Adapting Safety Perfomance Functions for signalized four-legged intersections

---

Losa M.
Professor - Department of Civil Engineering University of Pisa

Terrosi Axerio A.
Ph.D.Student - Department of Civil Engineering University of Pisa

Synopsis

Safety Performance Functions have several important uses in road safety analysis. These functions, also known as Accident Prediction Models, are equations able to give an estimate of the expected average number of accidents at similar entities, relating the annual accident experience of an entity to its features.
These safety tools can forecast the expected annual number of accidents for a given "past period" or "future period", in way to allow the assessment of safety performance of an entity and the safety effect of design changes for new road projects and treatments to existing road. Moreover the application of these models can avoid the problems closely related with the police-reported accidents which are influenced by reportable criteria, police procedures, lack of reporting and errors to report.
These models have to be properly calibrated, but this task is particularly hard due to the complexity to specify the mathematical form, the accommodation of the peculiarities of accident data and the transferability of models to other jurisdictions.

The aim of this study is to develop a Safety Performance Function for four-legged signalized intersections; some of these are located on two collector roads, crossing build-up areas, while the others are in urban areas, so that all selected intersections are characterized by urban environment, factor that can directly influence the count of the expected accidents.
The methodological approach used during this research lies on choosing an appropriate base model and on verifying its suitability to the real traits of the examined context.
In order to obtain this purpose, it has been chosen one of the model for four-legged signalized intersections proposed during a research conducted in Toronto. It calculates the expected accidents for these types of entities as the product of the intersection traffic demands raised to a power; exactly the Toronto Safety Performance Function relates accidents to the entering AADT of the major and minor roads, whose exponents change with intersection features and type of accident data (injury or all accident severities).

The selection of the more suitable model form has been based on the integral-derivative (ID) method. Basically the method consists of creating an empirical integral function (EIF), for each independent variable, and then to compare the EIF graph created with pre-established graphs of well-known functions (power, gamma, polynomial, etc…) in order to indicate the proper relationship between the dependent and independent variables.

In order to obtain the coefficients of the selected statistical model it has been implemented a calibration procedure that, by using the method of the maximum likelihood function, assesses the model parameters that make this function the largest.
The validity of the selected models and of its coefficients has been investigated with the Cumulative Residual (CURE) method: this method consists of plotting the cumulative residuals (the difference between the actual and fitted values for each intersection) for each independent variable. It is possible to assert that the selected Accident Prediction Model fits with good accuracy the data available as the cumulative residuals oscillate around the zero value and moreover lie between their two standard deviation boundaries ($62\sigma^*$).

Adapting Safety Performance Function for signalized four-legged intersection

INTRODUCTION

The management of highway safety requires a tool which enables highway agencies to estimate the safety performance of an existing or planned roadway and to assess the safety impacts of roadway design alternatives.

Highway engineers need to know not only the present or past safety of a roadway, but also how it will perform if particular proposed actions are taken. For this reason historical accident data are not suitable as indicators of safety performance of a roadway, as their major weakness is to be highly variable. This variability is due to the random nature of accidents which are very rare events; therefore locations with a high number of accidents could have a future experience with fewer accidents also if no improvement is made. This phenomenon, known as “regression to the mean”, produces problems both to identify potential issues of sites from the study of their historical accident data and to estimate the potential effectiveness of improvements made at such sites.

In order to overcome these problems, for many years, safety analysts have studied statistical techniques in way to develop models to predict the accident experience of roadway segments and intersections. In order to obtain these models, it is needed a large database of accidents and roadway characteristics; first, these data allow to select the functional form for the model and then, by a regression analysis, the parameters in the model are estimated. In the past, for this purpose, researchers have used the multiple linear regression analysis but recently it seems to be more suitable to use Poisson and negative binomial regression analyses. Although Poisson models are very accurate tools for predicting the expected total accidents at a site or a class of sites, they present technical difficulties which concern the phenomenon of “overdispersion” that could be overcome with the use of the negative binomial model.

Known as Accident Prediction Models (APM) or Safety Performance Functions (SPF), they are mathematical functions that estimate the accident frequency of a site (intersections or road sections) as a function of traffic flow and other site characteristics. Given that Accident Prediction Models make quantitative estimates of accident frequency, they have nowadays several important uses in safety analysis; for example they can be used by highway agencies in the identification of sites which need a possible safety treatment and in the comparison of the anticipated safety performance of two or more geometric alternatives for proposed treatments. These models are also used in the empirical Bayes method to reduce the random fluctuation in accident counts for estimating the number of accidents expected at a specific roadway site. In particular, this expected value is used in before – after studies.

This paper presents an Accident Prediction Model for urban four legged signalized intersections and the procedure used for model coefficient estimation, which is not a straightforward operation. The complexity of calibrating Accident Prediction Models is due, first, to the high quality of data required for a large enough sample of entities and accidents; second, in order to obtain a large enough sample of accidents, several years of data are used. This difficulty is increased by temporal trends in accident counts because of the influence of factors that change every year. Third, the specification of the mathematical form is not a trivial task.
The validity of the proposed model has been assessed by the statistical model validation; this warrants the transferability of models both over future time periods and at different geographic locations, and identifies where future model improvements might be made.

BACKGROUND OF ACCIDENT PREDICTION MODELS FOR INTERSECTIONS

James Bonneson and Patrick McCoy (Bonneson J. and McCoy P., 1993) developed a safety prediction model for stop-controlled rural intersections. They used a model structure based on the nonlinear relationship between accident frequency and traffic demand, to relate expected accident frequency to the product of the average daily traffic demands on the major and minor roadways at the junction.

In order to calibrate the accident prediction model the Authors adopted the generalized linear modelling approach; this analysis tool, described by Hauer et al. (Hauer et al. 1988), overcomes problems found when using the traditional least-square regression of accident data, which considers a normally distributed error structure and constant variance.

A.Vogt and J.Bared (Voght A. and Bared J., 1998) used advanced statistical models for modelling SPF for segments, three-leg and four-leg intersections stop-controlled on the minor legs. Models were of negative binomial and extended negative binomial form and they were developed through the study of data obtained from Highway Safety Information System (HSIS) files for the state of Minnesota and Washington.

Variables used in modelling intersection accidents were:

- Total number of accident in the given period (intersection accident and intersection-related accidents occurring within 76m (250ft) of the intersection);
- Injury accidents in the time period: INJACC;
- Average daily traffic on mainline: ADT_1 in veh/day;
- Average daily traffic on minor road: ADT_2 in veh/day;
- Degree of curvature for horizontal curves: HI_m
- Crest curve grade rate: VI_m;
- Posted speed on the main road, averaged if necessary: $SPDI_m$;
- Roadside rating within 76m (250ft) of the intersection on the major road: HRI;
- Number of driveways within 76m (250ft) of the intersection on the main road: ND;
- Canalization on the main road: RT;
- Intersection angle: α in degrees.

The Minnesota data gave plausible models, indicating that different variables are significant for three-legged and four-legged intersections.

The researchers demonstrated the suitability to use SPFs to calculate the expected accident frequency as a function of the product of traffic demands entering the junction, where traffic demands are raised to a power that usually is less than unity.

Later, the FHWA sponsored a research with the aim to validate statistical models and algorithms, in way to develop a reliable method to estimate safety performance of roadways. A report (Harwood et al.2000) documents an accident prediction algorithm for implementing a new approach for rural two lane highway sections that includes road segments and five types of intersections. The accident prediction algorithm consists of base models and accident modification factors (AMF); the base models give an estimate of the safety performance of sites for base conditions. The AMFs adjust the base model prediction in way to take into account the effects on safety, in the case of at grade intersections, of skew angle, traffic control, exclusive left and right turn lanes, sight distance and driveways.
The accident prediction algorithm has been developed for the incorporation in the Interactive Highway Safety Design Model (IHSDM) as the “Crash Prediction Module” (CPM), but it is also suitable for alone applications.

In the Harwood et al. report, separate base models have been formulated for three-leg intersections with STOP control, four-leg intersections with STOP control, and four-leg signalized intersections.
The effect of traffic volume on predicted accident frequency for at-grade intersections is incorporated trough the base models, while the effects of geometric and traffic control features are incorporated through the AMFs. Each of the base model for at-grade intersections incorporates separate effects for the AADTs on the major and minor road legs, respectively.

The base models presented in the Harwood et al. report for the three different types of intersections for base conditions are presented below:

$N_{bi} = \exp(-10.9 + 0.79 \ln AADT_1 + 0.49 \ln AADT_2)$ for Three-leg STOP-Controlled Intersections; (1)
$N_{bi} = \exp(-9.34 + 0.60 \ln AADT_1 + 0.61 \ln AADT_2)$ for Four-leg STOP-Controlled Intersections; (2)
$N_{bi} = \exp(-5.73 + 0.60 \ln AADT_1 + 0.20 \ln AADT_2)$ for Four-leg Signalized Intersections. (3)

where:

$AADT_1$ = annual average daily traffic volume (veh/day) on the major road;
$AADT_2$ = annual average daily traffic volume (veh/day) on the minor road.

The reliability of an APM estimate is enhanced if the APM is based on data for as many years as possible; so it is necessary to account the trend in accident count because of the influence of factors that change every year like weather, the economy, accident-reporting practices etc…So if a Generalized Linear Model (GLM) is used to calibrate a safety performance function, in order to incorporate trend in accident data in coefficients of the GLM, they must be calculated using the traditional maximum-likelihood method. However, the likelihood function can be very complicated to define and to solve; in way to overcome this difficulty, an alternative method known as the Generalized Estimating Equation (GEE) procedure was proposed by Liang and Zeger (Lord D. and Persaud B.N. 2000).

The same procedure was used to calibrate accident prediction models for urban three- and four-legged signalized and unsignalized intersections in Toronto (Persaud B., Lord D. and Palmisano J. 2002). Furthermore the authors of this report showed the need that models calibrated for one jurisdiction could be applied for another jurisdiction, in way to consider differences between the States in climate, animal population, driver population and accident reporting practices. So during this study the procedure proposed for the application in the IHSDM (Harwood et al.,2000) was applied to recalibrate the British Columbia and California models for Toronto conditions.
Latter analysis revealed that for all intersection classes, the relationship between accidents and each covariate could be described either by the power function or the gamma function; therefore the mathematical forms used for Toronto APMs were the following:

$E(K) = \alpha F_1^{\beta 1} F_2^{\beta 2} e^{(\beta 3 F2)}$ F1 = Power function form, F2 = Gamma function form; (4)

$E(K) = \alpha F_1^{\beta 1} F_2^{\beta 2} e^{(\beta 4 F1)}$ F1 = Gamma function form, F2 = Power function form; (5)

$E(K) = \alpha F_1^{\beta 1} F_2^{\beta 2}$ F1 = Power function form, F2 = Power function form. (6)

where:

E(K) = the expected annual number of accidents;
F_1, F_2 = entering AADT of the major and the minor roads;
α, β_1, β_2, β_3, β_4 = coefficients to be estimated.

Choosing the suitable model form and applying the GEE procedure the Authors obtained the estimate of four types of urban intersection models (Table 1):

Tab 1: Estimates of Coefficients for Toronto APMs

	Signalized 4-legged		Signalized 3-legged		Unsignalized 4-legged		Unsignalized 3-legged	
Parameters	All	Injury	All	Injury	All	Injury	All	Injury
N° of intersection	868	868	250	250	59	59	117	117
Accidents	54.989	16.339	7.214	2.074	1317	357	1690	472
Model Form	Eq. 1	Eq. 1	Eq. 3	Eq. 3	Eq. 2	Eq. 3	Eq. 3	Eq. 1
LN(α)	-8.424 (14.28)*	-10186 (15.18)	-11.232 (11.88)	-13.997 (11.76)	-11.025 (4.59)	-7.584 (2.92)	-7.566 (4.91)	-35.098 (3.64)
β_1	0.534 (13.02)	0.622 (12.44)	0.803 (10.16)	0.984 (10.14)	0.607 (3.63)	0.602 (3.31)	0.440 (3.03)	3.320 (2.88)
β_2	0.566 (13.16)	0.530 (11.52)	0.568 (13.21)	0.524 (8.19)	0.903 (3.07)	0.205 (1.42)	0.565 (10.22)	0.478 (7.84)
β_3	8.92E-6 (2.24)	6.94E-6 (1.64)						
β_4					-2.29E-4			

					(2.04)			
γ	6.91	5.64	4.51	4.35	3.52	4.08	4.75	7.00
R_{α}^{2}	0.77	0.75	0.70	0.78	0.41	0.58	0.63	0.83

*The numbers in brackets are the values of the t-statistics

The results of the calibration are reported in Table 2.

Tab 2: Model Forms, Prameter Estimates, and Calibration Factors for California and British Columbia Models

Parameters	Jurisdiction	Unsignalized 3-legged (All severities)	Unsignalized 4-legged (All severities)	Signalized 4-legged (Injury)	Signalized 4-legged (All severities)
Model form	California	Equation 3	Equation 3	Equation 3	Equation 3
	Vancouver	Equation 3	Equation 3	Equation 3	Equation 3
	Toronto	Equation 3	Equation 2	Equation 1	Equation 1
α	California	0.08615	0.000502	0.00106	0.00324
	Vancouver	0.0002464	0.0002585	n/a	n/a
	Toronto	0.0005177	0.0000162	0.0000376	0.0002195
β1	California	0.683	0.620	0.574	0.503
	Vancouver	0.4531	0.4489	n/a	n/a
	Toronto	0.440	0.607	0.622	0.534
β2	California	0.245	0.281	0.251	0.234
	Vancouver	0.5806	0.6475	n/a	n/a
	Toronto	0.56	0.903	0.530	0.566
Calibration Factor	California	1.238	1.790	1.214	1.633
	Vancouver	1.642	1.272	n/a	n/a
	Toronto	1.0	1.0	1.0	1.0
Root-mean-square error of prediction	California	1.495	2.083	1.920	6.009
	Vancouver	1.360	2.254	n/a	n/a
	Toronto	1.364	2.003	1.557	5.266

MODELING PROCEDURE

The aim of this study is to develop a Safety Performance Function for four-legged signalized intersections; some of these are located on two collector roads, both characterized to cross build-up areas, while the others are within urban areas, so that all the selected intersections are characterized by a urban environment that can directly influence the count of the expected accidents.

The methodological approach used during this research lies on choosing an appropriate base model and on verifying its suitability to the real traits of the examined context; therefore it is necessary to determine the types of data needed to calibrate the chosen Safety-Prediction Model.

For this purpose, it has been chosen one of the model for four-legged signalized intersections proposed during the research conducted by Persaud et al. with the Toronto data (Persaud B., Lord D.and Palmisano J. 2002). The functions presented in the paper calculate the expected accidents for these types of entities as the product of the intersection traffic demands raised to a power; exactly these Safety Performance Functions relate accidents to the entering AADT of the major and minor roads, whose exponents change with intersection features and type of accident data (injury or all accident severities).

Data Sources

The accident data utilized in this study have been obtained directly from the archives of Police Forces.

The accident database has been developed in order to select “intersection” or “intersection-related” police reported accidents, that are those occurred within 100m of the intersection; if an accident is not indicated as “intersection-related”, then it is used the criteria to consider “intersection-related” accident if it is:

- an accident where one vehicle involved is making a left turn or right turn prior to the crash;
- a multi-vehicle accident in which the accident type is either a sideswipe, rear end or broadside/angle.

The accident database consists of fatal and injury accidents (excluding accident with property damage only) at 14 four-legged signalized urban intersections in the Pisa Province for each of the years 1999 through 2002. The count of accidents is reported in Table 3.

Tab 3: Injury and fatal accident data

Intersection	1999	2000	2001	2002
1	3	3	2	5
2	0	4	1	0
3	2	1	2	2
4	1	1	2	1
5	0	0	0	3
6	1	6	3	3
7	1	3	0	4
8	6	1	3	4
9	1	1	3	2
10	1	2	6	7
11	6	9	9	6
12	1	2	3	4
13	1	2	1	1
14	1	0	1	0

In way to calibrate an APM for four-legged signalized urban intersection, the others data utilized are major and minor road flows of each intersection. These traffic flows are not available for all years at any given intersection, so the missing traffic counts are estimated by considering an yearly rate of traffic growth equal to 2%. The application of this procedure is needed to overcome biased estimates due to missing values, which are usually systematic because of the trend of counting affects more often the higher-volume intersections than low-volume ones. The traffic flows data for each intersection are presented in Table 4.

Tab 4: Traffic flows of major F_1 and minor F_2 roads (veh/day)

Intersection	F_1(99)	F_1(00)	F_1(01)	F_1(02)	F_2(99)	F_2(00)	F_2(01)	F_2(02)
1	17000	17300	17700	18000	6800	6900	7000	7200
2	10500	10700	10900	11100	9900	10100	10300	10600
3	10000	10200	10400	10600	8000	8200	8400	8500
4	9500	9700	9900	10100	3600	3700	3800	3800
5	14600	14900	15200	15500	6700	6900	7000	7200
6	15800	16100	16500	16800	4400	4500	4600	4700
7	19600	20000	20500	20900	8600	8800	9000	9200
8	10700	10900	11100	11400	5100	5200	5300	5400
9	15900	16200	16500	16900	3700	3800	3900	3900
10	16800	17200	17500	17900	6400	6500	6700	6800
11	14000	14300	14600	14900	5600	5700	5800	5900
12	13600	13800	14100	14400	6700	6800	7000	7100
13	12900	13100	13400	13700	2000	2000	2040	2100
14	10000	10200	10400	10600	5100	5200	5300	5400

Selection of the Model Form

The selection of the model form is based on the integral-derivative (ID) method proposed by Hauer and Bamfo (Persaud B., Lord D. and Palmisano J., 2002; Lord D. and Perseaud B.N., 2000). Basically the method consists of creating an empirical integral function (EIF), for each independent variable, which is separated into a series of bins in increasing order. For a given intersection, the left boundary of the bin is located halfway between the value of the independent variable for the current intersection and that for the previous one. The right boundary is located halfway between the value for the current intersection and that for the next one. The bin height is the number of accidents that occurred at that intersection. Hence the value of the EIF at the right boundary of the current bin is the sum of all bin heights from the lowest value up to that boundary. The aim of the method is to compare the EIF graph created with pre-established graphs of well-

known functions (power, gamma, polynomial, ect…) in order to indicate the proper relationship between the dependent and independent variables.
The cumulative probability graphs F(x) for the power, polynomial and gamma functions are on the right side of Figure 1, whereas the actual form and the shape of the graph f(x), which relates the dependent variable (accidents) to the independent variables (traffic flows), are indicated by the plots on the left side of Figure 1.

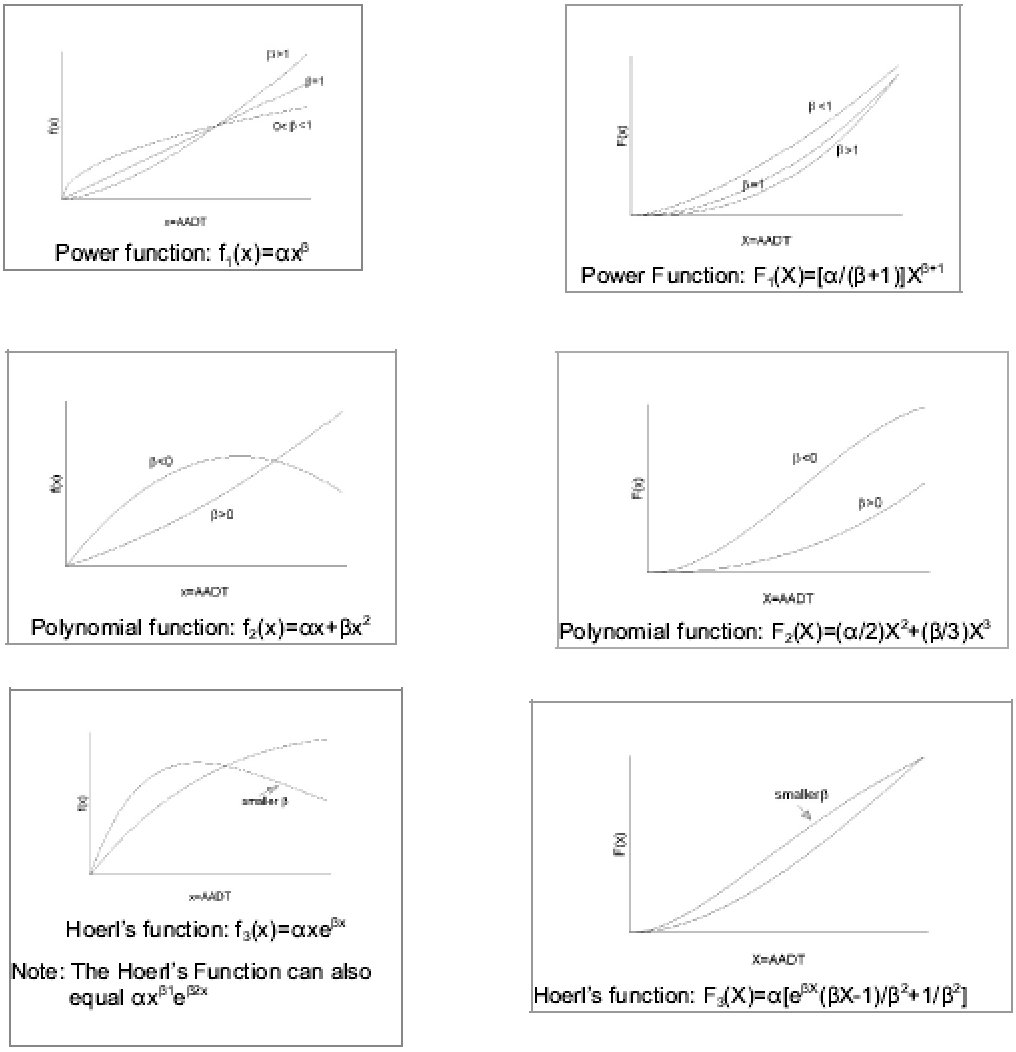

Figure 1: Corresponding f(x) and F(x) of power, polynomial and gamma(or Hoerl's) functions

The explanatory variables selected to create the EIF graphs are the annual average daily traffic (AADT) on the major road (F_1) and on the minor road (F_2). These functions can be considered as an assessment of the "integral function" for the most suitable functional form; for this reason, during this research, the functions able to fit both the EIF graphs have been detected. The derivatives of the found functions can be considered the most appropriate model forms which relate the dependent variable to the two independent ones.

The integral function that fits in a more suitable way the EIF graphs has the following mathematical form:

$$F(x)=\frac{1}{(-a)^{(n+1)}}\cdot\left[n\cdot\Gamma(n)-e^{ax}\cdot(-ax)^{n}-n\cdot\Gamma(n,-ax)\right] \qquad (7)$$

where:

F(x) = Integral Function

$$\Gamma(n)=\int_0^{\infty}t^{n-1}\cdot e^{-t}dt$$

$$\Gamma(n,-ax) = \frac{1}{\Gamma(-ax)} \int_0^n e^{-t} \cdot t^{-ax-1} dt$$

n,a = coefficients to be estimated

In Figure 2 and Figure 3, EIF graphs for the two flows, respectively F1 and F2, and the “integral functions” (7) that fit the graphs are shown.

Since it has been demonstrated that the more accurate distribution of the traffic accidents is the negative binomial one, the EIF variance becomes greater with an increasing of the traffic flow. For this reason, when the EIF graphs are compared with the “well-known” functions, the end of the curve would not be considered important for the study because they could generate an illusory effect for choosing the right functional form.

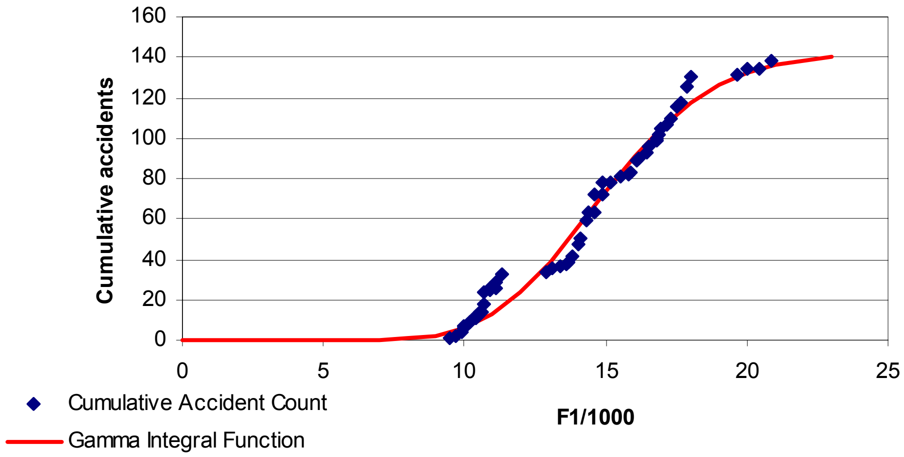

Figure 2: EIF of injury accidents (fatal plus nonfatal) at four-legged signalized intersections versus major road entering AADT/1000

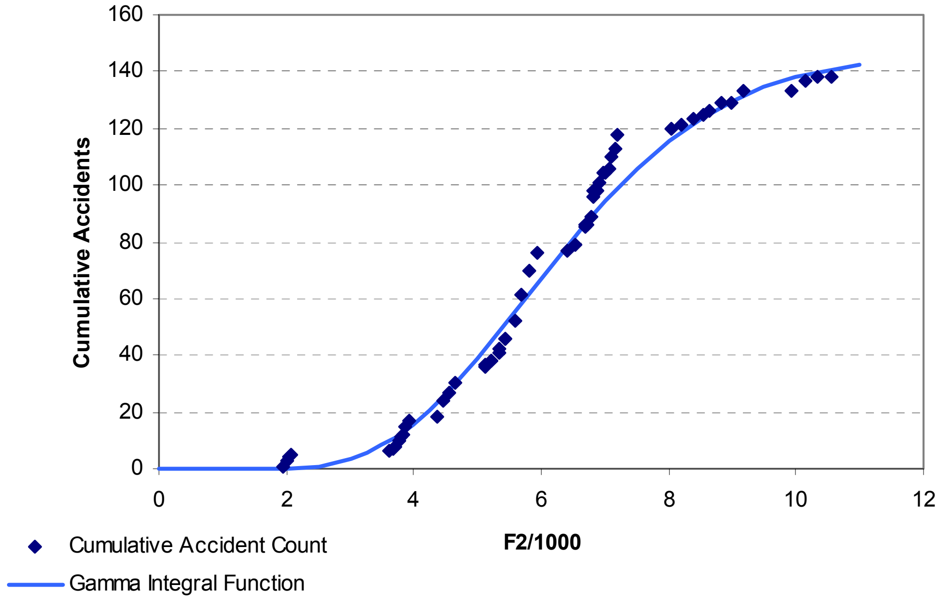

Figure 3: EIF of injury accidents (fatal plus nonfatal) at four- legged signalized intersections versus minor road entering AADT/1000

The derivative of the function (7) gives the model form which relates the dependent variable to the two independent ones. The resulting model form is expressed by the following relationship:

$$E(k) = \alpha F_1^{\beta_1} F_2^{\beta_2} e^{(\beta_3 F_1 + \beta_4 F_2)} \qquad (8)$$

where:

E(k) = the expected annual number of accidents;
F_1, F_2 = entering AADT/1000 of the major and the minor roads;
α, β_1, β_2, β_3, β_4 = coefficients to be estimated.

The Integral-Derivative (ID) method shows that for the examined data, the expected number of accidents is related to the flow F_1 and the flow F_2 by a gamma (or Hoerl's) relationship.

Calibration Procedure and Results

The selected equation (8) is a multivariate statistical model that links the expected accident frequency (E(k)) of an entity to its observed traits, also called "casual factors" (in this case: F_1 and F_2). The parameter α is used to capture the influence of all factors that change year by year (weather, economic condition, accident-reporting practices etc…) except the change in traffic flow on major and minor roads of intersections because they are included in the model equation. It is necessary to underline that the effect of a specific change from year to year influences all entities in the same manner; for this reason, each year has its own α which is able to capture the trend in accident data count (Mountain L., Maher M. and Fawaz B. 1998).

The approach taken in calibrating the accident prediction model is based on the procedure described by Hauer (Hauer E.1997). The calibration procedure consists in estimating parameter values (α, β_1, β_2, β_3, β_4) using data in way that the model equation fits the data as well as possible. These coefficients cannot be assessed by the traditional least-squares regression methods because they assume a normally distributed error structure and constant variance; two assumptions that do not follow the reality of the discrete and non-negative nature of accident count. For these reasons, the estimate of the APM coefficient values is made by using the method of maximum likelihood which expresses the probability to observe the accident counts in the data as a function of the only unknown parameters, when the covariate values are put in the maximum likelihood expression. So the task of this study is to set the model parameters that make this function the largest. The natural logarithm of the likelihood function utilized in this estimation is the following:

$$\ln(L/\cos t) = \sum_{i=1}^{R} \left(\left[\sum_{y=1}^{Y} K_{i,y} \ln(C_{i,y}) \right] + b \ln(b / E\{k_{i,1}\}) - (\sum_{y=1}^{Y} K_{i,y} + b) \ln(b / E\{k_{i,1}\} + \sum_{y=1}^{Y} C_{i,y}) + \right.$$

$$\left. + \ln(b) + \ln(b+1) + ... + \ln(b + \sum_{y=1}^{Y} K_{i,y} - 1) \right) \qquad (9)$$

where:

$C_{i,y}$ = (model equation of entity i for year y)/ (model equation of entity i for year 1);

$$b = \frac{(E\{k_{i,y}\})^2}{VAR\{k_{i,y}\}}; \qquad (10)$$

$K_{i,y}$ = accident counts on entities i=1, 2, …,R in years y=1, 2, …,Y;
$K_{i,1}$ = accident counts on entities i=1, 2, …,R in the year 1;
$E\{k_{i,y}\}$ = the expected annual number of accidents for entity i in the year y;
$Var\{k_{i,y}\}$ = variance of the expected annul number of accidents for entity i in the year y.

The identification of the set of coefficient values that maximizes the log-likelihood function is not a trivial task; so the parameters have been estimated by using the Matlab software which can solve iteratively the generalized nonlinear model (GNLM) of the selected model.
The important property of GNLMs is the flexibility in specifying the probability distribution for random components; so GNLMs are especially useful in the context of traffic safety, for which the distribution of accident counts in a population often follows the negative binomial distributions (McCullagh P. and Nelder J.A.1989). This type of distribution seems to be more accurate than the normal one, generally utilized in the regression studies of accident data until now.

The choice of the negative binomial distribution has allowed us to assume a negative binomial error structure, with the consequent estimate, during the model calibration process, of the overdispersion parameter *b* (10). This value can also be used to compare the goodness of fit of various models fitted to the same data. Indeed, as we can see by equation (11), the variance of the model decreases when b increases, showing that the model with the smaller variance hence the better one.

$$Var(k) = \frac{E(k)^2}{b} \tag{11}$$

The results obtained from the calibration procedure are reported in the Table 5.

Tab 5: Set of parameter estimates

	Estimate
α_1	0.0123
α_2	0.017
α_3	0.0174
α_4	0.02
β_1	0.83
β_2	3.64
β_3	0.0189
β_4	-0.63
b	5

In way to confirm this statement it was used another procedure, called Cumulative Residual Method (CURE), which is dealt in the following section.

Model Validation

The validity of the selected model and of its coefficients is investigated by the Cumulative Residual (CURE) method (Lord D., Hauer E., Bamfo J., 1999). This method consists of plotting the cumulative residuals (the cumulative difference between the actual and fitted values for each intersection) for each independent variable. If the chosen function concurs with good precision the data, the cumulative residuals R(n) will oscillate around the value 0 with an AADT increase; this fact demonstrates that the mean of cumulative residuals would tend to 0, so the variance of R(n) can be calculated immediately from the cumulative residuals raised to square ($R(n)^2$). Under these conditions it is possible the application of the central limit theorem in way to assert that R(n) follows roughly a normal distribution with mean equal to zero and variance equal to $R(n)^2$ and, in the same way, the cumulative residuals between the values n and N have the same features (where N is total number of available data and n is a number between 0 and N). So the probability density function is the product of the two density functions: one with mean equal to 0 and variance $\sigma^2(n)$ and the second one with mean equal to 0 and variance $\sigma^2(N)-\sigma^2(n)$. The probability density function has the following equation:

$$f(R) = \frac{1}{\sqrt{2\cdot\pi\cdot\sigma^2(n)}}\cdot e^{-\frac{s^2}{2\cdot\sigma^2(n)}} \times \frac{1}{\sqrt{2\cdot\pi\cdot(\sigma^2(N)-\sigma^2(n))}}\cdot e^{-\frac{s^2}{2\cdot\left[\sigma^2(N)-\sigma^2(n)\right]}} \tag{12}$$

the exponential part of the equation can be rewritten in the following way:

$$e^{-\frac{s^2}{2\cdot\sigma^2(n)}} \times e^{-\frac{s^2}{2\cdot\left[\sigma^2(N)-\sigma^2(n)\right]}} = e^{-\frac{s^2}{2}\cdot\frac{\sigma^2(N)}{\sigma^2(n)\cdot\left[\sigma^2(N)-\sigma^2(n)\right]}} = e^{-\frac{s^2}{2\cdot\sigma*^2}} \tag{13}$$

so the variance of the probability function can be calculated by the following relationship:

$$\sigma *^2 = \frac{\sigma^2(n) \cdot \left[\sigma^2(N) - \sigma^2(n)\right]}{\sigma^2(N)} = \sigma^2(n) \cdot \left(1 - \frac{\sigma^2(n)}{\sigma^2(N)}\right) \tag{14}$$

After the variance calculation, the cumulative residual graphs of the model are compared with their standard deviation boundaries (62σ*) (Lord D., Hauer E. and Banfo J., 1999); if the graph lies between these two curves, the selected model and the found coefficients could be considered good fitted for the chosen covariates (Figures 4 and 5).

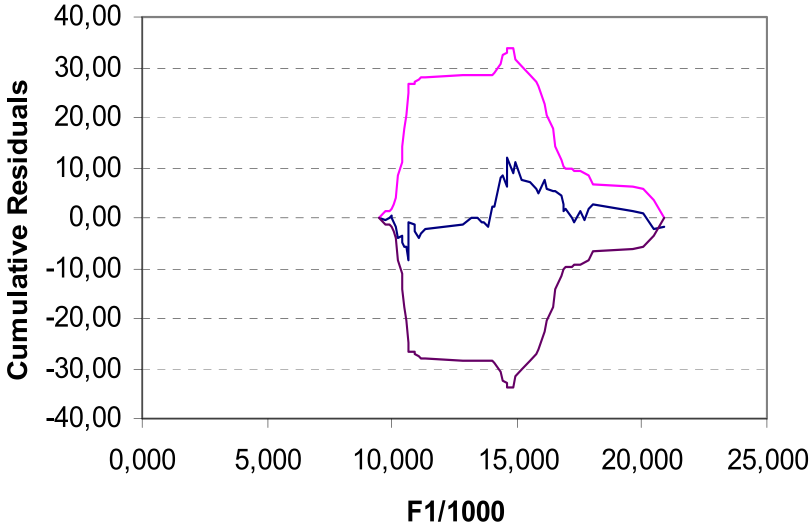

Figure 4: Cumulative residuals with ±2σ* of flows F1/1000

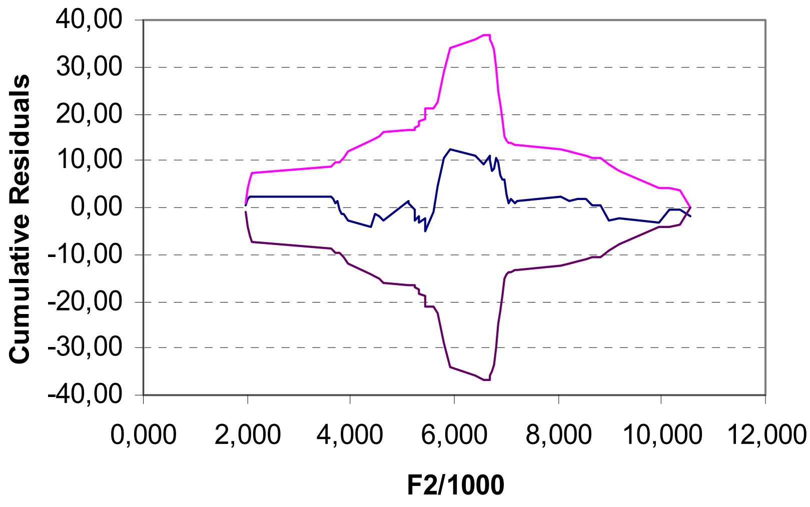

Figure 5: Cumulative residuals with ±2σ* of flows F2/1000

The CURE method shows that the chosen statistic model and its coefficients, determined by the use of the available data, has a good validity; indeed the cumulative residual curves oscillate closer to the value of zero and they exceed the two standard deviation boundaries only at the end of the curves for high values of traffic flows F1 and F2 respectively.

So the Safety Performance Function found has the following final equation:

$$E(k) = \alpha \cdot F_1^{0.83} \cdot F_2^{3.64} \cdot e^{(0.0189 F_1 - 0.63 F_2)} \tag{15}$$

with α changing every year that allows to consider the variation in accident occurrence from year to year.

CONCLUSIONS

The paper deals with an analysis procedure aimed to validate a statistical model able to provide an estimate of safety performance for signalized four-legged intersections by the assessment of accident frequency. The function presented in the paper calculates the expected accidents of an entity as the product of the intersection traffic demands raised to a power; exactly this Safety Performance Function relates accidents to entering AADT of the major and minor roads.

The model form of the SPF has been selected based on the integral-derivative method while the coefficient values have been estimated by using the method of maximum likelihood. The set of coefficient values that maximizes the log-likelihood function have been identified by using the Matlab software.

In way to assess the validity of the selected models and of its coefficients the Cumulative Residual method (CURE) was implemented. By this method it has been verified that the cumulative residuals oscillate around the value of 0 and the obtained graph lies between the two standard deviation boundaries (62σ*).

By this check it can be asserted that the selected model and the found coefficients could be considered good fitted for the chosen covariates.

ENDNOTES

The Authors wish to thank the Road Administration of the Pisa Municipality that has allowed to collect the accident and traffic flows data.

REFERENCES

HAUER E. (1997), *Observational Before-After Studies in Road Safety: Estimating the Effects of Highway and Traffic Engineering Measures on Road Safety,* Pergamon, Elvisir Science Ltd., Oxford, U.K.

PERSAUD B., LORD D. and PALMISANO J. (2002), *Calibration and Transferability of Accident Prediction Models for Urban Intersections,* Transportation Research Record 1784, Paper No. 02-3293, pp. 57 -64.

LORD D. and PERSAUD B.N. (2000), *Accident Prediction Models With and Without Trend: Application of the Generalized Estimating Equations Procedure,* Transportation Research Record 1717, Paper No. 00-0496,pp. 102 -108.

LYON C., OH J., PERSAUD B., WASHINGTON S. and BARED J. (2003), *Empirical Investigation of Interactive Highway Safety Design Model Accident Prediction Algorithm,* Transportation Research Board 1840, Paper No. 03-3718, pp. 78 -86.

HARWOOD D.W., COUNCIL M., HAUER E., HUGES W.E. and VOGT A. (2000), *Prediction of the Expected Safety Performance of Rural Two-Lane Highways,* Report FHWA-RD-99-207, FHWA, U.S. Department of Transportation.

BONNESON J.A., MCCOY P.T. (1993), *Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways,* Transportation Research Record 1401, pp. 83 -89.

LORD D., HAUER E., BAMFO J. (1999), *Application de Deux Nouvelles Méthodes pour Examiner la Relation entre les Accidents et les Variables Explicatives,* Routes et Transports, Vol.28, No.3, Association québécoise du transport et des routes, pp.11 -20.

MOUNTAIN I., MAHER M. and FAWAZ B. (1998), *The Influence of Trend on Estimates of Accidents at Junctions,* Accident Analysis and Prevention, Vol.30, No.5, pp. 641 -649.

MCCULLAGH P. and NELDER J. A. (1989), *Generalized Linear Model: Second Edition,* Chapman and Hall, Ltd., London.

VOGT A. and BARED J. (1998), *Accident Models for Two-Lane Rural Segments and Intersections,* Transportation Research Record 1635, Paper No.98-0294, pp.18 -29.

VAN EVERY B.E., *A Guide to the Economic Justification of Rural Grade Separations,* Australian Road Research, Vol. 12, No.3, pp.147 -154.

HAUER E. and LOVELL J. (1988), *Estimation of Safety at Signalized Intersections,* Transportation Research Record 1185, TRB, National Research Council, Washington, D.C., pp.48 -61.

MOUNTAIN L., MAHER M. and FAWAZ B. (1998), *The Influence of Trend on Estimates of Accidents at Junctions,* Accident Analysis and Prevention, Vol.30, No.5, pp. 641 -649.

MIAOU S.P.,LU A. and LUM H. S. (1996), *Pitfalls of Using* R^2 *to Evaluate Goodness of Fit of Accident Prediction Models,* Transportation Research Record 1542, pp.6 -13.