CRASH PREDICTION MODELS FOR ROADS INCLUDING RAINFALL AND HAZARDOUS POINTS

Ciro Caliendo
Department of Civil Engineering, University of Salerno, Italy

ABSTRACT

Predictive models for multilane roads including also the effects of rainfall and hazardous points on the crash frequency are presented in this paper. Accident data observed on a specific motorway during a period of eight years were used in statistical analysis. Negative Binomial Distribution, was applied for modelling the random variation of the crash data. Model parameters were estimated by Maximum Likelihood Method, and the Generalised Likelihood Ratio Test was used to detect the significant variables to be included in the final model. Separate prediction models for curves and tangents are proposed. For curves, it is found that the crash frequency is positively associated with the following variables: length (L), curvature (1/R), and the presence of hazardous points such as junctions (J) or tunnels (T); whereas for tangents with L and T. The effect of rain precipitation shows that with a wet pavement significant increases in crash frequency are expected, more especially on curves compared to tangents. This might be due to the combined effect of rain and centrifugal force in curve.

Key words: Negative Binomial model, Road crashes, Geometry, Rainfall, Hazardous points.


1. INTRODUCTION

Crash prediction models for the evaluation of safety interventions are continuing to be developed to provide a guide in road design to engineers and/or traffic planners with the aim to reduce vehicle crashes on roads. Generally speaking, these models are developed using appropriate statistical approaches in which the police-reported crashes are linked to traffic and geometric design variables. More especially, they link the crash frequency (crashes for year), which is observed along a specific road section, to the annual average daily traffic (AADT) and percentage of trucks; as well as to the length of road section investigated, the curvature radius, the number of lanes, the longitudinal slope, etc. However, the combined effect of these...
variables with the rainfall and the presence of both junctions and/or tunnels have been hitherto less investigated.

The statistical methodology which is generally applied in crash analysis is based on the assumption that the fluctuation of accident counts, say $Y_i$, which occur on a road section $i$ during the observational time interval, is a Negative Binomial (NB) random variable (r.v.) with $E[Y_i] = \lambda_i$ and $\text{Var}[Y_i] = \lambda_i (1 + \lambda_i / \varphi)$. Unlike the Poisson model, the NB model allows for the variance of accident counts to be greater than the mean, provided that $1/\varphi > 0$. For this reason $1/\varphi$ (or $\varphi$ itself) is often called the “over dispersion parameter”.

In terms of methodological approach, more recent research has also shown the potential of random-effects models and random-parameters models. Random-effects models are necessary when temporal correlations (data collected on the same road section over successive time periods) or spatial correlations (data collected from the same region) are suspected to be present. Random-parameters models are, instead, used to capture the possible unobserved heterogeneity (unobserved factors that may vary across accident observations), by allowing the regression model parameters to be random instead of being considered constant across observations. As regards crash data concerning to different severity levels, such as accidents involving damage only, injuries, and fatalities, it is to be said that they can be affected by significant correlations, so that analyses based on separate models might be inappropriate, and joint-probability models are needed. A large literature exists on the strengths and weaknesses both of traditional models and these new models; which however is not within the purpose of this paper. Some studies that deal with crashes modelling can be found more especially in: Hauer [1, 2], Caliendo et al. [3,4], Anastasopoulos and Mannering [5], Caliendo et al. [6], Caliendo and Guida [7], Caliendo et al. [8].

There is a general opinion that severe weather is associated with more risky driving conditions. This is probably due to a reduction in the pavement-tyre friction and a restricted sight distance for stopping during rainfall. But negative relationships can also be found between rain and road accidents. This might be due to: (i) a reduction in traffic flow and/or in operating speeds during wet weather; (ii) an analysis based on different accident types such as non-severe accidents (accident with damages only) or severe accidents (accidents with injuries and/or fatalities); (iii) a different amount of rain and/or data-sources.

Superficial characteristics of road pavement also play an important role on the occurrence of crashes. The knowledge of parameter values such as, for example, the sideways force coefficient, the international roughness index, the mean profile depth, the rutting, and the cracking due to fatigue, might help in understanding more about the occurrence of crashes associated with rain. Unfortunately in accident data-bases these parameters are not contained and only the condition “wet pavement” or “dry pavement” is reported. Some of the aforementioned pavement parameters are, instead, reported in other data-bases where, in contrast, accidents are not recorded. In general, it is to be said that these data-bases are used prevalently with the objective to assess the deterioration process of the pavement performance over time (e.g., in Caliendo [9]) or in the design process of pavements (e.g., in MEPDG [10], and in Caliendo [11, 12, 13]).

However, some papers showing that precipitation in the form of rain causes more accidents compared to dry conditions can be found more especially in Fridstrøm et al.[14], and in Caliendo et al. [3, 4].

Hazardous points such as intersections, junctions, access points, or the presence of tunnels along a road can also contribute to cause accidents. Although several studies have established
relationships between accidents, traffic and geometry of intersections, these papers focus on accidents occurred at intersections rather than on roadway segments or on the functionality of intersections (e.g., Vogt e Bared [15], Greibe [16], Oh et al. [17], Washington et al.[18], Wang [19], Caliendo [20, 21, 22]). From the perspective of statistical modeling, in order to take into account hazardous points, the analyst has to assume a structure model that is composed by two parts: multiplicative and additive. The multiplicative part of the statistical model represents the influence of variables that can be considered to be continuous along a specific road segment (e.g., longitudinal slope); while the additive part represents the effect of the aforementioned hazardous points. In this respect, let \( \mathbf{x} \) be a vector of \( k \) covariates and \( \mathbf{\beta} \) a vector of \( k \) (unknown) coefficients. A regression model of the expected number of accidents is defined by

\[
\lambda_i = g(\mathbf{x}_i ; \mathbf{\beta}) = \exp(\beta_0 x_{i0}) \left[ \exp\left( \sum_{j=1}^{k-1} \beta_j x_{ij} \right) + \sum_{j=k}^{n} \beta_j x_{ij} \right]
\]

in which as to the multiplicative component, the exponential choice appears to be a natural one, since it ensures that the expected number of accidents is always a positive number.

As regards crashes within tunnels, the statistical models applied are those used for open roads. However if the collisions among vehicles and/or against the tunnel wall are followed by fire, it is to be said that a quantitative risk analysis is needed (e.g., see Caliendo [23, 24, 25]).

In the light of the above considerations, the objective of this paper is to show the effects of rainfall and hazardous points on crash frequency. In this respect, the results of the predictive models, which had been developed in previous studies, are here disseminated and commented. Then reviews and discussion, with implications on the use of these models for safety interventions in road engineering, are also made.

2. DATA DESCRIPTION

The analysis was based on the data-base of an 8-year monitoring period on a four-lane median-divided motorway. Each carriageway was divided into segments with constant horizontal curvature and longitudinal slope. The number of crashes occurring on these homogeneous segments during the period monitored was assumed as dependent variable. The analysis was conducted separately for total accidents (accidents with damages, injuries and/or fatalities) and severe accidents (accidents with injuries and/or fatalities) which occurred on curves or tangents. The traffic flow expressed in terms of annual average daily traffic (AADT) was about 20,000 vehicles per day. Some 1,505 total accidents considered in this study, 551 of which were severe accidents. A number of 588 accidents occurred on curves (215 of which were severe accidents) and 917 accidents occurred on tangents (336 of which were severe accidents). Of all the 1,505 accidents observed in the monitoring period, 471 are reported as having occurred on "wet pavement".

3. METHODOLOGY

The Negative Binomial distribution, in its classical form, was used in order to account for “overdispersion” observed in accident counts. The likelihood function was maximized to obtain the estimations of model parameters. The Generalized Likelihood Ratio Test (GLRT) was used to decide which subset of the full set of potentially explanatory variables should be included in the model.
Junctions (J), and the presence of tunnels (T) were considered in the statistical analysis as dummy variables (J or T) assuming value equal to 1 if they were present in the examined trait, otherwise 0. An addition point hazard variable (SF), assuming value 1 or 0, was also considered for taking into account a major concentration of accidents that occurred in the entrance zone of the subsequent motorway.

In the following sections, the author first reported the final model in order to highlight the regression coefficient signs and their magnitude; recognizing that is difficulty a comparison with other studies of the literature for the multitude of variables, different data sources, and modeling approaches. This is followed by an application of models and discussion on the results obtained. Then possible links with safety interventions in road design are also made.

4. RESULTS

4.1. Model for Curves

Table 1 shows the parameter estimates for the Negative Binomial regression model, applied separately to total an severe accidents, with reference to curves. All the regression coefficients have the signs expected.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Total accidents</th>
<th>Severe accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.9941</td>
<td>1.1610</td>
</tr>
<tr>
<td>Log of the section length (km)</td>
<td>1.0345</td>
<td>1.4867</td>
</tr>
<tr>
<td>Curvature (km⁻¹)</td>
<td>0.7551</td>
<td>1.1131</td>
</tr>
<tr>
<td>Junctions (1 if present, 0 otherwise)</td>
<td>0.9254</td>
<td>N.S.</td>
</tr>
<tr>
<td>SF (1 or 0)</td>
<td>13.1005</td>
<td>11.3444</td>
</tr>
<tr>
<td>Overdispersion parameter (φ)</td>
<td>1.937</td>
<td>2.481</td>
</tr>
<tr>
<td>logLKH</td>
<td>1017.30</td>
<td>157.95</td>
</tr>
</tbody>
</table>

4.1.1. Total Accidents

The number of total accidents increases with the length $L$ (in a non-linear way), the curvature $1/R$, and the presence of joints $J$. Moreover also the $SF$ variable has a positive sign. All these variables were found to be statistically significant.

It is to be said, that the fraction of total variation explained by the regression model was found to be 27.4%. This might be due to the fact that other variables influencing accidents were not recorded in data-base, and/or information for quantifying the effects of driving behavior on accidents was not known.

As an example, with $L = 0.275$ km, $1/R = 2$ km⁻¹ ($R = 500$ m), one estimates 1.09 total crashes/year in absence of junctions ($J = 0$) and far to the entrance zone of the subsequent motorway (SF= 0). At junctions one estimates 1.94 total crashes/ year, and finally 13.09 total crashes/year in correspondence with the other motorway.

4.1.2. Severe Accidents

The model developed for severe accidents also shows quite similar results, with the variables statistically significant identified in $L$, $1/R$, and $SF$. However, the magnitude of the estimated coefficients are higher 1.4867 (against 1.0345) with reference to $L$, and 1.1131 (against 0.7551) for $1/R$; while the magnitude is lower 11.3444 (against 13.1005) for the $SF$ variable. In other words, the number of severe accidents increases more rapidly with an increase in $L$
and $I/R$ (the severity of accidents is higher on curves having smaller radii). The fraction of total variation explained by the model is higher (36.9%).

With $L = 0.275$ km, $1/R = 2$ km$^{-1}$ ($R = 500$ m), one estimates 0.54 severe crashes/year, and 5.07 severe crashes/year at the entrance zone of the subsequent motorway.

In reviewing some papers that provide estimates of the regression coefficients as a function of the horizontal curvature, one can find that they confirm the results obtained in the present study even if with a different magnitude. This might be due to the number and type of variables introduced in models developed, different characteristics of traffic and road geometry, different data-sources, etc. Abdel-Aty and Radwan [26] found, for example, that the coefficient estimate is positive and significant with a value of 0.124. Caliendo in a previous study [3] found a positive coefficient of 0.27 for total crashes and 0.34 for severe crashes. Labi [27] showed, with reference to fatal plus injury crashes, that the coefficient of regression is positive and that this varied from 0.0262 to 0.058. Finally Bauer and Harwood [28] also estimated a positive coefficient of 0.19 for curves.

### 4.2. Model for Tangents

Also the models developed for tangents have coefficients with the expected signs as reported in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Total accidents</th>
<th>Severe accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.142</td>
<td>2.0150</td>
</tr>
<tr>
<td>Log of the section length (km)</td>
<td>0.9401</td>
<td>0.8254</td>
</tr>
<tr>
<td>Tunnels (1 if present, 0 otherwise)</td>
<td>3.0824</td>
<td>3.9763</td>
</tr>
<tr>
<td>Overdispersion parameter ($\phi$)</td>
<td>1.407</td>
<td>1.933</td>
</tr>
<tr>
<td>logLKH</td>
<td>1675.84</td>
<td>240.75</td>
</tr>
</tbody>
</table>

#### 4.2.1. Total Accidents

The number of total accidents increases with the length ($L$) and the presence of tunnels ($T$). The fraction of total variation explained by the regression model is 20.7%.

As an example, with $L = 1.2$ km, one estimates 3.43 total crashes/year in absence of tunnels ($T = 0$) and 12.35 total crashes/year when a tunnel is present.

#### 4.2.2. Severe Accidents

The model developed for severe accidents also shows that the number of these accidents increases with $L$ and $T$. However, the magnitude of the estimated coefficients is lower 0.8254 (against 0.9401) with reference to $L$ and higher 3.9763 (against 3.0824) for the $T$ variable. In other words the number of severe accidents increases less rapidly with an increase in $L$ compared to total accidents; while it increases more with the presence of tunnels. The fraction of total variation explained by the model is higher 24.9%

With $L = 1.2$ km, one estimates 1.1 severe crashes/year in absence of tunnels ($T = 0$) and 4.82 severe crashes/year for the presence of tunnels.

Also for tangents one can find in the literature a different magnitude of the regression coefficients. In [3], for example, the regression coefficient associated with the length of tangent ($L$), was found to be of 0.86 for total crashes and 0.76 for severe crashes.
4.3. Models for Rain Effects
Wet pavement was also found to be a significant factor in increasing crashes. This was captured by means a dummy variable (0 if the status of pavement was "dry", otherwise 1).

4.3.1. Total Accidents
When road surface was wet, the expected number of crashes increased by a factor 3.72 for tangents and by a factor 6.11 for curves compared to dry surface conditions.

4.3.2. Severe Accidents
The expected number of severe accidents on wet pavement increased by a factor 4.16 for tangents and by a factor 5.16 for curves compared dry conditions.

The results obtained show that rain increases the number of total and severe accidents more on curves than on tangents. This is probably due to the combined effect of rain and centrifugal force in curve.

In [3] the magnitude of coefficients estimated was found to be, with reference to total crashes, of 2.32 and 2.70 for tangents and curves, respectively; and 2.81 and 3.26 for severe accidents. This means that the positive signs of coefficients associated with rain are here confirmed and that the magnitude can vary due to different factors.

5. DISCUSSION
The results show that engineers should pay more attention in road design when hazardous points such as junctions or tunnels are present. Junctions should be located on tangents rather than on curves, and appropriate acceleration and deceleration lanes should be designed. The length of these exclusive lanes should be sufficiently long enough to accommodate speed changes and to limit interactions among vehicles.

With reference to tunnels, these should not be preceded by curves with a small radius, and their cross section should be as that of open roads (i.e., including the presence of emergency lane or shoulder). The entrance zone of tunnels should be designed for having appropriate lighting and road signs. Moreover, tunnels should also be sufficiently distance from junctions or intersections.

There also evidence of the negative effect of rain. The risk of an accident on a wet pavement is higher, and the severity is also greater compared to dry pavement. In this respect, porous-asphalt pavements might be used. This solution might prevent many wet-skidding or hydroplaning accidents. Other measures might consist in lower speed limits in adverse weather and above all when the pavement is wet by the first rain (a low friction coefficient between the pavement and tyre is expected at the first rains).

6. SUMMARY AND CONCLUSIONS
The present paper was motivated first of all by the need to quantify the expected number of total and severe crashes as a function of road geometry, hazardous points and rain. In this respect, on the basis of a monitoring period, predictive models were developed.

Negative Binomial Distribution, was applied for modelling the random variation of the crash data. Model parameters were estimated by Maximum Likelihood Method, and the Generalised Likelihood Ratio Test was used to detect the significant variables to be included in the final model. Separate prediction models for curves and tangents were developed.
With reference to curves, the results show that crash frequency is positively associated with the length $L$ (in a non-linear way), the curvature $1/R$, and the presence of hazardous points such as junctions $J$ or tunnels $T$.

Also for tangents, crash frequency was found to be positively associated with $L$ and $T$ variables.

The influence of rain shows that with a wet pavement significant increases in the number of crashes are expected, more especially on curves compared to tangents. This might be due to the combined effect of rain and centrifugal force in curve.

All the models developed have the signs expected and in keeping with the most part of the international literature. However, the magnitude of the regression coefficients was found to be different compared to that of other models. This might be due to omitted variables and/or different modeling approaches. However, since the omitted variables may be considered as a part of the unobserved heterogeneity, the use of random-parameters models can mitigate the adverse impact of omitting significant explanatory variables, as a result a reduction in the difference of magnitude of estimates might be expected. In addition, it is to be stressed that a special case in the class of random-parameters models is the so called random-intercept negative binomial model. In this model only the regression intercept is random, and this model is known to be equivalent to random-effects models (e.g., see Greene [29], and Hilbe [30]). In other words the intercept-random model can be used to capture, for example, temporal correlations among crash data.

In the light of the aforementioned considerations, research needs to be addressed towards studies based on the random-parameters models in order to make additional developments possible in the field of unobserved heterogeneity. This appears to be a promising methodological direction in accident analysis (e.g., see Mannering and Bhat [31], Mannering et al.[32], for greater in depth knowledge).

REFERENCES


Ciro Caliendo


loads/MEPDG-1Frontmatter.pdf


Crash Prediction Models for Roads Including Rainfall and Hazardous Points


