Development of safety performance functions for Spanish two-lane rural highways on flat terrain

Garach Morcillo, Laura
TRYSE Research Group,
Department of Civil Engineering, University of Granada,
ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel.: +34 958 249455; e-mail: lgarach@ugr.es

De Oña López, Juan (corresponding author)
TRYSE Research Group,
Department of Civil Engineering, University of Granada,
ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel.: +34 958 249979; e-mail: jdona@ugr.es

López Maldonado, Griselda
TRYSE Research Group,
Department of Civil Engineering, University of Granada,
ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel.: +34 958 249450; e-mail: griselda@ugr.es

Baena Ruiz Leticia
TRYSE Research Group,
Department of Civil Engineering, University of Granada,
ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada (Spain)
Tel.: +34 958 249450; e-mail: lbaenar@ugr.es

ABSTRACT

Over decades safety performance functions (SPF) have been developed as a tool for traffic safety in order to estimate the number of crashes in a specific road section. Despite the steady progression of methodological innovations in the crash analysis field, many fundamental issues have not been completely addressed. For instance: Is it better to use parsimonious or fully specified models? How should the goodness-of-fit of the models be assessed? Is it better to use a general model for the entire sample or specific models based on sample stratifications? This paper investigates the above issues by means of several SPFs developed using negative binomial regression models for two-lane rural highways in Spain. The models were based on crash data gathered over a 5-year period, using a broad number of explanatory variables related to exposure, geometry, design consistency and roadside features. Results show that the principle of parsimony could be too restrictive and that it provided simplistic models. Most previous studies apply conventional measurements (i.e., R^2, BIC, AIC, etc.) to assess the goodness-of-fit of models. Seldom do studies apply cumulative residual (CURE) analysis as a tool for model evaluation. This paper shows that CURE plots are essential tools for calibrating SPF, while also providing information for possible sample stratification. Previous authors suggest that sample segmentation increases the model accuracy. The results presented here confirm that finding, and show that the number of significant variables in the final models increases with sample stratification. This paper point out that fully models based on sample segmentation and on CURE may provide more useful insights about traffic crashes than general parsimonious models when developing SPF.

Keywords: Cumulative residuals; Safety Performance Functions; two-lane rural highways; flat terrain; parsimonious models; fully models

1. INTRODUCTION

According to the World Health Organization, approximately 1.24 million people are killed every year on the world's roads, and another 20 to 50 million sustain nonfatal injuries as a result of road crashes (WHO, 2013). All efforts to reduce traffic crashes are therefore well justified. In Europe, approximately 60% of road accident fatalities occur on two-lane rural roads (Cafiso et al., 2010). Two major factors usually play an important role in traffic accident occurrence: the first is related to the driver; and the second is related to the roadway design (Abdel-Aty and Radwan, 2000).

Safety Performance Functions (SPF) make it possible to predict the number of crashes that may take place on a given stretch of roadway with certain characteristics. For many years this type of model was developed using simple or multiple linear regression techniques. However, Miaou (1994) showed that Poisson regression models —or, in the case of overly dispersed data, Negative Binomial (NB) regression models— are more appropriate. Later research showed that, in general, the number of crashes used when calibrating the prediction models presents over-dispersion, with a greater dispersion than would be consistent under a Poisson model (Hauer et al., 2002). Most studies nowadays therefore assume that the number of crashes follows an NB distribution (Persaud et al., 1999; Cheng and Washington, 2008; Montella et al., 2008; Cafiso et al., 2010; FHWA, 2010; Montella, 2010; Camacho-Torregrosa et al., 2013).

Although substantial research has been conducted on the development of crash models, there are issues still on the forefront regarding: generalized models; unobserved heterogeneity; confounding variables; variables to be considered in models and how to add them (parsimonious vs. fully specified models); overfitting of models; measures used in assessing the goodness of fit; and the appropriateness of stratifying a sample to get better models.

The generalized models are used by authorities to study the safety of other locations in a given region that have characteristics similar to those of the location used to build the model. Thus, models containing variables with highly significant parameters can predict accident frequencies at new locations not used in the model development. In addition, because explanatory variables that have statistically significant model parameters help explain the variability of the accident data, their inclusion in the model improves its fit with the data (Sawalha and Sayed, 2006).

As for the unobserved heterogeneity, the fact that crashes involve complex interactions among human, vehicle, roadway, traffic and environmental elements makes it impossible to take into account all factors influencing the likelihood of highway crashes. Crash databases contain a lot of information about road, vehicle and environment characteristics, yet many other elements remain unobserved, such as human behaviour, friction measurements, etc. These factors constitute unobserved heterogeneity and can introduce variation in the impact of the effect of observed variables on accident likelihood (Mannering et al., 2016).

Unobserved heterogeneity can be defined as variations in the effect of variables across the sample population that are unknown to the researcher. If this issue is ignored and the effects of observable variables is held to be the same across all observations, the model may be misspecified and the estimated parameters might be biased, leading to erroneous predictions (Mannering et al., 2016). Although relatively recent research has explored unobserved heterogeneity, allowing new insights to be extracted from crash databases, the model-estimation process involved becomes considerably more complex; the result obtained from methods such as random parameter models may not be easily transferable to other datasets or different locations since the individual parameter vector associated with each observation is unique to that observation (Lord and Mannering, 2010; Mannering et al., 2016).

A further issue that concerns researchers is that of confounding variables. In general, confounding variables are those that are not controlled in the model but may have a latent effect. A confounding factor can be defined as any variable —other than the cause of principal interest in a study— that can either (a) generate effects that may be mixed up with the effects of the causal variable, (b) distort the effects attributed to the causal variable, e.g. modifying their direction or strength, or (c) hide the effects of the causal variable (Elvik, 2011). Controlling for confounding factors is important in establishing causality, and poor control of confounding factors can seriously distort the findings of road safety studies and make them completely worthless (Elvik, 2008). However, the number of potentially confounding factors that are successfully controlled for is always limited due to the fact that most are unknown. Moreover, it is a fallacy to believe that if a model fits the data very neatly, this demonstrates that it includes all important factors and that factors not included in the model cannot have major effects (Elvik, 2011). Hence, this matter may be a limitation in most crash-frequency studies. In the models applied to all accidents, there is slight confounding owing to the mixture of different levels of accident severity (Elvik, 2011).

SPF are used for a variety of purposes. Most frequently they serve to estimate the expected crash frequencies from various roadway entities (highways, intersections, interstates, etc.) and to identify geometric, environmental, and operational factors that are associated with crashes. With respect to the selection of variables, the explanatory variables that are potentially relevant in SPF can be grouped in two main categories: (a) Variables describing exposure to crash risk; (b) Risk factors that influence the number of crashes expected to occur in a road.

In the first category, most studies include Annual Average Daily Traffic (AADT) and section length as exposure variables (Hadi et al., 1995; Anderson et al., 1999; Persaud et al., 1999; Pardillo and Llamas, 2003; Ng and Sayed, 2004; Pardillo et al., 2006; Dell'Acqua and Russo, 2008; Cafiso et al., 2010; Park and Abdel-Aty, 2015). Among the exposure variables, some authors moreover take into account the percentage of heavy vehicles (Fitzpatrick et al., 2000; Elvik, 2007; Ramírez et al., 2009; Montella, 2010; Hosseinpour et al., 2014).

In the second category, among risk factors that influence the number of crashes expected to occur on a highway, most authors consider explanatory variables included in one of the three following groups: geometric variables, consistency variables or context variables. A number of studies have attempted to quantify the effects of road geometric design variables and exposure variables on accident frequencies (Hadi et al., 1995; Persaud et al., 1999; Fitzpatrick et al., 2000; Anastasopoulos et al., 2008; Dell'Acqua and Russo, 2008; Cafiso et al., 2013; Park and Abdel-Aty, 2015). Some authors have looked into the influence of consistency variables —or a combination of geometric, environment and consistency variables— in SPF development for two-lane rural highways (Anderson et al., 1999; Ng and Sayed, 2004; Cafiso et al., 2010; De Oña et al., 2014). Others have developed consistency indexes that may be used as independent variables in SPF (Polus and Mattar-Habib, 2004; Camacho-Torregrosa, 2014; Garach et al., 2014). Some studies have attempted to relate crash frequency with environmental variables such as driveway density (Pardillo and Llamas, 2003; Pardillo et al., 2006; Cafiso et al., 2010).

Within this substantial body of research on SPF development, the vast majority of SPF studies include some kind of measure of exposure, such as AADT or segment length. Still, there is a lack of consensus regarding the number of variables that should be added in the model, and questions relating to parsimonious vs. fully specified models.

According to Sawalha and Sayed (2006), model generality requires that a model be developed in accordance with the principle of parsimony, which calls for explaining as much variability of the data as possible using the least number of explanatory variables. The notion behind the principle of parsimony is to avoid overfitting. If many variables are included in a model, a perfect fit could be obtained; but the developed model would not produce reliable predictions when applied to a different set of locations. In addition, as the data available to researchers is often limited, and many variables known to significantly affect the frequency of crashes may not be available, there is also a need to develop relatively simplistic models using only explanatory variables than can, in practice, be gathered and projected for use. Given these data limitations and the need to specify models with a few simplistic explanatory variables, parsimonious models are often estimated.

However, other authors disagree with the concept of parsimonious models. According to Mannering and Bhat (2014), the real problem with them is that models having a few simplistic explanatory variables might exclude significant explanatory variables; and the model-estimated parameter for the basic variables (like traffic volume) might be estimated with bias (omitted variables bias). The application of the model would be fundamentally flawed, because changes in the omitted variables cannot be captured and predicted crash frequencies will be incorrect. Mannering et al. (2016) indicated that if factors affecting the likelihood of an accident are not included (unobserved heterogeneity), these factors could introduce variation in the impact of the effect of observed variables on accident likelihood. Omission of important variables introduces bias in model parameters, and will lead to incorrect inference (Washington et al., 2010; Mitra and Washington, 2012).

Regarding model evaluation, many studies use statistical measures such as Akaike Information Criterion (AIC) or Pearson Chi-square statistics, among others. Few use cumulative residual analysis as a method to evaluate the calibrated prediction models. Hauer (2015) recommends analysing residual plots as an essential tool to calibrate crash models. Lord and Persaud (2000) applied cumulative residual analysis to evaluate prediction models showing the variation in the accident rate in consecutive years; and they ruled out the use of the conventional R^2.

Another issue to consider is that when a study uses data from highways covering a broad region, there may be very different characteristics in roadway sections of the same overall type. For instance, in the studies cited above, the models calibrated included a wide range of AADTs, from as little as 166 veh/day (Anderson et al., 1999) up to 25,000 veh/day (Park and Abdel-Aty, 2015). In such a situation, even if the models obtained are valid, they may leave room for improvement. This point was brought out by Vogt and Bared (1998): they concluded that their model could be improved if the sample had been divided on the basis of ranges of some of the explanatory variables. Vogt and Bared (1998) came to this conclusion after analysing models through a comparison of cumulative residuals plotted against leading variables, so as to check for systematic trends that might contradict the assumed model form or suggest model refinements. Pardillo et al. (2006) showed that the stratification of the model oriented by the results of the cumulative residuals analysis is a valid method to refine crash prediction models. According to Hauer (2004), when a model is used for prediction, it is important that it fit well throughout the range of each variable. He suggested the possibility of stratifying the models to overcome the lack of flexibility of the most common exponential functional forms.

The aim of this paper is to develop SPF analysing cumulative residuals for two-lane rural highways, using a high number of explanatory variables related to exposure, geometric design, design consistency and roadside features. In the process of adding variables to the model, two types of models are compared: parsimonious models vs. fully specified models. The paper is organized in four main sections. The first section has presented an introduction to the main concepts and previous crash models. In the second section, we describe the database and the methodology. The third section presents the results and discussion. Finally, in the last section the main conclusions of this study are given.

2. DATA AND METHODOLOGY

2.1. Road and accident data

This study was conducted on 972 km of two-lane rural highways over flat terrain in Andalusia (Spain). The roadway data were obtained from the General Direction of Roads under the Andalusian Regional Government and included roadway inventories with characteristics of the road and traffic volume. Urban segments, intersections[1] and passing or climbing lanes were removed because of their characteristics, as SPF used to predict crashes in these cases are very different from the SPF that would be obtained on conventional two-lane rural highways. Moreover, only those sections in which AADT was higher than 500 veh/day were included in the study, as it was assumed that when traffic volumes are lower, traffic conditions and safety problems are not representative of regular two-lane rural roads. Segments undergoing significant changes during the study period were excluded from the sample. As a result, 606 km of two-lane rural highways were involved in the analysis.

[1] A portion of the road was considered intersection if it had a stop and left turn lane on the main road.

Accident data were obtained from Spain´s "General Traffic-accident Directorate" (DGT) for a five-year period (2006-2010). The total number of crashes on the studied roads was 1,443.

2.2. Methodology

Initially, each road was divided into horizontal curves and tangents. The next step was to subdivide the sample into homogeneous road segments. The explanatory variables were then selected. Some of these could be obtained directly from the database, while others, related with the design consistency, were obtained from operating speed profiles in each homogeneous road segment. Once the variables had been selected, the prediction models were calibrated and evaluated by means of several statistical measures.

2.2.1. Homogeneous road sections

Several authors have pointed out the need to study segments with homogeneous characteristics to ensure coherent road safety studies (Resende and Benekohal, 1997; Fitzpatrick et al., 2000; Pardillo and Llamas, 2003; Cafiso et al., 2010, Garach et al., 2014). Following previous studies (Cafiso et al., 2010, Garach et al., 2014), in order to work with such homogeneous road segments, the following parameters were used: AADT, average paved width (P_w) and curvature change rate (CCR).

[Insert Table 1 here]

For AADT, a new segment was identified when there was a change of the intervals specified in Table 1 (AASHTO, 2010; Garach et al., 2014). For roadway width, the distribution of road widths was analysed and the ranges defined in Table 1 were used. The sections with constant CCR were identified on the basis of the section curvature change rate (CCRsect), defined as follows:

$$\mathrm{CCR_{sect}} = \frac{\Sigma_i|\gamma_i|}{L_{HS}} \qquad (1)$$

where CCRsect = section curvature change rate (gon=km); γ_i = deflection angle for a continuous element i (curve or tangent) (gon: centesimal degree); LHS = road segment length (km).

For each road segment, a diagram was drawn. The sum of the γ_i was represented in the y-axis and the distance in the x-axis. Road sections with homogeneous horizontal alignment were identified in this diagram by sections where the slope of the cumulative angle deviation curve (CCR_{sect}) was relatively constant. Based on the German procedure (RAS-L, 1995), a minimum section length of 2 km was adopted. A section was considered homogeneous when the three parameters discussed (AADT, road width, and CCR) were constant. Applying these criteria to all roads under study, 456 sections with homogeneous characteristics were identified.

2.2.2. Explanatory variables

Once the homogeneous sections had been defined, the variables considered for the model development were selected. Explanatory variables related to traffic volume, geometric characteristics, design consistency and roadside context were considered. A single value for each variable was assigned to every homogeneous road section.

Table 2 shows the variables initially considered, grouped by categories (exposure, geometry, consistency and context), along with the main statistics regarding the variables (mean, minimum, maximum and standard deviation).

Exposure and geometric variables

AADT and percentage of heavy vehicles were obtained directly from the road database. The length of the section is equivalent to the length of the homogeneous road segment as established above.

The variables lane width, shoulder width, platform width, longitudinal grade and radius were also taken directly from the road database. Thus, a value for each one of these variables was obtained for each homogeneous road section.

Other geometric and operational variables, such as the Curvature Ratio (CR) and Tangent Ratio (TR), were computed using the following equations:

$$\mathrm{CR} = \frac{\sum_{j=1}^{k} \mathrm{L}_{\mathrm{C}j}}{\mathrm{L}_{\mathrm{HS}}} \tag{2}$$

$$\mathrm{TR} = \frac{\sum_{j=1}^{k} \mathrm{L}_{\mathrm{T}j}}{\mathrm{L}_{\mathrm{HS}}} \tag{3}$$

where L_{HS} is the total length of the homogeneous section (km); L_{Cj} is the length of jth curve in the homogeneous section composed by k curves (km); and L_{Tj} is the length of jth tangent in the homogeneous section composed by k tangents (km).

[Insert Table 2 here]

Consistency variables

To obtain the consistency variables, it is necessary to know the operating speed (V_{85}) for each road element. To this end, the respective operating speed profiles were built using the criteria established by De Oña et al. (2014). To construct the speed profile, one must first define an operating speed on curves, an operating speed on tangents and an acceleration or deceleration between the two elements. Given the importance of using speed prediction models calibrated according to local conditions (Misaghi and Hassan 2005), the model of Camacho-Torregrosa et al. (2013) was applied in this study, adjusted for horizontal curves on two-way rural highways in Spain (Eq. 4-5).

$$V_{85} = 97.4254 - 3{,}310.94/\mathrm{R} \text{ for } 400 \text{ m} < R \leq 950\, m \quad (4)$$

$$V_{85} = 102.048 - 3{,}990.26/\mathrm{R} \text{ for } 70 \text{ m} < R \leq 400\, m \quad (5)$$

where R = radius of curvature (m).

To build the speed profile, a constant curve speed was considered. The tangent speed value considered was 110 km/h (desired speed according to Camacho-Torregrosa et al., 2013). Otherwise, the acceleration and deceleration rates proposed by Fitzpatrick and Collins (2000) for horizontal curves were taken into account.

The average operating speed from the speed profile (V_{85avg}) was computed on the basis of the operating speed profile as follows:

$$\mathrm{V_{85avg}} = \frac{\sum_{i=1}^{n} \mathrm{V_{85i} L_i}}{\mathrm{L_{HS}}} \text{ (km/h)} \quad (6)$$

where V_{85i} is the operating speed of the i_{th} geometric element (km/h) computed using the operating speed profile; L_i the i_{th} element length of the homogeneous section (km); and n is the number of geometric elements along a section.

The relative area bounded by the speed profile (R_a) and the average operating speed from speed profile (V_{85avg}) were also considered as design consistency variables, as well as the standard deviation (σ) of the operating speed profile. They were calculated by means of the following equations (Polus and Mattar-Habib, 2004):

$$R_a = \frac{\sum_{i=1}^{n} a_i}{L_{HS}} \text{ (m/s)} \tag{7}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(V_{85i} - V_{85avg})^2}{n}} \text{ (km/h)} \tag{8}$$

where a_i is the area bounded by the operating speed profile and the average operating speed line (m^2/s); V_{85i} is the operating speed of the i_{th} geometric element (km/h); V_{85avg} is the average operating speed along the entire homogeneous section of length L_{HS} (km/h); and n is the number of geometric elements in the homogeneous section.

Considering the operating speed profiles, two more indicators were derived:

- Ea10 (m/s) is a measurement of speed dispersion. Similar to R_a, it is the area bounded by the operating speed profile and the average operating speed profile plus and minus 10 km/h. The length of the road segment finally divides that area.
- Ea20 (m/s) is similar to the previous indicator, but considering 20 km/h.

Two other consistency indicators were also selected in light of the speed differentials between contiguous elements in the homogeneous section, using the following equations:

$$\Delta V_{10} = \frac{N(\Delta V > 10)}{L_{HS}} \text{ (km/h)} \tag{9}$$

$$\Delta V_{20} = \frac{N(\Delta V > 20)}{L_{HS}} \text{ (km/h)} \tag{10}$$

where $N(\Delta V>10)$ is the number of speed differentials (ΔV_s) higher than 10 km/h in the homogeneous section; and $N(\Delta V>20)$ is the number of speed differentials (ΔV_s) higher than 20 km/h in the homogeneous section.

One more consistency indicator was obtained with regard to speed differentials between contiguous elements in a homogeneous segment. The variable average speed reduction $\Delta(V_{85i} - V_{851+1})_{avg}$ was calculated as follows:

$$\Delta(V_{85i} - V_{851+1})_{avg} = \frac{\sum_{s=1}^{n} |V_{85i} - V_{85i+1}|}{n_{\Delta V}} \text{ (km/h)} \tag{11}$$

where $n_{\Delta V}$ is the number of speed differentials in the homogeneous section and V_{85i} is the operating speed of the i_{th} geometric element (km/h). Road segments are expected to be more inconsistent as this variable increases, because of the higher speed reductions.

The consistency variable $\Delta(V_{85i} - V_d)_{avg}$ was calculated as the difference between the operating speed from the speed profile and the design speed of the road.

$$\Delta(V_{85i} - V_d)_{avg} = \frac{\sum_{s=1}^{n} |V_{85i} - V_d|}{n_{\Delta V}} \text{ (km/h)} \quad (12)$$

Using R_a and σ, Polus and Mattar-Habib (2004) developed a consistency index (C_p) based on a negative exponential function.

$$C_p = 2{,}808 * e^{-0{,}278[R_a * (\frac{\sigma}{3{,}6})]} \text{ (m/s)} \quad (13)$$

Garach et al. (2014) developed an enhanced version of the Polus consistency model, indicating that the original consistency model equation was not ideal for consistency analysis. Thus, they developed the consistency index C_g, likewise dependent on Ra and σ:

$$C_g = \frac{195.073}{\left(\frac{\sigma}{3{,}6} - 5.7933\right)(4.1712 - R_a) - 26.6047} + 6.7823 \text{ (m/s)} \quad (14)$$

Polus and Mattar-Habib (2004) established some thresholds for C_p, Ra and σ. Accordingly, consistency could be considered as good, acceptable or poor (Table 3). The same limits as for the model of Polus and Mattar-Habib (2004) were proposed for the C_g index (Garach et al. 2014).

[Insert Table 3 here]

Camacho-Torregrosa (2014) developed another consistency index (C_c) that was defined as follows:

$$C_c = \sqrt[3]{\frac{V_{85avg}}{d_{85avg}}} \text{ } (s^{1/3}) \quad (15)$$

where V_{85avg} is the average operating speed from speed profile (m/s), and d_{85avg} is the average deceleration rate (m/s^2) defined as:

$$d_{85} = \frac{(v_{max}^2 - v_{min}^2)}{2 \times l} x \frac{1}{3.6^2} \text{ (m/s}^2\text{)} \quad (16)$$

where, in turn, v_{max} is the operating speed before the deceleration (km/h), v_{min} is the operating speed after the deceleration (km/h) and l is the length of the speed transition (m).

Context variables

As it has been demonstrated that direct accesses to roads can significantly increase crashes (Miaou et al., 1996), driveway density (DD) was considered relevant and gathered from the roadway database.

The percentages of existing shoulder (%SH) and of existing paved shoulder (%SH_p) in the homogeneous section were obtained in view of the shoulder width variable available in the roadway database. For each homogeneous segment the proportion of existence of shoulder was obtained. Speed limit (V_{limit}) was also taken from the roadway database.

2.2.3. Modelling traffic crashes

SPF are developed using the general linear regression (GLM) approach. The GLM approach has the advantage of overcoming the limitations associated with the use of conventional linear regression in modelling traffic collisions (Hauer and Lovell, 1988; Sawalha and Sayed, 2001, 2006). The model form used is shown below.

SPF form

The relationship between crash frequencies and selected variables related was modelled using loglinear regression models and Negative Binomial (NB) distribution. The NB and Poisson distributions are an appropriate choice since accident frequencies are integers, relatively small numbers, and necessarily non-negative. The Poisson distribution was not used because it is appropriate in those cases where mean and variance are equal. When this basic assumption is substantially violated, the NB distribution may stand to be an improvement over the Poisson distribution (Lord and Mannering, 2010).

According to Sawalha and Sayed (2006), the mathematical form used for any SPF should satisfy the following conditions: yield logical results (it must not lead to the prediction of a negative number of accidents and it must ensure a prediction of zero accident frequency for zero values of the exposure variables) and there must exist a known link function than can linearize the model for the purpose of coefficient estimation. The mathematical form generally accepted in the literature (Pardillo and Llamas, 2003; Sawalha and Sayed, 2006; Cafiso et al., 2010; Montella, 2010; De Oña et al., 2014) is:

$$\hat{E}(Y) = e^{\beta_0} * AADT^{\beta_1} * L^{\beta_2} * e^{\Sigma(\beta_i * x_i)} \quad (17)$$

where $\hat{E}(Y)$ is the estimated number of crashes; L is the length of the segment (km); AADT is the Average Annual Daily Traffic (AADT) (veh/day); x_i are the explanatory variables; and β_i are the model parameters. Hauer (2015) holds that the number of crashes depends on the amount of traffic and the segment length, which he considers to be intuitively obvious and empirically substantiated. It is therefore clear that a traffic variable and a segment length variable should be in the model equation. Intuition is, however, insufficient regarding other variables. The research perspective offers no consensual statistical procedure for adding or deleting variables from a model equation; the question of which procedures to use obeys "a great deal of personal judgment" (Draper and Smith, 1981). In some cases the parameter that accompanies the variables in the models proves incorrect and it is therefore deleted from the model equation. According to Hauer (2015), the purpose of adding a variable to the model equation is to increase the accuracy with which the number of crashes is estimated while reducing the magnitude of the standard deviation. According to Sawalha and Sayed (2006), inclusion of a large number of explanatory variables may cause model overfitting.

Model Evaluation

Four measurements were used here to assess the goodness-of-fit of the model. They are: the ordinary multiple correlation coefficient (R^2), Akaike´s Information Criterion (AIC), the generalized Pearson χ^2 statistic and the Scaled Deviance (SD). The AIC compares different models based on the balance between the bias and variance explained by them. The Pearson χ^2 statistic can be used for null hypothesis significance testing regarding the equivalence of the variance assumed in the modelling effort and the sample variance. The SD is useful for comparing the proposed model and the saturated model. However, again according to Hauer (2015), the goodness-of-fit measures describe only how the model fits overall; hence a single number is insufficient. The model estimation must be nearly unbiased for all variable values. For this reason, it is commonly recommended to plot Cumulative Residuals (CURE) to examine model fit in detail (Hauer, 2015). The residuals are equal to the difference between the observed and estimated values of the dependent variable.

Each variable in the model will have its own CURE plot to be used in examining the goodness-of-fit for each variable and to examine ways in which the fit for that variable could be improved. These residuals, calculated based on each one of the variables, should be within certain limits for the model to be considered well adjusted. The upper and lower limits, accordingly, would be given by $2 * \hat{\sigma}'_s(i)$, where $\hat{\sigma}'_s(i)$ has the following expression:

$$\hat{\sigma}'_s(i) = \pm\hat{\sigma}_s(i) * \sqrt{1 - \frac{\hat{\sigma}^2{}_s(i)}{\hat{\sigma}^2{}_s(n)}} \tag{18}$$

where $\hat{\sigma}'_s(i)$ is the limit of the residuals accumulated for the variable of analysis; $\hat{\sigma}_s(i)$ is the square root of the variance $\hat{\sigma}^2{}_s(i)$; $\hat{\sigma}^2{}_s(i)$ is the variance of the accumulated residuals up to the homogeneous section (i); and $\hat{\sigma}^2{}_s(n)$ is the variance of the accumulated residuals in the total homogeneous sections (n).

Selection of model variables

As previously mentioned, the variable selection problem has attracted attention in previous traffic crash research. If many variables are included in a model, a perfect fit to the data can be achieved. Yet the same model could be over-fitted and perform poorly when applied to a new sample. Sawalha and Sayed (2006), applying the principle of parsimony, found that using less but statistically significant explanatory variables can avoid overfitting and improve the reliability of a model. Still, as noted by Mannering and Bhat (2014), parsimonious models are not only biased, but are fundamentally flawed, and offer little practical value. To control the overfitting when fully specified models are developed, Hauer (2015) found that models whose CURE plot does not go beyond the 0.5σ´ limits are close to being unbiased, and that attempts to further "improve" such models court the danger of overfitting. With this guideline one can decide whether a model requires improvement or is good enough to be left alone. In this paper, parsimony models and fully specified models are developed and compared. The latter are referred to here as best-fit accident prediction models.

The steps followed in the selection of model variables were as follows:

- Step 1: Building a model with the variables AADT and length. The goodness-of-fit criteria shown above as well as the cumulative residuals of the model are analyzed. This provides the Basic Model.
- Step 2: Developing best-fit accident prediction models. Other predictive variables are subsequently introduced to the basic model, until all variables (and their combinations) are tested. Models with all possible combinations of the available variables are developed and analyzed. The decision to keep a variable in the model is based on four criteria. First, the t-statistic for each parameter had to be significant at the 95% confidence level. Second, engineering judgment deemed the variables´ sign to be

logical. Third, the variable exhibited a low correlation (i.e. <0.7) with other independent variables already in the model (Wei and Lovegrove, 2013). Fourth, it was verified that the cumulative residuals were within the established limits. In addition, according to Hauer (2015), to avoid model overfitting, it was verified that the model´s CURE plot did not surpass the 0.5σ´ limits. The order in which variables are added was based on their t-stat, from highest to lowest.

- Step 3: Verifying which of the models developed in step 2 actually meet the parsimonious criterion. Thus, in this step parsimonious accident prediction models are developed. A new variable introduced in the model in step 2 is kept if the addition of this new variable generated a significant drop in the SD for a 95% level (>3.84). Otherwise, the parsimonious criterion dictates that the variable should not be considered (Sawalha and Sayed, 2006).

Based on Sawalha and Sayed (2006), an outlier analysis was performed for all the models. First, potential outliers are detected and they are removed one by one. The drop in SD is observed after the removal of each point. Then, points causing a significant drop in SD are considered influential outliers, and thus they are eliminated.

Regarding to the correlation between the variables indicated in steps 2 and 3, according to Turner et al. (2012), identification of variable correlations is required to avoid having two or more significantly correlated variables in the same prediction model. In such cases the variability within one variable does, to a certain extent, predict the variability in the correlated variable. The authors further indicate that adding a variable correlated to those already in an existing model does not improve the fit of the model compared with the addition of important non-correlated variables. In the case at hand, the correlation matrix was previously calculated. Some variables, such as paved width and shoulder width were highly correlated (coefficient over 0.70). However, it was decided to keep both variables in the analysis, but imposing that two correlated variables were never in the same model.

3. RESULTS

Having identified the 456 homogeneous sections by means of the variables AADT, paved width and CCR, the values of the variables in each one of these sections were calculated (see Table 2). Below the models are developed.

3.1. Step 1 Results: base model

Following the process described in the methodology, the base model considers only two variables: AADT and length (Eq. 19).

$$\hat{E}(Y) = e^{-12.3248} * AADT^{0.7512} * L^{1.0083} \quad (19)$$

Figure 1 shows the residual analysis for the variables AADT (Fig.1a), length (Fig.1b) and fitted crashes (Fig.1c). Fig.1b and Fig.1c show satisfactory results.

However, the AADT cumulative residuals plot showed that the fit was not good (Fig 1a). On the one hand, in a range of AADT between 9,200 and 19,000 veh/day the values of the residuals surpass the limits of ±2σ; and on the other hand, after an AADT of approximately 4,000 veh/day, the curve begins to rise considerably and continuously. From an AADT of 5,000 veh/day onward the number of crashes observed is greater than the crashes estimated with the model (the accumulated sum of the differences between the crashes that occurred and those expected is positive, and therefore the curve is above the x axis).

This shows, as highlighted Hauer (2004), that usually it is not easy to find a relatively simple function that suits the data along its entire domain. For this reason, and according to other authors (Vogt and Bared, 1998; Hauer, 2004; Pardillo et al., 2006), the sample was stratified. A stratification of the sample based on splitting the sample by AADT ranges was explored.

The "Observed/Fitted" ratio was chosen for examining if fitted values are into line with observed values (Table 4).

[Insert Table 4 here]

As Table 4 shows, the AADT ranges in which there are greater differences between fitted and observed values are the 4,000-5,000 range (ratio 1.20) and the 5,000-6,000 range (ratio 1.27). Different stratifications of the sample considering the different thresholds for each range were explored:

1. AADT ≤ 4,000 and AADT > 4,000
2. AADT ≤ 5,000 and AADT > 5,000
3. AADT ≤ 6,000 and AADT > 6,000

The first strata (AADT≤4,000 and AADT>4,000) was selected because the models provided better overall results than the ones developed in the other stratifications. Thus, the sample was divided in two sub-samples (one in which all the AADT values were less than 4,000 veh/day and another in which all the AADT values were greater than 4,000 veh/day), and different models could be derived according to these different ranges of AADT.

[Insert Figure 1 here]

Table 5 (model 1) and Table 6 (model 1) show the basic models obtained for the two different AADT values. In both models AADT and length are significant. Moreover, their coefficients present the expected signs (positive): greater volume of traffic and greater section length are associated with more crashes. As for the overall goodness of fit, the R^2 values obtained were similar to those reported by previous authors (Abdel-Aty and Radwan, 2000; Camacho-Torregrosa et al., 2013).

[Insert Figure 2 here]

Figure 2 shows the residual analysis for the models calibrated for AADT$\leq$4,000 veh/day and for AADT>4,000 veh/day with regard to the variables AADT, length and fitted crashes. As can be seen, the residuals are substantially improved. Hence models will be created for different AADT values, as they will significantly enhance the base model.

Regarding the outliers, the difference between adjusted and observed values was calculated in the entire database and the data that had a large difference between the two were considered as possible outliers. Seventeen points (3.74% of the sample) were detected as potential outliers. None of them caused a significant drop in scaled deviance and therefore they were kept in the analysis (Sawalha and Sayed, 2006). The same outlier process was carried out in each of the databases (AADT<4,000 and AADT>4,000) and the same results were obtained; so all the possible outliers were kept in the analysis.

In addition, according to the outlier ignoring approach (El-Basyouny and Sayed, 2010), if few outliers are identified, representing a small percentage of the sample size (e.g., less than 5%), it is still acceptable to include them —especially if the analysts are not certain about whether or not they are outliers.

3.2. Step 2: Results of best-fit models

At this point the variables of Table 2 are added to the exponent part of the model of Eq. 17. These models are developed with all possible combinations of the available variables complying with all the criteria listed in step 2, related to t-statistic, logical sign, no correlation and cumulative residuals. Models are calibrated considering, separately, the AADT≤4,000 veh/day database (Table 5) and the AADT>4,000 veh/day database (Table 6). Table 5 presents parameter estimates, p-values, and the goodness-of-fit measures for the models with AADT≤4,000 veh/day.

Table 5 only shows models with four variables. Models with more (five and six variables) are included in the Appendix to simplify reading. These models give increasingly complex models without providing significant improvements. No model with more than six variables meets the conditions of step 2.

[Insert Table 5 here]

Table 5 shows that the variables AADT and length are significant and present the expected signs. The variables participating in the models built with a single variable in the exponent part are:

- The consistency index C_c
- The driveway density (DD).

The variables that participate in the models with two variables in the exponent part are:

- The DD combined with variables: percentage of shoulder; percentage of paved shoulder; consistency index C_c
- The longitudinal grade (LGr) combined with variables: average operating speed and consistency index C_g.

All the variables in Table 5 are significant ($p<0.05$). The two exposure variables AADT and length have positive signs, indicating that traffic volume and length increase crash occurrence. In the next section the coefficients obtained for the rest of the variables will be interpreted.

Model 5 in Table 5 presents the best goodness-of-fit values according to three of the four measurements of fit calculated (R^2=0.571; AIC=797.446; χ^2=263.985) and it includes the variables: AADT, length, percentage of paved shoulder in the section and driveway density.

Table 6 presents the parameter estimates, p-value, and goodness-of-fit measures for the models with AADT > 4,000 veh/day.

Table 6 only shows models with four variables. (Models with five variables are shown in the Appendix.) No model with more than five variables meets the conditions of step 2.

[Insert here Table 6]

All the variables of Table 6 are significant at the 95% confidence level. The variables AADT and length have, as in Table 6, positive signs. The significance of the rest of the variables is explained below.

In the case of a single variable in the exponent part, the variables that intervene are:

- Percentage of heavy vehicles
- Average operating speed
- Consistency index C_p
- Driveway density.

In the case of two variables in the exponent part, the variables intervening are:

- Percentage of heavy vehicles combined with the variables: CCR; average operating speed; consistency index C_p; consistency index ΔV_{10}
- The mean longitudinal grade combined with variables: CCR; average operating speed; consistency index C_p; consistency index C_g; consistency index ΔV_{10}; and consistency index C_c

The models with variables in the exponent part present very similar values for R^2, AIC, SD and χ^2 .

In the models developed in both databases (AADT≤4,000 and AADT>4,000 veh/day), explanatory variables that have statistically significant model parameters contribute to the explanation of the variability of crash data and allow predicting crash frequencies at new locations not used in the model development. In addition it is seen that no model is over-fitted, and therefore the results would be transferable to different locations. Still, the extrapolation of these results to the same type of roadway in other countries is a matter to be approached with caution.

3.3. Step 3: Parsimonious models

At this point it is necessary to confirm the variables that were added in Step 2 (meeting the criteria related to t-statistic, logical sign, no correlation and cumulative residuals), moreover generated a significant drop in the

SD at a 95% level. If a given variable does not generate a significant drop, it is not kept in the model. Models are calibrated considering, separately, the AADT<4,000 veh/day database and the AADT>4,000 veh/day database.

If the parsimony criterion is applied in the AADT≤4,000 veh/day database, only two models are obtained: model 1 (basic model) and model 2 in Table 5. The driveway density (DD) variable is the only one that should be retained in the model. None of the other variables should be added according to the parsimony criterion because none of them meets the above criteria (t-ratio of its estimated parameter is not significant at the 95% confidence level, the addition of the variable to the model does not cause a significant drop in the scaled deviance at the 95% confidence level, or it does not have a logical sign).

If the parsimony criterion is applied in the AADT>4,000 veh/day database, the only resulting model is model 1 (basic model) of Table 6. None of the other variables should be added according to this criterion.

In both databases, the parsimonious models have proved to be quite simplistic. This is a good solution if the data available to researchers is limited. Moreover, as underlined by Mannering and Bhat (2014), if a model is developed using only the volume of traffic and length as explanatory variables, it will exclude significant explanatory variables bias because there are clearly many other factors affect the frequency of crashes.

3.4. Analysis of variables in the models

In order to facilitate interpretation of the models obtained for AADT under and over 4,000 veh/day, following several authors (Osgood, 2000; Olmstead, 2001; Chin and Quddus, 2003), the coefficients are transformed to incidence rate ratios (IRR) —i.e., e^{β} rather than β. IRR can take on different values. If the IRR of a given variable is much less than 1.0, then an increase in the value of the variable is associated with a significant improvement in safety. Conversely, if the IRR is much greater than 1.0, an increase in the value of the variable is associated with a significant decline in safety. Otherwise, the variable has no effect on safety (Chin and Quddus, 2003).

[Insert Table 7 here]

Table 7 shows the final set of all the variables included in the models, their maximum and minimum coefficients, the models where they appear and the corresponding IRR. To facilitate interpretation, the

$IRR^{0.10}$ is given, indicating the effect that a 10% increase in the independent variable would have on the total number of crashes.

Models for AADT≤4,000 veh/day database

Of all the geometric variables considered in the models calibrated in the AADT≤4,000 veh/day database, the only ones kept in the models are the average longitudinal grade (LGr) and the average operating speed (V_{85avg}). LGr presents a negative sign, thus indicating that when the average longitudinal grade increases, the occurrence of crashes decreases. Several studies (Pardillo and Llamas, 2003; Pardillo et al., 2006; Montella et al., 2008; Montella, 2010; Cafiso et al., 2013) report similar results. The coefficients for LGr vary between -0.0171 and -0.0128 (Table 5 and Table 7), indicating that all other things being equal, an increase of 10% in longitudinal grade is associated with a 0.1%-0.2% reduction in total annual crashes ($IRR^{0.1}$ between 0.999 and 0.998). This value for IRR indicates that longitudinal grade has little effect on safety.

V_{85avg} shows a negative sign, indicating that if V_{85avg} increases, the occurrence of crashes decreases. This is logical if one considers (disregarding other factors) that higher speed on flat terrain could be indicative of good road design, hence fewer crashes. Hauer et al. (2004) found that the higher the speed limit, the fewer the expected crashes. It is likewise possible that roads where a low speed is posted may be considered to be of high risk. $IRR^{0.1}$ for V_{85avg} is 0.998, indicating that all other things being equal, a 10% increase in V_{85avg} is associated with a 0.2% reduction in total annual crashes.

C_g and C_c present a positive sign, indicating that the worse the section, the greater the number of crashes expected (Ng and Sayed, 2004; Cafiso et al., 2010; Camacho-Torregrosa et al., 2013; Garach et al., 2014). IRR for C_c is 1.000, meaning this variable has no effect on safety. The $IRR^{0.1}$ for C_g is 0.977, so that other things being equal, an increase of 10% in C_g is associated with a 2.3% reduction in total annual crashes.

Among the context variables, the percentage of shoulder and the driveway density variables are found to contribute to accident occurrence significantly. The estimated coefficients of the variable percentage of shoulder (paved or not paved) are highly significant.

The coefficient for the percentage of shoulder is -0.5116, indicating that, all other things being equal, an increase of 10% in the percentage of shoulder is associated with a 5% ($IRR^{0.1}$ is 0.950) reduction in total annual crashes. The variable percentage of paved shoulder has a similar effect, reducing the number of crashes by 6.3% when there is an increase of 10% for paved shoulder in the segment. The negative sign accompanying these variables has also been reported by other authors. Head and Kaestner (1956) concluded that total crashes increase with increasing shoulder width, except for roadways having AADT between 3,600 and 5,500 veh/day. Perkins (1956) found that all accident types decreased with increased shoulder width for AADT's between 2,600 and 4,500 veh/day. Stohner (1956) observed reductions in crashes as shoulder width increased, especially in the 2,000-6,000 AADT range. Hadi et al. (1995) found that increasing lane and shoulder widths decreased the accident rate. Fitzpatrick et al. (2000) reported that the number of crashes decreased when shoulder and lane width increased. Dell'Acqua and Russo (2008) concluded that accident frequency increases with lower roadway paved width. Anastasopoulos et al. (2008) also concluded that the number of crashes decreases when the shoulder width is greater.

Driveway density has a positive sign, indicating that higher driveway density increases the likelihood of accident occurrence. Other authors have arrived at similar results (Fitzpatrick et al., 2000, 2010; Pardillo and Llamas, 2003, Pardillo et al., 2006; Cafiso et al., 2010). This variable intervenes in the four models. In all of them the coefficient ranges from 0.1121 to 0.1145, thus indicating that a 10% increase in driveway density is associated with increase of 1.1%-1.2% in the number of crashes ($IRR^{0,1}$ is between 1.011-1.012).

Models for AADT>4,000 veh/day database

In the models obtained for AADT>4,000 veh/day, among the exposure variables, the percentage of heavy vehicles has a high influence on crashes (models 2, 6-9 in Table 6). The highest value for β is 2.0429 (Tables 6 and 7), which means that a 10% increase in the percentage of heavy vehicles would result in a 22.7% greater crash occurrence (IRR^{10} is 1.227). This variable has a positive sign: a higher number of crashes is associated with the higher percentage of heavy vehicles. Ramírez et al. (2009) demonstrated, with different roadway types, that a reduction in the total number of crashes would occur as a result of a drop in the number of heavy vehicles. Hosseinpour et al. (2014) presented similar findings.

CCR, average longitudinal grade, and average operating speed also contribute to accident occurrence. CCR has a high influence on crashes. The parameters maximum and minimum estimate for CCR are 2.1633 and 1.9699 (Table 7). These values show that a 10% increase in the percentage of CCR increases the number of crashes by an average of 24.2% (IRR^{10} is 1.242) or 21.8% (IRR^{10} is 1.218). The positive sign by this variable indicates that the greater the change in curvature, the more the expected crashes. Cafiso et al. (2013) obtained the same sign for this variable. The average longitudinal grade and the average operating speed variables have the same signs as in the models obtained for AADT≤4,000 veh/day. The values of IRR are also similar, although they have a lesser influence on crashes (they are associated with a 0.1%-0.2% reduction in total annual crashes).

The consistency variables that intervene in all the models are: indexes C_g and C_p, and ΔV10. Index C_g has a negative sign, as in the AADT≤4,000 veh/day database, indicating that the worse the road design, the greater the number of crashes expected. However, the coefficient that accompanies this variable in the AADT>4,000 database is lower, meaning that the variable is less influential with regard to crashes (IIR^{10} is 0.986, hence a 1.4% reduction in total annual crashes). Index C_p has a negative sign that leads to the same interpretation as for C_g. The IRR^{10} for C_p varies between 0.998 and 0.987, indicating that all other things being equal, an increase by 10% in C_p is associated with a reduction between 0.2%-1.3% in total annual crashes. ΔV10 presents a positive sign, indicating that more the differences in speed (over 10 km/h) among successive elements entail a greater probability of crash occurrence. This variable has little effect on safety, given that the IRR^{10} varies only from a minimum of 1.007 to a maximum of 1.008; a 10% increase in the variable ΔV10 is associated with an increase of 0.7%-0.8% in total crashes.

The only context variable that intervenes in the models is driveway density, with the same positive sign as seen for the models obtained in the AADT≤4,000 veh/day database. This variable affects crashes less in the AADT>4,000 day database than in the AADT≤4,000 database. In the latter, as commented earlier, the coefficients of the order of 0.11 would indicate that a 10% increase in the driveway density variable is associated with approximately 1% more crashes. In the database with AADT>4,000 the coefficient of 0.0524 implies an increase in crashes of 0.05%.

Comparison of the models obtained for AADT≤4,000 veh/day and for AADT>4,000 veh/day

A general comparative analysis of the models obtained in both databases shows that there are variables that have a great effect in one database but not in the other. For example, the variables percentage of heavy vehicles and curvature change rate (CCR) are included in the AADT>4,000 veh/day database and not in the other; whereas the variables percentage of shoulder (paved or not paved) and driveway density are in the AADT≤4,000 veh/day database but not in the other.

A detailed comparison of the models obtained in the two databases points to these noteworthy findings:

- In five models for AADT>4,000 veh/day there appears the variable percentage of heavy vehicles (not appearing for AADT≤4,000 veh/day). In the AADT>4,000 veh/day database, the percentage of heavy vehicles has, together with the variable CCR, the greatest relative effect on the crash frequency among all the independent variables. Thus, a 10% increase in %hv is thought to cause an increase of up to 22.7% (model 7) in the fatal crashes. It is logical that heavy vehicles influence crash statistics on roadways with high traffic volume more than they do on roadways with low traffic volume. A high volume of traffic usually translates as high light vehicle traffic, which could produce scenarios of even greater traffic conflicts caused by speed differences, resulting in overtaking maneuvers using the oncoming lane, thereby increasing the risk of crashes.
- CCR is included in roadways with AADT>4,000 veh/day but does not take part in any model when the database is AADT≤4,000 veh/day. This variable has a high effect on the crashes in roadways having AADT>4,000 veh/day, as a 10% increase in CCR is thought to cause an increase of up to 24.2% (model 10) in the crashes. Therefore, roadways with a volume of traffic over 4,000 veh/day should take special care regarding curvature changes. The high volume of traffic could produce a greater number of dangerous maneuvers in which a change in curvature would favor the occurrence of crashes.
- The percentage of shoulder (paved or not paved) participates in the models based on AADT≤4,000 veh/day, but in no model with the database AADT>4,000 veh/day. Roadways with a greater volume of traffic usually have a shoulder, and it is usually paved; whereas along roadways with less traffic this is generally not the case. Moreover, the effect of both these variables in the models with database AADT≤4,000 veh/day is considerable. Coefficients between -0.5111 and -0.6464 indicate that a 10% increase in this variable is associated with a reduction in total crashes between 5% and 6.3%.
- When AADT>4,000 veh/day, the driveway density appears in just one of the models, while this variable intervenes in four of the models when AADT≤4,000 veh/day. The coefficients show that this variable has more impact on crashes in the AADT≤4,000 veh/day database than in the AADT>4,000 veh/day database. In the former, the regression coefficient of the order of 0.11

indicates that an increase by 10% in the variable driveway density means an increase in crashes of 1.1%; in turn, in the database of roadway with AADT>4,000 veh/day the regression coefficients around 0.05 point to an increase of 0.5%. This could be due to the fact that roads with more traffic volume have more controlled access than roadways with less traffic. In addition, Spanish legislation allows left turns on roadways with AADT≤5,000 veh/day if they have a middle lane for waiting, but left turns are not permitted on roadways with AADT>5,000 veh/day.

- C_g intervenes in models of both databases and it presents the same effect as CCR: inconsistencies in the road's design with high traffic volumes can give rise to a great number of dangerous maneuvers, with an ensuing greater risk of crash occurrence.

4. CONCLUSIONS

This paper investigates the relationship between crash frequency and several variables related with exposure, geometry, consistency and context for Spanish two-lane rural highways on flat terrain. Cumulative residual analysis of the model built with only the variables AADT and length made it possible to identify regions where the model either under- or over-estimates crashes. The original sample was divided on the basis of ranges of the explanatory variable AADT. Stratification for AADT under and over 4,000 veh/day led to a significant improvement of the models generated.

The parsimonious models have proved to be quite simplistic in both databases. This is a good solution if the data available to researchers as limited. The problem is that the model will be excluding significant explanatory variables bias because there are clearly many other factors affecting the frequency of crashes.

The fully specified models show appreciable differences for the SPF obtained in each one of the databases. In the AADT>4,000 veh/day database, the percentage of heavy vehicles has a large effect on the crash frequency. A 10% increase in the percentage of heavy vehicles is determined to cause a 22% increase in the occurrence of crashes. The variable CCR is also highly significant for crashes on this roadway type, as a 10% increase in CCR means 24% more crashes. Neither of these variables is included in the models for AADT≤4,000 veh/day.

In the AADT≤4,000 veh/day database, the percentage of shoulder (paved or not paved) bears a high influence on crashes. According to the models generated, an increase of 10% in these variables is associated with around a 5% reduction in total crashes. Notwithstanding, this variable does not participate in any model generated for AADT>4,000 veh/day, as highways with a greater volume of traffic normally have a shoulder, most often a paved shoulder, whereas roadways with less traffic do not. The driveway density takes part in four models of the AADT≤4,000 veh/day database and in just one model based otherwise. In the first database an increase of 10% in the variable driveway density would give an increase of 1.1% in the occurrence of crashes, while in the AADT>4,000 veh/day database, there would be an increase of 0.5%. On roadways with greater volumes of traffic, the number of driveways is usually regulated and channeled through service roads. Furthermore, Spain´s regulations allow for left turns on roadways with AADT under 5,000 veh/day as long as there is a middle lane for waiting, whereas this is not allowed for roadways with AADT>5,000 veh/day.

In view of the results expounded here, Spain´s Highway Administration should pay special attention to the curvature changes and the percentage of heavy vehicles on two-lane rural highways with a volume of traffic exceeding 4,000 veh/day, as well as the percentage of shoulder and the driveway density on two-lane rural highways with a volume of traffic under 4,000 veh/day. Extrapolation of these results to this same type of roadway in other countries is a matter to be approached with caution.

As future work, different stratifications of the sample according to the different AADT values could be analysed.

An additional analysis could also be carried out using advanced techniques to deal with variation of the effectiveness of predictor. Some of these techniques might be: Generalized Additive Models (GAM) which offer more flexible functional forms than traditional generalized models and allow for more adaptable variable interactions (Li et al., 2010); or Multivariate Adaptive Regression Splines (MARS) which avoid the over-estimation problem through consideration of interaction impacts between variables (Park, 2015).

Furthermore, the developed crash prediction models predict crashes for all types of accidents and they do not distinguish crash severity levels. If enough data were available, it would be interesting to conduct analyses for different crash types and severity levels in future research efforts.

ACKNOWLEDGEMENTS

The authors would like to thank the ERDF of the European Union for financial support via project “Bases para un sistema experto que permita la identificación probabilística de Tramos de Concentración de Crashes (TCA)” under the “Programa Operativo FEDER de Andalucía 2007-2013”. We also thank the Public Works

Agency and the Regional Ministry of Public Works and Housing of the Regional Government of Andalusia. Griselda López wishes to express her acknowledgement of the regional ministry of Economy, Innovation and Science of the regional government of Andalusia (Spain) for their scholarship to train teachers and researchers in Deficit Areas.

REFERENCES

Abdel-Aty, M.A., Radwan, A.E. (2000). Modelling traffic accident occurrence and involvement. Accident Analysis and Prevention, 32 (5), 633-642.

Anastasopoulos, P. C., Tarko, A. P., Mannering, F. L. (2008). Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention, 40 (2), 768-775.

Anderson, I.B., Bauer, K.M., Harwood, D.W., Fitzpatrick, K. (1999). Relationship to safety of geometric design consistency measures for rural two-lane highways. Transportation Research Record, 1658, 43-51.

AASHTO. (2010). Highway Safety Manual. Washington, DC, 1500.

Cafiso, S., Di Graziano, A., Di Silvestro, G., La Cava, G., Persaud, B. (2010). Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. Accident Analysis and Prevention, 42, 1072–1079.

Cafiso, S., D' Agostino, S., Persaud, S. (2013). Investigating the influence of segmentation in estimating safety performance functions for roadway sections. Transportation Research Board, 92nd Annual Meeting, Washington.

Camacho-Torregrosa, F. J., Pérez-Zuriaga, A. M., Campoy-Ungría, J. M., García-García, A. (2013). New geometric design consistency model based on operating speed profiles for road safety evaluation. Accident Analysis and Prevention, 61, 33-42.

Camacho-Torregrosa, F.J. (2014). Desarrollo y calibración de un modelo global de consistencia del diseño geométrico de carreteras convencionales basado en el empleo de perfiles continuos de velocidad de operación. Ph.D. Thesis, pp. 837, University of Valencia.

Cheng, W., Washington, S. (2008). New criteria for evaluating methods of identifying hot spots. Transportation Research Record, 2083, 76-85.

Chin, H.C., Quddus, M.A. (2003). Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accident Analysis and Prevention, 35 (2), 253-259.

De Oña, J., Garach, L., Calvo Poyo, F., García-Muñoz, T. (2014). Relationship between predicted speed reduction on horizontal curves and safety on two-lane rural roads in Spain. Journal of Transportation Engineering, 140 (3), 1-12.

Dell'Acqua, G., Russo, F. (2008). Accident prediction models for road networks. Department of Transportation Engineering, University of Naples, Italy.

Draper, N., Smith, H. (1981). Applied regression analysis, 2nd ed. Wiley, New York, pp 407.

El-Basyouny, K., and T. Sayed (2010). A method to account for outliers in the development of accident prediction models. Accident Analysis and Prevention, 42 (4), 1266-1272.

Elvik, R. (2007). State-of-the-art approaches to road accident black spot management and safety analysis of road networks. Institute of Transport Economics, Norwegian Centre for Transport Research.

Elvik, R. (2008). Making sense of road safety evaluation studies. Developing a quality scoring system. Report, 984.

Elvik, R. (2011). Assessing causality in multivariate accident models. Accident Analysis and Prevention, 43(1), 253-264

Federal Highway Administration (2010). American Association of State Highway and Transportation Officials. Highway Safety Manual (HSM). First Edition.

Fitzpatrick, K., Collins, J. (2000). Speed-profile model for two-lane rural highways. Transportation Research Record: Journal of the Transportation Research Board, 1737, 42-49.

Fitzpatrick, K., Elefteriadou, L., Harwood, D. W., Collins, J. M., McFadden, J., Anderson, I. B., Krammes, R.A., Irizarry, N., Parma, K.D., Bauer, K.M., Passetti, K. (2000). Speed prediction for two-lane rural highways. FHWA-RD-99-171

Garach, L., Calvo, F., Pasadas, M., De Oña, J. (2014). Proposal of a New Global Model of Consistency: Application in Two-Lane Rural Highways in Spain. Journal of Transportation Engineering, 140 (8), 04014030.

Hadi, M. A., Aruldhas, J., Chow, L., Wattleworth, J. A. (1995). Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record, 1500, 169-177.

Hauer, E. (2004). Statistical road safety modeling. Transportation Research Record, 1897, 81-87.

Hauer, E. (2015). The art of regression modeling in road safety. Springer, pp. 231.

Hauer, E., Council, F.M., Mohammedshah, Y. (2004). Safety models for urban four-lane undivided road segments. Transportation Research Record, 1897, 96-105.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S. (2002). Estimating safety by the Empirical Bayes method: a tutorial. Transportation Research Record, 1784, 126-131.

Hauer, E., Lovell, J., (1988). Estimation of safety at signalized intersections. Transportation Research Record 1185, 48-61.

Head, J. A., Kaestner, N. F. (1956). The Relationship between Accident Data and the Width of Gravel Shoulders in Oregon, Proceedings, Highway Research Board.

Hosseinpour, M., Yahaya, A. S., Sadullah, A. F. (2014). Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: Case studies from Malaysian Federal Roads. Accident Analysis and Prevention, 62, 209-222.

Lord, D., Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, 44 (5), 291-305.

Lord, D., Persaud, B. N. (2000). Accident prediction models with and without trend: Application of the generalized estimating equations procedure. Transportation Research Record, 1717, 102-108.

Li, X., Lord, D., Zhang, Y., (2010). Development of Accident Modification Factors for Rural Frontage Road Segments in Texas Using Generalized Additive Models. Journal of Transportation Engineering, 137 (1), 74-83.

Mannering, F.L., Bhat, C., (2014). Analytic methods in accident research: methodologial frontier and future directions. Analytic Methods in Accident Research. 1, 1-22.

Mannering, F.L., Shankar, V., Bhat, C.R. (2016). Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1-16.

Miaou, S. P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention, 36, 471-482.

Miaou, S. P., Lu, A., Lum, H. S. (1996). Pitfalls of using R2 to evaluate goodness of fit of accident prediction models. Transportation Research Record, 1542, 6-13.

Misaghi, P., Hassan, Y. (2005). Modelling operating speed and speed differential on two-lane rural roads. Journal of Transportation Engineering, 10.1061/(ASCE) 0733-947X(2005)131:6(408), 408-417.

Mitra, S., Washington, S. (2012). On the significance of omitted variables in intersection crash modeling. Accident Analysis and Prevention, 49, 439, 448.

Montella, A., Colantuoni, L., Lambert, R. (2008). Crash prediction models for rural motorways. Transportation Research Record, 2083, 180-189.

Montella, A. (2010). A comparative analysis of hotspot identification methods. Accident Analysis and Prevention, 42, 571-581

Ng, J. C. W., Sayed, T. (2004). Effect of geometric design consistency on road safety. Canadian Journal of Civil Engineering, 31 (2), 218-227.

Olmstead, T. (2001). Freeway management systems and motor vehicle crashes: a case study of Phoenix, Arizona. Accident Analysis and Prevention, 33, 433-447.

Osgood, D. W. (2000). Poisson-Based Regression Analysis of Aggregate Crime Rates. Journal of Quantitative Criminology, 16 (1), 21-43.

Pardillo, J. M., Bojórquez, R., Camarero, A. (2006): Refinement of Accident Prediction Models for Spanish National Network. Transportation Research Record, 1950, 65-72.

Pardillo, J. M., Llamas, R. (2003). Relevant variables for crash rate prediction in Spanish two lane rural roads. Transportation Research Board, 82nd Annual Meeting. Washington DC.

Park, J. (2015). Exploration and development of crash modification factors and functions for single and multiple treatments (Doctoral dissertation, University of Central Florida Orlando, Florida).

Park, J., Abdel-Aty, M. (2015). Assessing the safety effects of multiple roadside treatments using parametric and nonparametric approaches. Accident Analysis and Prevention, 83, 203-213.

Perkins, E. T. (1956); Relationship of Accident Rate to Highway Shoulder Width, Bulletin 151, Highway Research Board, pp. 13-14.

Persaud, B., Lyon, C., Nguyen, T. (1999). Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. Transportation Research Record, 1665, 7-12.

Polus, A., Mattar-Habib, C. (2004). New Consistency Model for Rural Highways and its Relationship to Safety. Journal of Transportation Engineering, 130, 286-293.

Ramírez, B. A., Izquierdo, F. A., Fernández, C. G., Méndez, A. G. (2009). The influence of heavy goods vehicle traffic on accidents on different types of Spanish interurban roads. Accident Analysis and Prevention, 41(1), 15-24.

RAS-L. (1995). Guidelines for the Design of Roads "Forschungsgesellschaft für Strassen und Verkehrswesen Linienführung." Richtlinien für die Anlage von Strassen, Bonn, Germany.

Resende, P., Benekohal, R. F. (1997). Effects of Roadway Section Length on Accident Modeling. Traffic Congestion and Traffic Safety in the 21st Century: Challenges, Innovations, and Opportunities. American Society for Civil Engineers. Chicago, IL, 403-409.

Sawalha, Z., Sayed, T., (2001). Evaluating the safety of urban arterial roadways. Journal of Transportation Engineering ASCE, 127(2), 151-158.

Sawalha, Z., Sayed, T. (2006). Traffic accident modeling: Some statistical issues. Canadian Journal of Civil Engineering, 33(9), 1115-1124.

Stohner, W. R. (1956). Relation of Highway Accidents to Shoulder Width on Two-Lane Rural Highways in New York State. Highway Research Board Proceedings, 35, pp. 500-504.

Turner, S, R Singh and G Nates (2012). The next generation of rural road crash prediction models: final report. NZ Transport Agency research report 509. 98pp.

Vogt, A., Bared, J. (1998). Accident models for two-lane rural segments and intersections. Transportation Research Record, 1635, 18-29.

Washington, S., Karlaftis, M., Mannering, F., 2010. Statistical and Econometric Methods for Transportation Data Analysis, 2nd ed. Chapman and Hall, Boca Raton.

Wei, F., Lovegrove, G. (2013). An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis and Prevention, 61. 129-137.

World Health Organization (2013). Global status report on road safety. Available at: http://apps.who.int/iris/bitstream/10665/78256/1/9789241564564_eng.pdf.

List of Tables:

List of Figures: