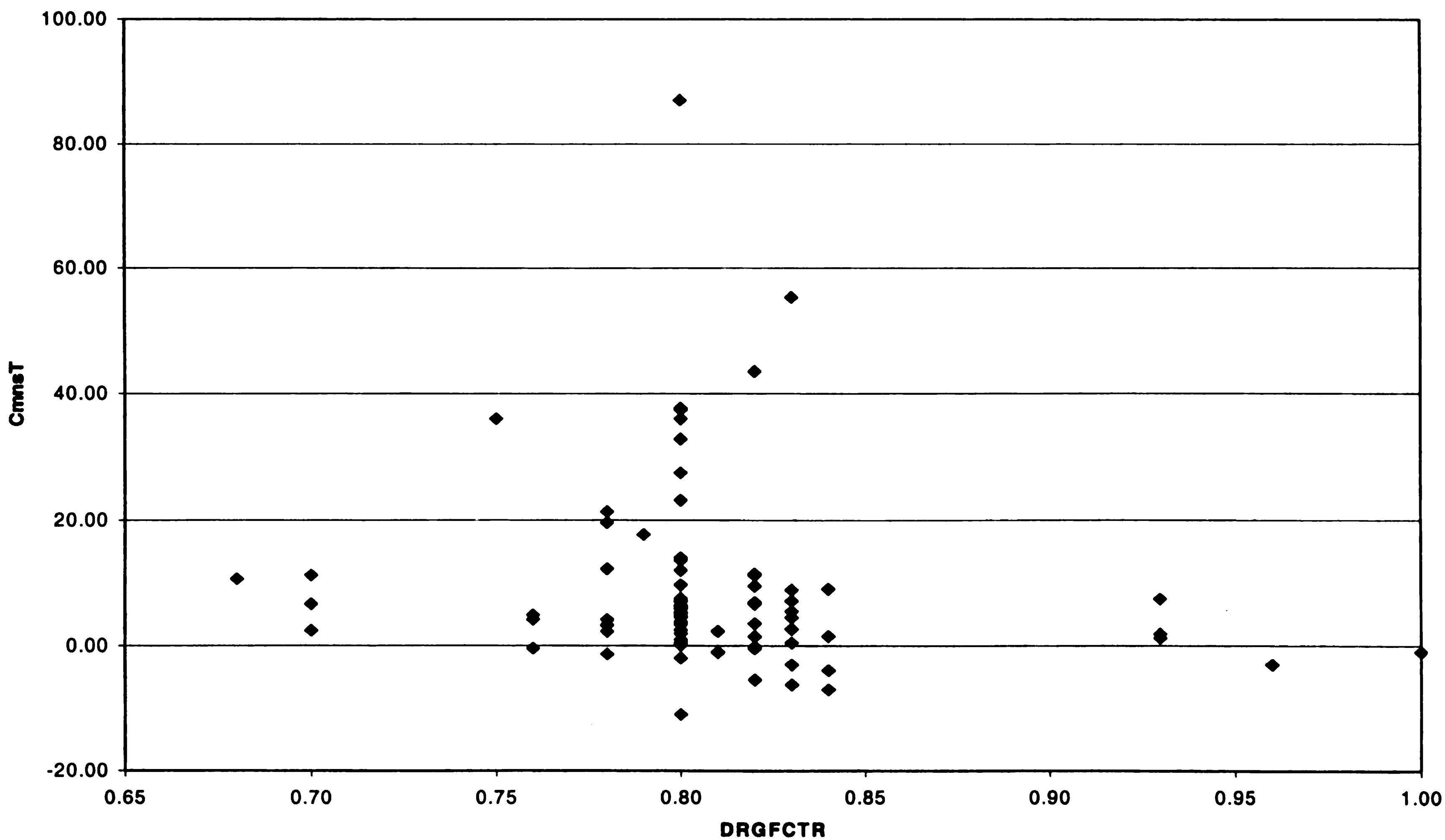


Figure 47 Curve crash rate minus tangent crash rate (CmnsT), for various values of drag factor (DRGFCTR)

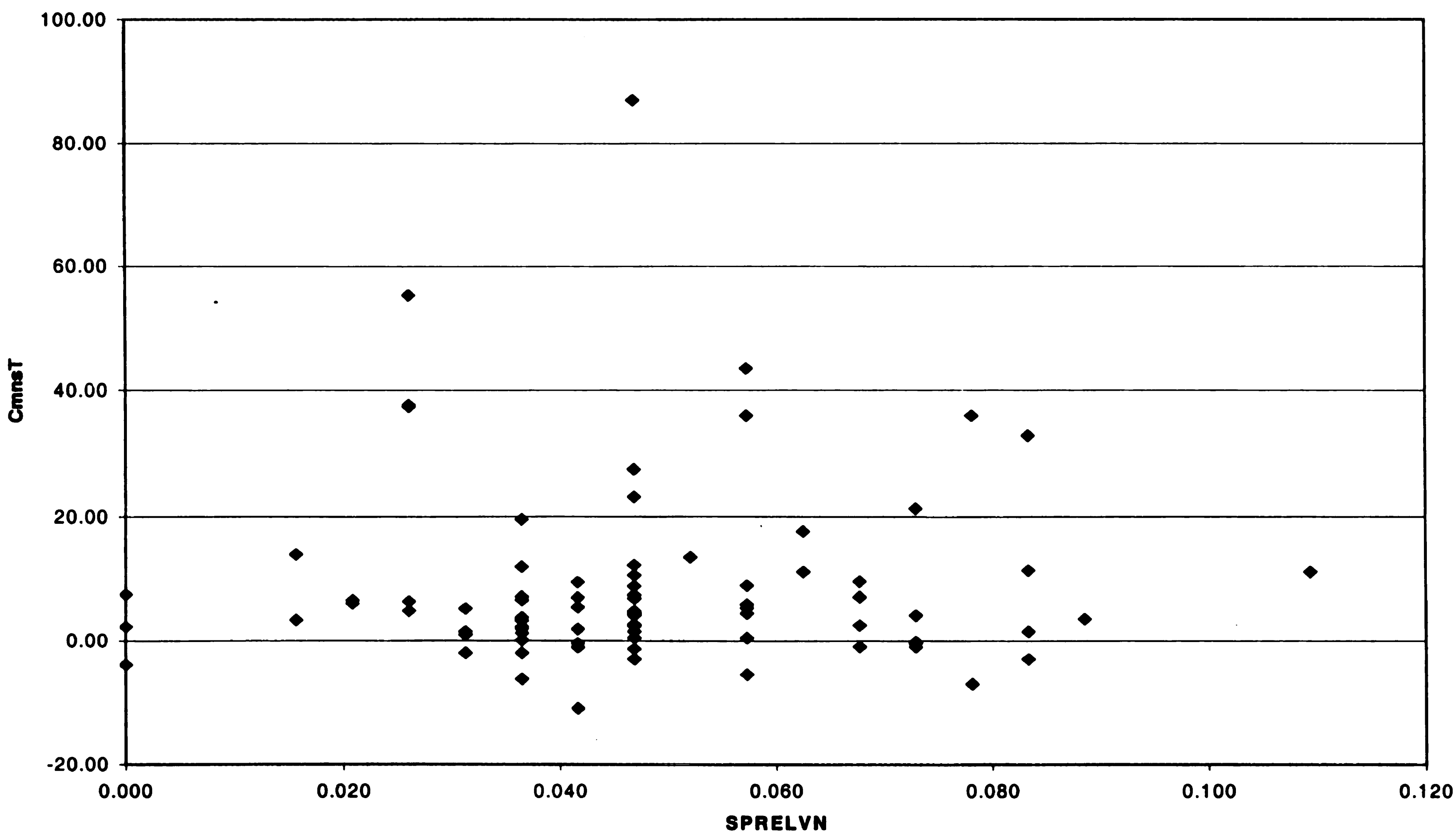


Figure 48 Curve crash rate minus tangent crash rate (CmnsT), for various values of superelevation high values (SPRELEV)

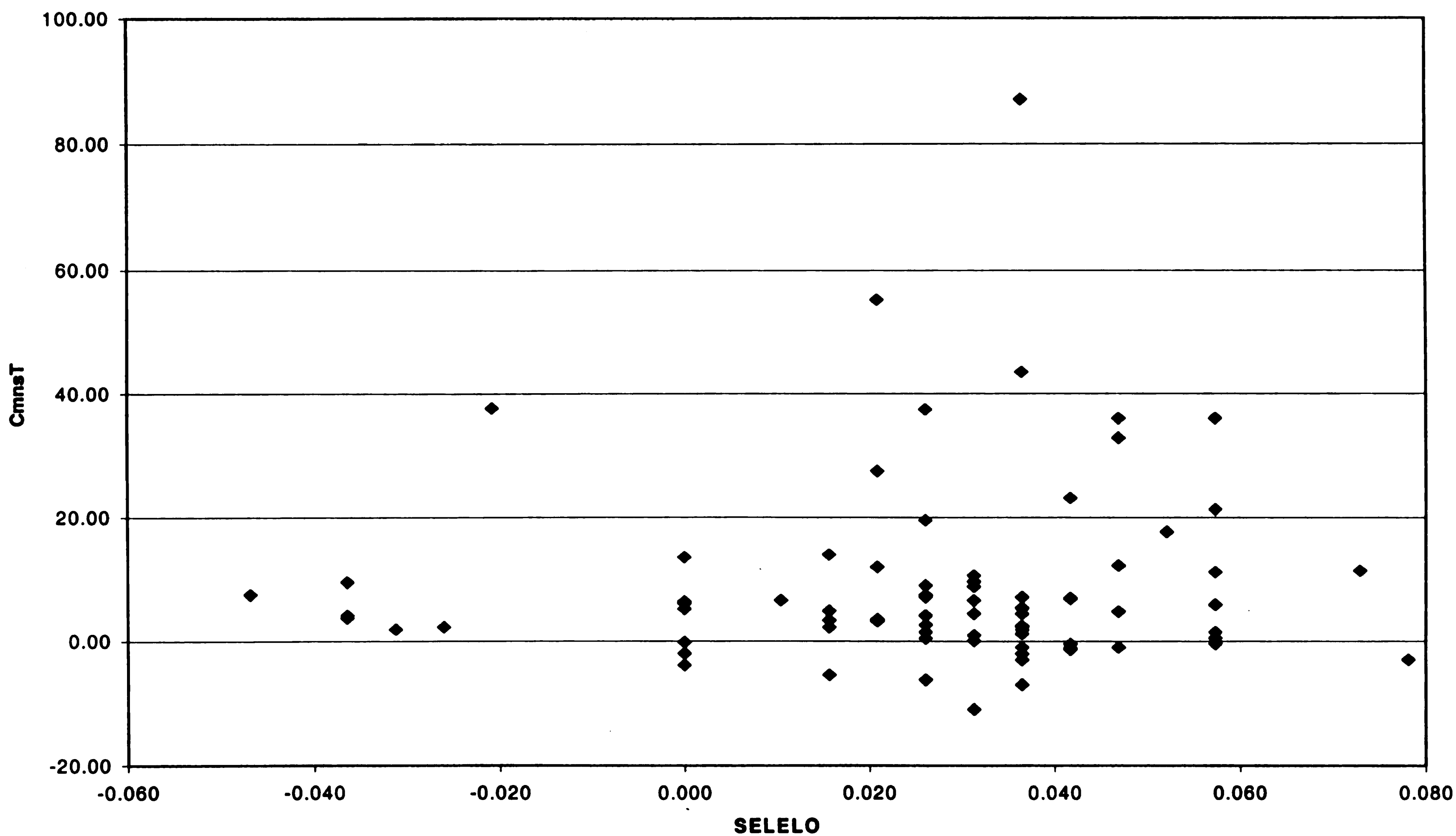


Figure 49 Curve crash rate minus tangent crash rate (CmnsT), for various values of superelevation low values (SELElo)

RESULTS

Discriminant analysis and cluster analysis provide information useful in meeting the objectives of this study. Specifically, they can be used to depict characteristics of low crash rate curves which distinguish them from high crash rate curves. Therefor curves with a high crash rate that should (based on their characteristics) have a low crash rate can be identified. These curves then could be studied for improvement by application of appropriate countermeasures.

Using the discriminant analysis results from the Cper380 analysis, sixteen curves fell in this category. The crash rate on these curves ranged from 7.13 to 21.71 when, based on their characteristics, they should have fallen in the group with a crash rate below 5.0. These curves are shown in Table 21, along with the value of some of the variables used in the analysis.

The significant characteristics of these curves include:

- Most do not have curve signs, target arrows and delineators
- There are no chevrons
- The distance at which the curve is first visible to the driver is usually short
- The radius is relatively large
- The tangent crash rate is low

An almost identical set of curves was identified when CmnsT was used as the grouping variable. This is consistent with the results shown in Table 21. Since most of these curves had a high value for Cper380 and low value for Tper380, they would fall in the high range of CmnsT values.

Table 21 Curves with a high curve crash rate (Cper380)

CRVno	CS	BMP	CT-sign	CHEV-RON	ARROW	DLNTR	OBS-DSTW	HCLFT	HCRFT	Tper380	Cper380
136	45012	5540	0	0	1	1	10	845	1042	0.00	7.13
14	5051	7280	0	0	0	0	40	264	2865	1.00	7.60
72	24011	4377	1	0	0	0	23	1056	2292	3.00	7.60
200	73131	0	0	0	0	0	0	264	2865	2.00	7.60
3	2021	15020	0	0	0	0	40	739	1910	1.00	8.14
39	12021	490	0	0	0	0	70	739	2292	3.00	8.14
33	10011	5620	1	0	1	0	33	475	2865	0.00	8.44
82	28052	5530	0	0	1	0	40	475	1910	1.00	8.44
81	28052	4790	1	0	0	0	50	634	2865	2.00	9.50
94	31013	5810	0	0	0	1	30	370	1910	3.00	10.86
117	38071	7490	1	0	1	0	10	1214	2865	8.00	13.22
87	30062	1640	1	0	0	1	10	1478	2456	0.00	13.57
156	51011	50	0	0	0	0	50	581	1146	1.58	13.82
19	8011	8990	0	0	0	1	80	211	1763	1.00	19.00
193	67011	2130	0	0	1	1	40	475	1637	4.00	21.11
172	58032	4150	0	0	0	0	80	370	2644	4.00	21.71

The results of the cluster analysis provide additional guidance that may be useful in establishing programs to reduce traffic crashes. Low crash rate curves are clustered with large radius and long length. The average radius for curves in this group (based on modified Cper380) is 398 meters (1305 ft). The average length for the same curves is 274 meters (900 ft). These curves tend to have target arrows but no chevrons.

High crash rates are clustered with short, sharp curves as expected. These curves tend to have both chevrons and target arrows in place, but still tend to experience crashes because of their limiting geometry. The countermeasures for these curves are likely to involve a change in alignment rather than traffic control measures.

The third cluster is the group of curves where traffic engineering countermeasures may be most effective. These curves have a crash rate over twice as high as the low crash rate curves, even though they have approximately the same radius. The primary geometric difference is that they are very short curves, averaging 95 meters (312 ft). These curves generally do not have chevrons or target arrows or advisory speed plates in place because the large radius results in a high design speed as determined from the AASHTO procedure.

Chevrons and target arrows are not intended for these types of curves according to the Michigan Manual of Uniform of Traffic Control Devices (MMUTCD), since they do not constitute a sharp change in alignment. However, based on this study, it would be appropriate to consider the use of these signs to increase the visibility of the curves.

These same curves were placed in this cluster whether the crash rate variable was Cper380, Modified Cper380, CmnsT, or modified CmnsT. There were approximately 70 curves that belong to this cluster. To provide additional guidance to the process of selecting curves where traffic engineering countermeasures have a high probability of reducind crashes, the curves in the cluster with the highest rates should be addressed first. Table 22 lists the curves for which both the Cper380 and CmnsT were significantly higher than the average for this cluster.

To provide additional guidance in selecting curves to be studied, the curves categorized in each of the three clusters were plotted in ascending order of the value of Cper380. Figure 50 shows these results. It is clear that within each cluster there is a significant range of values for the crash rate. The curves with the highest crash rate should be studied for possible countermeasure implementation. Table 23 lists these curves which have a crash rate equal to or greater than twice the average value of their cluster.

Table 22 Curves with a high curve minus tangent crash rate (CmnsT)

CRVno	CS	BMP	CTsign	CHEVRON	ARROW	DLNTR
39	12021	490	0	0	0	0
200	73131	0	0	0	0	0
68	23051	2220	1	0	0	0
177	61012	4910	0	0	0	0
14	5051	7280	0	0	0	0
4	2021	23640	0	0	0	1
3	2021	15020	0	0	0	0
82	28052	5530	0	1	1	0
81	28052	4790	1	0	0	0
94	31013	5810	0	0	0	1
33	10011	5620	1	1	1	0
12	5031	3900	1	0	0	0
214	81031	750	1	0	0	0
100	31051	9143	1	0	0	0
172	58032	4150	0	0	0	0
88	30062	2900	1	0	0	1
19	8011	8990	0	0	0	1
101	32011	3050	1	0	0	0
62	22021	499	0	0	0	0
140	45013	11700	1	0	1	0
215	81031	1370	1	1	0	0
71	23111	3670	1	0	0	0

Table 22 (Continued)

BSDST	HCLFT	HCRFT	Tper380	Cper380	CmnsT
70	739	2292	3.00	8.14	5.14
0	264	2865	2.00	7.60	5.60
20	845	2083	6.00	11.88	5.88
48	327	2292	12.00	18.39	6.39
40	264	2865	1.00	7.60	6.60
70	581	2865	0.00	6.91	6.91
40	739	1910	1.00	8.14	7.14
40	475	1910	1.00	8.44	7.44
50	634	2865	2.00	9.50	7.50
30	370	1910	3.00	10.86	7.86
33	475	2865	0.00	8.44	8.44
60	370	2292	2.00	10.86	8.86
10	317	2292	10.00	19.00	9.00
13	338	1910	1.00	11.88	10.88
80	370	2644	4.00	21.71	17.71
30	581	1719	3.00	20.73	17.73
80	211	1763	1.00	19.00	18.00
30	370	2292	4.00	27.14	23.14
49	306	1879	16.0	45.86	29.86
60	634	1910	2.00	34.83	32.83
70	370	2989	6.00	43.43	37.43
30	211	1910	3.00	47.50	44.50

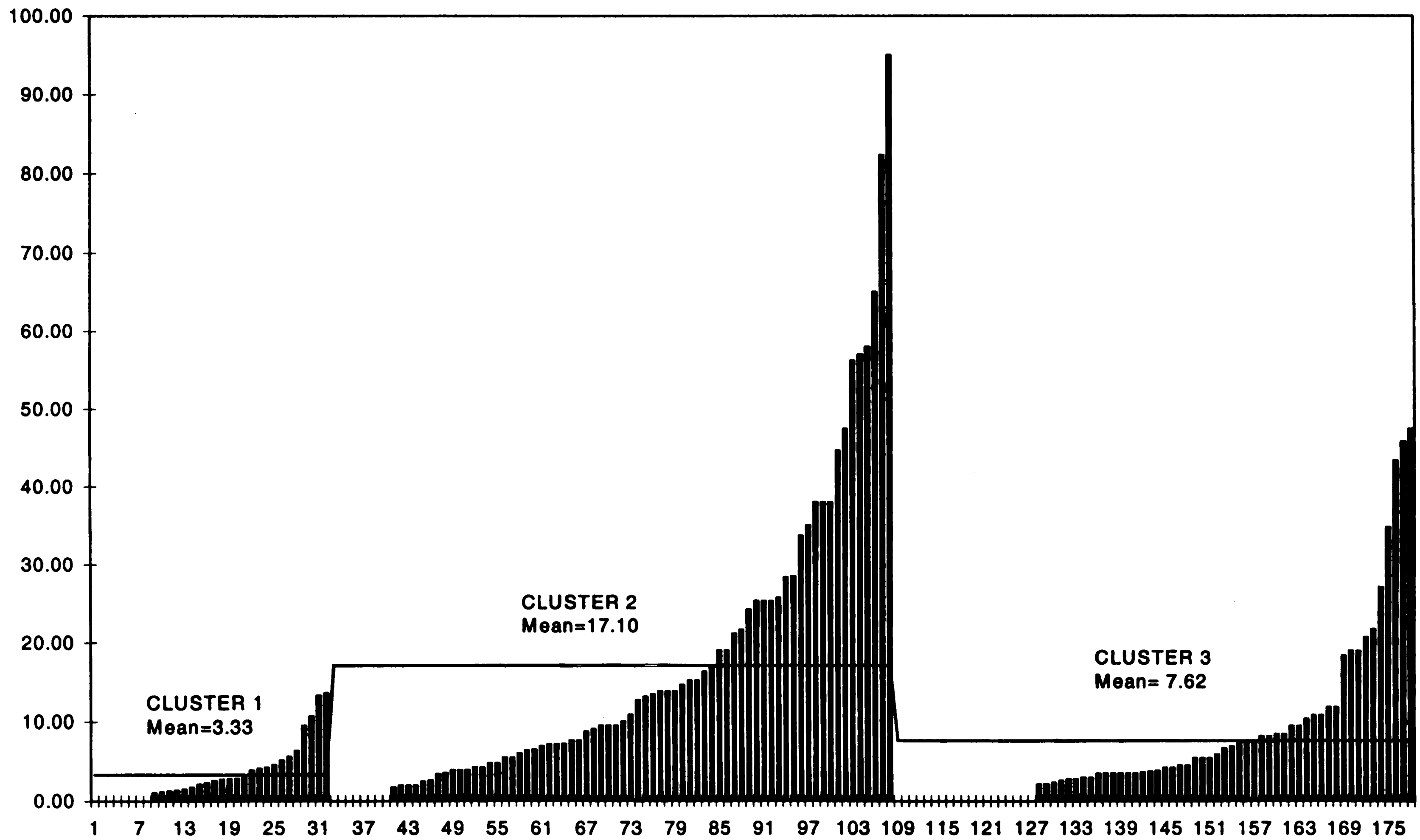


Figure 50 Curve crash rate (Cper380), for the three clusters, arranged in ascending order of Cper380 within each cluster

Table 23 Curves with a crash rate larger than twice the mean for their cluster

CRVno	CS	BMP	CTsign	CHEVRON	ARROW	DLNTR	OBSDSTW	HCLFT	HCRFT	Tper380	Cper380	C>2Mn
23	8031	2990	1	0	1	1	50	1267	1763	6	9.5	2.85
35	11052	14040	1	0	0	0	10	1320	2865	12	10.64	3.99
117	38071	7490	1	0	1	0	10	1214	2865	8	13.22	6.56
87	30062	1640	1	0	0	1	10	1478	2456	0	13.57	6.92
92	31012	4227	0	0	0	0	17	343	477	4	35.08	0.87
28	10011	7470	1	0	1	1	30	158	521	2	38	3.79
152	47041	21730	1	1	0	1	60	158	286	2	38	3.79
181	62031	3160	0	0	0	0	10	264	820	7	38	3.79
211	79081	8450	0	0	1	1	18	539	1008	7	44.71	10.5
18	8011	7100	1	0	0	0	30	211	229	4	47.5	13.29
196	72051	7673	0	0	0	0	10	143	1146	1	56.3	22.09
85	29042	6270	0	0	0	0	20	106	1146	4	57	22.79
168	56032	8814	1	0	0	0	34	380	1146	4	58.06	23.85
199	73061	3930	0	1	0	1	10	370	727	6	65.14	30.94
29	10011	8920	1	0	1	0	30	317	215	3	82.33	48.13
151	47041	19440	1	1	0	0	30	211	744	8	95	60.79
177	61012	4910	0	0	0	0	48	327	2292	12	18.39	3.14
19	8011	8990	0	0	0	1	80	211	1763	1	19	3.75
214	81031	750	1	0	0	0	10	317	2292	10	19	3.75
88	30062	2900	1	0	0	1	30	581	1719	3	20.73	5.48
172	58032	4150	0	0	0	0	80	370	2644	4	21.71	6.47
101	32011	3050	1	0	0	0	30	370	2292	4	27.14	11.9
140	45013	11700	1	0	1	0	60	634	1910	2	34.83	19.59
215	81031	1370	1	1	0	0	70	370	2989	6	43.43	28.18
62	22021	499	0	0	0	0	49	306	1879	16	45.86	30.62
71	23111	3670	1	0	0	0	30	211	1910	3	47.5	32.25

CONCLUSIONS AND GUIDELINES

Based on the results of this study, the following conclusions were reached.

1) The variation in the crash frequency or rate between horizontal curves with similar geometry is large, thus, regression analysis is not effective in identifying specific curves where countermeasures would be most effective in reducing curve crashes. The only studies that report high correlation coefficients are those that aggregate curves into groups with similar characteristics and then conduct the regression analysis on the group means.

   This information is used in the design of new highways, (i.e., setting the minimum curve radius as a function of the design speed, specifying the lane and shoulder width, etc.), but it is not useful in meeting the objectives of this study.
2) The predicted crash rate using existing models (Zegeer and Glennon) does not match the actual crash rates on Michigan two-way, two-lane rural trunklines, as shown in Figures 28-32. These models can not be used to identify curve locations where countermeasures could successfully be deployed to reduce crashes.
3) The addition of data on the distance on the approach at which the curve first becomes visible to the motorist is not highly correlated with the crash rates as a single variable, but it was found to be a contributor to some of the models that use multiple variables.
4) The superelevation and the drag factor showed a low simple correlation with the crash rate and the addition of these two variables contributed little to multiple variable analyses.
5) Discriminant analysis techniques, using the variables collected for this study, can successfully distinguish the characteristics of high crash rate curves from low

crash rate curves. This technique can also be used to identify outliers in each of the two categories (high and low) for both the absolute crash rate on curves (Cper380) or the difference in the crash rate between the curve and the tangent roadway segments (CmnsT).

6) Cluster analysis identified three distinct groups of curves. The group with a high crash rate (Cper380) is characterized by short radii and short curve lengths. These curves generally are marked with curve sign, advisory speed panels and chevrons or delineators. The group with a low crash rate is characterized by large radii and long curve lengths.
   The third group, with an intermediate crash rate, is characterized by large radii but short curve lengths. These results are shown in Figures 51 and 52.
   The high crash rate on the first group of curves is probably related to constraint the geometry imposes on drivers' ability to negotiate the curve at their approach speed.
   The intermediate crash rate curves may be related to the driver perception (or lack of perception) of the presence of a curve that does not require extraordinary driver input to negotiate safely. These are the curves that are likely to benefit the most from traffic engineering countermeasures.
7) The factor analysis results are more difficult to interpret, because the crash rate does not appear in each cluster. The results are consistent with the discriminant analysis and the cluster analysis results. In general, the variables significant in defining the factor groups are the same as those used to distinguish the group membership in these analyses.

The results of the study provide guidance on a cost effective approach to selecting curves to be studied:

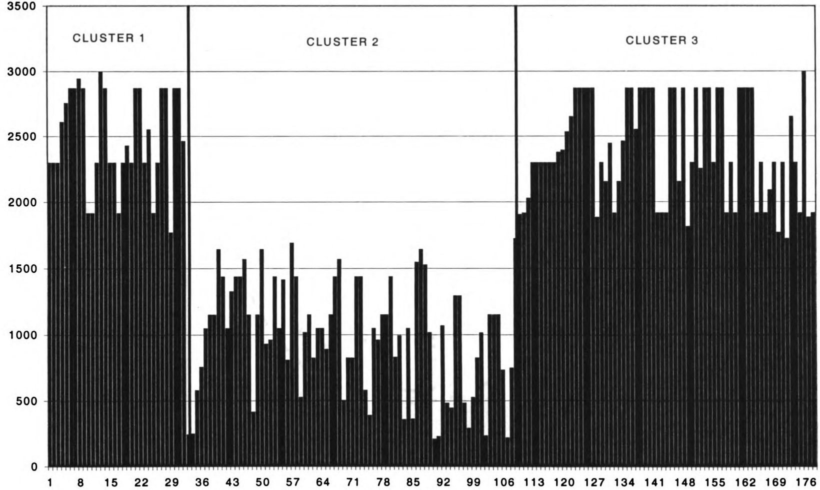


Figure 51 Curve radius in feet (HCRFT) for the three clusters, arranged in ascending order of Cper380 within each cluster

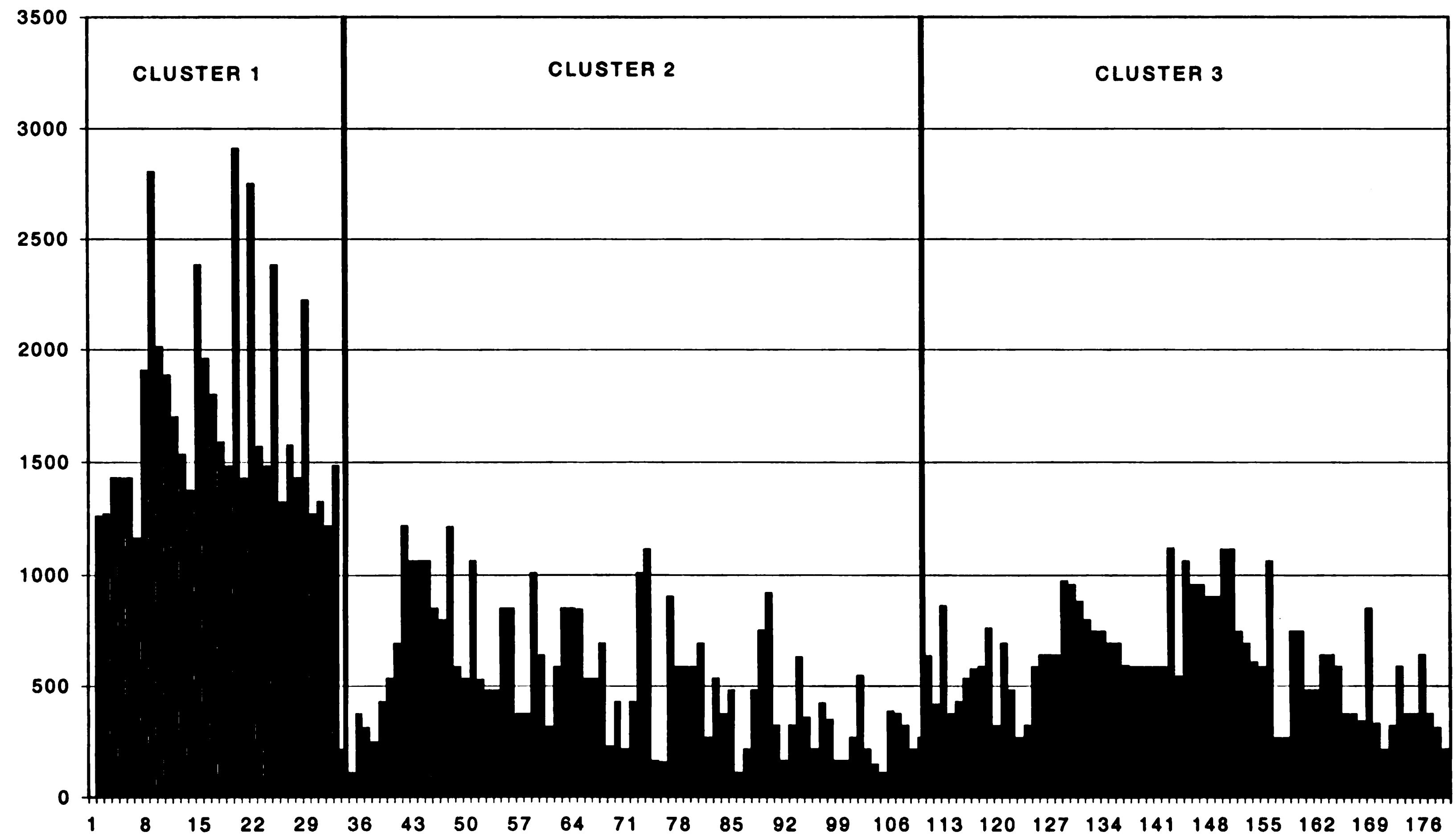


Figure 52 Curve length in feet (HCLFT), for the three clusters, arranged in ascending order of Cper380 within each cluster

1) The curves identified in Table 21 from the discriminant analysis results should be targeted for analysis and potential countermeasures implementation. These sixteen curves have the characteristics of low crash rate curves, but are experiencing a high rate of crashes.

2) The curves identified in Table 22 from the cluster analysis results should be targeted for analysis and potential countermeasure implementation. These segments have been identified as experiencing a high crash rate on its curve compared to the crash rate on the straight sections at the two ends of each curve, (CmnsT).

3) The curves identified in Table 23 from the cluster analysis results should be targeted for analysis and potential countermeasure implementation. These curves have been identified as experiencing a crash rate at least twice that of the average crash rate for curves in their cluster.

4) Curves characterized by a large radius and short curve length should be analyzed to determine if there are inexpensive countermeasures that could be applied at these curves to reduce the crash rate. These curves have been identified from the cluster analysis as having an intermediate crash rate which is not explained by the curve geometry.

5) Discriminant analysis and cluster analysis techniques should be used to analyze other sets of curves on state trunkline highways. These techniques have been useful in identifying specific curves that are candidates for countermeasures. It should be determined whether these techniques are equally valid for curves that are not screened for approach tangents and intersections and curves on four-lane cross sections.

6) Curves that are both in Table 21 and also Table 22 should, in particular, be considered for upgrade. These are curves: 3, 14, 19, 33, 39, 81, 82, 94, 172 and 200.
7) Similarly, curves that are both in Table 23 and also Table 21 or Table 22, should, in particular, be considered for upgrade. These are curves: 62, 71, 87, 88, 101, 117, 140, 177, 214, and 215.
8) Had there been any curves common to all the three Tables (21, 22 and 23), such curves would have had the highest priority for countermeasure implementation.

LIST OF REFERENCES

References

1. Roy Jorgenson Associates, Inc., MD, "Cost and Safety Effectiveness of Highway Design Elements", NCHRP Report 197, 1978.

2. Glennon, J.C., Neuman, T.r., Leisch, J.E. "Safety and Operational Considerations For Design of Rural Highway Curve", Jack E. Leisch & Associates, Evanston, IL, FHWA-RD-86/035, 1985.

3. Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L. and Hunter, W., "Safety effects of cross-section Design for two-lane Roads, Volume I," Report No. FHWA/RD-87/008, FHWA, Washington DC, December 1986.

4. Zegeer, C.V., and Deacon, J.A., "Effects of Lane Width, Shoulder Width, and Shoulder Type on Highway Safety."

5. Neuman, T.R., Zegeer, C., and Slack, K.L. "Design Risk Analysis Volume II - User's Guide", FHWA Contract No. DTFH 71-88-C-00011, U.S. Department of Transportation, April 1991.

6. McGee, H.W., Hughes, W.E., Daily, K., Bellomo-McGee, Inc., "Effect of Highway Standards on Safety", Final Report, prepared for National Cooperative Highway Research Program, Transportation Research Board, National Research Council, Vienna, Virginia.

7. Kach, Brett, Benac, Jack D., "Horizontal Curve Crash Model Assessment", Traffic and Safety Division, Michigan Department of Transportation, June 1992.

8. Fink, K.F., Krammes, R.A., "Tangent Length and Sight Distance Effects of Crash Rates at Horizontal Curves on Rural two-lane Highways", The 74th Annual Meeting, Transportation Research Board, January 1995.

9. Zegeer, C.V., Twomey, J.M., Heckman, M.L. and Heyward, J.C., "Safety Effectiveness of Highway Design Features", Volume II, Alignment. Report No. FHWA-RD-91-045. FHWA, U.S. Department of Transportation, 1992.

10. Datta T.K. et al., "Crash Surrogates for Use in Analyzing Highway Safety Hazards." Report No. FHWA/RD-82/103. Washington, D.C., 1983.

11. Terhune, K. W. and Parker, M.R., "An Evaluation of Crash Surrogates for Safety Analysis of Rural Highways." Volume 2. Technical Report. Report No. FHWA/RD 86/128. Federal Highway Administration, Washington, D.C., 1986.

12. Matthews, L.r. and Barnes, J.W., "Relation Between Road Environment and Curve Crashes." Proceedings, 14th Australian Road Research Board Conference, Vol. 14, Part 4, Victoria, Australia, 1988.

13. Zegeer, C., Stewart, R., Reinfurt, D. et al., "Cost-Effective Geometric Improvements for Safety Upgrading of Horizontal Curves." Report No. FHWA-RD-90-021. Federal Highway Administration, Washington, D. D., 1991.

14. Glennon, J.C., Neuman T.R. and Leisch, J.P. "Safety and Operational Considerations for Design of Rural Highway Curves." Report No, FHWA/RD-86/035. FHWA, U.S. Department of Transportation, 1985.

15. Fambro, D.B., Urbanik, T. II, Hinshaw, W.M., Hanks Jr., J.W., Ross, M.S., Tan, C.H. and Pretorius, C.J. "Stopping Sight Distance Considerations at Crest Vertical Curves on Rural two-lane Highways in Texas", (TX-90/1125-1F). Texas Transportation Institute, College Station, Texas, 1989.

16. Baker, Stannard J., Northwestern University Traffic Institute, The Traffic-Accident Investigation Manual, At-Scene Investigation and Technical Follow-Up, Evanston, IL.

17. Berry, William D., Feldmam, Stanley, University of Kentucky. Multiple Regression in Practice, Sage Publications, Newberry Park London New Delhi, 1990.

18. Bhatti, M. Ishaq, School of International Business Relations, Testing Regression Models Based on Sample Survey Data, Avebury, Aldershot Brookfield USA Hong Kong Singapore Sydney, 1995.

19. Hand, D.J., Biometrics Unit, London University Institute of Psychiatry, Discrimination and Classification, John Wiley and sons, Cichester New York Brisbane Toronto, 1981.

20. NATO Advanced Study Institute, Edited by T. Cacoullos, University of Athens, Athens, Greece. Discriminant Analysis and Applications, Academic Press, New York and London, 1973.

21. Jambu, M., Centre National d'Etudes des Te'le'communications, Issy-les-Molineaux, France and Lebeaux, M-O., Centre National de la Recherche Scientifique, Centre d'Etudes Sosiologiques, Paris, France. Cluster Analysis and Data Analysis, North-Holland Publishing Company. Amsterdam New York Oxford, 1983.

22. Tryon, Robert Choate, Associate Professor of Psychology, University of California. Cluster Analysis, correlation profile and Orthometric (Factor) Analysis. Edward Brothers, Inc., Ann Arbor, Michigan, 1939.

23. Van de Geer, P. John, University of Leiden, the Netherlands, Introduction to Multivariate Analysis for the Social Sciences, W. H. Freeman and Company, San Francisco, 1971.