given. The first variable entered is the number of poles, which explains 25.7 percent of the variation. Offset is then entered at step 2 and explains a further 0.6 percent of the variance. Road grade is entered at step 3, road path at step 4, and speed limit at step 5, and each explains an additional 0.5, 0.2, and 0.3 percent of the variance respectively. The remaining three steps given in the table each contributed another 0.1 percent to the total variation explained.

It is clear from this regression analysis that the overriding factor in predicting utility-pole accidents is the number of poles. Note that this variable not only identifies that a line of poles exists but also indicates average pole spacing since poles that were within 183 m (600 ft) of either side of the struck pole (or the rest position of the vehicle in run-off-road accidents) were counted. Furthermore, it is encouraging that offset is entered as step 2 because it complements the number-of-poles parameter by providing a more complete definition of pole placement.

The remaining parameters that are entered describe the type of road—i.e., road grade—or are related to the vehicle departure angle—i.e., road path and speed limit. This suggests that, if better measures of departure attitude were available—e.g., angle and speed—a higher proportion of variation might be explained.

REFERENCE

1. Accident Facts. National Safety Council, Chicago, 1976.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics.

Abridgment

Mathematical Models That Describe Lateral Displacement Phenomena

Ali A. Selim, South Dakota State University
James L. Josey, Clemson University

In this research, a unique technique was used to collect a reliable and permanent type of data (1). Data were collected by using two super 8-mm movie cameras to study the behavior of traffic in the right lane of freeways as it approaches a vehicle parked on the right shoulder. The general tendency of vehicles as they near a parked vehicle is to swerve away from it. The path of the average vehicle at the test location is expressed by a predictive model in terms of independent variables related to geometric parameters and traffic characteristics. By using the model, the magnitude of lateral displacement at any location can be determined as the difference between the paths of the average vehicle in the presence of a side obstruction (parked vehicle) and under normal conditions (no side obstruction).

In this research, vehicles of different sizes were used and placed on the right shoulder at various distances from the freeway edge of the pavement. Vehicles were used since they are the most common type of side obstruction. A full description of the process of data collection and methods used to extract different parameters is beyond the scope of this paper but is available elsewhere (1). A brief summary of the research methodology used is presented below.

For each experiment run, a vehicle of known width was placed on the right shoulder, and the clear distance between the most remote left point of the vehicle and the edge of the pavement was measured and recorded. Two observers, each operating a camera, were signaled by a third observer by way of portable CB units to start running approximately 7.6 m (25.0 ft) of film at a speed of 8 frames/s. Three minutes of filming were designed for each experiment (1). The camera speed of 8 frames/s permitted the running of two experiments with a 15.2-m (50.0-ft) roll of film. A digital stopwatch was placed about 15 cm (6 in) in front of each camera's objective lens; these stopwatches read to 1/100 of a second and appeared in the unused portion of the frame.

The first observer was stationed on a crossover (pedestrian or crossroad) and above the center of the right lane of the freeway. The observer's line of sight during filming was parallel to the traffic flow, and the edge of the pavement was ensured to be in view. The observer was completely concealed from motorists to ensure that lateral displacement did not occur because of any outside distraction but was a normal reaction of the driver when approaching the parked vehicle at the test section. A second observer, stationed evenly with the parked vehicle and on the other side of the highway, was generally outside the right-of-way; this allowed visual coverage of about 35 to 45 m (120 to 150 ft) of the roadway with the parked (test) vehicle in the middle of the observer's view.

Both films were later advanced simultaneously through stop-action projectors, and several parameters were extracted either by visual counting or by constructing special scales that were placed on the screen to measure distances. Time was read from the photographed stopwatches.

Movies taken by the first observer were used to extract parameters such as the total volume of vehicles in the right lane, including trucks and buses, and distance between the edge of the pavement and the center of a vehicle as it passed next to the parked (test) vehicle. The speeds of individual vehicles in the right lane and in the adjacent lane, headways in the right lane, and other parameters were extracted from the movie taken by the second observer.

Data from each experiment were classified as either geometric parameters (such as degree of curvature at the test location and grades in the direction of traffic flow) or traffic characteristics (such as those parameters extracted from movie films). Data were collected from two large metropolitan areas (St. Louis and Chicago) to study whether a general model could be developed that would apply to more than one met-

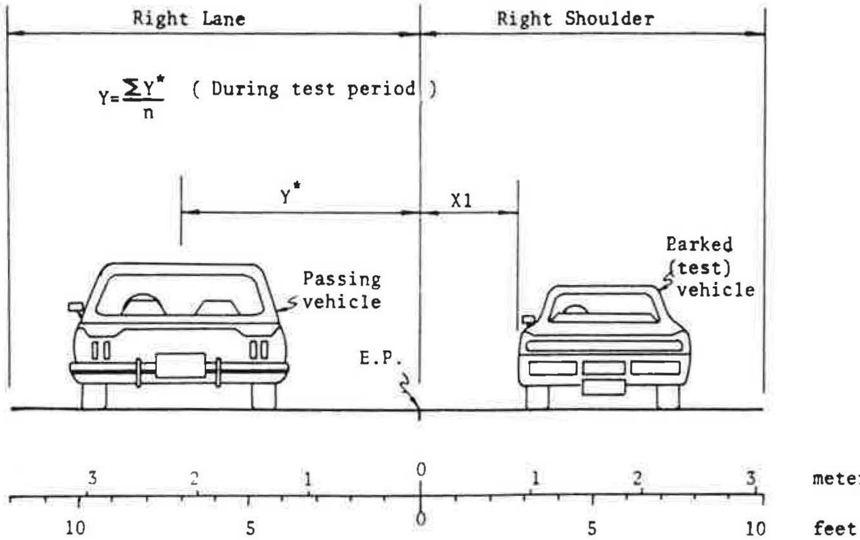

Figure 1. Path of average vehicle in relation to test vehicle.

Table 1. Numerical values of variables and their ranges in each metropolitan area.

Variable	St. Louis		Chicago		Unit of Measurement
	Maximum	Minimum	Maximum	Minimum	
Y*	2.18	1.64	2.17	1.85	Meters
X1	10.00	0.58	10.00	0.59	Meters
X2	1.60	-2.66	2.50	-0.50	Percent
X3	3.00	-2.00	0.00	-2.75	Degrees
X4	14.30	0.00	26.20	4.60	Percent
X5	11.10	0.00	27.30	0.00	Percent
X6	81	29	69	22	Vehicles per 3 min
X7	103.10	86.80	104.30	87.50	Kilometers per hour
X8	16	7	13	5	Vehicles per kilometer
X9	2.06	1.55	1.95	1.68	Meters
X10	0	0	1	1	Dimensionless

Note: 1 m = 3.3 ft 1 km = 0.62 mile
*Dependent variable

ropolitan area. Multiple regression analysis was then applied, and independently predictive models were obtained for each area as well as for the combined data to obtain a general model.

DEFINITION OF VARIABLES

One or more of the following variables appeared in the predictive models:

- Y = path of the average vehicle or mean distance from the edge of the pavement to the center of the vehicle as it crosses the test location (m) (Figure 1),
- X1 = distance between the most remote left point of the parked vehicle and the edge of the pavement (m) (Figure 1),
- X2 = highway grade in the direction of traffic flow (percent),
- X3 = degree of curvature in the direction of traffic flow,
- X4 = trucks in the right lane (percent),
- X5 = trucks in the adjacent lane (lane 2 in highway terminology) (percent),
- X6 = volume of traffic in the right lane (vehicles/3 min),
- X7 = average speed of traffic in adjacent lane (km/h),
- X8 = density of traffic in the right lane (vehicles/km),
- X9 = width of the test vehicle placed on the shoulder (m), and
- X10 = dummy variable (only used in the general model) = 0 for St. Louis area and 1 for Chicago area.

The Y variable is dependent, and all other variables listed are independent. X1, X2, and X3 are geometric variables, and the rest of the X variables are traffic variables. Other parameters, such as the total number of lanes, lane width, and volume of traffic in the median lane (except for four-lane divided highway), appeared to be insignificant.

Table 1 gives all variables involved in the analysis and their range of occurrence in each metropolitan area.

REGRESSION AND CORRELATION ANALYSIS

Several techniques in multiple regression analysis are widely used by statisticians and engineers. The following two techniques were used because of their proven worth in the field of transportation research and especially in traffic flow analysis: (a) stepwise regression procedure (2) and (b) maximum R^2 improvement (3). The final selection of the predictive models by either technique was based solely on obtaining the best value for the multiple correlation coefficient R^2. A summary of the value of R^2 obtained by both techniques for each metropolitan area as well as for the combined data is given below (in the general model, a dummy variable is used for area identification):

Model	Stepwise Procedure	Maximum R^2 Improvement
St. Louis	0.92	0.92
Chicago	0.88	0.91
General	0.74	0.82

St. Louis Models

The stepwise regression procedure yielded model 1, mathematically described by Equation 1:

$$Y = 1.7084 - 0.0123(X5) + 0.0168(X2)^3 - 0.0256(X9/X1)^2 + 0.5736(X9*X1)^{-0.5} \quad (1)$$

Model 2 was given by the maximum R^2 improvements technique as

$$Y = 2.5031 + X5[0.0014(X5) - 0.0284] + 0.016(X2)^3 - 0.1804(X9) - 0.1023(X1)^{0.5} \quad (2)$$

This model was the best five-variable model found by the technique. The best six-variable model has a higher value for R^2. As the number of variables in the model increases, the R^2 value also increases. The best five-variable model was chosen so that the number of observations is about four times the number of variables in the model (2).

Chicago Models

The stepwise regression procedure yielded model 3, which is expressed by the following equation:

$$Y = 2.5165 - 0.6188 * 10^{-4}(X7)^2 - 0.0034(X2)^3 - 0.1216 * 10^{-3}(X5)^2 + 0.2065(X9 * X1)^{-0.5} \quad (3)$$

The maximum R^2 improvement yielded model 4 for the Chicago area:

$$Y = 2.6165 + X2[0.0737 - 0.0194(X2)^2] - 0.0453(X1)^{0.5} - 0.1083(X6/X8) \quad (4)$$

This model was the best four-variable model found by the technique. All variables in the models above were found to be significant at the 0.1 level.

Regardless of the multiple regression technique used, the models obtained for each area were found to be different in nature either in the beta coefficients or in the set of independent variables involved (X_i).

DISCUSSION OF RESULTS

In our view, the variations were mainly attributed to the following causes:

1. Unequal sample sizes were collected from each metropolitan area because of restrictions on site selection (1). Twenty sites were tested in St. Louis whereas only 16 were tested in Chicago.
2. Each metropolitan area has its own geometric and traffic characteristics that make it different from others; for example, Chicago has the following traffic and roadway features that St. Louis does not have: (a) The percentage of trucks is much higher (see Table 1); (b) there are more kilometers of depressed freeways with retaining walls; and (c) local traffic regulations, enforced by the state of Illinois, forbid trucks from using the median lanes on some sections of freeways.

An attempt was made to develop a general model by combining the data from both locations. The multiple correlation coefficient obtained by using the stepwise procedure for the combined data was 0.694. Maximum R^2 improvements resulted in a multiple correlation coefficient of 0.805 for the best eight-variable model.

The significant reduction in the multiple correlation coefficient when data were combined (compared with the R^2 value for each area separately) was expected. The reduction was mainly attributed to combining data from two different metropolitan statistical areas that are not compatible in traffic and geometric characteristics. However, an appreciable increase in the multiple correlation coefficient was obtained by using dummy variables. Dummy variables are used to account for the fact that the various areas might have separate deterministic effects on the response (dependent variable). The dummy variable (X10) had a zero value when used with St. Louis data and a value one when used with Chicago data. When dummy variables were used, the following models were obtained:

$$Y = 3.4867 + 0.0354(X2) - 0.0050(X4) - 0.178(X7)^{0.5} + 0.1491(X10) + 0.3127(X9 * X1)^{-0.5} \quad (5)$$

Model 5 is given by the stepwise procedure, and model 6 is given by the method of maximum R^2 improvements:

$$Y = 1.9932 - 0.0070(X4) - 0.0518(X1)^{0.5} - 0.3986 * 10^{-4}(X7)^2 - 0.0221(X3) + 0.0058(X2)^3 + 0.4006(X1)^{-1} - 0.0117(X9/X1)^3 + 0.1728(X10) \quad (6)$$

In using these predictive models, the average path of vehicles under normal conditions (no side obstruction) can be determined by assuming a fictitious vehicle of average width [X9 = 1.68 m (5.5 ft)] placed at a large distance from the edge of the pavement [X1 = 10.0 m (3.3 ft)].

CONCLUSIONS

The findings reached in this research are based solely on data collected from the metropolitan areas of St. Louis and Chicago:

1. From the analysis of data, it appeared that general models are not recommended for the following reasons: (a) Each metropolitan area has different characteristics related to the type of local traffic regulations, location, size, land use, social and economical status, and so on; and (b) the multiple correlation coefficients for individual area models were higher than those for the general models because the assumption that all data came from the same population holds true only for individual models.
2. In comparison with other common methods, the data collection procedure used in this study is considered one of the most economical for collecting a reliable, permanent type of data (1).
3. In our view, the maximum $\bar{R}^2$ improvement technique was advantageous over the stepwise procedure in developing predictive models.
4. The developed models can be presented graphically through a series of nomographs to show the effect of each independent variable on the amount of lateral displacement. These nomographs can provide the designer with an additional tool for analysis and comparison of proposed alternative designs.

ACKNOWLEDGMENTS

We wish to express our sincere appreciation to the individuals who devoted their time to the data collection process and analysis that made the completion of this research possible. The Civil Engineering Department of the University of Missouri at Rolla is to be especially noted for its financial support of the research.

REFERENCES

1. A. A. Selim. A Study of the Lateral Displacement Phenomena, for Uninterrupted Traffic Flow in Freeways. Univ. of Missouri at Rolla, PhD dissertation, Dec. 1976.
2. N. Draper and H. Smith. Applied Regression Analysis. Wiley, New York, 1967.
3. A. J. Barr and J. H. Goodnight. A User's Guide to the Statistical Analysis System. Department of Statistics, North Carolina State Univ., Aug. 1972.

Publication of this paper sponsored by Committee on Operational Effects of Geometrics.