Standard Title Page -- Report on State Project

Report No.	Report Date	No. Pages	Type Report: Final	Project No. : 9242-062-940
	November 1990		Period Covered:	Contract No.:

Title and Subtitle Traffic and Geometric Characteristics Affecting the Involvement of Large Trucks in Accidents Vol. II: Linear, Poisson and Logistic Regression Models

Author(s) Nicholas J. Garber and Sarath C. Joshua

Performing Organization Name and Address

Virginia Transportation Research Council
Box 3817, University Station
Charlottesville, Virginia 22903-0817

Sponsoring Agencies' Names and Addresses

Va. Dept. of Transportation
1221 E. Broad Street
Richmond, Virginia 23219

University of Virginia
Charlottesville
Virginia 22903

Key Words

Truck Accidents
Highway Type
Traffic Characteristics
Geometric Characteristics

Supplementary Notes This project was sponsored by The Department of Motor Vehicles
P.O. Box 27412
Richmond, VA 23269

Abstract

Recent trends indicate that vehicle miles traveled for large trucks is increasing at a higher rate than for other vehicles. The resulting competition between large trucks and other vehicles for highway space can be expected to result in more multivehicle collisions involving large trucks. The likelihood of these collisions causing severe injuries to vehicle occupants will also increase with the trend towards the use of smaller automobiles and heavier and larger trucks. In order to develop countermeasures that will alleviate this problem, it is first necessary to identify the characteristics of large-truck accidents and the role of traffic and geometric variables in such accidents. This study investigated the major factors associated with large truck accidents including the effect of highway facility type and highway geometry, and the development of mathematical models relating the factors with accident rates and probability of occurrence.

This second volume gives a detailed description of the development of the regression and logistic models.

FINAL REPORT

TRAFFIC AND GEOMETRIC CHARACTERISTICS AFFECTING THE INVOLVEMENT OF LARGE TRUCKS IN ACCIDENTS

VOLUME II

LINEAR, POISSON, AND LOGISTIC REGRESSION MODELS

Nicholas J. Garber
Faculty Research Engineer

and

Sarath C. Joshua
Graduate Research Assistant

(The opinions, findings, and conclusions expressed in this report are those of the authors and not necessarily those of the sponsoring agencies.)

Virginia Transportation Research Council
(A Cooperative Organization Sponsored Jointly by the
Virginia Department of Transportation and
the University of Virginia)

Charlottesville, Virginia

January 1991
VTRC 91-R18

1700

SAFETY RESEARCH ADVISORY COMMITTEE

W. H. LEIGHTY, Chairman, Deputy Commissioner, Department of Motor Vehicles

C. W. LYNN, Executive Secretary, Research Scientist, VTRC

J. L. BLAND, Chief Engineer, Department of Aviation

R. J. BREITENBACH, Director, Transportation Safety Training Center, Virginia Commonwealth University

MAJ. J. K. COOKE, Asst. Chief of Law Enforcement, Department of Game and Inland Fisheries

S. H. COOPER, Director of Rail and Public Transportation, VDOT

M. L. EDWARDS, Executive Assistant, Office of the Secretary of Transportation

W. S. FELTON, JR., Administrative Coordinator, Commonwealth's Attorneys' Services and Training Council

P. D. FERRARA, Ph.D., Director, Division of Forensic Sciences, Dept. of General Services

D. R. GEHR, Asst. Chief Engineer, VDOT

LT. COL. L. A. GRAHAM, Director, Bureau of Field Operations, Dept. of State Police

J. T. HANNA, Special Assistant to the Commissioner, Department of Motor Vehicles

C. P. HEITZLER, JR., Program Manager, Department of Information Technology

T. A. JENNINGS, Safety/Technology Transfer Coordinator, Federal Highway Administration

B. G. JOHNSON, Supervisor, Driver Education, Department of Education

SGT. P. J. LANTEIGNE, Operations & Tactics Bureau, Virginia Beach Police Dept.

W. T. McCOLLUM, Executive Director, Commission on VASAP

S. D. McHENRY, Director, Division of Emergency Medical Services, Dept. of Health

CAPT. R. P. MINER, Commander, Traffic Division, Fairfax County Police Dept.

COMM. S. E. NEWTON, Patrol Division, Albemarle County Police Department

J. T. PHIPPS, Director, Roanoke Valley ASAP

F. F. SMALL, Asst. State Traffic Engineer, VDOT

J. A. SPENCER, ESQ., Asst. Attorney General, Office of the Attorney General

E. W. TIMMONS, Director of Public Affairs, Tidewater AAA of Virginia

A. R. WOODROOF, ESQ., Manakin-Sabot, Virginia

1701

TABLE OF CONTENTS

LIST OF TABLES v

LIST OF FIGURES vii

ABSTRACT ix

INTRODUCTION 1

METHODOLOGY 2

Establishment of the Accident Data Base 2

Identification of Sites for Data Collection 3

Data Collection 3

Surrogates of Roadway Alignment 5

DEVELOPMENT OF MODELS 6

Model Selection Criteria 8

Components of AIC 8

Multiple Linear Regression Model 9

Poisson Regression Model 12

Multiple Logistic Regression Model 13

RESULTS 15

Multiple Linear Regression Models 15

Poisson Regression Models 17

Logistic Regression Model for Environment I (Primary Highways, Undivided 4-Lane and 2-Lane) 23

Logistic Regression Model for Environment II (Primary Highways, Divided 4-Lane, AADT $\leq$ 15,000) 25

Logistic Regression Model for Environment III (Interstate and Primary Highways, Divided 4-Lane, AADT > 15,000) 26

GUIDELINES FOR SELECTING TRUCK ROUTES 29

CONCLUSIONS 30

Multiple Linear Regression Models 30

1702

Poisson Regression Models 30

Logistic Regression Models 30

Measures for Complexity of Highway Alignment 31

REFERENCES 33

APPENDIX I: List of Study Sites 35

APPENDIX II: Sample Field Data From Two Study Sites 39

LIST OF TABLES

Table 1. Linear Models for TRATE — Environment I 15
Table 2. Linear Models for TRATE — Environment II 16
Table 3. Linear Models for TRATE — Environment III 17
Table 4. Poisson Regression Model — Environment I 17
Table 5. Poisson Regression Model — Environment II 19
Table 6. Poisson Regression Model — Environment III 21
Table 7. Logistic Regression Model — Environment I 23
Table 8. Logistic Regression Model — Environment II 26
Table 9. Logistic Regression Model — Environment III 27

1705

LIST OF FIGURES

Figure 1. Components of Highway Horizontal Alignment 7

Figure 2. Components of Highway Vertical Alignment 7

Figure 3. Effect of SCR and TPERCENT on TINVOL — Environment I 18

Figure 4. Effect of SCR and TPERCENT on TINVOL — Environment I 19

Figure 5. Effect of SCR and TPERCENT on TINVOL — Environment II 20

Figure 6. Effect of SCR and TPERCENT on TINVOL — Environment II 21

Figure 7. Effect of CCR and Speed Difference on Truck Involvements — Environment III 22

Figure 8. Effect of CCR and Speed Difference on Truck Involvements — Environment III 22

Figure 9. Effect of Lane and Shoulder Widths - Environment I 24

Figure 10. Effect of SCR - Environment I 25

Figure 11. Effect of TPERCENT - Environment II 27

Figure 12. Effect of CCR - Environment III 28

Figure 13. Effect of TPERCENT - Environment III 29

ABSTRACT

Vehicle miles traveled for large trucks, which are defined here as trucks having six or more wheels in contact with the road and having a gross weight greater than 10,000 lb, have been steadily increasing during the past few years. On some sections of Virginia interstate routes, for example, the proportion of large trucks is as high as 50 percent. The resulting competition between large trucks and other vehicles for highway space can be expected to result in more collisions involving large trucks. An analysis of large-truck accidents in Virginia also indicated that driver-related factors are the primary associated factors for truck crashes. For example, driver error is associated with over 50 percent of fatal accidents involving large trucks, and fatal crashes for which driver error is listed as the primary factor occur predominantly on stretches of highways with vertical or horizontal curves and/or grades. In order to develop countermeasures that will alleviate this problem, it is necessary to identify the specific traffic and highway geometric characteristics that significantly affect the occurrence of large-truck crashes.

This study was therefore conducted by the Virginia Transportation Research Council with the objective of identifying appropriate countermeasures for highway geometrics to reduce large-truck crashes. The major factors associated with large-truck accidents, including the effect of highway facility type and highway geometry, were investigated. This study is reported in two volumes.

This volume presents mathematical relationships obtained through linear, Poisson, and multiple logistic regression analyses relating the probability of a large truck being involved in an accident with a set of associated traffic and highway geometric variables. These models indicate that lane width, shoulder width, percentage of trucks in the traffic stream, and changes in the vertical and horizontal alignments have some influence on the probability of a truck being involved in an accident on a given stretch of highway.

FINAL REPORT

TRAFFIC AND GEOMETRIC CHARACTERISTICS AFFECTING THE INVOLVEMENT OF LARGE TRUCKS IN ACCIDENTS

VOLUME II

LINEAR, POISSON, AND LOGISTIC REGRESSION MODELS

Nicholas J. Garber

and

Sarath C. Joshua

INTRODUCTION

The incidence of large-truck accidents in general has been observed to be less than for passenger cars, light trucks, and vans, but the relative proportion of fatal accidents has been observed to be much higher for large trucks. In Virginia, for example, although the total number of fatal accidents for all vehicles has been increasing, the fatal accident rate based on accidents per million vehicle miles of travel (VMT) has been decreasing in recent years. However, large trucks have been experiencing almost twice the fatal accident rate per miles traveled than passenger cars.[1] Accident experience of large trucks in the future will be influenced by the amount of truck VMT and the changing vehicle mix. If VMT for large trucks in the next decade continues to increase at the present rate and the rate of fatal large-truck accidents based on VMT remains constant, then the number of fatal accidents that involve large trucks can be expected to rise significantly.

Recent trends indicate that VMT for large trucks is increasing at a higher rate than for other vehicles. The resulting competition for highway space by more vehicles can be expected to result in more frequent collisions. The likelihood of these collisions causing severe injuries to vehicle occupants has increased with the trend toward the use of smaller and more fuel economical vehicles on the highways. This problem is made worse by the fact that average truck sizes and weights are also increasing. The passage of the Surface Transportation Assistance Act of 1982, for example, permitted wider and larger trucks on the interstate and designated primary highways across the country. Clearly, as the truck sizes increase and the roadway environment remains relatively the same, the likelihood of incompatibilities between the two will increase, and this will result in higher accident rates for trucks. Therefore, in order to ensure the desired safety levels on the highways, it is essential that the incompatibilities between large trucks and highway and traffic characteristics that lead to hazardous situations be clearly understood. Unfortunately, only limited information is available on the quantitative relationship between the risk of occurrence of a large-truck accident and the associated traffic and geometric characteristics. This study was therefore conducted to identify the traffic and geometric variables that significantly affect the involvement of large trucks in accidents. As part of this study, linear, poisson, and multiple logistic regression models were developed relating the probability of a large truck being involved in an accident with a set of associated traffic and geometric variables. These models are presented here.

METHODOLOGY

Establishment of the Accident Data Base

Accident data files compiled by the Virginia Department of Transportation (VDOT) for all of Virginia highways for 1984 through 1986 were obtained, and subfiles were created from them for use on the CDC computer. From among the data available in the main file, the following data fields were extracted and included in the subfiles for this part of the study:

- county/city
- route number
- section number
- sequence number
- type of collision
- traffic control
- alignment
- severity
- major factor
- begin terminal
- end terminal
- average daily traffic (ADT)
- highway type.

The county/city, route number, section number, and the sequence number identify a specific roadway segment between the begin and end terminals.

Identification of Sites for Data Collection

Two sources were used to identify suitable test sites for data collection. The first source was the subfiles created from the VDOT files, and the second was the Department of State Police. The subfiles created were sorted by the first four fields that define the sites. Once this sorting was accomplished, all the accident records occurring on the same roadway segment appeared adjacent to each other. This highlighted segments with a large number of accidents. These files were one of the sources used for identifying sites with high accident rates from which test sites were selected. The Department of State Police was also asked to provide a list of sites with frequent large-truck accidents. This list was also used to select additional test sites. The selection of sites with high truck accident rates as the test sites for the study was based on the premise that these sites will have traffic and geometric characteristics that are associated with truck accidents. Control sites with similar highway configurations were also identified adjacent to test segments, either upstream or downstream from the locations with high accident rates. In selecting the test sites from the accident files, the following criteria were used:

- The length of the highway segment must be 2 miles or less.
- Most of the accidents occurring on the highway segment must be of the mobility type.
- There must be a consistent pattern of accident occurrence on the segment of highway.

Accident sites from four different highway categories were selected. These were: (1) interstate routes, (2) primary routes (divided with four or more lanes), (3) primary routes (undivided with four or more lanes), and (4) primary routes (undivided with two lanes). Once the sites were selected, it was necessary to determine whether it would be feasible to collect data at the site. The reason for this is that sometimes as a result of temporary highway construction or maintenance activities in the vicinity of a site, data collected on traffic variables, such as speed, may not represent the true conditions. A list of the study sites is shown in Appendix I.

Data Collection

This task consisted of two subtasks, namely, compilation of historical data relevant to each site and collection of on-site traffic and geometric data.

Historical Accident Data

The historical accident data were obtained from the main VDOT accident files using a computer program. The relevant data on all accidents occurring at each site selected for the study were extracted and stored in the subfiles created.

Traffic and Geometric Data

All the sites selected for the study were visited, and the following kinds of data were collected:

- highway type, number of lanes, lane width, and shoulder width
- 24-hour count of vehicle volume
- 24-hour count of vehicle classification
- 24-hour count of speed classification
- 24-hour spot speed samples of trucks and nontrucks
- measurements of the horizontal and vertical alignment using an automated highway geometry data collection technique
- sight distance
- speed limit or advisory speed limit.

Traffic data, such as vehicle and speed classifications, were obtained using Streeter Amet traffic counters. On highways with ADT less than about 15,000, pneumatic tubes were used for vehicle detection. On highways carrying higher volumes of traffic, it was not possible to use tubes because of the likelihood of tube damage. On these routes, the data collection was performed utilizing the nearest permanent induction loops that the VDOT uses for the traffic count program.

Spot speeds of trucks and nontrucks were observed using radar speed detectors of the type used by the state police officers. This type of speed detector sends out the radar signal as a pulse instead of a continuous signal, thereby reducing the chances of being detected by radar detectors frequently used by truck drivers and many other motorists. These radar units enabled the discreet collection of speed samples at each site. These traffic data were collected on weekdays, except Mondays and Fridays, so that the influence of weekend traffic could be eliminated.

Geometric data such as number of lanes and lane and shoulder widths were obtained by direct measurements at each site.

The alignment data, however, required horizontal and vertical roadway profiles of segments of roadways, each between 1 and 2 miles in length. One possible way of obtaining these data was by consulting the as-built plan sheets available from VDOT; however, that would have required the identification of the location of accident sites sufficiently accurately to identify all the relevant plan sheets and, if these were up to date, the obtainment of alignment data from these sheets. Data for sites that could not be located or for which up-to-date plans were not available would have to be obtained by field measurements. An automated highway geometry data collection technique, which avoids such problems by providing a consistent method for obtaining alignment data, was therefore developed for this study. The technique uses the Slopemaster electronic ball bank indicator, which consists of two main components: (1) a ball bank indicator with a digital display of the angle of displacement and (2) a data acquisition unit (DAU), which collects and later prints out data on ball bank indicator readings.

The electronic ball bank indicator provides two modes of measurements: degrees and percentage of grade. When the vehicle is being driven on a roadway with the indicator test switch turned on, all the readings of the indicator are stored in the DAU memory together with the distance from the starting point (in miles to two decimals places) and the speed of the vehicle (to the nearest mph). Once the test run is complete, the DAU prints out all of the readings pertaining to that particular test. The vertical alignment of any segment of highway in terms of percent grades and length of grades can be obtained directly from the output. Actual degrees or radii of horizontal curves cannot, however, be obtained directly because of the effect of the vehicle body roll and the rate of superelevation of the horizontal curve. A special procedure using the DAU was therefore developed for this study. It used two sets of data collected on angular readings of the ball bank indicator on each segment of highway with the vehicle being driven at two different speeds. The radius of each horizontal curve could then be computed using a model developed for the technique and the two sets of data. A detailed description of the technique and the models used to determine the radius at each horizontal curve is given in reference 2. The lengths of horizontal and spiral curves at each test segment were also obtained. A sample of traffic and geometric data is given in Appendix II.

Surrogates of Roadway Alignment

Three surrogate measures of horizontal and vertical alignment were used in this project: the curvature change rate (CCR) used widely in Germany to describe horizontal alignment of a roadway[3,4]; the slope change rate (SCR), which is analogous to CCR; and the absolute mean slope (AMS). CCR and AMS are directly proportional to the degree of variation in horizontal and vertical alignments, respectively. CCR is defined as the absolute sum of the angular changes in horizontal alignment divided by the length of the highway segment, and it is given as

$$\mathrm{CCR} = \left[\sum_{i=1}^{n}\left|\frac{L_i}{R_i}\right| + \sum_{j=i}^{n}\left|\frac{L_s}{2R_i}\right|\right]\frac{(57.3)(5280)}{L} \quad \text{deg/mile} \tag{1}$$

where L_i = length of circular curve i (ft) (see Figure 2)

L_s = length of transition curve s (ft) (see Figure 2)

R_i = radius of circular curve i (ft) (see Figure 2)

L = total length of section (ft) (see Figure 2).

The absolute mean slope is the sum of the absolute grade changes in the vertical alignment divided by the length of the highway segments. It is given as

$$\mathrm{AMS} = \frac{\sum_{i=1}^{k}\left|\frac{G_j + G_{j+1}}{2}\right| l_{j,j+1} + \sum_{i=1}^{k}|G_j L_j|}{L} \quad \text{percent} \tag{2}$$

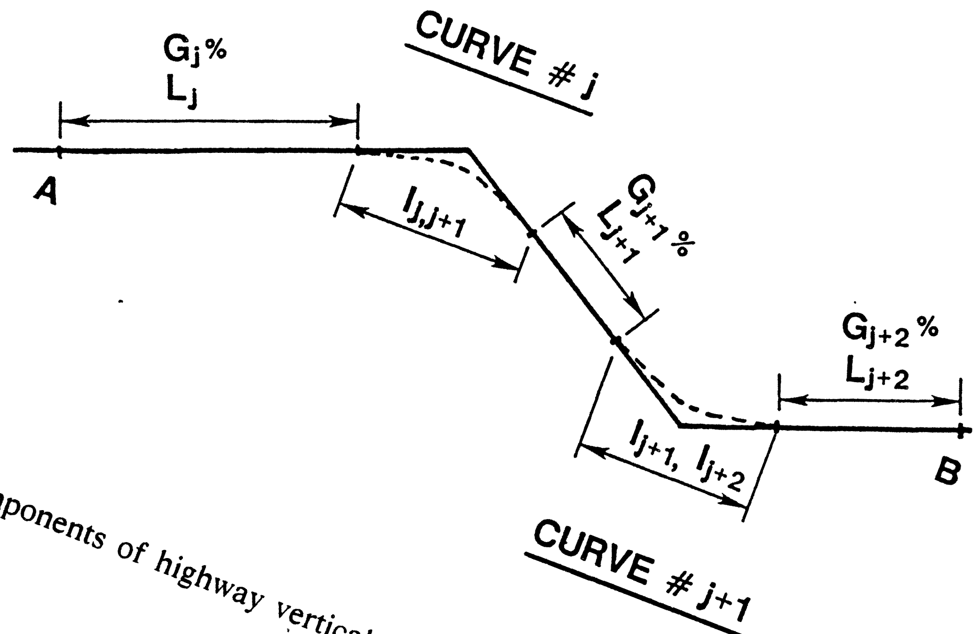
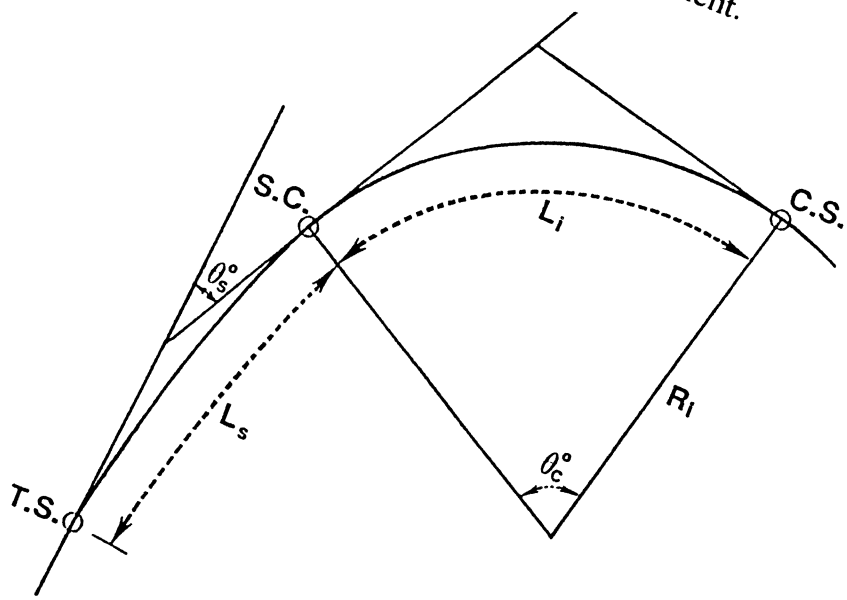


Figure 1. Components of highway vertical alignment.



ure 2. Components of highway horizontal alignment.

where R_i = radius of horizontal curve i (see Figure 1)

L_i = length of horizontal curve i (see Figure 1)

Θ_{ci} = deflection angle of horizontal curve i = 57.3 L/R (see Figure 1)

L_s = length of spiral curve s (see Figure 1)

Θ_s = deflective angle of spiral curves = $L_s \dfrac{(57.3)}{2R_i}$ (see Figure 1)

G_j = slope of jth grade percent (see Figure 1)

L_j = length of jth slope (ft) (see Figure 1)

L = length of entire segment (ft) (see Figure 1)

$l_{j,\ j+1}$ = length of curve between jth and $(j + 1)$th slopes (ft) (see Figure 1).

The slope change rate is given as

$$\text{SCR} = \left[\sum_{j+1}^{k} |G_{j=1} - G_i| \right] \left(\frac{5280}{L} \right) \text{ percent per mile} \tag{3}$$

where G_j = slope of jth grade percent (see Figure 1)

L = length of entire segment (ft) (see Figure 1).

DEVELOPMENT OF MODELS

The development of models entailed the development of mathematical relationships between the probability of any given accident being one that involves a large truck and a set of independent variables that are related to large-truck accidents.

The accident-related independent variables considered were:

- roadway geometry
  - number of lanes
  - lane width (LNWD)
  - shoulder width (SHLDWD)
  - curvature change ratio (CCR)
  - absolute mean slope (AMS)
  - slope change rate (SCR)
  - segment length (SEGLEN)

- traffic variables
  - average annual daily traffic (AADT)
  - mean speed (all vehicles)
  - speed variance (all vehicles)
  - mean speed (trucks)
  - speed variance (trucks)
  - mean speed (nontrucks)
  - percentage large trucks (TPERCENT)
  - difference in mean speeds between truck and nontruck (SPDIF).

The accident-related dependent variables were

- TINVOL (number of large-truck accident involvements in 1 year)
- P_i (probability of a large-truck accident involvement).

A list of the locations of the sites used and sample data are shown in Appendix I. Models were developed for three highway environments grouped by roadway configurations and traffic volumes:

- Environment I: primary highways (undivided 4-lane and 2-lane)
- Environment II: primary highways (divided 4-lane) (AADT $\leq$ 15,000)
- Environment III: interstate/primary highways (divided 4-lane) (AADT > 15,000).

Model Selection Criteria

The selection of the best model to fit the observed data is a problem that must be contended with in any statistical modeling task. To assist in this endeavor, various model selection criteria have been used in the past, such as adjusted multiple correlation, Mallow's C_p, prediction sum of squares (PRESS), etc. Akaike's information criterion (AIC), which was used in this study, is a criterion that is increasingly becoming popular for its versatility. The process of selecting the best model is usually a complex task in the absence of suitable criteria for that purpose. AIC, however, fulfills that role and has proven to be an effective criterion.[5]

Components of AIC

An objective measure of the distance between the true model and the hypothesized model is Boltzmann's generalized entropy, or the negentropy. This measure is also known as the Kullback-Liebler information quantity.[6] Model selection using AIC is based on the concept of entropy maximization or, in other words, the minimization of the negentropy. In AIC, the Kullback-Liebler information quantity is estimated by the mean log likelihood for the model. A complete explanation of the derivation of this criterion and its extensions is given by Bozdogan.[7] For the purpose of this analysis, the criterion is described in its final form.

Definition: If $\{M_k : k \in K\}$ is a set of competing models indexed by $k = 1, 2, \ldots, K$, then the criterion AIC is given by

$$\mathrm{AIC}(k) = -2 \log L(\hat{\Theta}_k) + 2k \tag{4}$$

where $\log L(\hat{\Theta}_k) = \log_e[\text{maximized likelihood}]$

k = number of free parameters in the model.

AIC is minimized to choose the model M_k over the set of competing models.

The first term in equation 4 is a measure of the badness of fit, or bias, when the maximum likelihood estimates of the parameters are used. The second term, $2k$, is a measure of the complexity of the model and compensates for the first term.

Using this criterion, the model yielding the minimum AIC was selected as the best model. The resulting model is the one with the least complexity and highest level of information.

Multiple Linear Regression Model

This is the simplest case of generalized linear models and can be expressed as

$$\underline{Y} = \underline{X}\beta + e \tag{5}$$

where $\underline{Y}$ is an N x 1 response vector

$\underline{X}$ is an N x p matrix of explanatory variables

β is a p x 1 vector of parameters, i.e., an N x 1 random vector whose elements are independent identically and normally distributed, i.e.,

$e_i \approx N(O, \sigma^2)$ for $i = 1, \ldots, N$.

AIC Derivation for Multiple Regression Models

A least square estimation of parameters was carried out using procedures available in the statistical software package SAS. The parameters, β_0, β_1, ..., β_p were estimated through the following minimization.

Minimize $S(\beta_0, \beta_1, \ldots, \beta_p)$

$$= \sum_{i=1}^{n} \left[y_i - E\left\{ \frac{y_i}{x_{i1},\ x_{i2},\ \ldots,\ x_{in}} \right\} \right]^2 \tag{6}$$

$$= \sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1 x_{i1} - \ldots - \beta_p x_{ip}\}^2. \tag{7}$$

In matrix form

$$S(\beta) = (\underline{Y} - \underline{X}\beta)'\ (\underline{Y} - \underline{X}\beta). \tag{8}$$

The vector of partial derivatives of $S(\beta)$ with respect to β elements is

$$\frac{\partial S(\beta)}{\partial \beta} = -2\underline{X}'\underline{Y} + 2\underline{X}'\underline{X}\beta = 0 \tag{9}$$

which gives

$$\hat{\beta} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y} \tag{10}$$

assuming that

$$y_i \approx N(x_i\beta, \sigma^2) \tag{11}$$

$$\varepsilon_i \approx N(0, \sigma^2) \tag{12}$$

for $i = 1, \ldots, n$.

The likelihood function of the dependent variable observation vector $\underline{Y}$ is

$$L(\beta, \sigma^2) = F(Y, \beta, \sigma^2) \tag{13}$$

$$= \left[\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \right] \exp \left[\left(-\frac{1}{2\sigma^2} \right) (\underline{Y} - \underline{X}\beta)'(\underline{Y} - \underline{X}\beta) \right]. \tag{14}$$

The likelihood function is

$$L(\beta, \sigma^2) = \log L(\beta, \sigma^2) \tag{15}$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\underline{Y} - \underline{X}\beta)'(\underline{Y} - \underline{X}\beta). \tag{16}$$

The maximum likelihood estimates $\hat{\beta}$ of β is obtained by the following partial derivatives

$$\frac{d}{d\beta} L(\beta, \sigma^2) = 0 \tag{17}$$

$$\frac{d}{d\sigma^2} L(\beta, \sigma^2) = 0 \tag{18}$$

resulting in

$$\sigma^2 = \frac{(\underline{Y} - \underline{X}\beta)'(\underline{Y} - \underline{X}\beta)}{n} = \frac{\underline{ee'}}{n} = \frac{SSE}{n}. \tag{19}$$

Now the maximized log likelihood becomes

$$\log L(\theta_k) = (\beta, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\frac{SSE}{n} - \frac{n}{2} \tag{20}$$

but

$$\text{AIC} = -2\log(\hat{\Theta}_k) + 2k \tag{21}$$

which gives

$$\text{AIC} = n\log 2\pi + n\log\left\{\frac{SSE}{n}\right\} + n + 2k \tag{22}$$

where $k = (p + 1) + \ell$

$p + 1$ = number of β coefficients estimated

ℓ = estimated parameters, σ^2.

Identification of Best Multiple Linear Regression Models

Using AIC as the criterion for model selection, the best model was the model yielding the minimum AIC value. A stepwise procedure was used in the selection of the best models. The best subset of singleton variables was first determined by using a backward elimination procedure. The model that included all variables and its AIC was first determined. The best combination of variables to yield the minimum AIC at the next lower level (i.e., with the number of variables reduced by one) was then determined. At each level (i.e., number of variables) of the variable selection process, all the terms of the expression for AIC remain the same except for SSE. Therefore, the best submodel would be the model yielding the minimum SSE, which is also the model with the maximum R^2 at that level. Using PROC RSQUARE of SAS, the best subset of variables at each level was thus determined. The AICs for all levels are then compared with each other, and the model with the minimum AIC is then selected as the best model.

Poisson Regression Model

The occurrence of highway accidents can be described by the nonstationary Poisson process.[6,7] The basic assumption of Poisson processes is that the numbers of accidents occurring within each observed time interval (1 year in our model) are independent with the expectation defined as in the following equation

$$E(y_{ij}) = f(x_i, \beta) \qquad (23)$$

$$i = 1, \ldots, n$$

$$j = 1, \ldots, m_i$$

where $x_i = x_{i1}, x_{i2}, \ldots, x_{i,p-1}$ is the ith set of values for the $p - 1$ independent variables

m_i = number of replications of ith experimental condition

$\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1})$ is a p-dimensional vector of parameters

Y_{ij} = a particular realization of the experiment.

It is assumed that a general form of the model exists, and $f(x_i, \beta)$ is a differential function of β. The experiment yields n values of the independent variables, where n is supposed to be sufficiently greater than p to ensure the estimability of the β parameters.

The probability of k accidents occurring during each interval (1 year) can be represented as

$$P_i(k) = \frac{e^{-y_i} y_i^k}{k!}. \qquad (24)$$

In the estimation procedure, we will determine the parameter vector β

$$y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} \ldots x_i^{\beta} + \epsilon, \qquad (25)$$

where y = expected accident frequency per year

x_i = traffic and highway geometry variables.

Three different methods can be used to estimate the parameters of the Poisson regression model. These are the maximum likelihood principle, weighted least squares analysis, and the minimum chi-square estimation. The weighted least squares method of parameter estimation was used in this analysis since it is more efficient in terms of computing time.

AIC Derivation Poisson Regression Model

The AIC derivation for least squares estimation of the parameters of this model is basically the same as that for the linear model described earlier.

Identification of the Best Poisson Regression Model

The model investigated was the multiplicative model described by equation 25 for the expected accident involvement. The parameter estimation was carried out in two steps. First, the model was linearized through a logarithmic transformation and parameters were estimated treating the transformed function as a linear model. Using the backward procedure described earlier, the best model at each level was then determined based on the AIC values. Then, starting out with these parameter estimates and the corresponding models, a more precise estimation of the nonlinear model was carried out using the PROC NLIN procedure in SAS. This produced least squares estimates of the parameters obtained through the Marquardt iterative method, in which the residuals are regressed on to the partial derivatives of the model with respect to the parameters until the iteration converges.

Multiple Logistic Regression Model

This is a commonly used generalized linear model based on the binomial distribution $Y_i \approx b(n, \pi_i)$. It is obtained by taking the natural paramenter as the link function

$$\log \frac{\pi_i}{(1 - \pi_i)} = [X_i]^{T_\beta} \tag{26}$$

$$\pi_i = \frac{e^{(X_i)^{T_\beta}}}{1 + e^{(X_i)^{T_\beta}}} \tag{27}$$

where $$(X_i)^{T_\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_N x_N. \tag{28}$$

For each site, the truck and nontruck accident involvements that have taken place during each year of analysis were determined from accident history. Thus, the probability of any given accident involvement being one that involved a truck was given by

$$\pi_i = \frac{\text{Total truck involvements}}{\text{Total (truck + nontruck) involvements}}.$$

The fundamental assumption in logistic regression analysis is that the logit or the logarithm of the odds is linearly related to the independent variables. No assumptions are made regarding the distributions of the x variables. An advantage of this model is that the explanatory or independent x variables may be discrete, continuous, or categorical.

AIC Derivation for Logistic Regression Model

From the basic definition of the binomial model, where the probability of an outcome (in this analysis a truck involvement) is (π_i) for such random variables $Y_1, \ldots Y_n$, their joint probability is

$$\prod_{j=1}^{n} (\pi_j) y_j (1 - \pi_j)^{1-y_j} = \exp\left[\sum_{j=1}^{n} y_j \log\left(\frac{\pi_j}{(1-\pi_j)}\right) + \sum_{j=1}^{n} \log(1-\pi_i)\right]. \tag{29}$$

Taking into consideration the number of possible ways in which N such outcomes can take place, the resulting $\log_e$ of the likelihood function is

$$l(\pi_1, \ldots, \pi_N;\ r_1, \ldots, r_N) = \sum_{i=1}^{N}\left[r_i \log\frac{\pi_i}{(1-\pi_i)} + n_i \log(1-\pi_i) + \log\binom{n_i}{r_i}\right]. \tag{30}$$

For the linear logit model this is equivalent to

$$l = \sum_{i=1}^{N}\left[r_i\left\{x_i^{T\beta}\right\} - n_i \log\left(1 + e^{x_i^{T\beta}}\right) + \log\binom{n_i}{r_i}\right]. \tag{31}$$

The maximum likelihood estimates of parameters β and of the probabilities π_i are obtained by maximizing the above function.

For the definition previously given in equation 4

$$\mathrm{AIC}(k) = -2\log L(\hat{\Theta}_k) = 2k$$

the PROC CATMOD with NOGL and ML options available in SAS were used to determine maximum likelihood estimates of the parameters β and then the first term of the expression for AIC. The value of AIC for each model was then determined by adding the value of the second term ($2k$).

A stepwise logistic regression procedure and AIC were then adopted for the selection of the best subset of variables. The best subset of singleton variables was first determined by using a backward elimination procedure. Starting out with a model that includes all variables, the best combination of variables to yield the minimum AIC at the next lower level (i.e., with the number of variables reduced by one) was determined. This procedure was continued until the model with one variable was reached. The resulting list of models represents the best models with singleton variables at each level. Once the best subset of variables was determined, interaction terms were then determined for each baseline model and added to improve on the AIC value.

RESULTS

Multiple Linear Regression Models

Linear regression models were developed for the annual truck accident involvement rate (TRATE) (i.e., accidents per 100 million truck VMT) as a linear function of highway- and traffic-related variables. The results obtained are given for the three different environments described earlier.

Environment I (Primary Highways, Undivided 4-Lane and 2-Lane)

Table 1 gives the best model for TRATE at each level for this environment. It can be seen that the model with eight variables has the minimum AIC of 288.46 and is therefore the best model. This model is given as

$$\begin{aligned}\text{TRATE} &= 289.233 - 27.249(\text{LNWD}) + 6.6192(\text{SHLDWD}) - 26.413(\text{AMS})\\ &+ 1.9328(\text{SCR}) + 0.0269(\text{CCR}) + 0.00492(\text{ADT}) - 1.3822(\text{TPERCENT})\\ &+ 0.2362(\text{SPDIFSQ}) \qquad (32)\end{aligned}$$

$R^2 = 0.6832$

Table 1

LINEAR MODELS FOR TRATE — ENVIRONMENT I

Level	SSE	AIC	Variables
1	5251.91	307.89	ADT
2	5079.73	308.59	SCR, ADT
3	4851.08	308.79	SCR, ADT, TPERCENT
4	4475.27	307.65	AMS, SCR, ADT, SPDIFSQ
5	3454.94	299.56	SHLDWD, SCR, ADT, TPERCENT, SPDIFSQ
6	2532.25	289.44	LNWD, SHLDWD, AMS, SCR, ADT, SPDIFSQ
7	2369.44	288.85	LNWD, SHLDWD, AMS, SCR, CCR, ADT, SPDIFSQ
8	2228.90	288.46*	LNWD, SHLDWD, AMS, SCR, CCR, ADT, SPDIFSQ, TPERCENT

*Best model.

Environment II (Primary Highways, Divided 4-Lane, AADT $\leq$ 15,000)

The best model at each level is shown in Table 2, which indicates that the best overall model is that consisting of three variables with an AIC of 245.21. It is given as

$$\text{TRATE} = -12.196 + 0.0231(\text{CCR}) + 0.00077(\text{ADT}) + 0.6444(\text{TPERCENT}) \quad (33)$$

$R^2 = 0.2187$

Table 2

LINEAR MODELS FOR TRATE — ENVIRONMENT II

Level	SSE	AIC	Variables
1	295.01	250.93	SCR
2	274.10	248.96	CCR, TPERCENT
3	246.38	245.21*	CCR, ADT, TPERCENT
4	243.30	246.53	CCR, ADT, TPERCENT, SPDIFSQ
5	241.22	248.06	AMS, SCR, CCR, ADT, TPERCENT
6	239.75	249.73	AMS, SCR, CCR, ADT, TPERCENT, SPDIFSQ
7	239.15	251.60	LNWD, AMS, SCR, CCR, ADT, TPERCENT, SPDIFSQ
8	239.08	253.58	LNWD, SHLDWD, AMS, SCR, CCR, ADT, TPERCENT, SPDIFSQ

*Best model.

Environment III (Interstate and Primary Divided Highways, AADT > 15,000)

Table 3 shows that the best model for this environment is that consisting of only two variables and is given as

$$\text{TRATE} = 5.416 + 0.1119(\text{CCR}) - 0.1580(\text{TPERCENT}) \quad (34)$$

$R^2 = 0.2317$

Using R^2 as a simple measure to compare how well each model represented the variation in data, only the model for Environment I, with an R^2 value of 0.68, demands any consideration. This model implies that an increase in SCR, CCR, ADT, SHLDWD, or SPDIFSQ increases the TRATE, whereas an increase in LNWD, AMS, or TPERCENT decreases the TRATE. An increase in the variable SHLDWD resulting in an increase in TRATE is contrary to expectations. However, this may be the result of observed correlation between ADT and SHLDWD. Also there is no logical explanation for an increase in either AMS or TPERCENT resulting in a decrease in TRATE. Because of this and the low R^2 values obtained, it can be concluded that linear models do not adequately describe the relationship between TRATE and traffic and geometric variables.

Table 3

LINEAR MODELS FOR TRATE — ENVIRONMENT III

Level	SSE	AIC	Variables
1	1470.14	871.12	TPERCENT
2	1414.37	866.39*	CCR, TPERCENT
3	1408.07	867.61	CCR, ADT, TPERCENT
4	1402.29	868.89	CCR, ADT, TPERCENT, SPDIFSQ
5	1398.58	870.43	AMS, SCR, CCR, TPERCENT, SPDIFSQ
6	1396.51	872.18	AMS, SCR, CCR, ADT, TPERCENT, SPDIFSQ

*Best model.

Poisson Regression Models

Environment I (Primary Highways, Undivided 4-Lane and 2-Lane)

The best nonlinear model at each level is shown in Table 4. The best overall model for this environment is the model at the third level with variables SCR, ADT, and TPERCENT. The parameters for these variables were then used as the starting values for the iterative least squares estimation. The resulting model is

$$\mathrm{TINVOL} = .015237(\mathrm{SCR})^{0.0577}(\mathrm{ADT})^{0.5024}(\mathrm{TPERCENT})^{0.5731} \tag{35}$$

$$\mathrm{AIC} = 62.06$$

This model indicates that ADT, SCR, and TPERCENT are variables that are significant in predicting the expected number of truck accident involvements. The low

Table 4

POISSON REGRESSION MODEL — ENVIRONMENT I

Level	SSE	AIC	Log Variables
1	10.29	64.71	LNWD
2	9.14	62.09	LNWD, AMS
3	8.67	62.06*	SCR, ADT, TPERCENT
4	8.36	62.60	LNWD, SHLDWD, AMS, SEGLEN
5	7.88	62.32	LNWD, SHLDWD, SEGLEN, TPERCENT, SPDIFSQ
6	7.80	63.94	LNWD, SHLDWD, CCR, SEGLEN, TPERCENT, SPDIFSQ
7	7.77	65.76	AMS, SCR, CCR, SEGLEN, ADT, TPERCENT, SPDIFSQ
8	7:70	67.41	LNWD, SHLDWD, AMS, CCR, SEGLEN, ADT, TPERCENT, SPDIFSQ
9	7.67	69.28	LNWD, SHLDWD, AMS, SCR, CCR, SEGLEN, ADT, TPERCENT

*Best model.

1726

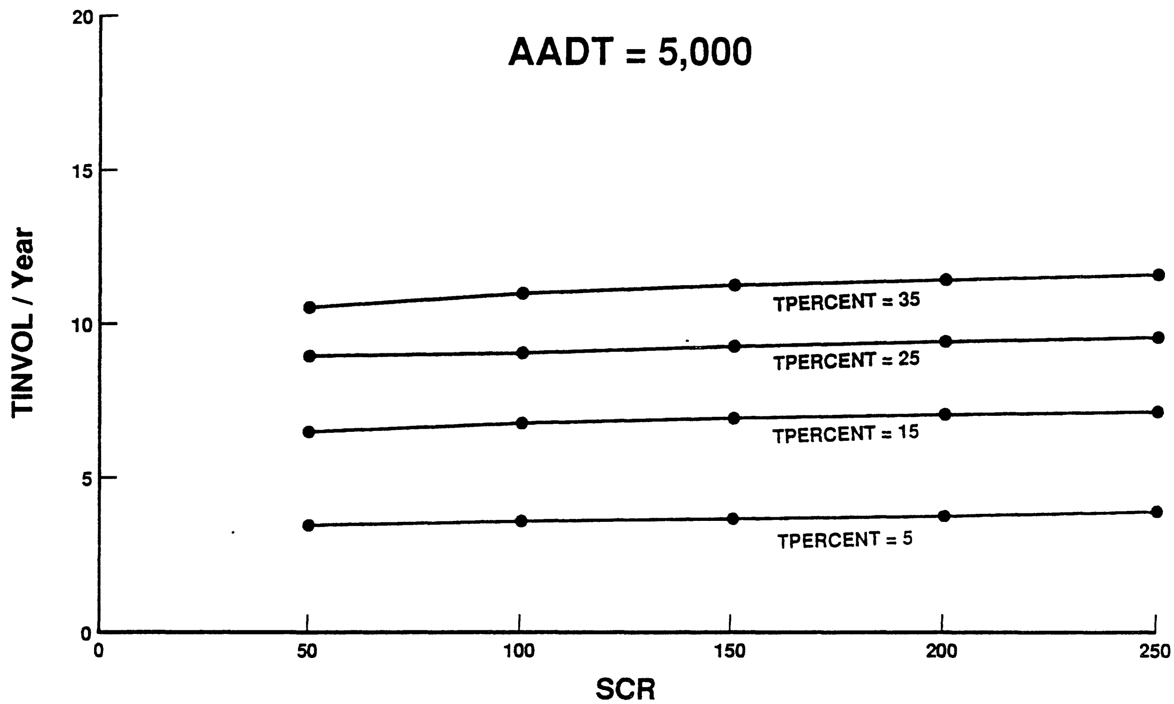


Figure 3. Effect of SCR and TPERCENT on TINVOL — Environment I (segment length equals 2 miles).

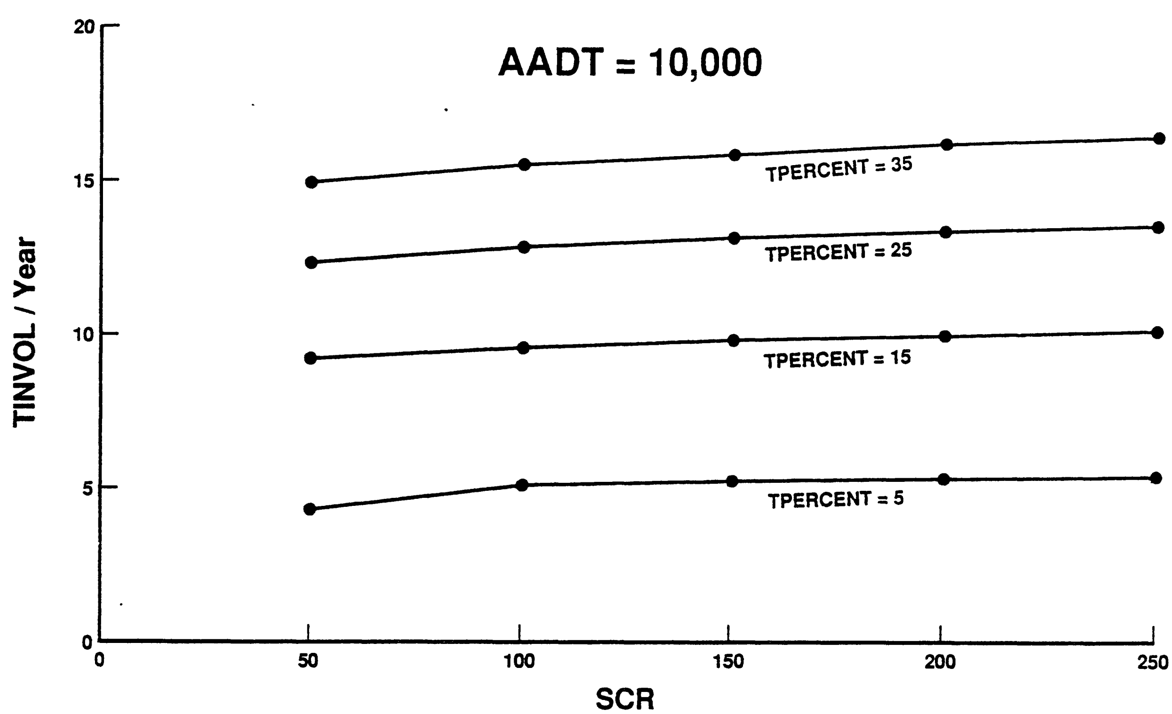


Figure 4. Effect of SCR and TPERCENT on TINVOL — Environment I (segment length equals 2 miles).

AIC value obtained for this model indicates that it is a good predictor for the number of truck involvements. Intuitively, one would expect to see the variables SEGLEN and LNWD included in the final model. SEGLEN was, however, not included in this model since all of the sites in this environment were about 2 miles in length. The lane width (LNWD) is also not included since the analysis showed some correlation between ADT and LNWD.

Figures 3 and 4 show how the number of truck involvements vary with the large-truck percentage (TPERCENT) and SCR for this model. The number of involvements shown is the expected total for 1 year. These figures indicate that, according to this model, the number of truck involvements is almost doubled as the percentage of trucks increases from 5 to 15 percent.

Environment II (Primary Highways, Divided 4-Lane, AADT $\leq$ 15,000)

The best nonlinear models for this environment at different levels are shown in Table 5. From these models, the best overall model is the model at the fourth level, with variables SCR, ADT, TPERCENT, and SEGLEN. The parameters for these variables were used as the starting values for the iterative least squares estimation. The resulting model is

$$\text{TINVOL} = 9 \times 10^{-8}(\text{SCR})^{0.0471}(\text{ADT})^{1.4358}(\text{TPERCENT})^{1.5232}(\text{SEGLEN})^{0.3826} \quad (36)$$

$$\text{AIC} = 83.56$$

This model indicates that SCR, ADT, TPERCENT, and SEGLEN are the best descriptors of the TINVOL for a particular segment of highway.

Table 5

POISSON REGRESSION MODEL — ENVIRONMENT II

Level	SSE	AIC	Log Variables
1	15.57	92.10	ADT
2	13.09	84.71	ADT, TPERCENT
3	12.58	84.59	SCR, ADT, TPERCENT
4	11.90	83.56*	SCR, SEGLEN, ADT, TPERCENT
5	11.88	85.48	SCR, CCR, SEGLEN, ADT, TPERCENT
6	11.82	87.21	AMS, SCR, CCR, SEGLEN, ADT, TPERCENT
7	11.79	89.09	SHLDWD, AMS, SCR, CCR, SEGLEN, ADT, TPERCENT
8	11.72	90.77	LNWD, SHLDWD, AMS, SCR, CCR, SEGLEN, ADT, TPERCENT
9	11.64	92.40	LNWD, SHLDWD, AMS, SCR, CCR, SEGLEN, ADT, TPERCENT

*Best model.

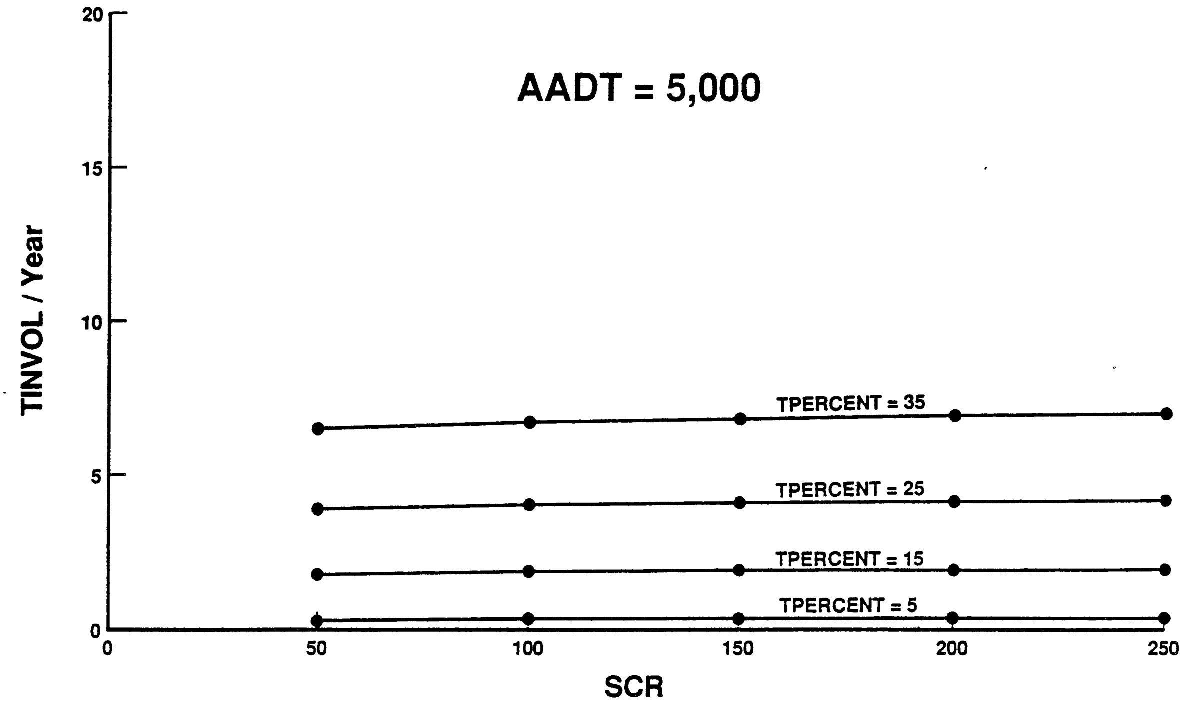


Figure 5. Effect of SCR and TPERCENT on TINVOL — Environment II (segment length equals 2 miles).

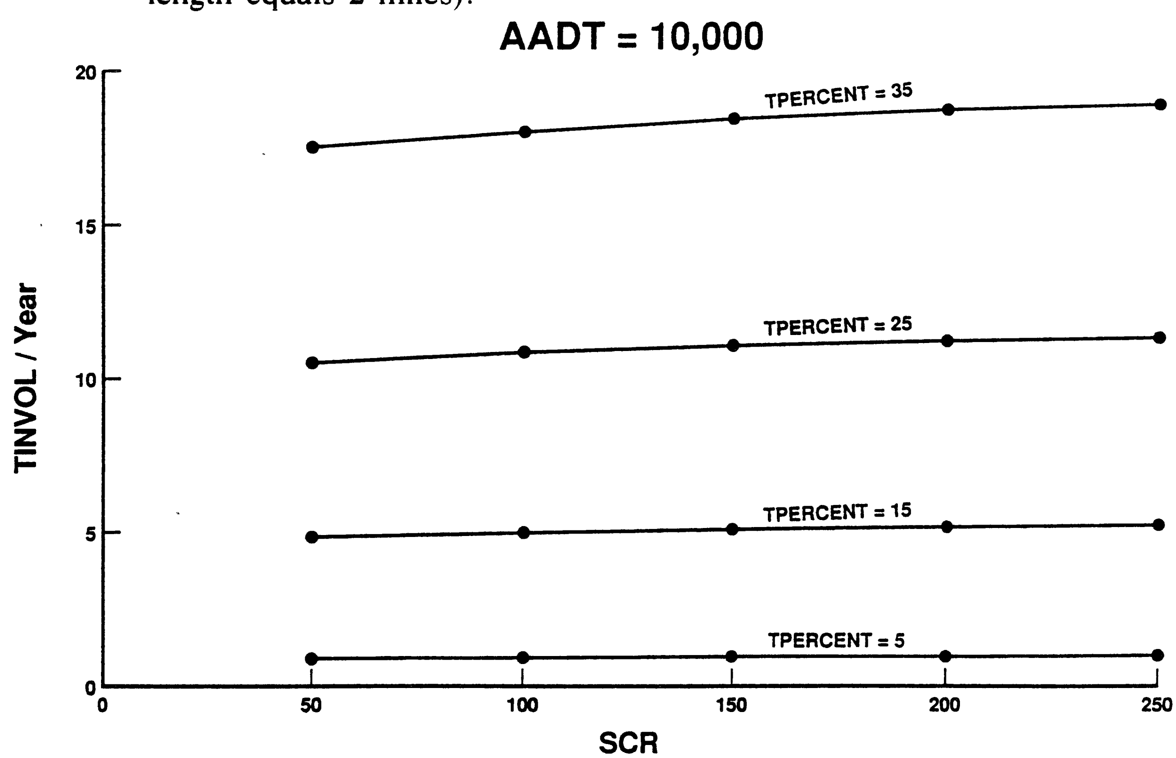


Figure 6. Effect of SCR and TPERCENT on TINVOL — Environment II (segment length equals 2 miles).

Figures 5 and 6 show how the number of truck involvements vary with TPERCENT and SCR using this model and a segment length of 2 miles. This model indicates a relationship between truck percentages, truck involvement, and SCR similar to that for Environment I.

Environment III (Interstate and Primary Divided Highways, AADT > 15,000)

The best nonlinear models for this environment at each level are shown in Table 6. It can be seen that the AIC values for the models at the fourth and fifth levels are almost identical. However, the model at the fifth level was selected since it included TPERCENT, which has been shown to have some effect on TINVOL. The other variables are CCR, ADT, SEGLEN, and SPDIFSQ. The parameters for these variables were then used as the starting values for the iterative least squares estimation. The resulting model is

$$\text{TINVOL} = .001465(\text{CCR})^{0.0336}(\text{SEGLEN})^{0.3318}(\text{ADT})^{0.7086}(\text{TPERCENT})^{0.2064}(\text{SPDIFSQ})^{0.0475} \quad (37)$$

$$\text{AIC} = 407.48$$

Table 6

POISSON REGRESSION MODEL — ENVIRONMENT III

Level	SSE	AIC	Log Variables
1	107.23	415.55	ADT
2	101.59	408.15	SEGLEN, ADT
3	100.07	407.54	CCR, SEGLEN, ADT
4	98.89	407.47*	CCR, SEGLEN, ADT, SPDIFSQ
5	97.76	407.48	CCR, SEGLEN, ADT, TPERCENT, SPDIFSQ
6	97.56	409.13	SCR, CCR, SEGLEN, ADT, TPERCENT, SPDIFSQ
7	97.27	410.59	AMS, SCR, CCR, SEGLEN, ADT, TPERCENT, SPDIFSQ

*Best model.

The model indicates that the difference between the average speeds of trucks and nontrucks has some effect on large-truck involvement in accidents on highways within this environment. Figures 7 and 8 show how the number of truck involvements vary with the CCR and speed difference. These figures indicate that both increasing the speed difference and CCR tend to increase the number of truck accidents.

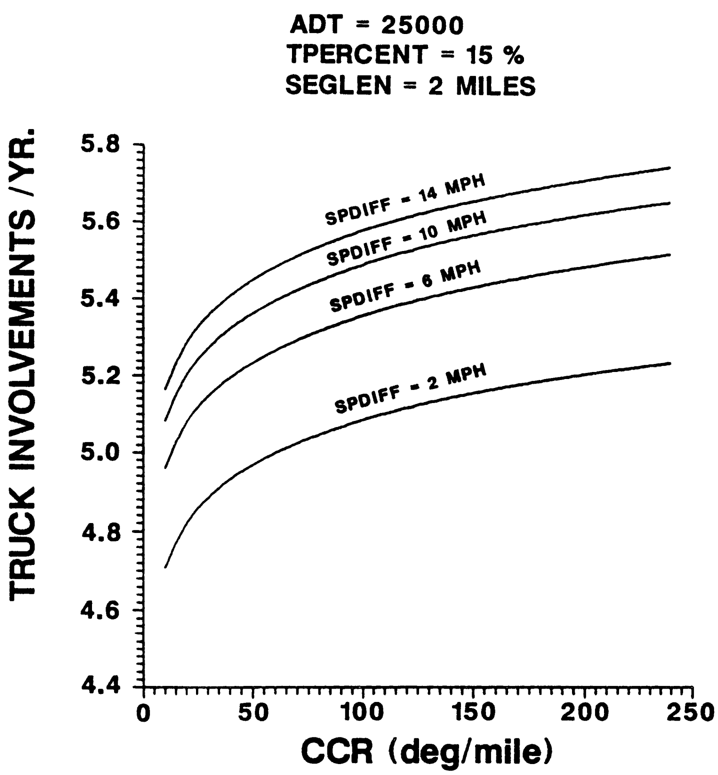


Figure 7. Effect of CCR and speed difference on truck involvements — Environment III.

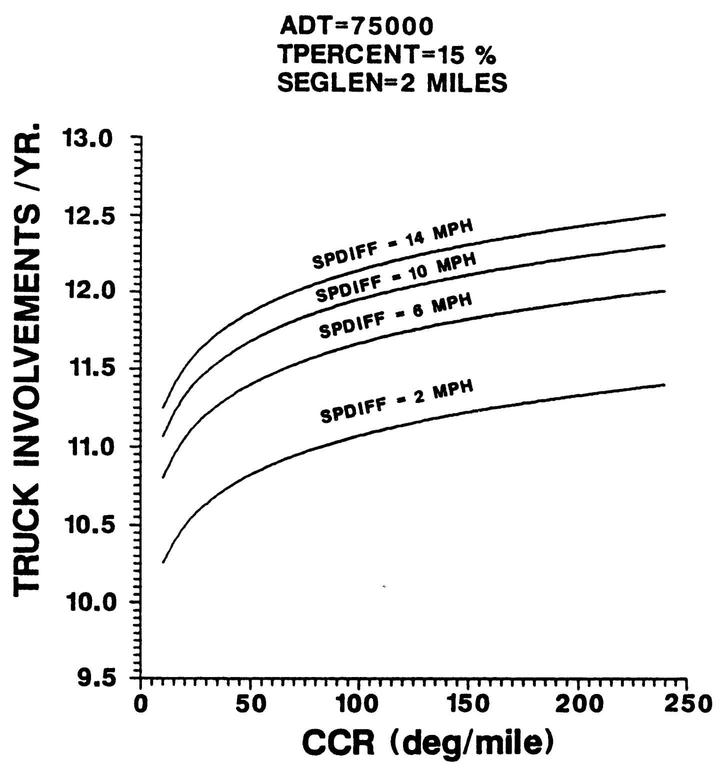


Figure 8. Effect of CCR and speed difference on truck involvements — Environment III.

Logistic Regression Model for Environment I
(Primary Highways, Undivided 4-Lane and 2-Lane)

Table 7 shows a partial list of the best set of singleton variables at each level with the corresponding AIC values for each model.

Based on these AIC values, the model with seven variables is the best model with singleton variables for Environment I. This model has an AIC value of 801.808. The following interaction terms were identified as the best possible interaction terms for the saturated model constructed with the seven variables determined earlier:

1. (SHLDWD)(AMS)
2. (SHLDWD)(SCR)
3. (LNWD)(AMS)
4. (AMS)(SCR)
5. (LNWD)SCR)
6. (SPDIFSQ)(CCR).

Table 7

LOGISTIC REGRESSION MODEL — ENVIRONMENT I

Lvl	L(QK)	AIC	Variables
1	822.298	826.298	LNWD
2	812.435	818.435	LNWD, CCR
3	811.101	819.101	LNWD, CCR, SPDIFSQ
4	800.648	810.648	LNWD, CCR, SPDIFSQ, AMS
5	797.176	809.176	LNWD, CCR, SPDIFSQ, AMS, SCR
6	791.030	805.030	LNWD, CCR, SPDIFSQ, AMS, SCR, SHLDWD
7	785.808	801.808*	LNWD, CCR, SPDIFSQ, AMS, SCR, SHLDWD, ADT
8	785.749	803.749	LNWD, CCR, SPDIFSQ, AMS, SCR, SHLDWD, ADT, TPERCENT
9	785.741	805.741	LNWD, CCR, SPDIFSQ, AMS, SCR, SHLDWD, ADT, TPERCENT, SEGLEN

*Best model.

These variables were then entered into the base model in a stepwise manner, and the resulting best interaction terms were determined as the ones that yielded the minimum AIC value. In this particular case, there was only one such term (SHLDWD x SCR) resulting in a final AIC of 797.47. Further attempts to refine this model by reducing the number of singleton variables resulted in the following model

$$P_i = \frac{1}{1 + e^{-\beta_x}} \tag{38}$$

where P_i = probability of a large-truck accident involvement

$$\beta_x = 13.648 - 1.164(LNWD) - 0.9095(SHLDWD) - 0.1969(SCR) + 0.0501(SHLDWD)(SCR) \quad (39)$$

AIC = 794.759.

Although the variables AMS and SPDIFSQ were initially selected as significant for this model, inclusion of the interaction terms have eliminated them from the best model describing the observed data. The effect of the variables on the probability of a truck involvement was investigated through this model. Figure 9 shows that lane width has a significant effect on the occurrence of truck accidents. For instance, the increase of lane width from 9.5 feet to 10.5 feet on a segment of roadway about 2 miles in length with a shoulder width of 4.0 feet and an SCR of 20 degrees per mile will reduce the likelihood of a truck involvement from 0.39 to 0.17, a reduction of about 56 percent. It can also be shown that on a segment of road with twice the above SCR value, this reduction will amount to about 50 percent. Figure 10 indicates that increasing the slope change rate of a roadway tends to increase the probability of truck involvements.

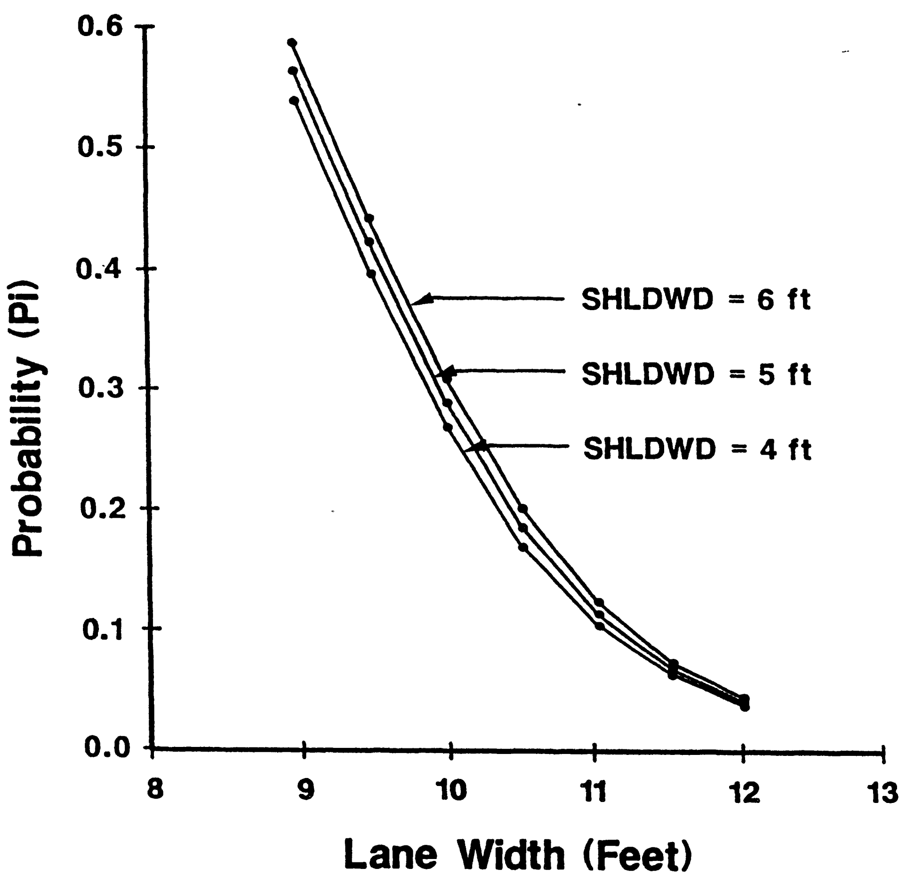


Figure 9. Effect of lane and shoulder widths — Environment I (SEGLEN = 2 miles, SCR = 20 deg/mile).

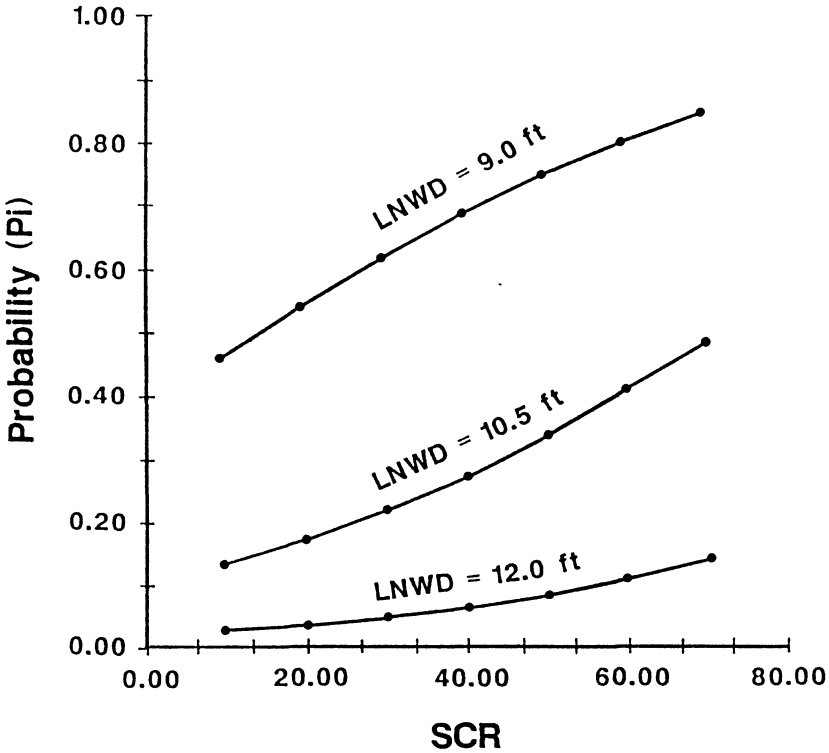


Figure 10. Effect of SCR — Environment I (4-ft shoulder width) (SEGLEN of 2 miles).

Logistic Regression Model for Environment II (Primary Highways, Divided 4-Lane AADT $\leq$ 15,000)

The best logistic models with singleton variables for this environment are shown in Table 8. The best model from among these is the model with TPERCENT as a single variable. However, the AIC values for 1st through 5th order models are separated only by small differences, indicating proximity of these models to each other. Therefore, in searching for the best model with interaction terms, the first five models were considered. The resulting best model has an AIC value of 500.83, which is an improvement on the AIC value of 507.403 for the best model with singleton variables. The best model for this environment is given as

$$P_i = \frac{1}{1 + e^{-\beta_x}} \tag{40}$$

where $\beta_x = -10.956 + 0.999(\text{TPERCENT}) + 0.1356(\text{SHLDWD}) + 1.2143(\text{SEGLEN}) - 0.0611(\text{TPERCENT})(\text{SHLDWD}) - 0.2503(\text{TPERCENT})(\text{SEGLEN}) + 0.3684(\text{SHLDWD})(\text{SEGLEN}).$

Table 8

LOGISTIC REGRESSION MODEL — ENVIRONMENT II

Lvl	L(QK)	AIC	Variables
1	503.403	507.403*	TPERCENT
2	502.800	508.800	TPERCENT, LNWD
3	500.318	508.318	TPERCENT, LNWD, SHLDWD
4	498.500	508.500	TPERCENT, LNWD, SHLDWD, SEGLEN
5	496.886	508.886	TPERCENT, LNWD, SHLDWD, SEGLEN, CCR
6	495.462	509.462	TPERCENT, LNWD, SHLDWD, SEGLEN, CCR, SPDIFSQ
7	495.330	511.330	TPERCENT, LNWD, SHLDWD, SEGLEN, CCR, SPDIFSQ, ADT
8	495.285	513.285	TPERCENT, LNWD, SHLDWD, SEGLEN, CCR, SPDIFSQ, ADT, AMS
9	495.284	515.284	TPERCENT, LNWD, SHLDWD, SEGLEN, CCR, SPDIFSQ, ADT, AMS, SRC

*Best model.

This model indicates that the most significant variables are the percentage of trucks, shoulder width, and the length of the roadway segment considered, with interactions among all three variables. An investigation of the model suggests that although increasing truck percentages increases the likelihood of truck accidents as predicted by

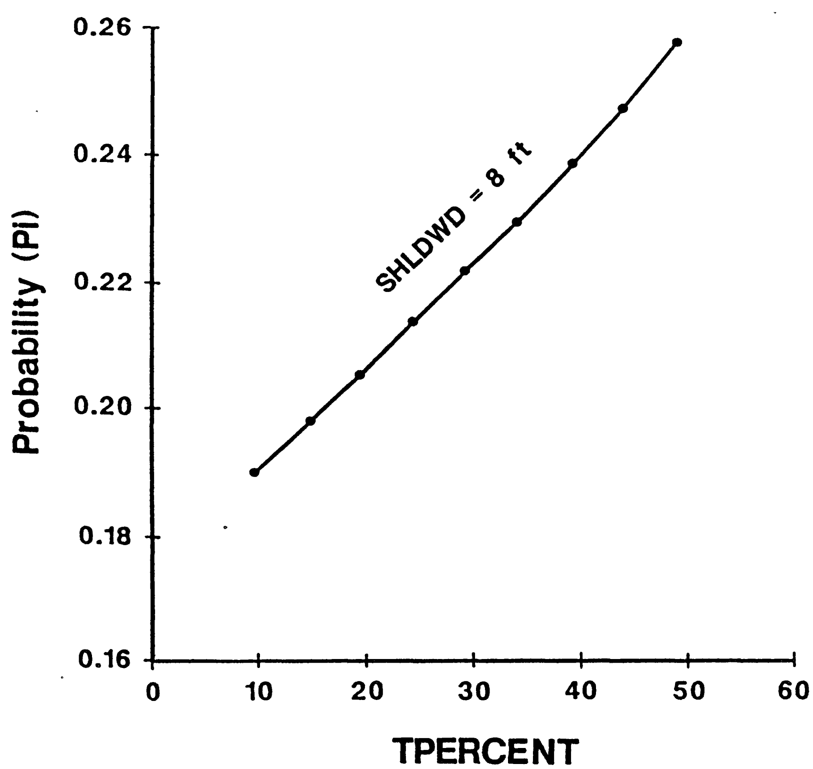


Figure 11. Effect of TPERCENT — Environment II.

this model, this increase is not very high. For instance, on a 2-mile highway segment with a shoulder width of 8 feet, an increase of TPERCENT from 16 to 50 increased the probability of a truck accident from .20 to .27. The indication that an increase in SHLDWD could increase truck involvements does not have any physical interpretations. However, an explanation for this irregularity in the model may be the result of an observed correlation of 0.50 between SHLDWD and ADT. None of the geometric variables appears in the final model, thereby indicating the overriding influence of variables such as TPERCENT, SHLDWD, and SEGLEN in the best model for this environment. Figure 11 shows the influence of TPERCENT on the probability of a truck accident.

Logistic Regression Model for Environment III
(Interstate and Primary Highways, Divided 4-Lane, AADT > 15,000)

The best set of models with singleton variables for this highway environment is shown in Table 9. Among these, the model with the minimum AIC value of 3836.76 is the model with the following four singleton variables: TPERCENT, SEGLEN, CCR, and SCR.

Table 9

LOGISTIC REGRESSION MODEL — ENVIRONMENT III

Level	L(OK)	AIC	Variables
1	3842.15	3846.15	TPERCENT
2	3839.28	3841.28	TPERCENT, SEGLEN
3	3830.17	3841.28	TPERCENT, SEGLEN, CCR
4	3826.76	3836.76*	TPERCENT, SEGLEN, CCR, SCR
5	3826.10	3838.10	TPERCENT, SEGLEN, CCR, SCR, ADT
6	3826.06	3840.06	TPERCENT, SEGLEN, CCR, SCR, ADT, AMS
7	3826.06	3842.06	TPERCENT, SEGLEN, CCR, SCR, ADT, AMS, SPDIFSQ

Further improvements to this model were sought through the identification of significant interaction terms of the saturated model. This resulted in an improvement of the AIC value to 3827.17, with the following logistic model

$$P_i = \frac{1}{1 + e^{-\beta_{xi}}} \tag{41}$$

where $\beta_x = -2.7736 + .014(CCR) + .201(SEGLEN) + .041(TPERCENT)$
$- .009(CCR)(SEGLEN) + .00022(CCR)(SEGLEN)(TPERCENT)$

AIC = 3827.17.

This model indicates that increasing CCR, TPERCENT, or SEGLEN could lead to an increase in the likelihood of truck involvements. It is evident that CCR and TPERCENT have a significant effect on the involvement of trucks in accidents. Figure 12 shows the effect of both CCR and TPERCENT on p_i. This figure clearly indicates that for this environment the likelihood of a truck being involved in an accident steadily increases with CCR. It is also evident that increasing truck percentages lead to higher values of p_i. The effect of CCR at low and high TPERCENT values are shown in Figure 13.

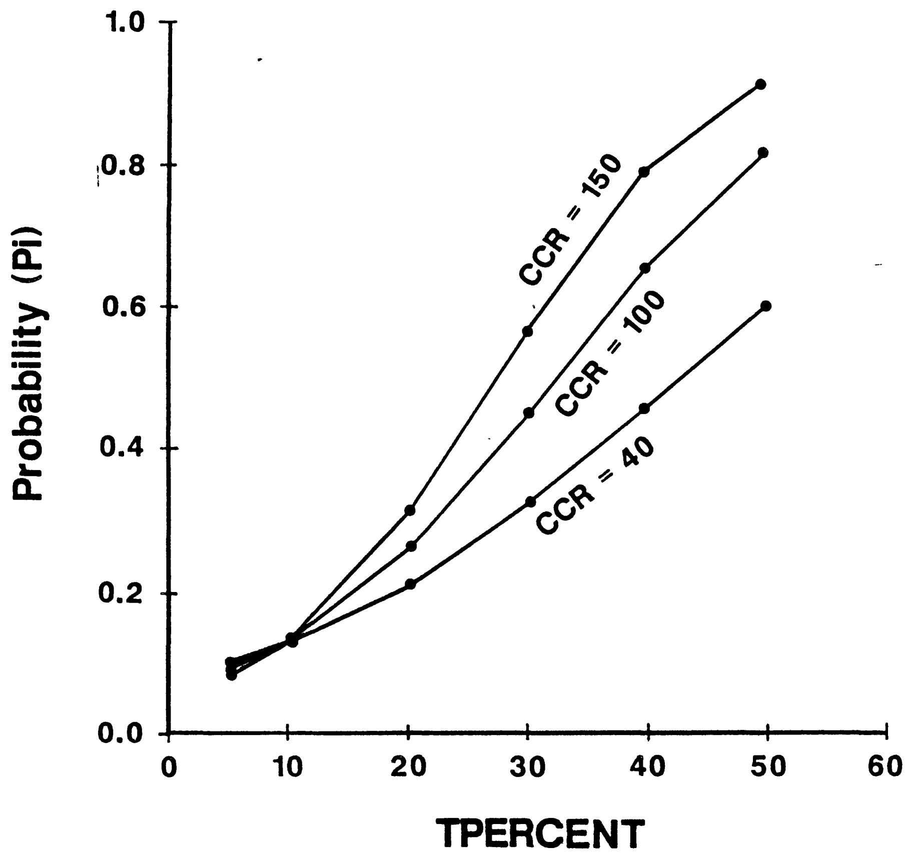


Figure 12. Effect of CCR — Environment III.

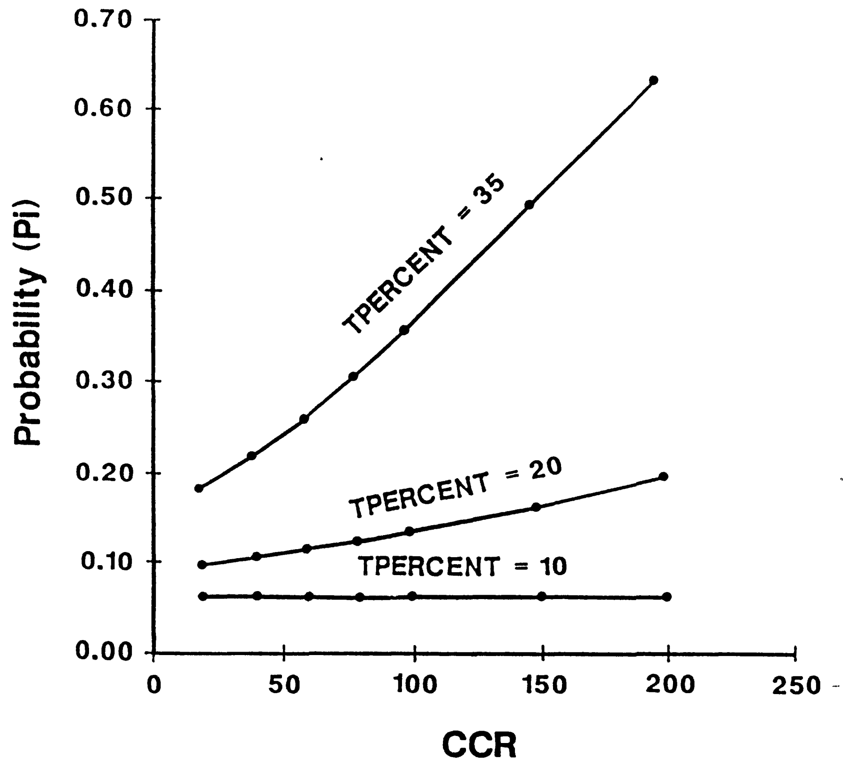


Figure 13. Effect of TPERCENT — Environment III.

GUIDELINES FOR SELECTING TRUCK ROUTES

It was originally intended to develop guidelines for selecting routes for large-truck traffic using the models developed. The models obtained, however, indicate that several traffic and geometric factors affect large-truck accident rates and the probability of an accident involving a large truck occurring. It is therefore extremely difficult to give guidelines because of the interaction of these different factors.

The authors are therefore developing an interactive, user-friendly computer program based on the models developed, which can be used to determine the accident rate or the probability of a truck accident occurring for different traffic and geometric conditions. The user of the program will then be free to determine an acceptable accident rate or probability level for his or her specific area.

The program will then be used to determine the accident rate or probability of a truck accident for the specific traffic and geometric conditions being considered. If the accident rate and/or probability is less than the acceptable value, then that section of road is suitable for large-truck operation; and if higher, it is not suitable. The user will also be able to use the program to determine the effect of traffic and geometric changes on the large-truck accident rate on a given stretch of road. This program is now being developed as a part of ongoing work on large-truck safety.

CONCLUSIONS

Multiple Linear Regression Models

Multiple linear regression models do not seem to describe adequately the relationship between large–truck involvement in accidents and associated traffic and geometric variables.

Poisson Regression Models

The Poisson models developed seem to describe adequately the relationship between large-truck involvement in accidents and associated traffic and geometric variables.

Environment I

- The most significant traffic variables contributing to large-truck involvement in accidents on highways within this environment are the ADT on the highway segment and the percentage of trucks.
- The most significant geometric variable is the SCR. Increasing SCR on a highway segment tends to increase the truck involvements, indicating that the vertical alignment is more critical for trucks in this environment.

Environment II

- The most significant traffic variables contributing to large-truck accidents are the ADT and the percentage of trucks on the highway segment.
- The most significant geometric variable is the SCR. Increasing SCR also increases the involvement of large trucks in accidents.

Environment III

- The most significant traffic variables contributing to large-truck involvement in accidents are the ADT and the speed difference between trucks and nontrucks.

- The most significant geometric variable is the CCR. Highway segments with high CCR tend to experience increased large-truck involvements.

Logistic Regression Models

Three logistic regression models were developed to describe the observed probabilities of truck involvements as a linear logistic function of traffic and highway variables. An analysis made using these models led to the following conclusions:

Environment I

- Lane width seems to have the greatest effect on the probability of a truck accidents. A decrease in lane width results in increased probability of truck accident. Shoulder widths also show a similar effect, but it is less pronounced.
- Increasing SCR tends to increase the incidence of truck accidents.
- It seems that shoulder width by itself does not have a significant effect on truck accidents. However, on segments with a high SCR value, an increase in shoulder width can create a reduction in the probability of a truck accident.

Environment II

- The most significant variable contributing to the probability of a truck accident for highways within this environment is the percentage of trucks on the highway segment.
- A secondary effect of ADT on truck accidents was also observed.

Environment III

- The probability of a truck accident increases steadily with the percentage of trucks using the facility.
- The probability of a truck accident increases with an increase of complexity in the horizontal alignment as measured by CCR.

Measures for Complexity of Highway Alignment

The measures of the complexity of highway alignment (such as CCR, SCR, and AMS) introduced in this study have made a significant contribution to the quantification of geometric complexity. Without such measures, any work on the impact of highway geometry on safety or even operations can be carried out only for geometrically homogeneous highway segments. The role of these variables in the models developed further verifies the relevance of such an approach.

REFERENCES

1. Garber, N. J., and Joshua, S. C., "Characteristics of Large Truck Crashes in Virginia." *Transportation Quarterly,* Vol. XLVIII, No. 1, January 1989.

2. Joshua, S. C., "A Causal Analysis of Large Vehicle Accidents Through Fault Tree Analysis and Statistical Modeling." Ph.D. Dissertation, Civil Engineering Department, University of Virginia, Charlottesville, August 1989.

3. Lamm, R., and Choueiri, E., "Recommendations for Evaluating Horizontal Design Consistency." *Transportation Research Record* 1122, Transportation Research Board, Washington, D.C., 1987.

4. Leisch, J. P., and Leisch, J. E., "New Concepts in Design Speed Application." *Transportation Research Record 631,* Transportation Research Board, Washington, D.C., 1977.

5. Gilchrist, W., *Statistical Modelling.* John Wiley & Sons, Inc., 1985.

6. Kullback, S., and Liebler, R. A., "On Information and Sufficiency." *Annals of Mathematical Statistics,* Vol. 22, 79-86, 1951.

7. Bozdogan, H., "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions." *Psychometrika,* Vol. 52, No. 3, 345-370, 1987.

174[illegible]

APPENDIX I

LIST OF STUDY SITES

LIST OF STUDY SITES

Site Number	Route Number	From	To
1	64-I	E.E. C & O OP	Rte 60
2	64-I	Rte 60	.39 ME Rte 60
3	64-I	.51 ME Rte 60	Mallory St.
4	64-I	115 MW ECLHP	ECL Hampton
5	81-I	Rte 679 UP	Rte 659 UP
6	58-P	12 MWWB3370P	EE Eliz Tunn
7	64-I	Rte 564 NB	22 ME NB 564
8	264-I	EE Rte 460 OP	EE NW RR OP
9	95-I	Rte 85 NB OP	19 MN Rte 85
10	58-P	ECL Ports	.15 MW WE Tunn
11	220-P	.09 MNNE 419 OP	SE Rte 419 OP
12	581-I	EB Rte 460 OP	.40 MS Rte 460 OP
13	81-I	SCL Salem	SE Rte 112
14	64-I	Ind RI Rd	49 ME Ind RI Rd
15	81-I	Rte 77 MP 1570	80 MN Rte 77
16	81-I	33 MS Rte 220	Rte 220
17	220-P	16 MN Rte 81 UP	Rte 31 SB UP
18	29-P	Rte 683	.26 MN Rte 683
19	29-P	.20 MN Rte 683	.70 MN Rte 683
20	1-P	Rte 150 OP	Rte 1479
21	360-P	.08 ME Rte 733	.48 MW Rte 733
22	360-P	Rte 652	.14 ME Rte 702
23	360-P	.03 MEE Rte 603	.45 NW W Rte 603
24	29-P	41 ME Rte 66 UP	Rte 28
25	66-I	ECL Vienna	10 ME Rte 650 UP
26	95-I	.22 MN Rte 644	16 MS Rte 495
27	95-I	.16 MN Rte 646	Rte 30
28	301-P	Rte 3	.24 MN Rapp RB
29	85-I	.31 MN Rte 630	.28 MS Rte 642
30	81-I	11 MN Rte 232	.54 MN Rte 665
31	13-P	SE TRBR Sec B	NE TRBR SECA
32	29-P	51 MN Fang CL	Rte 15
33	95-P	SE Gov. Rd OP	09 MS Rte 619 OP
34	81-I	17 MNNE 100 OP	40 MN Rte 600 UP
35	81-I	Montgomey CL	Rte 647
36	81-I	NE Rte 11 OP	79 MS Rte 710
37	81-I	EB Rte 64	SB Rte 11
38	23-P	Rte 58 & 421	Rte T-1112
39	81-I	1.31 MN Rt 11	.24 MS Rte 681
40	95-I	.08 M Rte 17 BP	.46 MN Rte 1WB
41	95-I	.24 MNNE 630 OP	Rte 610
42	81-I	.09 MN Rte 91 OP	Rte 11 & 751
43	77-I	NE Rte 620 OP	.60 MN Rte 608

APPENDIX II

SAMPLE FIELD DATA FROM TWO STUDY SITES

DATA SHEET - TRUCK STUDY

Date :..8-17-87...

Time:..3:30...am / pm

Site #:...905-3....

Direction of Travel : N S E W , bound ; Number of Lanes : 1 2 3

Location : Route #.460-3... From: .07 mE W Rt. 608..to: .92 mW Rt 695.....

Posted Speed Limit :..55......MPH

Recommended Speed at Curves (if posted) :..........MPH

HIGHWAY TYPE :

2 - lane Undivided   3 -lane   4 -lane Divided

Median - Barrier glass

Median - Traversible

For 2 lane highways : Passing Zone......... No Passing Zone.........

Lane Width :....11........ft;

Shoulders : Width.....0...ft; Depressed: Yes...., No....

Sight Dist : Adequate ..✓....; Inadequate........;

GEOMETRICS :

Straight - Level

Curve - Level

Up / Down Grade - Straight

Up / Down Grade - Curve

Hillcrest - Straight

Hillcrest - Curve

Dip - Straight

Dip - Curve

SLOPEMETER READINGS:

Test #'s :........;.........;.........;.........;.........;

REMARKS:..

ROUTE NUMBER = 460, Dir. of Travel = EASTBOUND, Site Number =905-3

Posted Speed Limit = 55

Begin Terminal =.07ME W RT 608, End Terminal =.92 MW RT695

Truck Speeds

Speed MPH	# of Observations
62	2
61	1
60	3
59	2
58	1
57	3
56	3
54	2
53	1
52	2
51	2
48	2
47	2

Mean = 55.27 MPH

Var = 21.80

Number of Obs. = 26

Non-Truck Speeds

Speed MPH	# of Observations
68	2
65	2
64	5
63	3
62	5
61	3
60	7
59	10
58	13
57	12
56	14
55	14
54	4
53	3
52	5
51	1
50	1
49	1
48	2
47	5
44	1
43	1
42	2
41	1

Mean = 56.39 MPH

Var = 26.98

Number of Obs.= 117

1752

TEST BEGINS

LOCATION 462 EB
Site ~~4600~~ Rt. 460 E
Station 5
TEST # 2 45 MPH
DATE: 08/17/87
START TIME
15:09:04
L 2.00 DEGREES
00.02 MILES
46 MPH

L 1.00 DEGREES
00.04 MILES
46 MPH

L 2.00 DEGREES
00.10 MILES
46 MPH

L 2.50 DEGREES
00.12 MILES
46 MPH

L 4.00 DEGREES
00.13 MILES
46 MPH

L 2.50 DEGREES
00.14 MILES
46 MPH

L 1.00 DEGREES
00.17 MILES
46 MPH

L 2.00 DEGREES
00.20 MILES
46 MPH

L 3.00 DEGREES
00.21 MILES
46 MPH

L 2.50 DEGREES
00.26 MILES
44 MPH

L 1.00 DEGREES
00.27 MILES
44 MPH

L 2.00 DEGREES
00.30 MILES

TEST BEGINS

LOCATION EB
Site ~~4600~~ 462
Rt. 460 E Station 5
TEST # 1 55 MPH
DATE: 08/17/87
START TIME
14:58:06
L 1.00 DEGREES
00.03 MILES
55 MPH

L 2.00 DEGREES
00.05 MILES
53 MPH

L 2.50 DEGREES
00.08 MILES
53 MPH

L 1.00 DEGREES
00.13 MILES
53 MPH

L 2.00 DEGREES
00.14 MILES
53 MPH

L 2.50 DEGREES
00.16 MILES
53 MPH

L 3.00 DEGREES
00.17 MILES
53 MPH

L 2.50 DEGREES
00.30 MILES
53 MPH

L 3.00 DEGREES
00.33 MILES
55 MPH

L 2.50 DEGREES
00.36 MILES
55 MPH

L 1.00 DEGREES
00.61 MILES
55 MPH

L 2.00 DEGREES
00.62 MILES
55 MPH

L 2.50 DEGREES
00.96 MILES
51 MPH

SLOPE METER READING 1753

TEST BEGINS

LOCATION Site 905-3
~~Site~~ Rt 46 E
Station 5
TEST # 2/4 Sl pt
DATE: 08/15/87
START TIME
15:29:41
L 2.00 % GRADE
00.03 MILES
46 MPH

R 2.00 % GRADE
00.14 MILES
44 MPH

L 0.00 % GRADE
00.21 MILES
46 MPH

R 2.00 % GRADE
00.22 MILES
46 MPH

R 3.00 % GRADE
00.71 MILES
46 MPH

L 2.00 % GRADE
00.80 MILES
44 MPH

L 3.00 % GRADE
00.81 MILES
44 MPH

R 3.00 % GRADE
01.06 MILES
46 MPH

END TIME
15:31:33
READY

1754

DATA SHEET - TRUCK STUDY

Date : 10-07-87 .. Time: 12:30 ...am / pm

Site #: 13-3 ......

Direction of Travel : N S E W , bound ; Number of Lanes : 1 2 3

Location : Route # 254 ...... From: WCL Staunton ... to: Rt 708 ....

Posted Speed Limit : 55 ......MPH

Recommended Speed at Curves (if posted) :..........MPH

HIGHWAY TYPE :

2 - lane Undivided	3 -lane	4 -lane Divided
	Median - Barrier	
	Median - Traversible	

For 2 lane highways : Passing Zone ✓...... No Passing Zone.........

Lane Width : ... 10 ....ft;
Shoulders : Width... 5 ...ft; Depressed: Yes ✓.., No....
Sight Dist : Adequate ✓.....; Inadequate........;

GEOMETRICS :

	Straight	-	Level
	Curve	-	Level
Up / Down	Grade	-	Straight
Up / Down	Grade	-	Curve
	Hillcrest	-	Straight
	Hillcrest	-	Curve
	Dip	-	Straight
	Dip	-	Curve

SLOPEMETER READINGS:

Test #'s :........;.........;........;........;........;

REMARKS:...

Tube Velocity Statistical Hourly Data Graph

```
Data File   : 02258202.IRF                      Position : 6
Station     : 133                               Ident    : 254
Start Date  : Oct 8 ,1987                       End Date : Oct 8 ,1987
Start Time  : 00:00                             End Time : 13:11
Location    : RT. 254 EB.; STAUNTON

  Speed          Count   0        100       200       300       400       500
-------------    -----   +---------+---------+---------+---------+---------+
                         |
  1 thru  44      353    |***********************************
 45 thru  46      122    |************
 47 thru  48      129    |************
 49 thru  50      132    |*************
 51 thru  52      151    |***************
 53 thru  54      102    |**********
 55 thru  56       79    |*******
 57 thru  58       38    |***
 59 thru  60       41    |****
 61 thru  62       12    |*
 63 thru  64       10    |*
 65 thru  99       15    |*
                         |
Total            1184    |
                         |
                         +---------+---------+---------+---------+---------+

                         0        100       200       300       400       500
```

Total	Mean	50%ile	50%ile	85%ile	>55	%>55	>60	%>60	>65	%>65
1184	43.6	46.8	46.8	53.4	116	9.8	[illegible]	[illegible]	0	0.0

1756

Tube Velocity Statistical Hourly Data Graph

Data File : 02258902.IRF
Station : 133
Start Date : Oct 7 ,1987
Start Time : 12:24
Location : RT. 251 EB.; STAUNTON

Position : [illegible]
Ident : 254
End Date : Oct [illegible] ,1987
End Time : 21:00

Speed	Count
1 thru 44	362
45 thru 46	98
47 thru 48	148
49 thru 50	162
51 thru 52	149
53 thru 54	100
55 thru 56	75
57 thru 58	56
59 thru 60	36
61 thru 62	8
63 thru 64	8
65 thru 99	11
Total	1213

0 100 200 300 400 500

Total	Mean	50%ile	50%ile	85%ile	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1213	43.4	47.0	47.0	53.3	11[illegible]	[illegible].8	[illegible]	[illegible]	0

WCL Staunton
708

TEST BEGINS 50 mph

LOCATION
Rte 254 EB

TEST # 4

DATE: 10/08/87
START TIME
12:08:22

L 2.00 DEGREES
00.02 MILES
44 MPH

R 2.50 DEGREES
00.06 MILES
48 MPH

R 1.00 DEGREES
00.08 MILES
49 MPH

L 1.00 DEGREES
00.09 MILES
49 MPH

L 2.50 DEGREES
00.11 MILES
51 MPH

L 3.00 DEGREES
00.12 MILES
53 MPH

L 2.50 DEGREES
00.14 MILES
53 MPH

R 1.00 DEGREES
00.15 MILES
53 MPH

R 2.00 DEGREES
00.17 MILES
51 MPH

L 0.00 DEGREES
00.18 MILES
49 MPH

WCL Staunton
Rte 708

TEST BEGINS

LOCATION 55 mph
Rte 254 EB

TEST # 2

DATE: 10/08/87
START TIME
11:49:44

L 3.00 DEGREES
00.04 MILES
51 MPH

R 2.00 DEGREES
00.07 MILES
53 MPH

L 0.00 DEGREES
00.08 MILES
53 MPH

L 2.00 DEGREES
00.10 MILES
53 MPH

L 2.50 DEGREES
00.13 MILES
55 MPH

L 3.00 DEGREES
00.14 MILES
55 MPH

R 2.50 DEGREES
00.18 MILES
56 MPH

L 2.50 DEGREES
00.21 MILES
56 MPH

L 1.00 DEGREES
00.26 MILES
56 MPH

L 2.00 DEGREES
00.29 MILES
55 MPH

L 3.00 DEGREES

1758

SLOPE METER READINGS

WCC Staunton
708

TEST BEGINS

LOCATION
Rte 254 EB

TEST # 6

DATE: 10/08/87
START TIME
12:28:28

L 2.00 % GRADE
00.02 MILES
46 MPH

L 3.00 % GRADE
00.06 MILES
46 MPH

R 2.00 % GRADE
00.24 MILES
49 MPH

R 3.00 % GRADE
00.26 MILES
49 MPH

R 4.00 % GRADE
00.28 MILES
48 MPH

R 3.00 % GRADE
00.33 MILES
42 MPH

L 2.00 % GRADE
00.35 MILES
42 MPH

L 3.00 % GRADE
00.37 MILES
42 MPH

L 4.00 % GRADE
00.69 MILES
46 MPH

L 3.00 % GRADE
00.70 MILES
46 MPH

L 4.00 % GRADE
00.73 MILES
46 MPH

L 3.00 % GRADE