Relation of Flow, Speed, and Density of Urban Freeways to Functional Form of a Safety Performance Function

Jake Kononov, Craig Lyon, and Bryan K. Allery

Constructive discussion of the appropriate choice for the functional form of safety performance functions (SPFs) is generally absent from research literature on road safety. Among researchers who develop SPFs, there appears to be a consensus that the underlying randomness in accident counts is well described by the negative binomial (NB) distribution. The underlying phenomenon itself, however, is not well understood and is rarely discussed. The choice of the regression equation is usually not explained or documented. Researchers most commonly use the power function, possibly because most generalized linear modeling (GLM) statistical packages can accommodate the power function with little effort. The modeling process, however statistically rigorous, at times seems disconnected from the physical phenomenon that it is trying to describe. The disconnect, however, has attracted only limited interest from researchers to date. Accidents on an urban freeway are a by-product of traffic flow; therefore, changes in the flow parameters may give clues about the probability of accident occurrence and changes in accident frequency. This study related traffic flow parameters, such as speed and density, to the choice of the functional form of the SPF. It compared SPF models for urban freeways developed with sigmoid and exponential functional forms with the use of data from Colorado and California and contrasted the cumulative residual (CURE) plots of the models. SPFs developed around a sigmoid functional form through the use of neural network (NN) methodology suggested underlying relationships between safety and traffic flow characteristics. CURE plots for NN-generated SPFs generally showed a better-quality model fit when compared with power-function SPFs, which were developed in the GLM framework with an NB error structure.

A review of extant literature on the development of safety performance functions (SPFs) suggests that the focus of most modeling efforts is on the statistical technique and underlying probability distribution without much consideration given to the systemic component of the phenomenon. Abdel-Aty and Radwan, and many others, have observed that most of the accident data are overdispersed, which points to the need for a correction to Poisson assumptions. They have correctly concluded that the negative binomial (NB) formulation is superior to the more restrictive Poisson formulations (*1*). Clearly there is a consensus among researchers that underlying randomness is well described by the NB distributions. The underlying phenomenon itself, however, is not well understood and is rarely discussed. Lord, Washington, and Ivan suggested that it may be preferable to begin to develop models that consider the fundamental process of an accident rather than strive for best-fit models in isolation (*2*). Lord, Manar, and Vizioli observed that traffic flow characteristics have a direct influence on the likelihood and severity of a crash. The effect of these characteristics on freeway safety has not been clearly established, however, nor properly modeled (*3*). Hauer noted that the art of choosing the regression equation is seldom transparent, reasoned, or documented (*4*) and observed that there is no reason to think that the underlying phenomenon follows any simple, mathematical function (*5*).

Selection of the functional form is heavily influenced by the choice of functions available in the software package used by the modeler. The process consists of trying to fit a preselected function available in the statistical software to a set of data and then use statistical techniques to estimate regression parameters of the chosen function. Such a process, however statistically rigorous, seems disconnected from the phenomenon it is trying to describe, yet this disconnect has attracted only limited interest from researchers to date. Hauer summarized the situation as follows: if the functional form that is used is inappropriate, the regression coefficients obtained have no clear meaning, and there is little interest in their estimated value or precision (*5*). Accidents on an urban freeway are a by-product of traffic flow. It is reasonable, therefore, to expect that the observation of changes in the flow parameters may give clues about the probability of accident occurrence and changes in accident frequency.

In this study, neural networks (NNs) were first used to explore the underlying relationship between accidents and other variables for urban freeway segments. The results were then compared with SPFs calibrated by using these same data with generalized linear modeling (GLM) and an NB error structure. NNs are not constrained by the underlying distributional assumptions. They learn by example and infer a model from training data. The functional shape generated through the training of the NNs suggests that a sigmoid may be a reasonable representation of the physics of accident occurrence on urban freeways. In this study, traffic operation parameters, described by the 2000 edition of the *Highway Capacity Manual,* were used to provide a possible explanation for the functional form (*6*).

The quality of fit was examined with the cumulative residual (CURE) method described in Hauer and Bamfo (*7*). This method consists of plotting the CURE for each independent variable. The goal is to graphically observe how well the function fits the data set. To generate a CURE plot, sites are sorted in ascending order by the independent variables of interest. For each site, the residual (observed–predicted accidents) is calculated, and then the CUREs are determined and plotted for each value of the independent variable. Because of

J. Kononov and B. K. Allery, Colorado Department of Transportation, 4201 East Arkansas Avenue, Denver, CO 80222. C. Lyon, 663 Gainsborough Avenue, Ottawa, Ontario K2A 2Y9, Canada. Corresponding author: J. Kononov, Jake.Kononov@dot.state.co.us.

Transportation Research Record: Journal of the Transportation Research Board, No. 2236, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 11–19.
DOI: 10.3141/2236-02

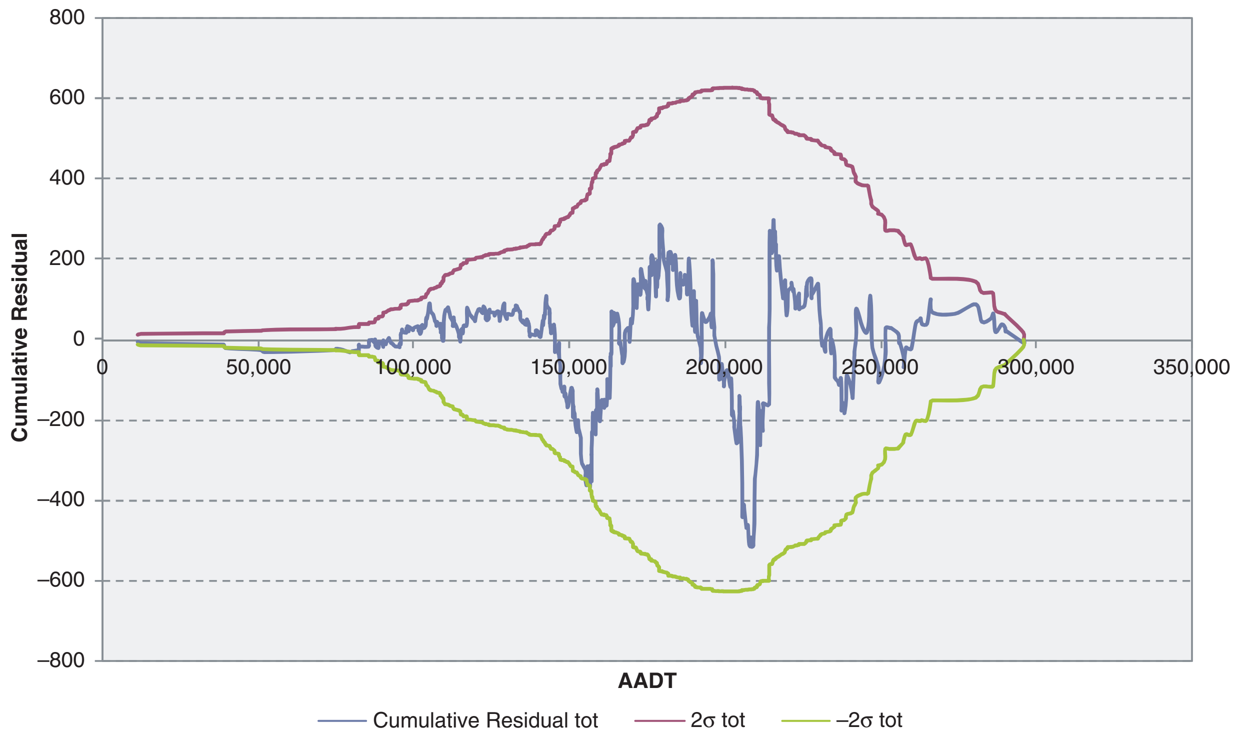

FIGURE 1 CURE plot for California eight-lane NN SPF (tot = total).

the random nature of accident counts, the CURE line represents a so-called random walk. For a model that fits well in all ranges, the CURE plot should oscillate around zero. If the CURE value steadily increases within a range of values of the independent variable, it means that, within that range, the model predicts fewer accidents than have been observed. Conversely, a decreasing CURE line indicates that, in that range, fewer accidents have been observed than are predicted by the model. A frequent departure of the CURE line beyond two standard deviations of a random walk indicates a presence of outliers, or signifies an ill-fitting model. Figure 1 shows a CURE plot for the independent variable of annual average daily traffic (AADT), which reflects the model fit of a NN-generated SPF for eight-lane urban freeways in California. Because the CURE residual line lies within the two standard deviation lines and generally oscillates around zero, it can be concluded that the functional form fits the data well.

DATA SET PREPARATION AND MODEL DEVELOPMENT

Five years of accident data from Colorado and California were used to develop SPFs for selected, multilane, urban freeways. California data were obtained from the Highway Safety Information System, while Colorado data were provided by the Colorado Department of Transportation. All of the accidents that occurred on ramps and crossroads were removed before the models were fitted, which left only accidents that occurred on a freeway mainline itself. Two kinds of SPFs were calibrated for Colorado and California: one for the total number of accidents and the other for injury plus fatal accidents.

Use of NNs

SPFs were developed with NNs, which were a subset of a general class of nonlinear models. NNs were used to analyze the data, which consisted of observed, univariate responses Y_i that were known to depend on corresponding, one-dimensional inputs x_i. NNs are not constrained by a preselected functional form and specific distributional assumptions. For this application, Y_i = accidents per mile per year and x_i = AADT. The model became

$$Y_i = f(x_i, \theta) + e_i$$

where

$f(x_i, \theta)$ = nonlinear function that relates Y_i to the independent variable x_i for the ith observational unit,
θ = p-dimensional vector of unknown parameters, and
e_i = sequence of independent random variables.

The goal of the nonlinear regression analysis was to find the function f that best reproduced the observed data. A form of the response function used in many engineering applications is a feed-forward NN model with a single layer of hidden units. The form of the model is

$$f(x, \theta) = \beta_0 + \sum_{k=1}^{K} \beta_k \varphi(x\,\gamma_k + \mu_k)$$

where

$\varphi(u)$ = $e^u/(1 + e^u)$, a logistic distribution function;
β_k = connection weights;
$\beta_0, \beta_1, \gamma_1, u_1$ = parameters to be estimated;
μ_k = biases, according to Ripley (*8*); and
K = number of hidden units.

The function f is a flexible, nonlinear model, which was used in this application to capture the overall shape of the observed data. The function $\varphi(u)$ is a logistic distribution function. When $K = 1$, there is one hidden unit. In this case, the function performed a linear transformation of the input x and then applied the logistic function $\varphi(u)$, which was followed by another linear transformation. The overall result was a flexible, nonlinear model.

The parameters β_0, β_1, γ_1, u_1 for each data set were unknown and were estimated by nonlinear least squares. The complexity for this application was the number of hidden units K in the model. $K = 1$ was chosen on the basis of the general understanding of the underlying physical phenomenon. In addition, a model's complexity is most often chosen on the basis of generalized cross validation model-selection criterion. Cross validation is a standard approach for selecting smoothing parameters in nonparametric regression, as described by Wahba (*9*). The residuals exhibited a pattern of increased variance as the AADT values increased. This was to be expected, given the overall pattern of the data. The overall model fit to the data was quite good.

Generalized Linear Modeling

For the GLM, which followed most SPFs available, power models of the following forms were used to obtain the best fit:

$$\left.\begin{aligned} E\{y\} &= x^{\beta_0} e^{\beta_1 x + \beta_2 x^2 \ldots} \\ E\{y\} &= \beta x^{\beta_1} \end{aligned}\right\} \rightarrow \text{power family}$$

The popularity of the power function in modeling safety is derived perhaps less from its suitability to describe the underlying processes that result in accidents than from the fact that most GLM statistical packages accommodate it with little effort, and it is a flexible shape. In this case, $E\{y\}$ was the annual number of accidents expected to occur on a segment of road, x was the independent variable (here AADT), and β was the parameters to be estimated.

The accident counts for segments of urban freeways generally exhibit overdispersion as compared with a Poisson distribution of accident counts. Although geometric characteristics of the freeways themselves are fairly uniform because they are designed to interstate standards, the overdispersion was consistently present in the data used for this study. The explanation may be the influence of ramp flows and spacing on safety performance. The influence of ramps was not introduced as an independent variable but was reflected by the number of accidents on the mainline. The β parameters for the urban freeways were estimated by maximizing the log likelihood function of the NB distribution

where

- μ = estimated number of accidents on a freeway segment over 1 year,
- y_i = observed number of accidents on freeway segment over 1 year,
- $L(\alpha, \mu)$ = NB likelihood function,
- β = estimated regression parameters, and
- α = overdispersion parameter, which was estimated by maximizing the NB log likelihood function.

$$\mu \in \text{negative binomial} \therefore \text{Var} > \mu$$

$$\text{Var}(y) = \mu(1 + \alpha\mu) = \mu + \alpha\mu^2 \therefore \sigma = \sqrt{\mu + \alpha\mu^2}$$

$$L(\alpha, \mu) = \prod_{i=1}^{n} \frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1}) y_i!} \left(\frac{\alpha\mu_i}{1 + \mu_i}\right)^{y_i} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\alpha^{-1}}$$

$$\ln(L(\alpha, \mu)) = \sum_{i=1}^{n} \left[\ln\left(\frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1}) y_i!}\right) + y_i \ln\left(\frac{\alpha\mu}{1 + \alpha\mu_i}\right) + \alpha^{-1} \ln\left(\frac{1}{1 + \alpha\mu_i}\right)\right]$$

Model Comparison

Figures 2 through 5 provide plots of the data, SPF predictions, and CURE plots for six- and eight-lane urban freeways in Colorado and California. In each figure, two graphs correspond to the GLM power function, and two graphs correspond to the NN models with a sigmoid functional form.

The CURE plots for the SPFs, which were generated by using NN sigmoid, consistently showed a model fit superior to that of the GLM-generated SPFs with power functional form. NN sigmoid SPFs were significantly less biased than GLM SPFs throughout the entire range of AADT; they showed a moderately amplified random walk virtually without excursions outside of the 2σ limits. In contrast to the NN sigmoid models, the GLM-generated SPFs showed more bias within certain ranges of AADT and frequently departed outside of the 2σ boundaries.

For both the data from Colorado and California, the sigmoid functional forms generated by the NN method provided reasonably good estimates of expected accident frequencies at different levels of AADT in California and Colorado. The sigmoid shape reflects a relationship similar to that of a dose–response curve, found in medicine and pharmacology, as well as other sciences.

Relating Changes in Flow, Speed, and Density to Changes in Crash Rates Reflected by Shape of SPF

Accident rates change with AADT, and the shape of an SPF reflects how these changes take place. Higher crash rates within the same SPF mean less safety than lower rates. Any accident frequency derived from the SPF expressed in accidents per mile per year can be easily converted into accident rates measured in accidents per million vehicle miles traveled (VMT). For instance, the Colorado SPF, which was calibrated for six-lane urban freeways (Figure 6) at AADT of 120,000, was expected to produce on average 58 accidents per mile per year; this figure could be directly converted to the accident rates as follows:

$$\frac{(58 \text{ accidents/mi/year})1{,}000{,}000}{120{,}000 \text{ vpd } (1 \text{ mi}) \text{ } 365 \text{ days/year}} = 1.32 \text{ accidents/million VMT}$$

The ordinate of point C represents the expected number of crashes per year on a six-lane freeway at AADT of 120,000 vehicles per day (vpd). In this representation, the expected accident rate for AADT = 120,000 vpd is proportional to the slope of the line that joins the origin and point C.

Figure 7 shows changes in the crash rate within the Colorado six-lane SPF for all crashes, for which the rate increased from 0.64 accidents per million VMT to 1.56 accidents per million VMT and then began to decrease to 1.4 accidents per million VMT. In this representation, the change in the slope of the line that connects the origin with the point on the SPF reflects an increase or decrease in the accident rate.

The sigmoid functional shape in Figure 7 has two critical points, B and D, where the rate of change in the gradient of the function is significantly altered. These points were located by using a sliding interval analysis in the framework of the numerical differentiation technique described by Rao (*10*).

In an effort to relate freeway flow parameters, such as speed and density during peak period, associated with the changes in the shape of the SPF, the methodology of the *Highway Capacity Model* was

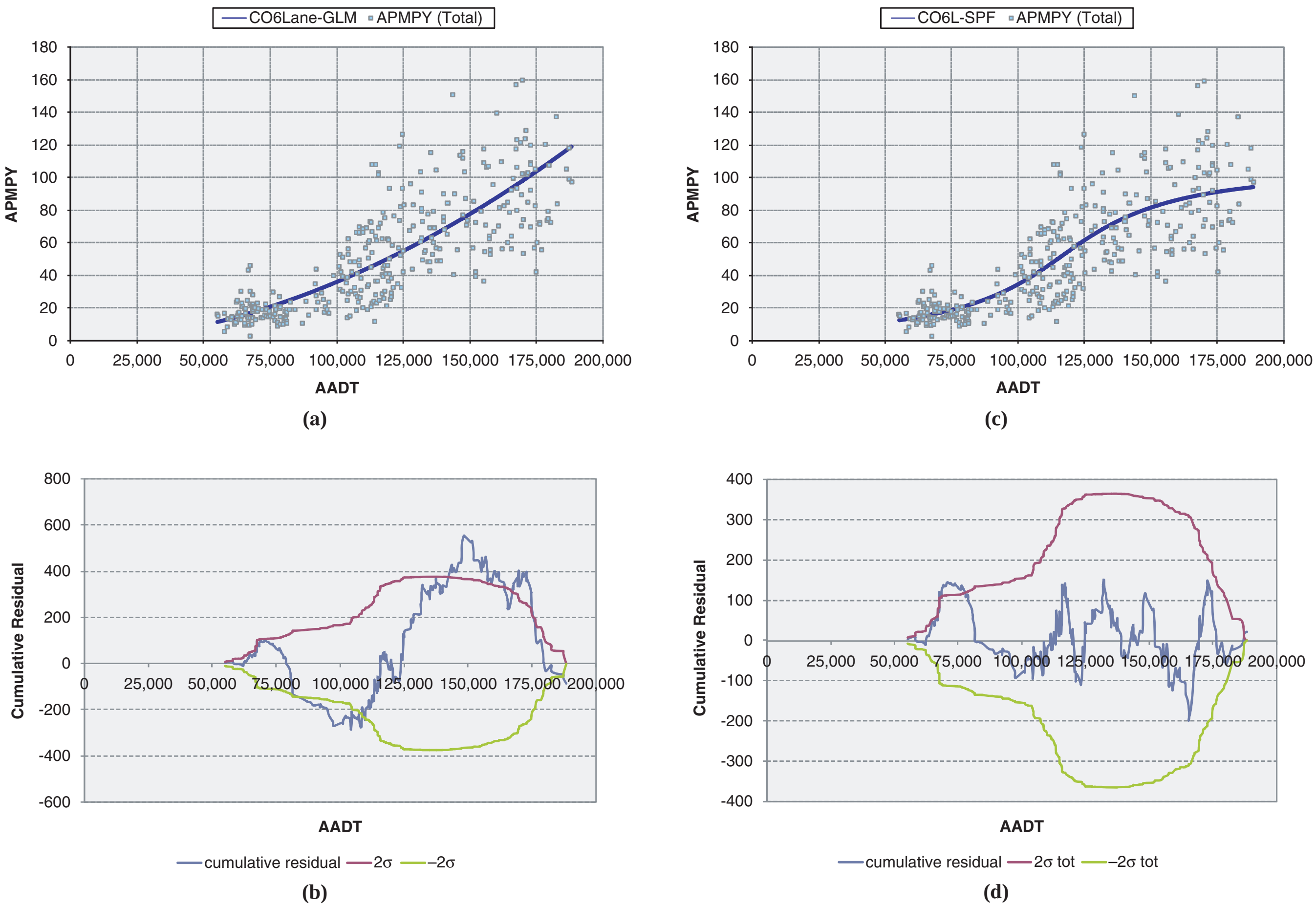

FIGURE 2 Comparison of Colorado six-lane urban freeway SPFs for all crashes with (*left*) GLM power function and (*right*) NN sigmoid (APMPY = accidents per mile per year).

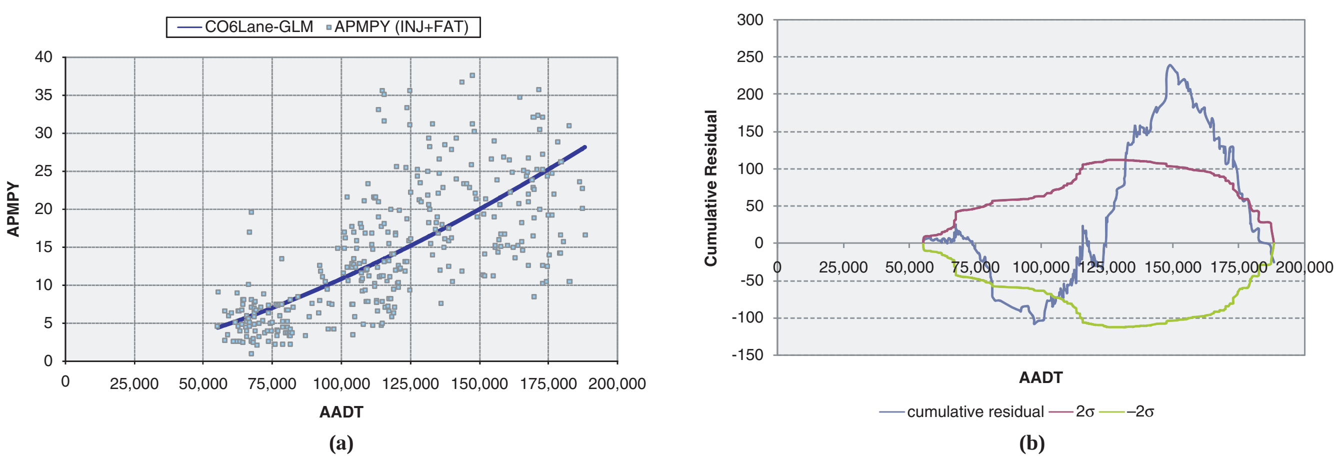

FIGURE 3 Comparison of Colorado six-lane urban freeway SPFs for injury and fatal crashes only with (*a* and *b*) GLM power function.

(continued)

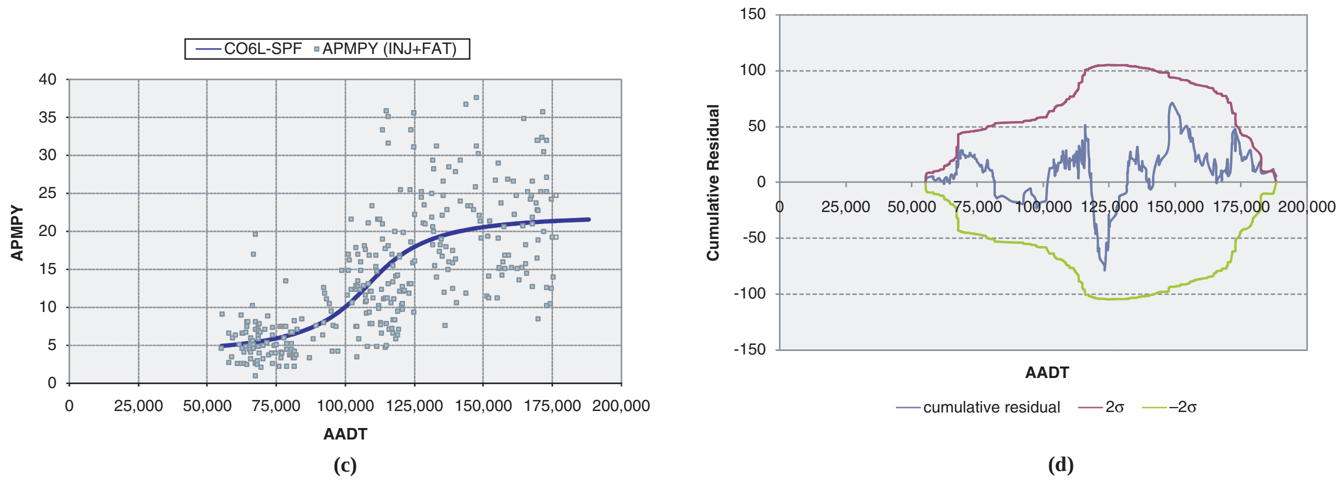

FIGURE 3 *(continued)* Comparison of Colorado six-lane urban freeway SPFs for injury and fatal crashes only with (*c* and *d*) NN sigmoid.

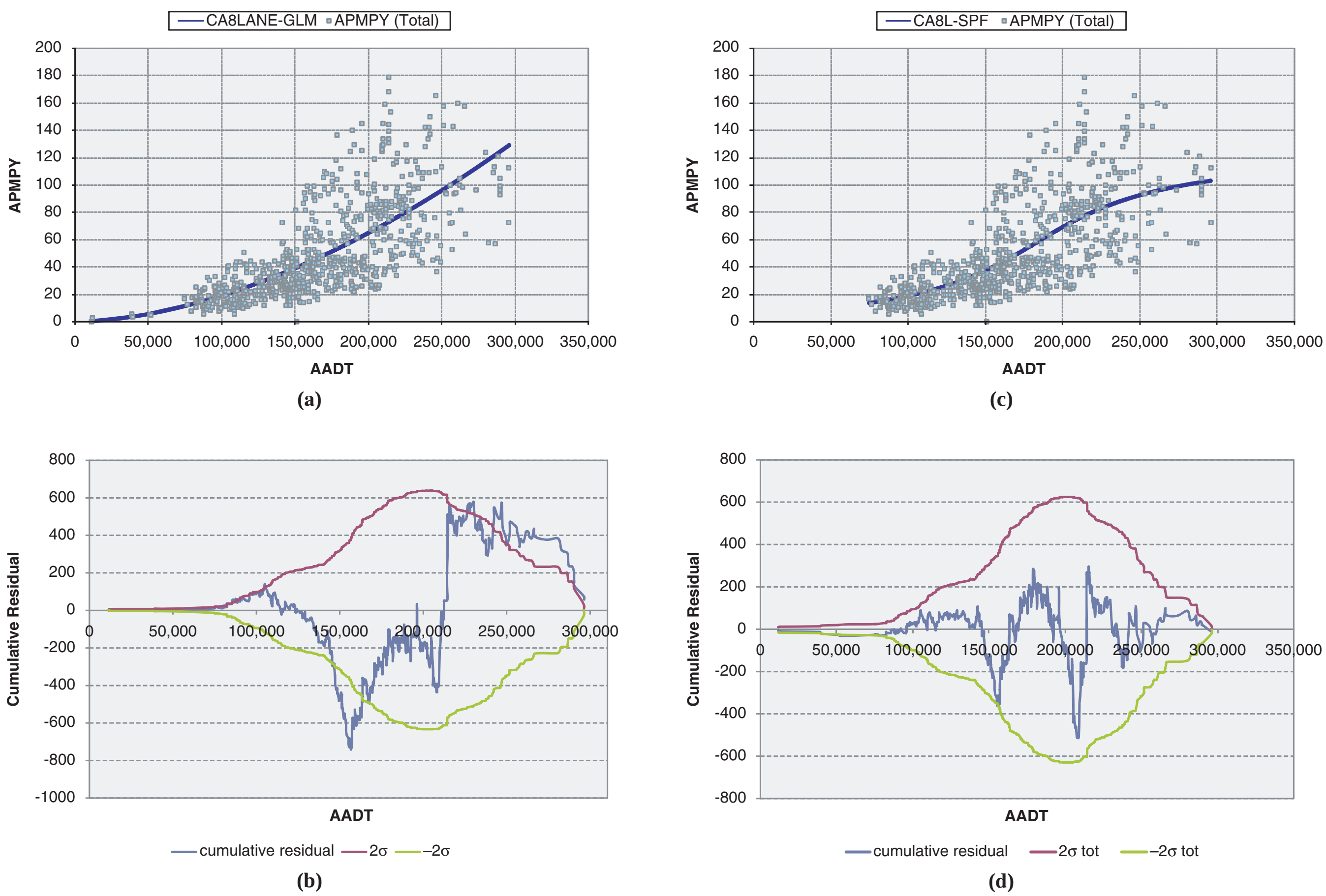

FIGURE 4 Comparison of California eight-lane urban freeway SPFs for all crashes with (*left*) GLM power function and (*right*) NN sigmoid.

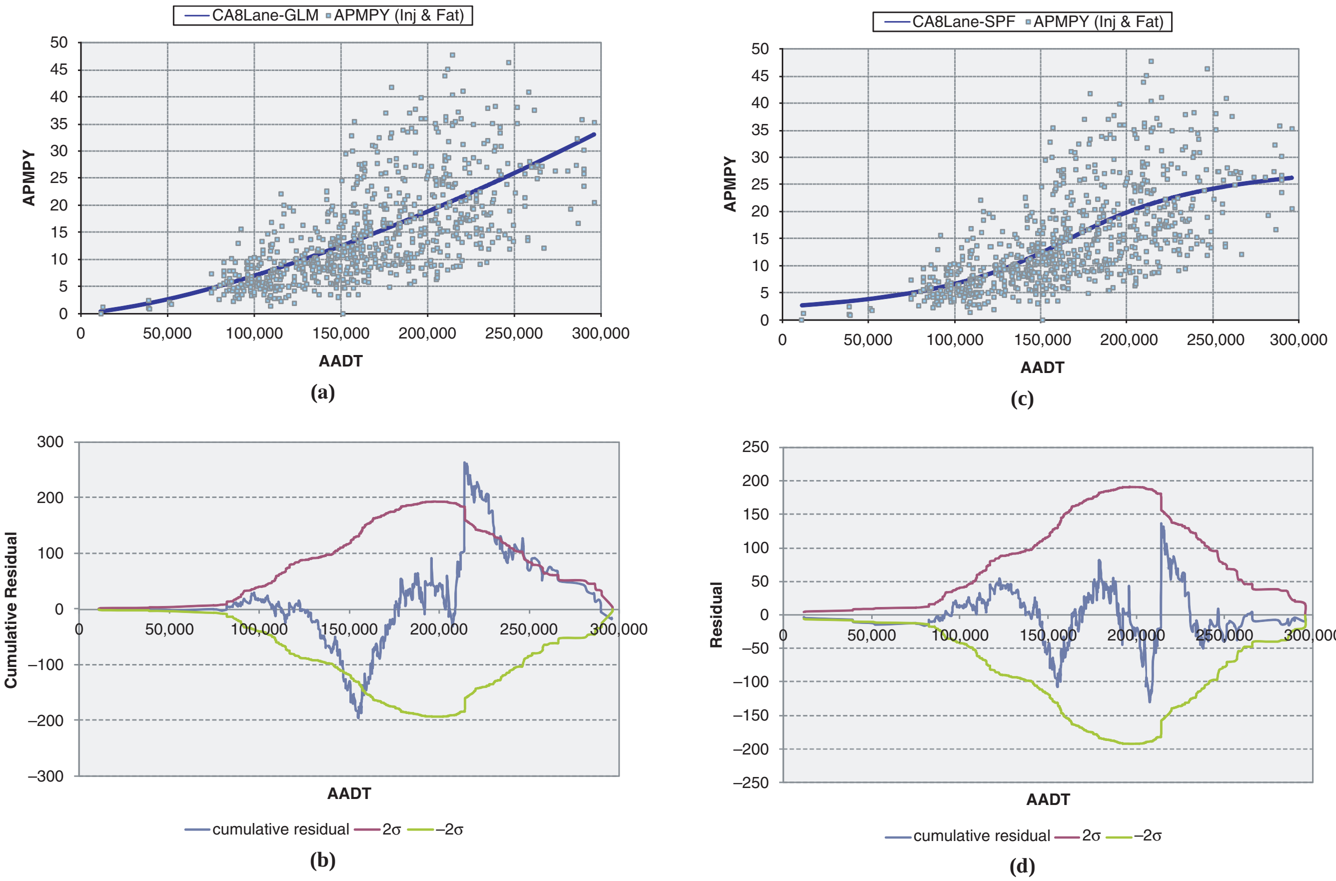

FIGURE 5 Comparison of California eight-lane urban freeway SPFs for injury and fatal crashes only with (*left*) GLM power function and (*right*) NN sigmoid.

used (*6*). The assumptions typical of the urban freeway environment were as follows:

- Design hourly volume = 10% of AADT for AADT < 130,000, 8% of AADT for AADT > 130,000;
- Peak hour factor = 0.9;
- Percentage of trucks during peak period = 2%;
- Terrain = level;
- Lane width = 12 ft;
- Shoulder width > 6 ft; and
- Interchange spacing = 1 interchange per mile.

Results of Analysis

The results of the *Highway Capacity Model* analysis were superimposed onto the SPF graph and are presented in Figure 8. Traffic density at 90,000 AADT (identified previously as a critical point on the SPF) can be viewed as a critical density, beyond which accidents increase at a faster rate. A portion of the SPF to the left of critical density can be viewed as a subcritical zone, where accidents increase gradually with AADT. Traffic density at 150,000 AADT can be viewed as a super-critical density, beyond which accidents increase gradually with AADT, and accident rates decline. A portion of SPF to the right of super-critical density can be viewed as a super-critical zone. A portion of the SPF between critical and super-critical densities can be termed a transitional zone.

As AADT increased from 60,000 to 90,000, traffic density increased by 50% (from 16 to 24 passenger cars per mile per lane), while operating speeds remained almost the same (70 and 69 mph). When traffic density increases by 50% and perception reaction times remain unchanged, performance characteristics of vehicles are constant, and operating speeds remain high, it is not unreasonable to expect that accident probability also increases. With an increase in traffic density at freeway speeds, the urban freeway environment becomes much less forgiving of driving error and road rage. The SPF reflects that, past the AADT of 90,000, the number of crashes increases at a much faster rate with an increase in AADT. A possible explanation is that traffic has reached some critical density, beyond which notably higher accident rates are observed. This increase in the rates is made manifest by the steeper gradient of the SPF.

Examination of the SPF in concert with traffic operations parameters suggests that, when freeways are not congested and traffic density is low, the number of crashes increases only moderately with an increase in traffic. That is why, initially, the slope of the SPF is relatively flat. Once critical density is reached, however, the number of crashes begins to increase at a much faster rate with an increase in traffic. Attainment of critical density can be viewed as similar to a critical mass in physics. The mixture of density and speed of traffic is such that the probability of a crash increases substantially, and thus a steep

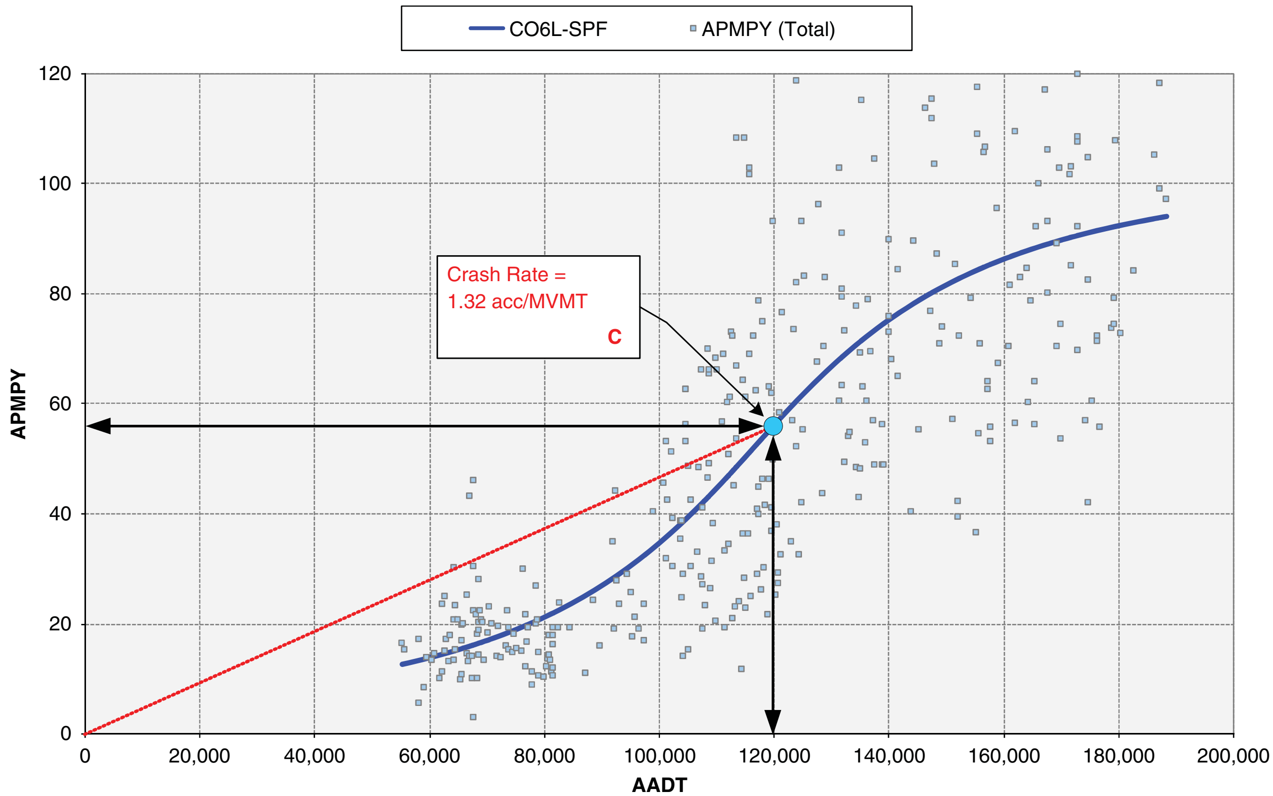

FIGURE 6 Colorado SPF six-lane urban freeways for all crashes: accident rates and frequency per mile per year (acc/MVMT = accidents per million vehicle miles traveled).

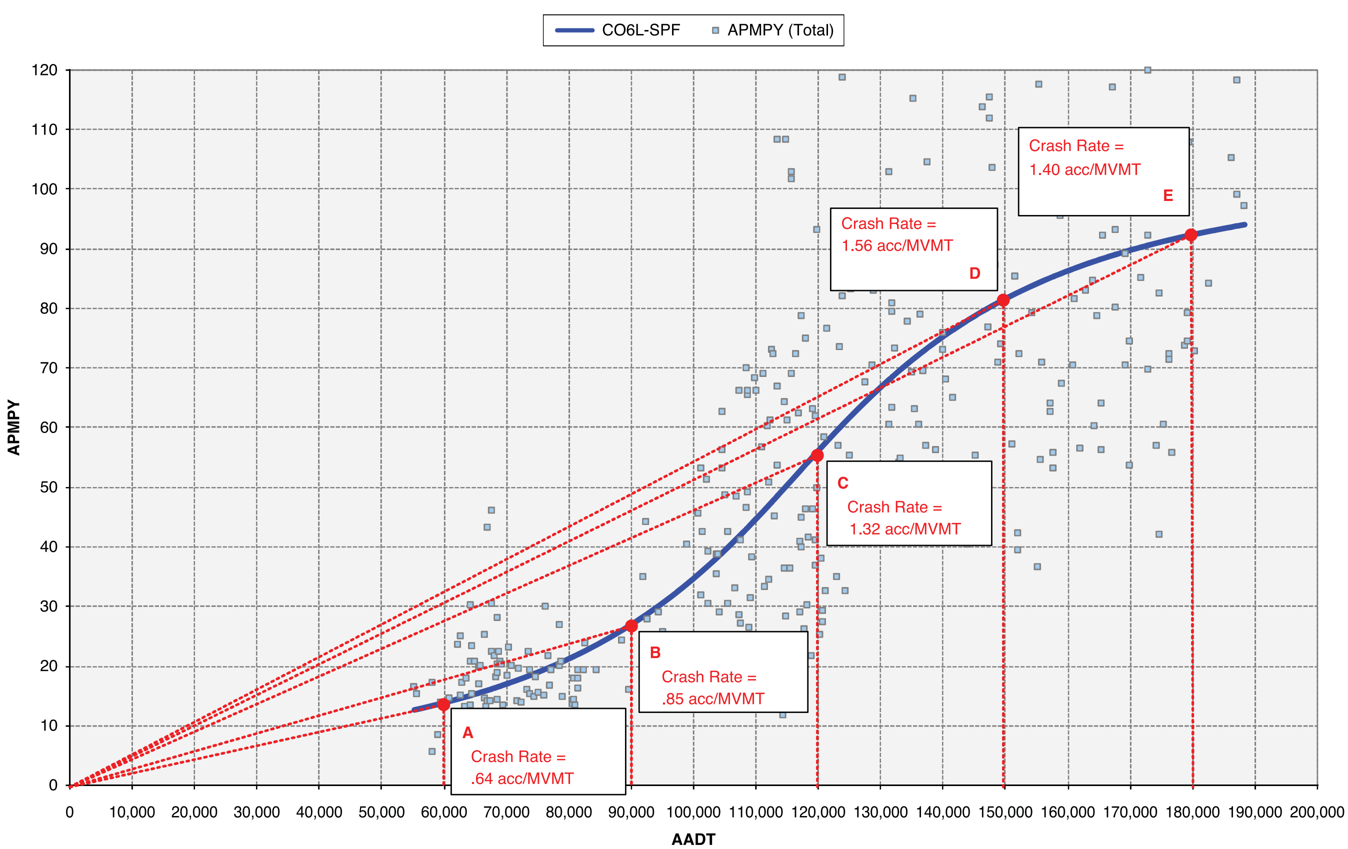

FIGURE 7 Changes in the crash rate with Colorado six-lane SPFs for all crashes.

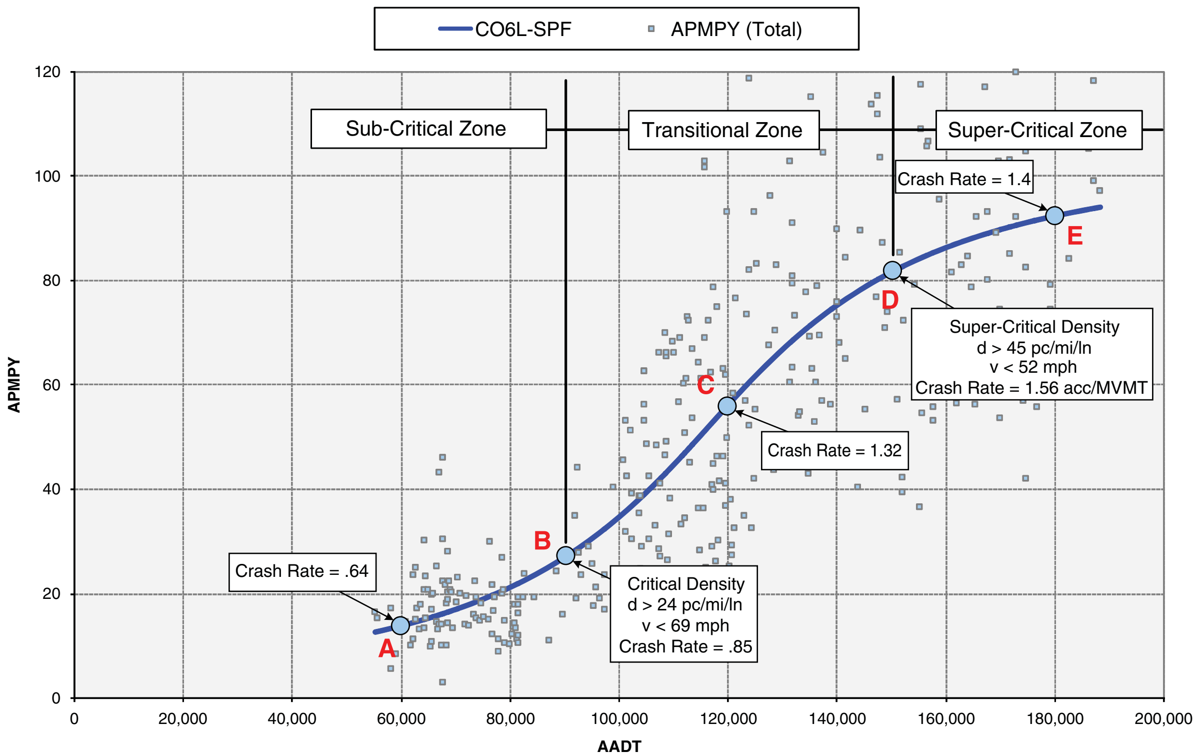

FIGURE 8 Six-lane freeway SPF: relating changes in traffic speed and density to crash rates (pc/mi/ln = passenger cars per mile per lane).

reach of the SPF. If perception–reaction time, vehicle characteristics, and roadway characteristics and speed remain constant while there are 50% more cars in the same space, it is highly plausible to expect an increased probability of crash occurrence.

Further examination of the SPF suggests that past the point of super-critical density (AADT of 150,000) the function begins to level off, which reflects only moderate increases in accidents and decreases in accident rates related to a high degree of congestion and a significant reduction in operating speeds. At this point, density exceeds 45 vehicles per mile per lane and speeds are well below 50 mph. Figure 9 graphically illustrates the idea that an accident is more likely at higher densities when the operating speed is virtually the same.

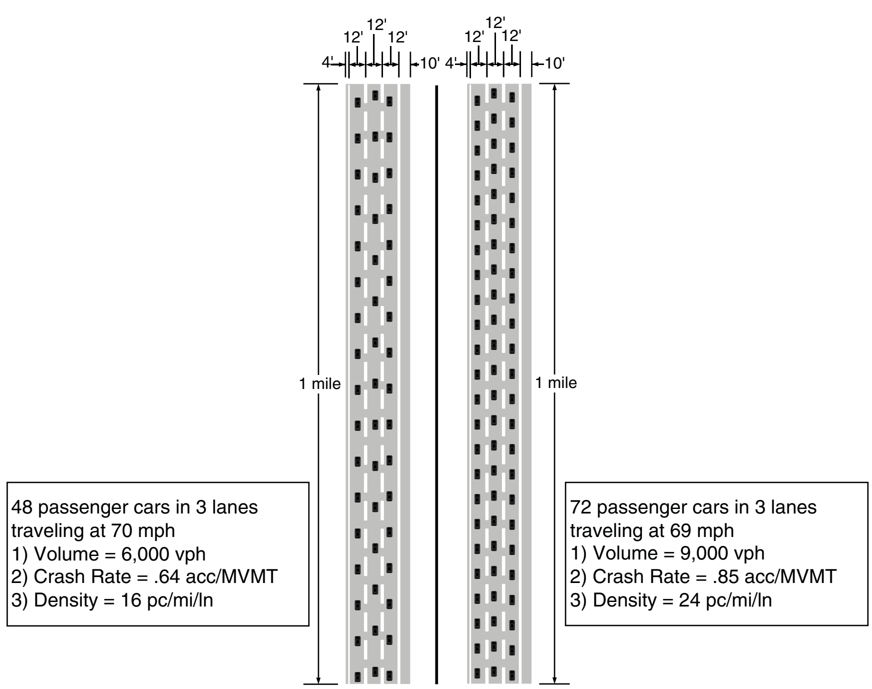

FIGURE 9 Crash rates at different densities at similar speeds (vph = vehicles per hour).

SUMMARY

According to Hauer, to do applied research without providing the corresponding theory is like attempting to build the roof of a house with no foundation (*4*). This paper offers a possible connection between the functional form of the SPF and the physics of accident occurrence on urban freeways by examining changes in speed and density and their effect on accident rates. The functional shape of SPF generated through the training of NNs suggests that a sigmoid may be a reasonable approximation of the physics of crash occurrence on urban freeways.

It was observed that, on uncongested freeway segments, the number of accidents increased only moderately with an increase in traffic. Once some critical traffic density was reached, however, the number of accidents began to increase at a much faster rate with an increase in traffic. This phenomenon was reflected in a steeper gradient of the SPF. High-density traffic in the high range of AADT is associated with approaching a super-critical density of flow. A leveling off of the SPF, accompanied by the reduction of accident rates, reflects a high degree of congestion and a significant reduction in operating speeds.

CURE plots of sigmoid SPFs generated by training NNs consistently showed better-quality model fit when compared with power function SPFs developed in the GLM framework with NB error structure.

The sigmoid represents only a simplified approximation of what actually happens on urban freeways, because driver focus on the driving task increases when the task becomes more demanding. Yet the sigmoid functional form offers a reasonably good estimate of the relationship between safety and exposure.

REFERENCES

1. Abdel-Aty, M. A., and A. E. Radwan. Modeling Traffic Accident Occurrence and Involvement. *Accident Analysis & Prevention,* Vol. 32, No. 5, 2000, pp. 633–642.
2. Lord, D., S. Washington, and J. Ivan. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention,* Vol. 37, No. 1, 2005, pp. 35–46.
3. Lord, D., A. Manar, and A. Vizioli. Modeling Crash-Flow-Density and Crash-Flow: V/C Ratio for Rural and Urban Freeway Segments. *Accident Analysis and Prevention,* Vol. 37, No. 1, pp. 185–199.
4. Hauer, E. Cause, Effect and Regression in Road Safety: A Case Study. *Accident Analysis and Prevention,* Vol. 42, No. 4, pp. 1128–1135.
5. Hauer, E. Statistical Road Safety Modeling. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 81–87.
6. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 2000.
7. Hauer, E., and J. Bamfo. Two Tools for Finding What Function Links the Dependent Variable to the Explanatory Variables. *Proc., International Cooperation on Theories and Concepts in Traffic Conference,* Lund, Sweden, November 1997.
8. Ripley, B. D. *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, United Kingdom, 1996.
9. Wahba, G. Spline Models for Observational Data. *Proc., Society for Industrial and Applied Mathematics,* Philadelphia, Pa., 1990.
10. Rao, S. S. *Applied Numerical Methods for Engineers and Scientists.* Prentice Hall, Upper Saddle River, N.J., 2002.

The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.