Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 160 (2014) 332 – 341

Procedia
Social and Behavioral Sciences

XI Congreso de Ingenieria del Transporte (CIT 2014)

Using Decision Trees for Comparing Different Consistency Models

Laura Garach Morcillo[a]*, Francisco Javier Calvo Poyo[a], Griselda López Maldonado[a]

[a] *TRYSE Research Group, University of Granada, ETSI Caminos, Canales y Puertos, c/Severo Ochoa, s/n, 18071 Granada, Spain*

Abstract

One technique used to improve highway safety from the point of view of the infrastructure is to examine the consistency of the design. Design consistency refers to if highway geometry is conformance to driver expectancy. When the consistency of the road is inadequate, the more likely it is that drivers will be startled and a crash will occur. The consistency, based on operating speed, has been calculated in Spanish two-lane rural highways. This consistency has been evaluated using a local method, to measure the consistency of each element of the road and using a global method, to measure the consistency of a segment of the road. Different models of consistency have been compared using Decision Trees (DTs). DTs are a Data Mining Techniques which can be used to solve classification problems. The results show that DTs are a suitable technique to compare consistence models and they permit to establish limits between the different models analyzed.

Peer-review under responsibility of CIT 2014.

Keywords: consistency; crashes; new global model; comparison consistency model; decision trees; CART.

1. Introduction

Design consistency is understood as the conformance of highway geometry with driver expectancy, or the relationship between the geometric characteristics of a highway and the conditions the driver expects to encounter (Castro et al., 2008). Many authors agree that operating speed is the form most commonly used to evaluate consistency, as it reflects driver behavior (Gibreel et al., 1999; Fitzpatrick et al. 2000; Ng and Sayed, 2004; Camacho-Torregosa et al., 2013). Operating speed is the most representative parameter of real driving performance (Dell`Acqua et al., 2013). The operating speed is defined as the 85th percentile of the distribution (V_{85}) of speeds by drivers under free-flow conditions on a particular location of the road alignment (Bella, 2007).

* Corresponding author. Tel.: +34 958 249455.
E-mail address: lgarach@ugr.es

Peer-review under responsibility of CIT 2014.
doi:10.1016/j.sbspro.2014.12.145

Models based on operating speed to calculate road consistency can be local, applied to a specific geometric element of one road segment; or global, producing a consistency value for the whole road segment. Different local models have been used for several authors. Babkov (1968) concluded that consistent and safe designs could be produced when the difference in the operating speed between two consecutive elements did not exceed 15% of the speed in the preceding element. Leisch and Leisch (1977) recommended a revised design speed concept that included guidelines on V_{85} reductions and differentials between the design speed (V_d) and V_{85}. Kanellaidis et al. (1990) suggested that a good design is achieved when the difference between V_{85} on the tangent and the following curve does not exceed 10 km/h. However, of all the local methods based on operating speed to determine the degree of consistency, the best known local criterion is that by Lamm et al. (1999) based on mean crash rates. They presented two design consistency criteria related to operating speed, consisting of the difference between V_d and V_{85} (criterion I) and the difference in V_{85} of successive elements (criterion II, named C_1 in this paper). Table 1 shows the consistency thresholds for both criteria:

Table 1.Thersholds for a determination of design consistency quality. (Lamm et al., 1999)

Consistency	Criterion I (km/h)	Criterion II (km/h):C_1
Good	$\lvert V_{85} - V_d \rvert \leq 10$	$\lvert V_{85i} - V_{85i+1} \rvert \leq 10$
Acceptable	$10 < \lvert V_{85} - V_d \rvert \leq 20$	$10 < \lvert V_{85i} - V_{85i+1} \rvert \leq 20$
Poor	$\lvert V_{85} - V_d \rvert > 20$	$\lvert V_{85i} - V_{85i+1} \rvert > 20$

As regards global models, Polus and Mattar-Habib (2004) developed a consistency model C_2 (Eq 1) to assess the consistency of whole road segments. Their model is based on two new consistency measures. The first is the relative area bounded between the operating speed profile (representing the V_{85} for each element of the road segment) and the average weighted operating speed (Ra) (Eq 2). The second one is the standard deviation of the operating speeds at every element of the road segment (σ) (Eq 3).

$$C_2 = 2.808 \cdot e^{-0.278 \cdot [R_a \cdot (\sigma/3.6)]} \tag{1}$$

where:

C_2 = Global consistency model according Polus and Mattar-Habib (2004) (m/s)

R_a= Relative area measure of consistency (m/s)(Eq 2)

σ = Standard deviation of operating speed (km/h) (Eq 3)

$$R_a = \frac{\sum_{i=1}^{n} a_i}{L} \tag{2}$$

where:

a_i= Area, between the speed in each element of profile and average speed (m^2/s)

L = Road segment lenght (km)

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (V_{85i} - \bar{V}_{85})^2}{n}} \tag{3}$$

where:

n = Numberof elements along a road segment

V_{85i} = Operating speed on each element i (tangent or curve) (km/h)

$\bar{V}_{85}$= Average operating speed (km/h):

$$\overline{V}_{85} = \frac{\sum_{i=1}^{n} V_{85i} L_i}{L} \quad (4)$$

where:

L_i =i element length of the road segment (km)

The consistency thersholds established by Polus and Mattar-Habib (2004) are shown in Table 2.

Table 2.Thersholds for the Determination of Design Consistency Quality (Polus and Mattar-Habib, 2004)

Design consistency quality		
Good	Acceptable	Poor
$C_2 > 2$ m/s	$1 < C_2 \leq 2$ m/s	$C_2 \leq 1$ m/s

Currently other authors have developed global models of consistency but are less well known. For example, Camacho-Torregrosa et al. (2013) put forth a new model of consistency based on data acquired by continuous speed profiles. For each road segment, they relate the speed profile with the geometric variables of the road segment and with crash statistics for the segment. The proposed consistency index relates the average operating speed with the average speed reduction on the road segments. García et al. (2013) developed a new consistency model for evaluating the performance of tangent-to-curve transitions on two-lane rural highways, based on the Inertial Consistency Index (ICI) defined for each transition. This was calculated at the beginning point of the curve, as the difference between the average operating speed of the previous 1 km road segment (inertial operating speed) and the operating speed at this point. The thresholds set for this new index were considered good when ICI was lower than 10 km/h; poor when ICI was higher than 20 km/h; and fair in between.

In this paper, the best known local consistency model, the model of Lamm et al. (1999), is compared with the best known global consistency, the model of Polus and Mattar-Habib (2004). To present the comparison between consistency models and discover underlying relationships, DTs are used.

DTs methods can be used to uncover pre-defined underlying relationships between the target (dependent) variable and the predictors (independent variables); and they are very popular due to their simplicity and transparency.

DTs, and more specifically the Classification and Regression Tree (CART) methods, are frequently applied in the field of safety analysis (Kuhnert et al., 2000; Karlaftis and Golias, 2002; Park and Saccomanno, 2005; Chang and Wang, 2006; Harb et al., 2009; Kashani and Mohaymany, 2011; De Oña et al., 2012; Montella et al., 2012; Abellán et al., 2013; De Oña et al., 2013). A main advantage of DTs is that their structure permits the extraction of Decision Rules (DRs) of the type “if-then”. Such DRs may reveal behaviors that occur within a specific dataset. Aside from being practical, they are easy to interpret from the perspective of safety analysis.

The paper is organized in four mayor sections. Section 1 provides an introduction and the description of different consistency models based on operating speed, and DTs are introduced. Section 2 presents the data and methodology, while in section 3 the results and discussion are expounded. Finally, the last section briefly offers the main conclusions of the study.

2. Data and methodology

2.1 Data

Data were obtained from the General Direction of Roadways, governed by the Andalusian Regional Government. Two-lane rural highways in the province of Granada (Spain) were analyzed. Portions of the roadway within small towns or speed zones, or in the vicinity of intersections with stop signs or traffic signal control on the major road were discarded, as were intersections with major changes in Annual Average Daily Traffic (AADT), and passing or climbing lanes.

The road length obtained after eliminating all these sections was 978 km (1,956 km if considering both directions of circulation), with a minimum AADT of 210 veh/day and a maximum of 8,681 veh/day.

2.2 Methodology

First, all roads under study were divided into two types of elements: horizontal curves and tangents. Next, the road data were pre-processed in order to derive homogeneous road segments. The need to study sections with homogeneous characteristics, for the sake of simplicity and coherency in road safety studies, has been demonstrated by a number of authors (Resende and Benekohal, 1997; Pardillo and Llamas, 2003; Cafiso et al., 2008). Taking into account the AADT, Curvature Change Rate (CCR) and average paved Width (W), (Cafiso et al., 2008), 506 homogeneous road segments were obtained. Table 3 shows the minimum, maximum, average values and standard deviation of the length of the sections, as well as the variables used to divide the sample into 506 homogeneous sections:

Table 3. Summary of the characteristics of homogeneous road sections

	Min	Max	Mean	Standard Deviation
Length Sections (km)	0.151	17.141	3.864	3.376
AADT (veh/day)	210	8,681	188.881	2,027.3
CCR (gon/km)	4.7	1,098.05	351.28	310.91
Road Width (m)	5	10	6.2	1.9

Having identified the 506 homogeneous road segments, the operating speed profile for each was built (V_{85} in each road element). This called for establishing a speed on curves and tangents, and deceleration or acceleration rates.

A constant curve speed was adopted. 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 adjusted for horizontal curves in two-way rural highways in Spain. The model thus made it possible to obtain the V_{85} in terms of the radius of the curve.

The tangent speed value taken was 110 km/h (derided speed according to Camacho-Torregrosa et al., 2013).

The acceleration and deceleration rates proposed by Fitzpatrick et al. (2000) for horizontal curves were selected, which are also determined as a function of the radius of the curve.

2.2.1 Calculation of the consistency models

After establishing the speed profile for each road segment, the consistency was determined. To derive the local consistency criterion II of Lamm et al. (1999) based on the difference in operating speed between successive elements (C_1) (Table 1) was applied. For global consistency, previously defined model C_2 was used.

Model C_2 carry out estimations of global consistency, making it possible to obtain a single consistency value per road segment. Model C_1 is used for local estimations; hence, for every road segment, there would be numerous C_1 consistency values, as many as the particular road elements of a given road segment. In order to compare the C_1 local model with the three models of global consistency, for each road segment the proportion of poor, acceptable and good elements present were calculated with respect to the total number of road elements for that road segment.

The percentage of good elements in each road segment (N_{10}) is calculated as:

$$N_{10} = \frac{N(\Delta V \leq 10)}{N_T} \times 100 \quad (5)$$

where:

$N(\Delta V \leq 10)$ = Number of elements in the road segment that present operating speed differences lesser than or equal to 10 km/h

N_T = Number de road elements en el road segment

Similar indices are proposed for the categories of acceptable ($N_{10\text{-}20}$) and poor (N_{20}). They are calculated in the same way as the N_{10} of Eq 6, but considering in the numerator, respectively, the speed differences between 10 km/h and 20 km/h, and those over 20 km/h.

In this way, for each road segment, as representative measures of C_1, the values of N_{10}, $N_{10\text{-}20}$ and N_{20} were calculated. The sum of these three indices represented 100% of the elements of the road segment.

After calculation of the four consistency models (C_1, by means of its three measurements N_{10}, $N_{10\text{-}20}$ and N_{20}, and C_2, C_3 and C_4), the values were compared amongst themselves by pairs, by means of DTs, considering one as the dependent variable and the other as an independent variable.

2.2.2 CART Method

A Decision Tree (DT) is an oriented graph made up of a finite number of nodes departing from a root node. DTs are built recursively, following a descending strategy, starting with the full data set constituted by the root node. Using specific criteria, the full set of data is then split into smaller subsets, and each subset is split recursively until all of them are pure (when the cases in each subset are all of the same class) or their “purity” cannot be increased. This is how the tree’s terminal nodes are formed, in view of the answer values of the class variable (De Oña et al., 2013).

Depending on the nature of the dependent variable, CART develops a classification tree when the value of the target variable is discrete, whereas a regression tree is developed for the continuous target variable. Because this study aims to explore categorical variables, classification trees were developed. This process consists of finding a model (or function) that describes and distinguishes data classes or concepts, to use the model to predict the classes of objects (Han and Kamber, 2006).

The CART method can be used to build binary trees, in which each parent node is linked to just two children nodes namely the left node and the right node. A branch of the tree constitutes a sub-tree, obtained by pruning the tree at a given internal node. By definition, the terminal nodes present a low degree of impurity compared to the root node. In the growing tree, predictors generate candidate partitions (or splits) at each internal node of the tree, so that a suitable criterion needs to be defined in order that the best partition (or the best split) of the objects can be selected.

The split criterion to measure the impurity of nodes used in CART is based on the Gini Index of diversity (the diversity of classes in a tree node being used). For a variable C, it is defined as:

$$gini(C) = 1 - \sum_j p^2\,(C = c_j), \quad (6)$$

In this way, we can define the split criterion based on the Gini Index as:

$$\mathrm{GI}x(C, X) = gini(C|X) - gini(X) \quad (7)$$

where $gini\,(C|X) = \sum_t p(x_t) gini\,(C|X = x_t)$, and X another known variable.

Thus, the best split is the one that minimizes $G\,x(C,X)I$.Following this process, a saturated tree is obtained. To decrease its complexity, the tree is pruned using the so-called cost-complexity algorithm. A more detailed description of the CART method can be found in Breiman et al. (1984).

2.2.2.1 Validation and evaluation of the model

In order to obtain more reliable results, and in light of previous studies (Kashani and Mohaymany, 2011; Abellán et al., 2013; Montella et al., 2013; De Oña et al., 2013), the dataset was randomly split into two different sets: the training set (70% of the data) used to build the model, and the testing check (remaining 30%) used to check the model.
To evaluate the goodness of a classification method, in particular a DT, the parameter accuracy may be used (De Oña et al., 2011, 2013; Mujalli and De Oña, 2011). Accuracy is defined as the percentage of cases correctly classified by the classifier (Delen et al., 2013; Hema Rajini and Bhavani, 2013):

$$\mathrm{Accuracy} = \frac{\sum_{i=1}^{I} \mathrm{TP_i}}{\sum_{i=1}^{I} \mathrm{TP_i+FN_i}} \times 100 \tag{8}$$

where

TP_i = True Positive, that is, instances observed to be from class i are classified (predicted) correctly as belonging to class i

FN_i = False Negative, that is, instances observed to be from class i are classified incorrectly as belonging to a class other than i

2.2.2.2 Rules extraction

The DT structure can be transformed into Decision Rules. A Decision Rule is a logical conditional structure of the type “IF A, THEN B”, A being the antecedent of the rule, and B the consequence. Thus, each rule starts at the root node, and each variable that intervenes in tree division makes an IF of the rule, which ends in leaf nodes with a value of THEN. A priori, the number of rules can be related with the number of terminal nodes in the tree.
In order to extract significant rules, two parameters and a minimum threshold were applied in this study (Abellán et al., 2013; De Oña et al 2013). These parameters are: Population (P_0) and Probability (P).
P_0, is the percentage of the cases on the terminal node from the entire sample:

$$P_0 = \frac{n_T}{N} \times 100 \tag{9}$$

where

n_T = Number of instances that reach the terminal node T

N = Number of cases that compose the whole sample

P is the percentage of the casescorrectly predicted among the proportion of instance reaching the terminal node.

$$P = \frac{n_{Ti}}{n_T} \times 100 \tag{10}$$

where

n_{Ti} = Number of cases that are correctly predicted in a terminal node T

Selection of the minimum thresholds may depend on the nature of the data (balanced or unbalanced) or sample size (small or large datasets). In this study, the thresholds used were P_0>3% (due to the nature of the dataset, though studies such as De Oña et al., 2013 and Abellán et al.,2013 used a lower threshold) and P≥60% (De Oña et al., 2013 and Abellán et al., 2013). Therefore, only rules in which the minimum thresholds are verified were extracted.

3. Results and discussion

Consistency was calculated using models C_1 and C_2 of the 506 homogeneous road segments of study. In this case, in order that the local consistency C_1 might be compared with the global consistencies, C_1* was calculated, representing the percentage of elements that were good, acceptable and poor in all the roads studied. The results of the calculations are shown in Table 4.

Table 4. Values of consistency measures C_1, C_2 and C_3

Consistency Model	Units	Good (%)	Acceptable (%)	Poor (%)	Mean	Min	Max	Standard Deviation
C_1^*	m/s	64.64	20.67	14.69	-	-	-	-
C_2	m/s	5.54	16	78.46	0.69	0.02	2.81	0.62

In Table 4, as shown by local consistency C_1^*, most of the elements (curves and tangents) in the roads of study are good (64.64%). However, analysis of consistency according to the global values for C_2, shows opposing results: most of the road segments (78.46%) are poor and the mean value of C_2 is 0.69, clearly poor. A closer look at C_1^* and C_2, reveals the reason for this: although within a given road segment there may be many locally good elements (many more than the poor ones), if the road segment contains a certain percentage of poor elements, an analysis of global consistency C_2 will show the road segment to be poor. Therefore, although C_1* and C_2 are not directly comparable (the former being related with local consistency and the second with global consistency), our study aimed to discern the cut-off determining whether a road segment will be considered globally poor due to its locally poor elements. The comparison between different consistency models is brought by means of DTs. Classification trees will then be developed.

The means of measuring the consistency of each model varies substantially. Using classification trees, rules are extracted, which will make it possible to compare the models.

The software used to build the DTs was Weka (Witten and Frank, 2005), which is an open source freeware, available at: http://www.cs.waikato.ac.nz/ml/weka/.

For each pairwise comparison, a tree is generated and the rules are extracted (rules in which P and P_0 lie within the thresholds considered, that is P≥60% and P_0≥3%).

- Comparison of C_1 with C_2

C_2, categorized as shown in Table 2, is the class variable, and C_1, with its three measurements N_{10}, $N_{10\text{-}20}$, and N_{20} (percentage of elements in the road segment that are good, acceptable or poor, respectively, according to criterion II of Lamm et al., 1999), is the independent variable (Figure 1).

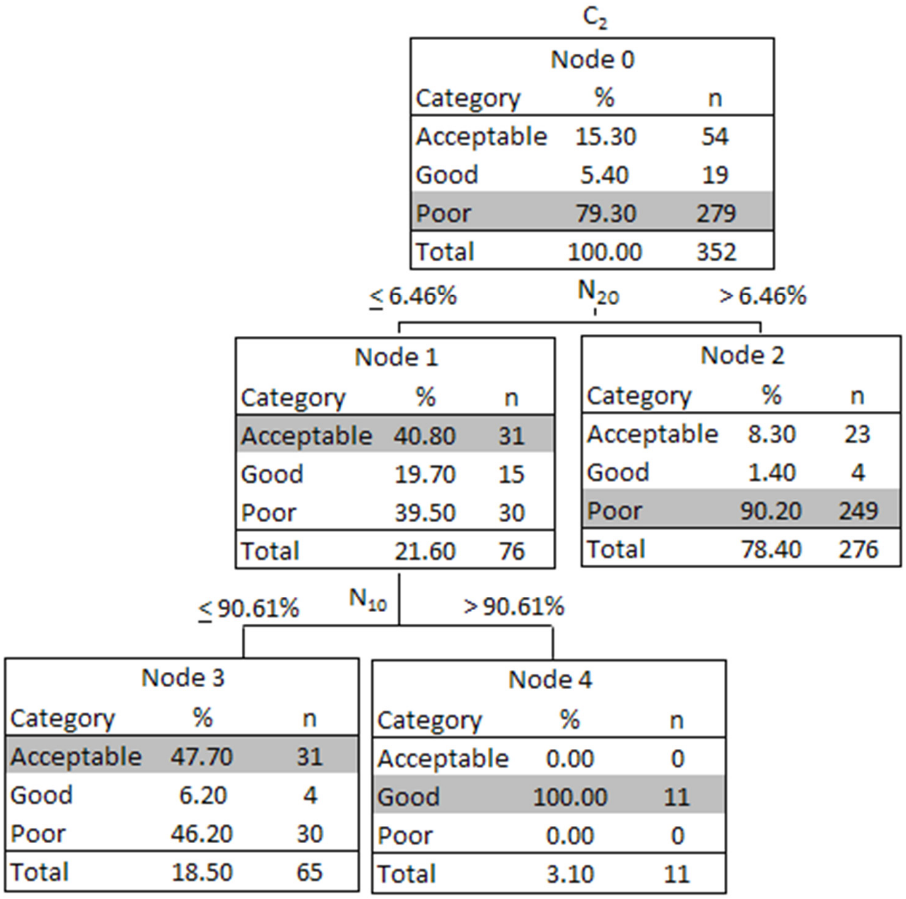

Figure 1. Decision Tree explaining C_2 by means of C_1

Figure 1 shows the tree built with the training set (70% of the data) and validated with the testing set (the remaining 30%). The precision of this tree is 83.60% in the training set and 78.00% in the testing set, a high value in the context of other road safety studies.

Table 5 shows a description of the 2 rules (of 3) identified in the tree (Figure 1), which verify the minimum values of parameters P_0 and P, both in the training set and in the test set. The rule identified by node 3 is not accepted as a definitive rule, since it gives a P value of 47.69% (under the 60% minimum established). The root node that generates the tree is N_{20} (Figure 1). This node is split into two branches: nodes 1 and 2. Nodes 3 and 4 are obtained with different values of N_{10}.

Table 5. Description of the rules obtained comparing C_1 with C_2

NODE	RULES CART	THEN C_2	Po (%)	P (%)
2	IF $N_{20} > 6.46$	Poor	78.41	90.22
4	IF $N_{20} \leq 6.46$ y $N_{10} > 90.61$	Good	3.13	100.00

The tree illustrates that for a segment to be good, it must have a combination of very few poor elements (N_{20}<6.46%) along with many good ones in the segment (N_{10}>90.61%). It is also seen that a segment is classified as poor when it has a small percentage of poor elements within (N_{20}> 6.46%).

The tree obtained in Figure 1 was built with a data set made up of 70% of all the sample data, and randomly chosen. Due to the instability of the trees, if another set of data, representing another 70% of the sample data, is also randomly chosen, the tree obtained will be different, as will the thresholds of the rules derived from it.

To try to resolve this instability, the authors of this paper propose as future lines research, construct different trees using different training sets. Also the authors propose compare more consistency models to each other.

4. Conclusions

Consistency was calculated, for 506 road segments of two-lane rural highways in the province of Granada (Spain), according to a local model, C_1 based on comparing the operating speed of successive elements, following Lamm et al. 1999; and according to the global model developed by Polus and Mattar-Habib, 2004, C_2.

The comparison of the two consistency models by means of DTs reveals, on the one hand, that for a road segment to be considered globally good (according to C_2) it must contain a large percentage of locally good elements according to C1 (around 90%) together with a small percentage of local elements that are poor (some 7%). In contrast, a road segment is considered globally poor if it has just a small percentage of locally poor elements (around 7%).

Comparison of the consistency models by means of DTs made it possible to determine which rules can be established among them. This entails two advantages. The first is that, given that local consistency is easier to calculate than global consistency, one can assess the local consistency of elements of a road segment; and then, with the rules obtained in the DTs, the global consistency can be derived, with a certain degree of probability, with no need to further calculate. The second advantage is that, given the relationship between the models, it can be determined if a road segment is consistent or not according to each one of the models, even if just one is actually calculated. This makes it possible to prioritize interventions in road segments susceptible of improvement for road safety undertaken by the Administration, with its generally limited resources. Thus, intervention could initially be directed towards those road segments that give inconsistent results in simultaneous evaluations using different consistency models. The posterior objective would be to treat the road segments that are classified as inconsistent by some models, but consistent by others.

Acknowledgements

Support from Spanish Ministry of Economy and Competitiveness (Research Project TRA2012-37823), co-funded with FEDER, is gratefully acknowledged.Griselda López wishes to express her acknowledgement to 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

Abellán, J., López, G., De Oña, J. (2013). Analysis of traffic accident severity using Decision Rules via Decision Trees. Expert Systems with Applications, 40, 6047-6054.

Babkov, V.F. (1968). 'Road design and traffic safety, Traffic Engrg, and Control, London, 9, 236–239.

Bella, F. (2007). Parameters for evaluation of speed differential: Contribution using driving simulator. Transportation Research Record, 2023, 37-43.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and Regression Trees. Chapman and Hall, Belmont, CA.

Cafiso, S., Di Graziano, A., Di Silvestro, G. and La Cava, G. (2008). Safety performance indicators for local rural roads: A comprehensive procedure from low-cost data survey to accident prediction model. Transportation Research Board 87th Annual Meeting. Washington, D.C.

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

Castro, M., Sánchez, J.A., Vaquero, c.m., Iglesias, L.and Rodríguez-solano, R. (2008). Automated GIS-based system for speed estimation and highway safety evaluation. Journal of Computing in Civil Engineering 22, 325-331.

Chang, L.Y. And Wang, H.W. (2006). Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accident Analysis and Prevention 38, 1019–1027.

DE oña, J., López G. and Abellán J. (2013). Extracting decision rules from police accident reports through decision trees. Accident Analysis and Prevention 50, 1151-1160.

DE Oña, J., Mujalli, R.O. And Calvo, F.J. (2011). Analysis of traffic accident injury on Spanish rural highways using Bayesian networks. Accident Analysis and Prevention 43, 402–411.

Delen, D., Kuzey, C. and Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. Expert Systems with Applications 40 (10), 3970-3983.

Dell´Acqua, G., Busiello, M., Russo, F. (2013).Safety data analysis to evaluate highway alignment consistency.Transportation Research Record, 2349, 121-128.

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. and Passetti, K. (2000). Speed prediction for two-lane rural highways. FHWA-RD-99-171.

García, A., Llopis-Castelló, D., Camacho-Torregrosa, F.J. and Pérez-Zuriaga, A.M. (2013). New consistency index based on inertial operating speed. Transportation Research Board 92nd Annual Meeting. Washington, D.C.

Gibreel, G. M., Easa, S.M., Hassan, Y. and El-Dimeery, I.A. (1999). State of the art of highway geometric design consistency. Journal of Transportation Engineering, 125, 305-313.

Harb, R., Yan, X., Radwan, E. and Su, X. (2009). Exploring precrash maneuvers using classification trees and random forests. Accident Analysis and Prevention 41, 98-107..

Hema Rajini, N. and Bhavani, R. (2013). Computer aided detection of ischemic stroke using segmentation and texture features. Measurement: Journal of the International Measurement Confederation 46 (6), 1865-1874.

Kanellaidis, G., Golias, J., and Efstathiadis, S. (1990). Driver's speed behaviour on rural road curves. Traffic Engineering and Control, 31, 414–415.

Kashani, A. and Mohaymany, A. (2011). Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. Safety Science 49, 1314–1320.

Karlaftis, M.G. and Golias, I. (2002). Effects of road geometry and traffic volumes on rural roadway accident rates. Accident Analysis and Prevention 34, 357-365.

Kuhnert, P.M., Do, K.-A. and McClure, R. (2000). Combining non-parametric models with logistic regression: an application to motor vehicle injury data. Computational Statistics and Data Analysis 34 (3), 371–386.

Lamm, R., Psarianos, B., and Mailaender, T. (1999). Highway design and traffic safety engineering handbook. McGraw-Hill Companies, Inc., New York, N.Y.

Leisch, J. E., and Leisch, J. P. (1977). New concepts in design-speed application. Transportation Research Record: Journal of Transportation Research Board, Vol. 631, 4–14.

Misaghi, P. y Hassan, Y. (2005). Modelling operating speed and speed differential on two-lane rural roads. Journal of Transportation Engineering, 131,408–417.

Montella, A., Aria, M., D'Ambrosio, A. and Mauriello, F. (2012). Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. Accident Analysis and Prevention, 49, pp. 58-72.

Mujalli, R.O. y de Oña, J. (2011). A method for simplifying the analysis of traffic acci-dents injury severity on two-lane highways using Bayesian networks. Journal of Safety Research 42, 317–326.

Ng, J. and Sayed T. (2004). Effect of Geometric Design Consistency on Road Safety. Canadian Journal of Civil Engineering, 31, 218–227.

Pardillo J.M. and Llamas R. (2003). Relevant variables for crash rate prediction in Spain's two-lane rural roads. Transportation Research Board 82nd Annual Meeting.

Park, Y.-J. and Saccomanno, F. (2005). Collision frequency analysis using tree based stratification. Transportation Research Record, 121–129.