Road safety evaluation through automatic extraction of road horizontal alignments from Mobile LiDAR System and inductive reasoning based on a decision tree

José Antonio Martín-Jiménez[1], Santiago Zazo[2], José Juan Arranz Justel[3], Pablo Rodríguez-Gonzálvez[2,4], Diego González-Aguilera[1*]

[1] University of Salamanca. Department of Cartographic and Land Engineering. Hornos Caleros, 50, 05003, Ávila, Spain. Email corresponding author: daguilera@usal.es

[2] University of Salamanca. TIDOP Research Group. Hornos Caleros, 50, 05003, Ávila, Spain.

[3] Technical University of Madrid. Department of Topographic Engineering and Cartography. Camino de la Arboleda, s/n Campus Sur, Autovía de Valencia Km 7, 28031, Madrid, Spain.

[4] Universidad de León. Department of Mining Technology, Topography and Structures, Astorga, s/n, 24401, Ponferrada (León), Spain.

Abstract

Safe roads are a necessity for any society because of the high social costs of traffic accidents. This challenge is addressed by a novel methodology that allows us to evaluate road safety from Mobile LiDAR System data, taking advantage of the road alignment due to its influence on the accident rate. Automation is obtained through an inductive reasoning process based on a decision tree that provides a potential risk assessment. To achieve this, a 3D point cloud is classified by an iterative and incremental algorithm based on a 2.5D and 3D Delaunay triangulation, which apply different algorithms sequentially. Next, an automatic extraction process of road horizontal alignment parameters is developed to obtain geometric consistency indexes, based on a joint triple stability criterion. Likewise, this work aims to provide a powerful and effective preventive and/or predictive tool for road safety inspections. The proposed methodology was implemented on three stretches of Spanish roads, each with different traffic conditions that represent the most common road types. The developed methodology was successfully validated through as-built road projects, which were considered as "ground truth."

Key words: Road safety, decision tree, geometric design consistency, horizontal alignment parameters, Mobile LiDAR System.

1. Introduction

According to Camacho-Torregrosa et al., (2013), annually, 1.20 million people die and another 20-50 million people are injured in traffic accidents; furthermore, road crashes involve important and high social costs (da Costa et al., 2016). Although the influence of

the human factor is well known (Siskind et al., 2011), roadway design also contributes to the occurrence of accidents (Garach et al., 2016). It should be noted that approximately 30% of road accidents are attributable to infrastructures, and this collision trend is focused on specific road segments (Camacho-Torregrosa et al., 2013; López et al., 2016). For that, the binomial human-road factor plays a crucial role (López et al., 2016). Thus, obtaining safe roads and reducing accidents are challenges that any society should address.

In this framework, the European Union adopted the Directive 2008/96/EC, which is based on the principle of prevention (EU, 2008). This directive establishes different procedures with the final goal of detecting roadway deficiencies and reducing Trans-European Transport Network (TETN) crashes along all its phases: from planning and design to operation of the road infrastructure (Sitran et al., 2016). This directive introduces relevant aspects, such as (a) Road Safety Impact Assessments (RSIA), (b) Road Safety Audits (RSA), (c) Safety Ranking and Management (SRM) and (d) Road Safety Inspections (RSI).

RSIA introduces road safety considerations in the initial planning stage. Through RSA, the road characteristics are checked in the design stage. These two procedures are carried out during the stage of planning. Already during the in-service road stage, SRM provides the ranking of high accident concentration sections and establishes the road infrastructure safety management; by means of RSI, road hazards and safety issues are detected. In this sense, RSI can be understood as an effective preventive tool for the road network. EU (2008) is mandatory on TETN; however, it can also be applied to any national road transport infrastructure as a "good practice guide."

Another approach to road safety, applicable to infrastructures already in service, is the one presented in the European Road Assessment Program (EuroRAP). In this program, a Risk Index (IR) of the road is obtained, based on the accident statistics and traffic intensity (Average Daily Traffic-ADT), which is complemented through an inspection protocol with images and a safety score using stars (EuroRAP, 2018). It should be noted that, in this EuroRAP approach, the geometrical aspects of the road are considered in a very generic manner.

From a dual research-engineering perspective, the concept of geometric consistency in road design has a direct influence on road safety (Ng and Sayed, 2004). The studies developed have been based fundamentally on aspects such as purely geometric, models

of the speed of operation, vehicle stability and the workload of the driver (Andrasik and Bil, 2016, Eftekharzadeh and Khodabakhshi, 2014, Pérez-Zuriaga et al., 2013). However, we must highlight the works developed by Lamm et al., (1991, 1995, 1999, 2001) due to their proposal of the simultaneous triple criterion of stability: (i) in the design (Criterion I), (ii) in the speed of operation (Criterion II) and (iii) in the driving dynamics (Criterion III), which today continue to be a benchmark in the field of road safety.

With respect to data acquisition, over the last few decades, geomatic advances in the Global Navigation Satellite System (GNSS), Light Detection and Ranging (LiDAR), the acquisition of radiometric information, and the hybridization between passive-active sensors, techniques and platforms have revolutionized the massive acquisition of survey data, as well as the inspection and monitoring techniques. A clear example of this is Mobile LiDAR System (MLS) (Bitenc et al., 2011; Gonzalez-Jorge et al., 2013; Mc Elhinney et al., 2010; Puente et al., 2013a).

MLS is principally composed of a navigation system and one or more LiDAR sensors. In this manner, positional data and intensity information of the environment are acquired. Additionally, MLS can be completed by photographic systems (RGB digital cameras) and/or other sensors, such as thermal cameras, ground penetrating radar (GPR) and profilometers. A navigation system is integrated via a set of sensors, such as GNSS, Inertial Measurement Unit (IMU), and Distance Measuring Indicator (DMI) (Holgado-Barco et al., 2015). Moreover, MLS is an accurate and efficient system for data acquisition in complex environments, such as urban and road corridors over large areas, which provides important time reduction for the collection and processing of data (Castro et al., 2016; Holgado-Barco et al., 2015; Varela-Gonzalez et al., 2014).

In the civil engineering field, MLS is beginning to be consolidated among researchers and engineers. Recently, considerable efforts and progress have been mainly made in tasks such as point cloud segmentation, road alignment extraction and automatic object detection. Varela-Gonzalez et al., (2014) proposed a novel method to automatically remove vehicles from mobile LiDAR datasets. Holgado-Barco et al., (2015) developed a method based on segmentation, parameterization and filtering LiDAR point clouds from MLS to extract, semi-automatically, road centrelines and determine horizontal road parameters and their alignment (i.e., straight lines, circular arcs and clothoids). For their part, Riveiro et al., (2015) described an algorithm for the automatic detection of zebra

crossings by means of the standard Hough Transform, which is applied over intensity images. More recently, a cost-effective traffic sign inventory method was proposed in Ai and Tsai, (2015). In parallel, Cabo et al., (2016) applied an automatic algorithm to detect road asphalt edge limits for road maintenance and safety assessment. An adequate review of the scientific literature can be found in Yang et al., (2013).

Alternatively, the automatic extraction of road alignment/markings and geometric design consistency are active research lines for traffic safety. In this sense, Marinelli et al., (2017) compared several methodologies and strategies to obtain the road alignment and geometric parameters of existing roads. Kumar et al., (2014) presented an automated algorithm for extracting road markings from MLS data sets. For their part, Kumar et al., (2013) provided a new approach to road edge extraction based on a parametric active contour model. In Camacho-Torregrosa et al., (2013), a new consistency model was presented based on continuous operating speed profiles, and Montella and Imbriani, (2015) clearly demonstrated the role that design inconsistencies play in road safety. In all these research fields, MLS has already produced significant contributions; however, due to its large research potential, still to be explored, the application of MLS could suppose a paradigm shift, especially over geometric assessment and road safety audits (Gargoum and El-Basyouny, 2017).

On the other hand, it is also worth noting other approaches, applied to in the road field to extract road geometries, which do not employ fundamentally terrestrial LiDAR sensors to data acquisition. Hatger and Brenner (2003) showed, through the combination of existing databases with aerial laser scanner (ALS) data, that it was possible to derive geometrical properties of roads such as height, longitudinal and transversal slope, curvature, and width. In Clode et al., (2004) is presented an approach to extract roads from ALS point clouds with a point density of 1 point m^{-2}, based on a progressive hierarchical classification technique using a digital terrain model created from the last pulse and intensity information of LiDAR. After that, in Clode et al., (2005) this approach to extract roads was improved by a building detection technique that also allowed the detection of existing bridges within the road network, improving the extraction of longitudinal and transverse road profiles. For its part, Alexander et al., (2010) suggested to apply backscatter coefficient versus discrete return data to classify roads from ALS data. In urban areas, Zhou and Vosselman (2012) addressed the problem of detection of

road edges by detecting curbstones in three steps. They found very similar values between ALS and MLS techniques. For instance, in Marinelli et al., (2017) it is shown a novel mobile mapping (MM) vehicle that integrated set of low-cost sensors (GNSS receivers, IMU system and high definition webcams), and where is found that reliability of terrestrial MM is highly dependent on the accuracy of GNSS sensors. For their part, Javanmardi et al., (2017) propose an automatic methodology to extract road features from high resolution airborne images using adaptive thresholding, and whose accuracy is decametric. In Azimi et al., (2018) a novel pixel-wise method is developed that semantically segments high resolution aerial images in order to detect lane markings. This was done through a combination of fully convolutional neural networks with discrete wavelet transform. The images were acquired by three low cost cameras Canon Eos 1Ds Mark III model, ground sampling distance was 13 cm approximately to a flight height of about 1,000 meters above ground level. This method reported high pixel classification accuracy, around 99%. Regarding high-resolution satellite images, recent studies have focused on road-centerline extraction. Sujatha and Selvathi, (2015) present an algorithm to segment and connect road region and remove non-road pixels using morphological operation, with a high average value of completeness-correctness-quality (90%, 96%, and 87%, respectively). Alshehhi and Marpu, (2017) presented a new approach based on hierarchical graph-based image segmentation to extract roads, indicated in urban areas especially, which displayed over 90% effectiveness in road network detection.

Last but not least, there are issues related to road safety and risks. Traditionally, safety studies have focused on factors such as the probability of crashes, types of drivers and roads, etc.; decision tree (DT) techniques have been successfully proved, individually or in combination with decision rules, either as predictive models or as a tool for searching patterns that can explain accident causes. This is also due to their simplicity, the hierarchical structure and the ease of interpretation of results (de Oña et al., 2013; López and de Oña, 2017). For that reason, DTs can be characterized as an effective and adequate tool for the decision-making process. Some examples are the studies developed by (Chang and Wang, 2006; Chang and Chien, 2013; Jung et al., 2016; Kwon et al., 2015; López et al., 2016). However, thus far, there are no studies that address how the road horizontal geometric alignment contributes to road safety by means of a categorization of its inherent geometric risk.

To this end, this work aims to provide a novel and efficient method to assess road safety by means of MLS and the estimation of a potential risk assessment (PRA). This PRA is exclusively derived from a coarse-to-fine approach using point clouds as input data: from the automatic segmentation of roads and the extraction of its horizontal alignment parameters; to the estimation of PRA and the road safety classification based on a decision tree that is an inductive reasoning applied for the first time on geometric parameters exclusively.

The proposed methodology has been successfully implemented and validated on three road stretches that represent the most common types of existing roads in Spain and that present different traffic conditions.

The remainder of this paper is organized as follows: after this introduction, a description of the study cases, the MLS technique and the proposed methodology to assess road risk are shown in Section 2. Section 3 presents the main experimental results drawn from the research. Lastly, in Section 4, the proposed algorithms and the decision support tool are discussed, and the general conclusions from the study are shown.

2. Materials and Methods

2.1 Case studies

90% of Spanish roads are secondary roads (MFOM, 2018) which connect population centres of minor importance in urban and rural areas. These roads present a greater accident rate, despite their density of traffic being less than that of highways and multilane roads (DGT, 2017). For this reason, in this research, three of the four real case studies are secondary roads, whereas the other case is a main road. According with the current Spanish geometric design standards (MFOM, 2016), there are two road groups: (i) highways/multilane roads and (ii) other roads, classified as conventional roads. Within conventional roads, the difference between main and secondary roads refers to the importance of the population centres that the road crosses or connects.

The first, second and fourth case studies are located on the road LU-722, which is situated in the northwest of Spain (Lugo province, in the region of Galicia). These non-consecutive case studies comprise a horizontal sinuous stretch, between kilometre points 4.0 and 121. Their lengths are 1,776.6, 3,160.6 m and 1,427.6 m, respectively. The average cross-section consists of a 6.50-m-wide roadway and a 0.75-m hard shoulder on

either side. The Average Annual Daily Traffic (AADT) is low at 918 vehicles per day (veh/d), (Xunta de Galicia, 2016).

N-640 is the third case. This is also located in Lugo province, and it is characterized by a succession of linked curves. The stretch covers a distance of 2.6 km between kilometre points 84 and 86. The roadway has a single road with two 3.50-m lanes with outer hard shoulders of 1.00 m, roughly. AADT varies between 8,261 and 8,764 veh/d (MFOM, 2017).

By means of the first, second and third case studies will be carried out both the extraction of the horizontal geometric parameters and the training process of the inductive reasoning through a decision tree. The last case study will be applied to risk validation process exclusively.

2.2 Mobile LiDAR System (MLS) and data sets

Data acquisition was carried out by means of Lynx Mobile Mapper by Optech. This system acquires a LiDAR point cloud and RGB imagery simultaneously. The system is composed of two LiDAR sensors, four RGB cameras, a GNSS system, an IMU and a DMI. In this research it did not use camera data but only point clouds. A complete description of the platform and sensors applied is provided in (Holgado-Barco et al., 2015; Puente et al., 2013b). Table 1 shows the main technical characteristics.

MLS sensors	Parameter	Value
GNSS	X,Y coordinates	0.020 m (1)
	Z coordinate	0.050 m (1)
IMU	Roll-Pitch	0.005 ° (1)
	Yaw	0.015 ° (1)
LiDAR	Measuring principle	Time of Flight (ToF)
	Maximum range	200 m
	Precision range	8 mm (1σ)
	Ranging accuracy	± 10 mm (1σ)
	Laser measurement rate	75-500 kHz
	Measurement per laser	Up to 4 simultaneous
	Scan frequency	80-200 Hz
	Laser wavelength	1,550 nm (near infrared)
	Angular resolution	0.001 ° (1)

Table 1. Lynx Mobile Mapper Optech technical characteristics. Accuracy (1).

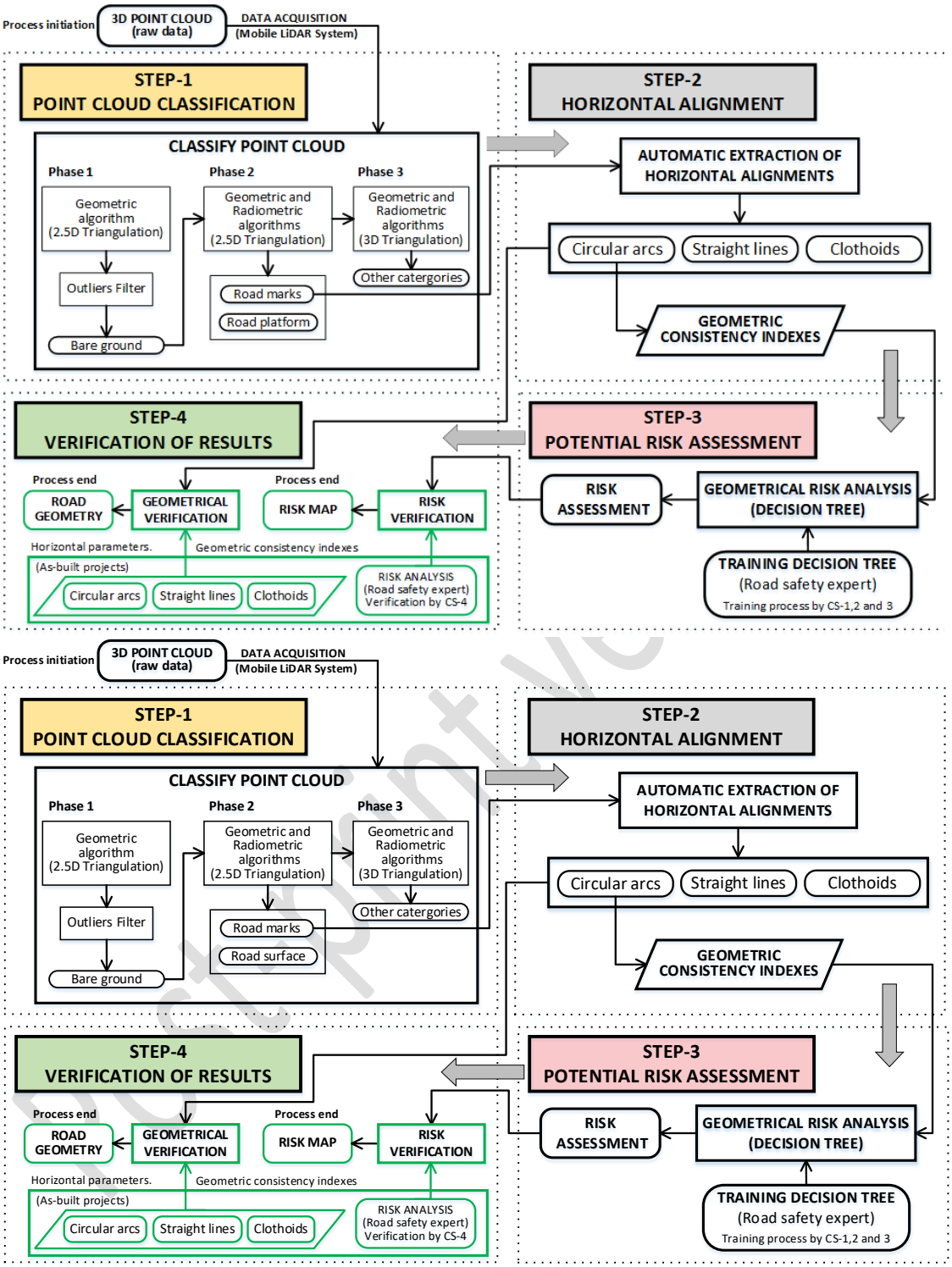
210 To avoid hiding areas, LiDAR sensors were set up at a 45° angle to the platform's
211 trajectory. Regarding three data sets, Table 2 presents a summary of acquired point clouds
212 of this research.

Case study	Length (m)	Point cloud / Point density
LU-722. Stretch 1	≈ 1,800	27,017,955 points / ≈ 121 points m^{-2}
LU-722. Stretch 2	≈ 3,200	43,705,509 points / ≈ 110 points m^{-2}
N-640	≈ 2,200	31,017,623 points / ≈ 35 points m^{-2}
LU-722. Stretch 3	≈ 1,430	18,373,715 points / ≈ 185 points m^{-2}

213 Table 2. Point clouds data sets.

214 **2.3 Methodology**

215 The methodology developed comprises four main steps once the data have been acquired
216 by MLS (Fig. 1). First, an alternative approach was implemented for Mobile LiDAR point
217 cloud classification based on a hierarchical geometric and radiometric analysis of the
218 original 3D MLS point cloud. Second, the horizontal alignment and its main road
219 parameters were automatically extracted together with the computation of geometric
220 design consistency indexes. Third, the Potential Risk Assessment of the road was
221 estimated by a new predictive tool based on a tree induction algorithm. Fourth, the results
222 obtained were compared and verified with those obtained through a road safety and
223 surveyor expert, which were considered as "ground truth."



224 Fig. 1. General methodology developed for the automatic evaluation of road safety based on the road
225 alignment. Note: CS: Case study.

226 **2.3.1 Step 1. Point cloud classification**

227 The acquisition of information from MLS of a roadway is characterized as being highly
228 dense; usually, methods are required to turn the original point cloud into a surface or

229 volume (Arranz Justel, 2013). Typically, this process has been approached fundamentally
230 by means of: (i) taking advantage geometric criteria (based on thresholds for the scan
231 angles of the laser sweep, or extracted points by height difference between trajectory data
232 and road surface, among others) or (ii) from the radiometric characteristics of the points
233 (fixed or adaptative thresholds for the intensity values) (Diaz-Vilarino et al., 2016; Yan
234 et al., 2016; , Holgado-Barco et al., 2014, 2015; Kumar et al., 2014; Riveiro et al., 2015).
235 However, this process is addressed here by an alternative, incremental and sequential
236 algorithm (Arranz Justel, 2013; see Fig. 1 Step-1), following a threefold approach: (i) a
237 first phase of detection of points belonging to the bare ground, (ii) a second phase where
238 the road surface together with the road marks are determined from bare ground points and
239 (iii) a third phase where all remaining elements around road surface and its environment
240 (e.g. vertical signals, protection elements, vegetation, etc.) are classified.

241 The first phase is essential because establishes the reference from which the road surface
242 and road marks (phase 2), as well as its remaining elements (phase 3) are detected. In this
243 first phase, a 2.5D Delaunay triangulation is used (Isenburg, 2006). Considering the huge
244 number of points, the process is based on a sequential algorithm, the so-called “divide
245 and conquer” strategy (Isenburg, 2006), where the cloud is divided by zones following a
246 quadtree scheme, allowing an efficient use of the computer memory.

247 The second phase determines which points, of those previously classified as bare ground,
248 are considered as road surface and road marks. To determine the points belonging to the
249 road surface, a geometrical approach based on slopes and height differences allows us to
250 find the edges of the asphalt and thus to determine the road surface. Regarding road
251 marks, a radiometric algorithm based on intensity values is applied to classify these
252 features of the road surface.

253 Finally, the third phase classifies the remaining points (e.g. protection elements, vertical
254 signals, vegetation, etc.). In order to obtain optimal results and considering the complexity
255 of the road environment, a 3D Delaunay triangulation (Cavendish, 1985) is used to
256 classify these elements, since points having the same planimetric location could have
257 different height. In particular, this 3D approach considers the 3D coordinates of a point
258 and thus takes advantage of the geometric relationships of the objects in the space.

259 It should be noted, that this third phase is not required for extracting road marks (see step
260 1 in Fig.1); however we perform a whole classification of the road and its environment

for other road safety studies related with visibility and protection elements that goes beyond the scope of this paper.

2.3.2 Step 2. Horizontal alignment and geometric consistency indexes

Once the road marks were classified, an iterative procedure using central lines exclusively, based on the Random Sample Consensus algorithm (RANSAC) (Fischler and Bolles, 1981) is utilized to automatically extract the horizontal geometric road elements/parameters (i.e., curves, straight lines and clothoids). It should be noted that in the case of conventional roads the geometric road axis is defined by the central horizontal road mark, which delimits each driving direction (MFOM, 2016).

Alternatively, a joint adaptive thresholding of the RANSAC algorithm is performed based on the geometric features of the road marks and the horizontal alignment parameters according to the type of road and the current regulation norm in Spain (MFOM 1987, 2016). This parameterization process is performed in three sub-steps (Fig. 2), which are supported by an open source point cloud library (PCL) based on the C++ language. First, each circular arc is obtained. Second, every straight line is found, and lastly, the remaining elements are classified as clothoids. Finally, all parametric elements are connected sequentially.

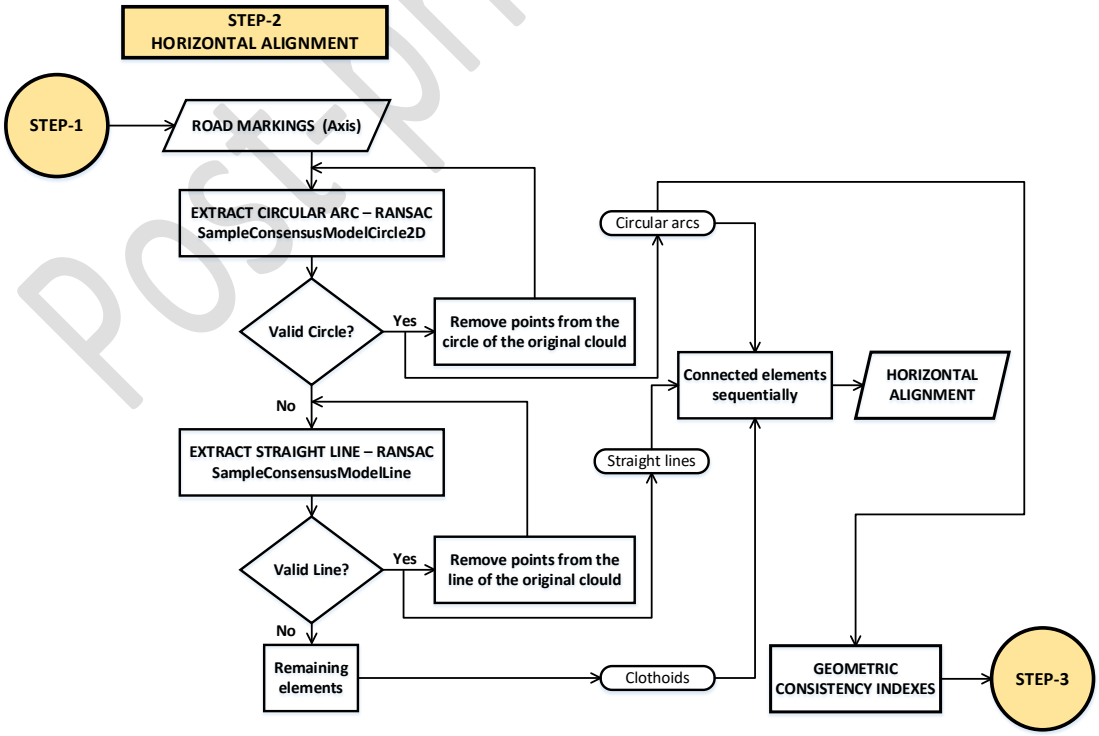


Fig. 2. Algorithm applied to extract the horizontal alignment parameters.

Regarding geometric parameters, the lack of consistency in road design has a negative direct impact on the increase in accident rate (Lamm, et al. 2001). In this sense, geometric consistency comprises three joint stability indexes that allow road safety to be assessed. These are (i) design consistency (Criterion I), (ii) operating speed consistency (Criterion II) and (iii) consistency in driving dynamics (Criterion III) (Lamm et al., 1999, 2001).

Furthermore, considering that the risk of accidents increases when the radius (*R*) of the curve decreases (Rasdorf et al., 2012; You et al., 2012) and that the Curvature Change Rate (*CCR*) is a key parameter due to its influence in the operation speed (Lamm et al., 2001), this study is focused on just circular alignments (Andrasik and Bil, 2016; Misaghi and Hassan, 2005). In this sense, it is noteworthy that the geometric consistency indexes are obtained exclusively from *R* and *CCR* parameters, highlighting the relevance that they have in road safety (Montella and Imbriani, 2015).

First, according to the Spanish geometric design standard (MFOM, 2016), the geometric consistency indexes are determined individually for each curve (Eq. (1)), as well as globally for the entire stretch analysed, excluding those straight lines (Eq. (2)) (Criterion I):

$$CCR_i = \frac{400^g}{2 \cdot \pi \cdot R_i} = \frac{63.6620}{R_i}[gon/m] \approx \frac{63{,}700}{R_i}[gon/km]\ CCR_i = \frac{63{,}700}{R_i} \quad (1)$$

$$\overline{CCR_S} = \frac{\sum_{i=1}^{n}(CCR_i \cdot L_i)}{\sum_{i=1}^{n} L_i} \quad (2)$$

where CCR_i is the curvature change rate corresponding to the *i* curve and expressed in gons/km, R_i is the radius of the *i* curve in metres, and L_i is the length of the element analysed in metres. CCR_S is the global curvature change rate (in gons/km) as a function of the weighted average of the elements considered. It should be noted that *CCR* = 0 in straight lines because $R = \infty$ (Lamm et al., 1999).

Next, the operating speed of each curve is estimated (Eq. (3) and Eq. (4)) via the eighty-fifth percentile of the speed (V_{85}) (Criterion II). This parameter represents the speed at which 85% of the drivers operate on a road in service, which is internationally accepted as a suitable measure of the operating speed (Fitzpatrick et al., 2000):

$$V_{85} = e^{(4.561 - 0.0058 \cdot DC_i)} \quad (3)$$

$$DC_i = \frac{360°}{2 \cdot \pi \cdot R} = \frac{57.295}{R} \approx \frac{5730}{R}[°/100\ m] \quad (4)$$

yielding V_{85} in km/h (Morrall and Talarico, 1994), with *DCi* being the degree of curvature expressed in degrees for every 100 metres.

Frequently, the value of the design speed (V_D) is unknown. For this reason, V_D has been estimated (Eq. (5)) according to the procedure described in (Lamm et al., 1999) for roads in service.

$$V_D \approx V_{85\overline{CCR_S}} = \frac{10^6}{8270 + 8.01 \cdot \overline{CCR_S}} \tag{5}$$

with V_D expressed in km/h.

Finally, the coefficients of lateral friction (Montella and Imbriani, 2015) considered in the design step (f_{RA}) (Eq. (6) and Eq. (7)) and demanded (f_{RD}) (Eq. (8)) according to the operating speed V_{85} (Lamm et al., 2001) are calculated (Criterion III):

$$f_{RA} = 0.925 \cdot n \cdot f_T \tag{6}$$

$$f_T = 0.59 - 4.85 \cdot 10^{-3} \cdot V_D + 1.51 \cdot 10^{-5} \cdot (V_D)^2 \tag{7}$$

$$f_{RD} = \frac{{V_{85}}^2}{127 \cdot R} - e \tag{8}$$

where 0.925 is a reduction coefficient related to tires, n is the utilizing factor (0.60 for roads in service), f_T is the tangential friction factor, and e is the superelevation expressed in %/100 (Lamm et al., 2001); in this case, the theoretical superelevation is considered according to the Spanish regulation norm (MFOM, 2016). Table 3 shows the three stability criteria considered together with their thresholds.

Criterion I. Stability in designs		
$\overline{\boldsymbol{CCR_S}}$ $(\boldsymbol{gon/km})$	$\boldsymbol{\Delta V(km/h)}$	**Feature**
$\lvert CCR_i - \overline{CCR_S}\rvert \leq 180$	$\lvert V_{85(i)} - V_D\rvert \leq 10$	Correct
$180 < \lvert CCR_i - \overline{CCR_S}\rvert \leq 360$	$10 < \lvert V_{85(i)} - V_D\rvert \leq 20$	Acceptable
$360 < \lvert CCR_i - \overline{CCR_S}\rvert$	$\lvert V_{85(i)} - V_D\rvert > 20$	Incorrect
Criterion II. Stability in operating speed		
$\overline{\boldsymbol{CCR_S}}$ $(\boldsymbol{gon/km})$	$\boldsymbol{\Delta V(km/h)}$	**Feature**
$\lvert CCR_i - \overline{CCR_S}\rvert \leq 180$	$\lvert V_{85(i)} - V_{85(i+1)}\rvert \leq 10$	Correct
$180 < \lvert CCR_i - \overline{CCR_S}\rvert \leq 360$	$10 < \lvert V_{85(i)} - V_{85(i+1)}\rvert \leq 20$	Acceptable
$360 < \lvert CCR_i - \overline{CCR_S}\rvert$	$\lvert V_{85(i)} - V_{85(i+1)}\rvert > 20$	Incorrect
Criterion III. Stability in driving dynamics		
$\overline{\boldsymbol{CCR_S}}$ $(\boldsymbol{gon/km})$	$\boldsymbol{\Delta f_R = f_{RA} - f_{RD}}$	**Feature**
$\lvert CCR_i - \overline{CCR_S}\rvert \leq 180$	$\Delta f_R \geq +0.01$	Correct
$180 < \lvert CCR_i - \overline{CCR_S}\rvert \leq 360$	$+0.01 > \Delta f_R \geq -0.04$	Acceptable
$360 < \lvert CCR_i - \overline{CCR_S}\rvert$	$\Delta f_R < -0.04$	Incorrect

Table 3. Road safety based on horizontal geometry consistency. Applied Criteria. $V_{85(i)}$ for an individual element. $V_{85(i)}$ - $V_{85(i+1)}$ between consecutive elements.

For a better understanding of the geometrical parameters considered, the reader is referred to Fig. 3.

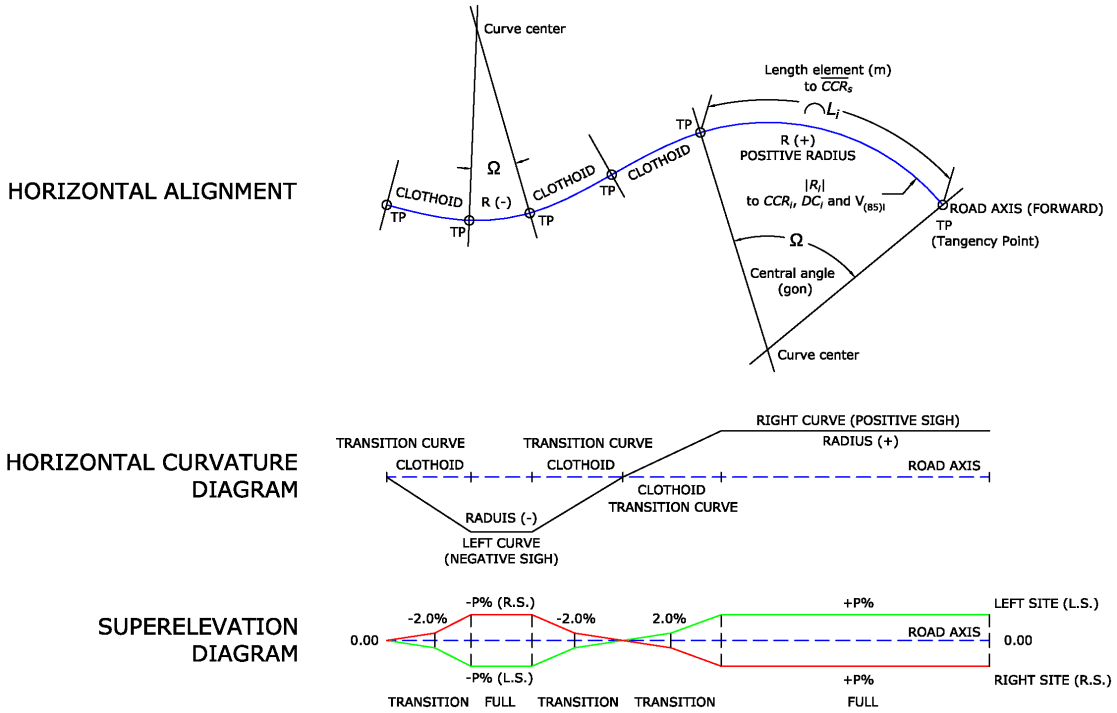


Fig. 3. Parameterization scheme considered for the formulation.

2.3.3 Step 3. Potential Risk Assessment by decision tree

This step is crucial in the proposed methodology. The potential risk assessment (PRA) of the road stretch is determined by means of a triple stability criterion because of the influence that the lack of consistency has on the increase in the accident rate on roads (Lamm et al., 2001). This triple criterion comprises geometric consistency indexes that will define the three attributes/variables of inductive process. To achieve this, a data mining process is performed based on geometrical parameters exclusively, which is supported by the decision tree (DT) inductive algorithm.

DTs, also known as identification trees, are among the nonparametric methods more widely applied to supervise inductive learning (Soler Flores, 2014). Moreover, the implementation of an approach by DT has the advantage that it does not require prior probabilistic knowledge of the study phenomena (de Oña et al., 2013). In this case and according to Information Theory (Quinlan, 1996), a process for categorizing the analysed attributes/variables is performed.

In this framework, the key is to establish a classification model that minimizes the uncertainty regarding the risk predictions. To this end, the uncertainty for the information content of a discrete and random variable *X*, or self-information, can be adequately measured by the entropy function *H(X)*, as an appropriate indicator of the associated average uncertainty of a process (Cover and Thomas, 1991; Molina et al., 2016; Pearl, 1988), which is expressed as:

$$H(X) = -\sum_{x} p(x) \cdot log_2 p(x) \tag{9}$$

where *P* is the probability mass function of *X*. The entropy measure enables an assessment of the additional information required to specify a particular alternative (Barton et al., 2008), and therefore, reducing *H(X)* by acquiring information is interpreted as reducing the uncertainty regarding *X* (Molina and Zazo, 2018; Molina et al., 2016).

On the other hand, to build a DT, it is necessary: (i) to establish a node sequences using attributes/variables (Oña et al., 2013), in this case the three stability criterion and (ii) a node splitting criterion to form a tree. This latter condition leads on the one hand, to reduce complexity through removing the sections that provide little power to classify instances (Galathiya et al., 2012), and on the other hand to reduce classification errors, due to specialization in the training set (Körting, 2006). In this manner, node uncertainties are reduced (Singh and Gupta, 2014), overfitting phenomenon is avoided (Breiman et al., 2017; Kang and Choi, 2000) and a better predictions are achieved (Galathiya et al., 2012).

Overfitting the training data is a negative phenomenon of machine learning process as consequence of an excessive adaptation of the algorithm to the training data (Chicco, 2017; Domingos, 2012). This leads to erroneous classifications on unseen data, although DT may correctly perform on the training data (Kang and Choi, 2000). Here, as splitting or pruning criteria, it is essentially applied information gain, which is defined by the difference between the entropy of the node before (parent) and after (child) splitting respectively (de Oña et al., 2013; Singh and Gupta, 2014). It is worth highlighting that the highest information gain involves the highest reduction in entropy (Zhang et al., 2004).

In essence, this general framework (based on node sequences and splitting criterion) define the decision rules by which a DT, by itself and automatically, decides data split and draws its boundaries. For a complete background on DT theoretical construction

process, please refer to benchmark works such as Breiman et al., (2017) and Quinlan (1993).

Alternatively, the inductive reasoning process is performed by open-source freeware WEKA data mining through the iterative J4.8 algorithm (de Oña et al., 2013; WEKA, 2018; Witten and Frank, 2005). J4.8 is WEKA's implementation of a decision tree learner (Witten and Frank, 2005), which is based on C4.5 algorithm (Al-Turaiki et al., 2016; Quinlan, 1993). Please note that these algorithms are inspired by entropy (de Oña et al., 2013), and the gain ratio "normalizes" the information gain (Quinlan, 1993), and besides, the overfitting is avoided by post-pruning process after the tree-creation, because of this is more effective method than pre-pruning to address overfitting problems; in this way the training data are suitably classified (Kang and Choi, 2000). J4.8 algorithm divides the dataset according to the best informative attribute/variable, selecting in every iteration the attribute/variable with the maximum gain ratio or highest reduction in entropy (Al-Turaiki et al., 2016; de Oña et al., 2013; Witten and Frank, 2005). This classification approach has easily interpretable results and comparable accuracy to other classification models as its main advantages (Al-Turaiki et al., 2016; Quinlan, 1993).

Finally, the inductive process of DT is exclusively trained with the circular alignments of the first three case studies (please see section 3.3; Fig. 6). Previously, these circular alignments were categorized by a road safety expert into three level risks (high, medium and low). It is worth to highlight that safety expert provides the reference data that will be used into the DT training.

2.3.4 Step 4. Verification of results

A twofold process of verification, both in terms of geometric results and road safety is carried out. First geometrically through "as-built" horizontal alignment and secondly by risk validation, both provided by an expert road surveyor using manual delineation and design works from airborne images. Note the use of term "as-built", because it is very common that some changes can affect the original road design project during the construction phase (e.g. unpredicted and specific terrain conditions, etc.).

Therefore, we evaluate the accuracy of the geometry by comparing the horizontal alignment detected by the RANSAC algorithm in this work with the as-built horizontal alignment obtained manually by an expert road surveyor.

Secondly, risk validation is performed via a road safety expert, who also discretizes the risk levels of the circular alignments of fourth case study according to three levels (high, medium, low). In order to validate the achieved risk levels through the DT inductive process they will be compared with the safety expert ground truth.

On the other hand, DT process verification is carried out by means of: (1) Overall Accuracy (*OA*) and (2) Kappa concordance coefficient (*K*). *OA* is defined as the probability that an instance will be correctly classified according to the following expression:

$$OA_{(\%)} = \left(\frac{TP + TN}{N}\right) \cdot 100 \tag{10}$$

where *TP* and *TN* and the true positives and true negatives respectively, and *N* is the total number of instances considered. For its part, *K* coefficient is a statistic that measures pairwise agreement between a set of categorized data, correcting for expected chance agreement (Carletta, 1996; Garcia-Rodenas et al., 2017). The kappa coefficient is expressed as:

$$K = \frac{P(A) - P(E)}{1 - P(E)} \tag{11}$$

where *P(A)* is the observed concordance proportion and *P(E)* is the expected concordance proportion. In this equation, the numerator is the observed proportion, while the denominator is the maximum value that the numerator can take. *K* is defined in the range [-1, 1]. *K*=1 is produced only when there exists concordance in 100% of the observations. *K* = 0 implies no agreement. Negative values indicate no agreement, but they are unlikely in practice.

This final step provides a full comprehensive reliability assessment, for both the defined geometric algorithms and the methodology developed as the decision-making process on road safety.

3. Results

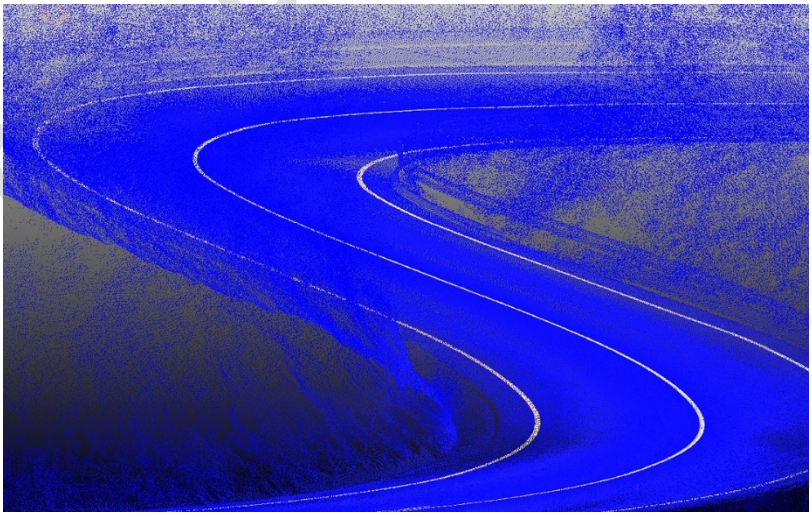
3.1 Road segmentation

The effectiveness of the developed algorithm for the classification of the data acquired by MLS is shown in Table 4 and Fig. 4. The high level of reduction obtained from the original 3D point cloud acquired (indicated in points m^{-2}) can be observed. In particular,

425 a first classification of 2D points (expressed in points per metre or spacing m^{-1})
426 corresponding to road marks is obtained, which represents approximately 1% of the
427 original 3D point cloud. Next, a second classification applied over road marks is applied
428 to extract only the road axis that represents between 4.8% and 8.7% of the road mark
429 points with a spacing between points of 10 cm. As commented in section 2.3.1, to increase
430 the quality of classification process an outliers filter is applied to the point cloud, in this
431 case Statistical Outlier Removal (SOR) filter (PCL, 2018). The SOR parameters were a
432 20 points neighbourhood and a standard deviation multiplier threshold of 1.0. These
433 values were determined on the basis of empirical tests. The high spacing achieved in the
434 classification of points has allowed a more efficient development of the subsequent
435 process of automatic extraction of the route (Fig. 2). This road classification approach has
436 allowed us to apply a more efficient process in the automatic extraction of the horizontal
437 alignment (Fig. 2).

Case study	3D Initial Point cloud		2D Road mark clasification		Road axis		
	Points	Density (point m^{-2})	Points	(%)	Points	Spacing (point m^{-1})	(%)
LU-722. Stretch 1	27,017,955	≈ 121	320,858	1.2	17,782	≈ 10	5.5
LU-722. Stretch 2	43,705,509	≈ 110	447,295	1.0	31,624	≈ 10	7.1
N-640	31,017,623	≈ 35	216,546	0.7	18,894	≈ 9	8.7
LU-722. Stretch 3	18,373,715	≈ 185	297,913	1.6	14,293	≈ 10	4.8

438 Table 4. Summary of road axis segmentation.



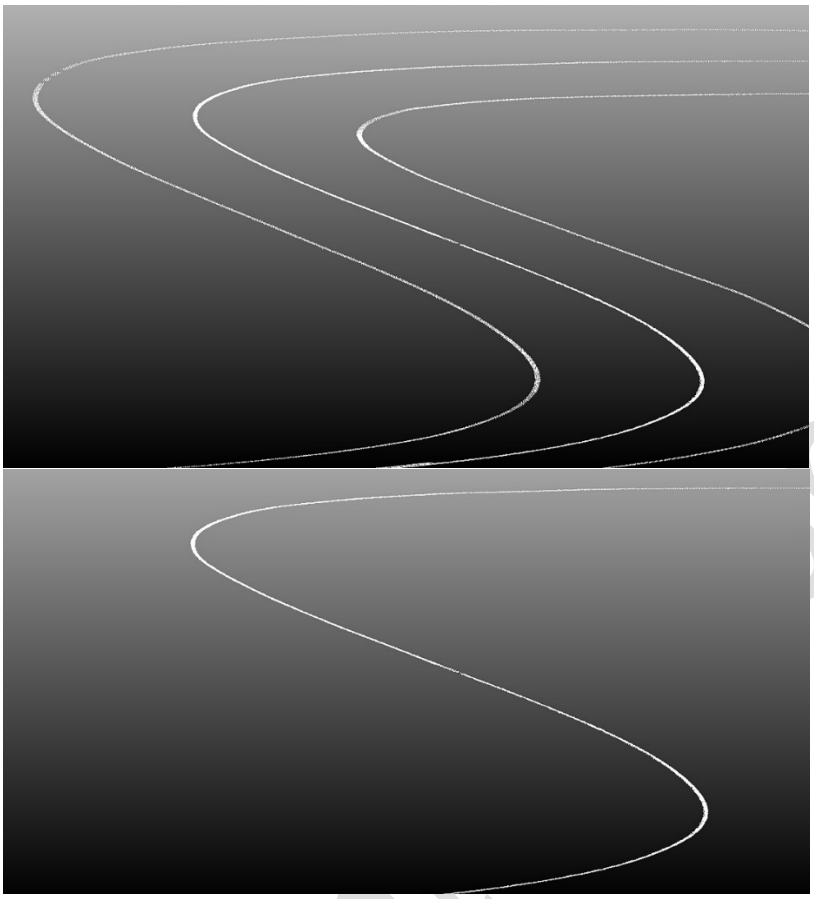


Fig. 4. Scheme resulting from the process of classification using the algorithm developed.

3.2. Geometric verification

This step involves the first verification of the proposed methodology by as-built road projects, which were considered as "ground truth." Applying Eq. (1) to Eq. (8) (see Step 2.3.2), the geometric consistency indexes and values of three joint criteria are calculated. The following tables (Tables 5 through 7) and Fig. 5 summarize the main results obtained, which are strictly focused on circular alignments.

Ground truth **General Geometric parameters**				**Obtained values by the RANSAC algorithm** **General Geometric parameters**			
Length: 1,776.60 m **Maximum/Minimum radius:** 177.40 / 29.08 (m) **Number of curves:** 14 $\overline{CCR_S}$: 789 (gon/km) V_D: 68.5 ≈ 70 km/h				**Length:** 1,776.35 m **Maximum/Minimum radius:** 180.32 / 29.78 (m) **Number of curves:** 14 $\overline{CCR_S}$: 788 (gon/km) V_D: 68.6 ≈ 70 km/h			
Horizontal alignment. Circular arc				**Horizontal alignment. Circular arc**			
Curve	**GC [2] Indexes**			**Curve**	**GC [2] Indexes**		
***R* [3] / Ω [4]**	***CCRi***	***DCi***	***V85***	***R* [3] / Ω [4]**	***CCRi***	**DCi**	***V85***
91.45 / 33.19	697	63	66	91.67 / 39.56	695	63	66
127.64 / 70.18	499	45	74	127.81 / 73.65	498	45	74
115.50 / 33.86	552	50	72	120.19 / 26.47	530	48	72
69.99 / 51.86	1044	94	56	75.72 / 56.30	841	76	62
115.00 / 35.01	554	50	72	174.75 / 13.84	365	33	79
170.00 / 79.39	375	34	79	170.84 / 83.43	373	34	79
31.50 / 59.69	2022	182	33	29.78 / 55.49	2139	192	31
109.00 / 74.78	584	53	70	109.89 / 74.09	580	52	71
29.08 / 58.44	2191	197	31	30.17 / 73.72	2111	190	32

29.92 / 140.84	2129	192	31	29.94 / 152.75	2128	191	32
41.84 / 20.34	1522	137	43	41.22 / 32.37	1545	139	43
49.56 / 75.17	1285	116	49	49.77 / 79.21	1280	115	49
29.60 / 90.41	2152	194	31	30.04 / 124.70	2120	191	32
177.40 / 69.17	359	32	80	180.32 / 88.22	353	32	79

Table 5. Geometric verification. LU-722 stretch 1 (Case study 1). (2) GC: Geometric consistency. (3) R: Absolute value of radius (m). (4) **Ω**: Central angle (gon). Please note that central angle refers to the azimuth variation between the ends of the circular alignment exclusively.

Ground truth **General Geometric parameters**				**Obtained values by the RANSAC algorithm** **General Geometric parameters**			
Length: 3,160.62 m **Maximum/Minimum radius:** 852.60 / 30.00 (m) **Number of curves:** 23 $\overline{CCR_S}$: 713 (gon/km) V_D: 71.5 ≈ 70 km/h				**Length:** 3,160.73 m **Maximum/Minimum radius:** 947.42 / 30.25 (m) **Number of curves:** 23 $\overline{CCR_S}$: 737 (gon/km) V_D: 70.6 ≈ 70 km/h			
Horizontal alignment. Circular arc				**Horizontal alignment. Circular arc**			
Curve	**GC (2) Indexes**			**Curve**	**GC (2) Indexes**		
***R* (3) / Ω (4)**	***CCRi***	***DCi***	***V85***	***R* (3) / Ω (4)**	***CCRi***	***DCi***	***V85***
367.60 / 10.37	156	16	87	385.38 / 11.05	165	15	88
40.30 / 14.65	1422	142	42	45.21 / 29.57	1409	127	46
309.90 / 18.74	185	18	86	310.53 / 23.57	205	18	86
30.00 / 97.42	1910	191	32	30.25 / 128.21	2106	189	32
852.60 / 5.11	67	7	92	947.92 / 5.57	67	6	92
81.25 / 23.50	705	71	63	82.70 / 23.08	770	69	64
43.93 / 102.45	1304	130	45	44.02 / 119.92	1447	130	45
178.00 / 30.84	322	32	79	177.44 / 43.04	359	32	79
798.50 / 8.64	72	7	92	900.44 / 7.92	71	6	92
150.60 / 61.83	380	38	77	151.00 / 60.70	422	38	77
99.30 / 32.87	577	58	68	103.31 / 38.81	617	55	70
51.50 / 84.92	1113	111	50	51.41 / 94.67	1239	111	50
66.39 / 37.04	863	86	58	62.16 / 33.28	1025	92	56
113.05 / 34.53	507	51	71	98.89 / 40.54	644	58	68
73.64 / 27.99	778	78	61	75.02 / 28.84	849	76	62
40.45 / 43.75	1417	142	42	42.13 / 51.35	1512	136	43
59.80 / 24.37	958	96	55	60.22 / 42.26	1058	95	55
44.57 / 38.17	1286	129	45	45.63 / 78.09	1396	126	46
217.14 / 15.59	264	26	82	246.63/ 16.77	258	23	84
45.00 / 21.68	1273	127	46	40.36 / 23.64	1578	142	42
30.40 / 63.97	1885	188	32	33.11 / 71.03	1924	173	35
56.60 / 55.75	1012	101	53	56.29 / 85.90	1132	102	53
104.38 / 21.09	549	55	70	105.08 / 30.89	606	55	70

Table 6. Geometric verification. LU-722 stretch 2 (Case study 2).

Ground truth **General Geometric parameters**				**Obtained values by the RANSAC algorithm** **General Geometric parameters**			
Length: 2,149.41 m **Maximum/Minimum radius:** 599.33 / 171.27 (m) **Number of curves:** 3 $\overline{CCR_S}$: 183 (gon/km) V_D: 102.7 ≈ 100 km/h				**Length:** 2,149.55 m **Maximum/Minimum radius:** 601.20 / 171.49 (m) **Number of curves:** 3 $\overline{CCR_S}$: 181 (gon/km) V_D: 102.9 ≈ 100 km/h			
Horizontal alignment. Circular arc				**Horizontal alignment. Circular arc**			
Curve	**GC (2) Indexes**			**Curve**	**GC (2) Indexes**		
***R* (3) / Ω (4)**	***CCRi***	***DCi***	***V85***	***R* (3) / Ω (4)**	***CCRi***	***DCi***	***V85***
368.51 / 45.18	173	16	87	367.07 / 56.09	174	16	87
171.27 / 36.61	372	33	79	171.49 / 35.20	371	33	79

599.33 / 22.12	106	10	90	601.20 / 22.53	106	10	90

453 Table 7. Geometric verification. N-640 (Case study 3).

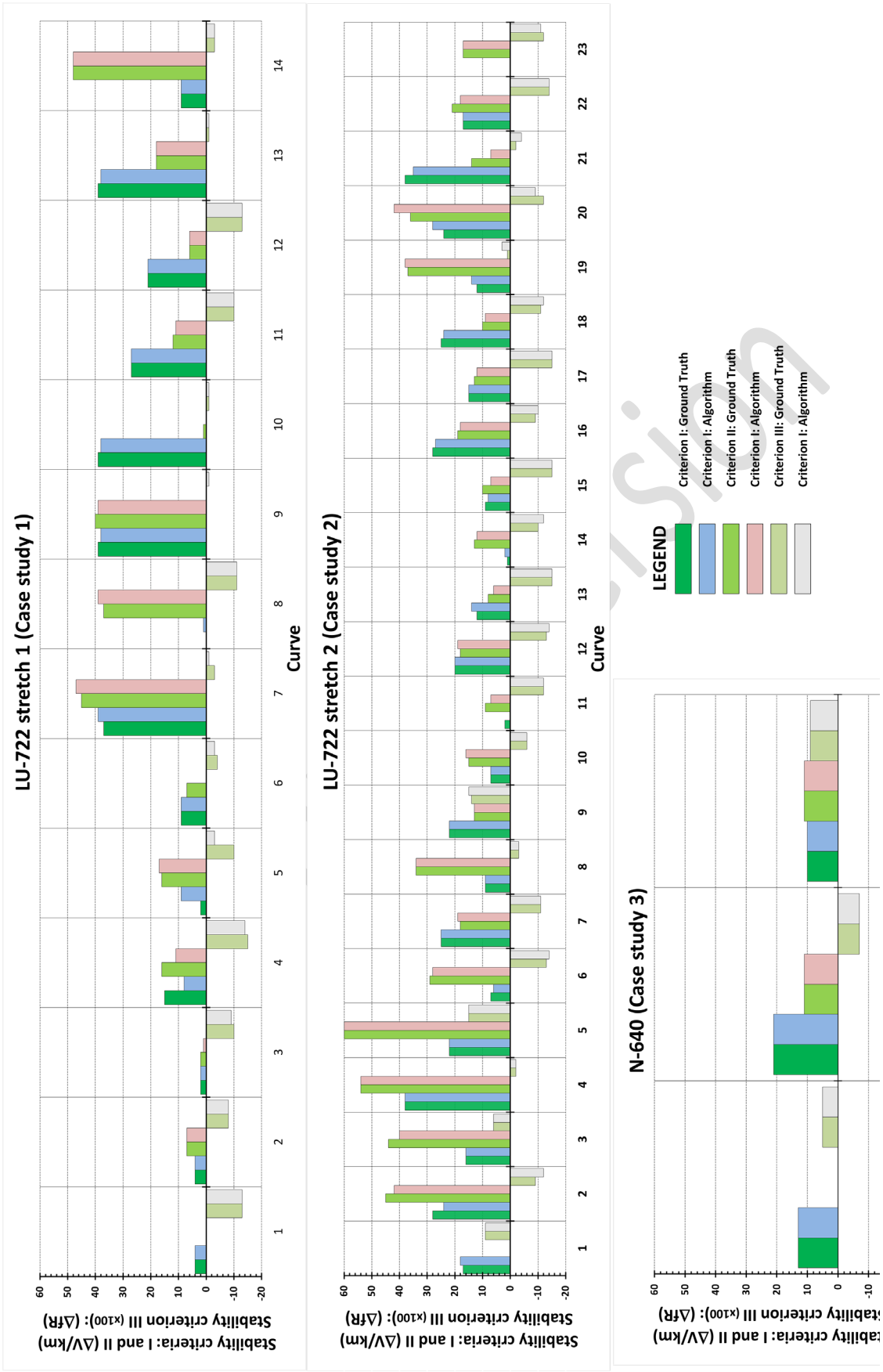


Fig. 5. Stability Criteria. Ground truth versus algorithm developed.

Generally, the main geometric parameters (R and CCR_i), did not present significant discrepancies, exhibiting discrepancies of approximately 3.6% and 4.9%, respectively. From a quantitative point of view, these discrepancies are the average absolute value of the percentage of variation of the analysed geometrical parameter with respect to the ground truth (in these cases as-built projects). In this sense, it is worth mentioning that the minimum absolute discrepancies obtained for the lengths of the sections were between 0.11 m and 0.25 m, for a total of 7,086.6 m, as well as a practical coincidence for the global parameter $\overline{CCR_S}$ (Tables 5 and 7). In addition, the correct identification of the total number of circular alignments for each study case should be noted.

Alternatively, the maximum radius (R_{max}) and the minimum radius (R_{min}) exhibit values similar to those obtained during the verification, with relative discrepancies of 4% for R_{max} and 1% for R_{min}, and with absolute discrepancies between 1.87 m and 2.92 m for R_{max} and between 0.22 m and 0.70 m for R_{min}. Although an absolute maximum discrepancy between radiuses of 94.82 m was observed (852.60 m versus 947.42 m), this type of discrepancy is common in the case of circular alignments with central angles lower than 6 gons (MFOM, 2016). In addition, this absolute difference has not provided significant changes in the geometric consistency indexes and thus in the stability criteria, as seen in Table 6. It is also necessary to outline the correct detection of the different circular curves and even circular curves linked consecutively.

In more detail and from a statistical point of view, for a maximum relative discrepancy of 10%, the success rates for the geometric parameters R, CCR_i and DC_i were 89%, 78% and 87%, respectively; even if the admissible relative discrepancy is reduced by up to 5%, the success rates for each parameter could be perfectly acceptable, reaching success rates of 81%, 67% and 78%, respectively. In contrast, if the maximum relative discrepancy is fixed at 15%, the success rates increase to 98%, 91% and 95%, respectively. In particular, analysed parameter success rate was calculated as the average percentage of the values that comply with the fixed threshold discrepancy.

3.3. Risk validation

To establish the relationships, thresholds and hierarchy among the different criteria established, as well as the expert classification, an inductive learning process based on a decision tree was applied, as described in subsection 2.3.3. Forty (40) instances were used

to train the DT (total number of circular alignments for the first three case studies considered). Fig. 6 shows the results obtained.

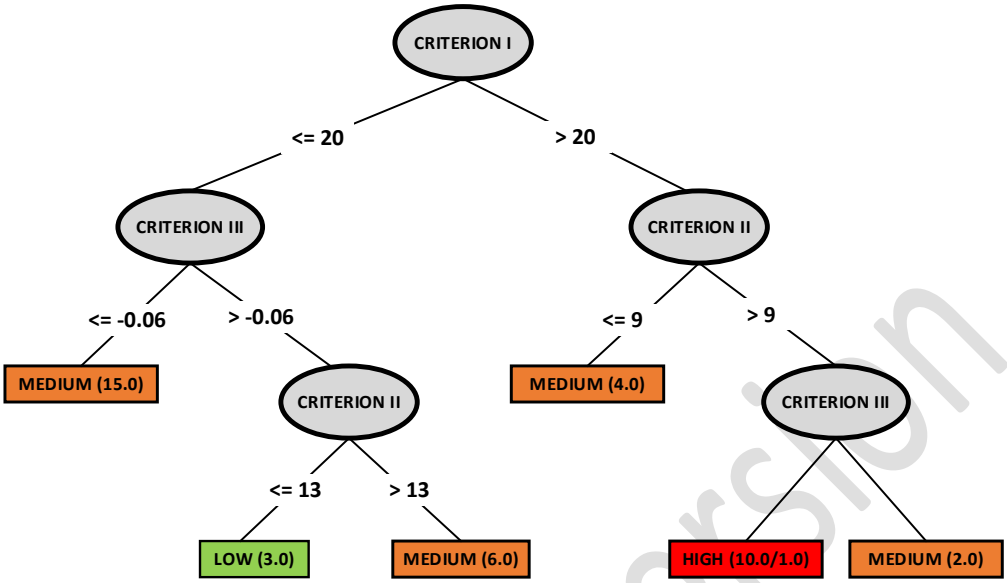


Fig. 6. Decision Tree. Hierarchical structure developed. K = 0.946. Note: (A) number format indicates the total number of instances that reached the leaf. In the (A/B) format case, B is the total number of misclassified instances.

The DT automatically establishes the discretization of attributes/variables (in this case the three stability criterion) based on the highest information gain (highest reduction in entropy) as a node splitting criterion to form a tree. In particular, determining the information gain and according to the gain ratio "normalizes," Criterion I of stability was identified as the most determining of the three criteria considered (gain = 0.415 bits). This criterion is the one that provides more information about the process and therefore the central node (root node) of the tree.

From the identified central node (Criterion I), the instances are split into child nodes, and recursively in each iteration, the attribute/variable with the maximum gain ratio is selected as node splitting criterion. It should be noted that, in the next levels, both criteria II and III are applied, which implies that there is not a significant entropy improvement between them. Alternatively, the final tree may not necessarily be symmetric because of the application of pruning algorithms that reduce the complexity of the tree, keeping the final accuracy. By the post-pruning process the tree subsections that do not improve the classification results are removed. In this sense, it is worth to highlight that a tree subsection is reconverted to a leaf, whose output is defined only if this operation does not get worsen the prediction accuracy. In the case of J4.8 algorithm this is done in post-

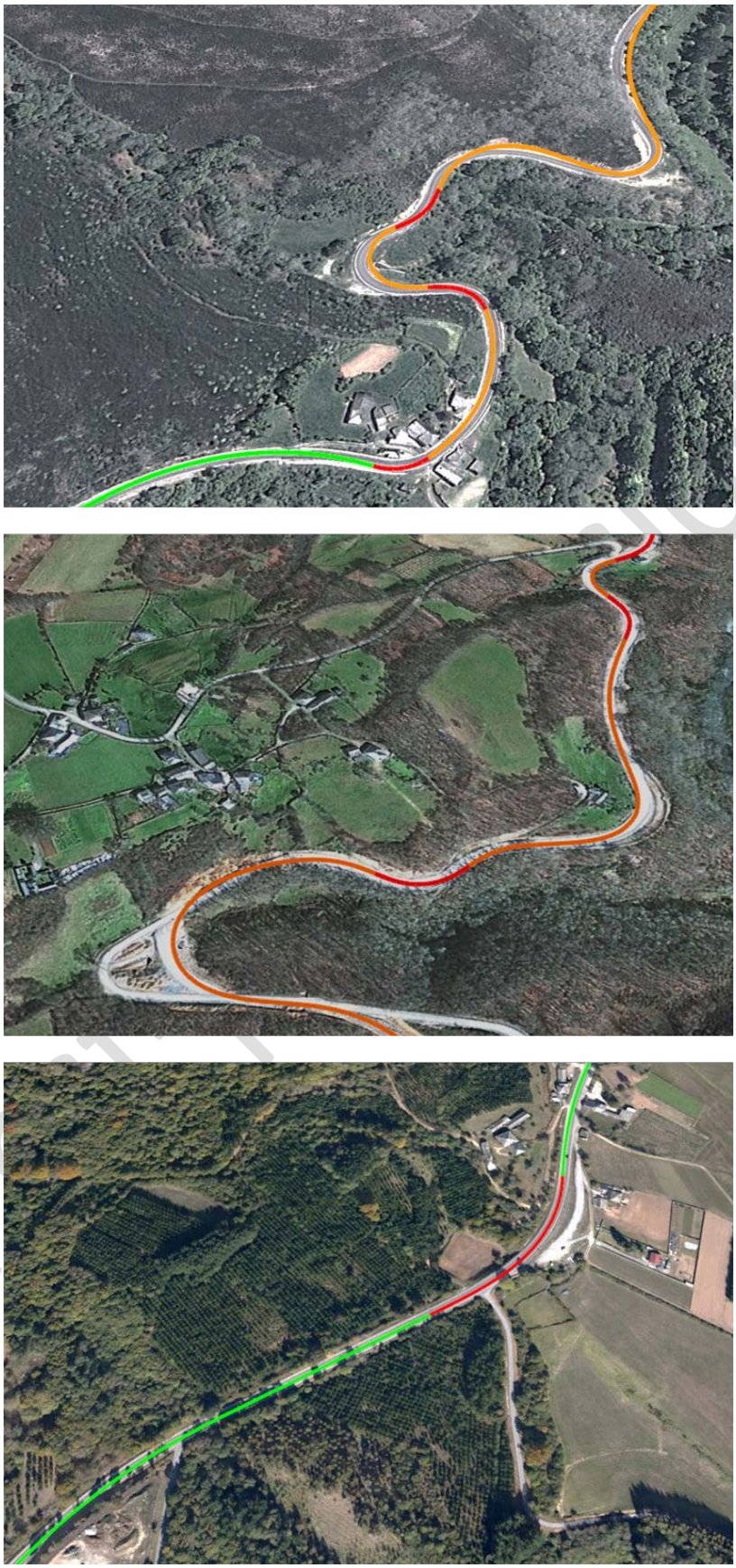
processing, once the final structure has been obtained, the full training data is back-fitted against the structure.

The final kappa coefficient obtained from the trained DT was 0.946, which entails a high degree of agreement. Then, this trained DT was applied to circular alignments of fourth case study, as a form of validating the classification obtained through the DT. The achieved overall accuracy (*OA*) and kappa coefficient (*K*) were 90.9% and 0.8553 respectively. These values show a high degree of agreement, and they validate the DT and its geometrical risk classification results. Table 8 shows confusion matrix results achieved. In this sense, the DT provided a suitable parameterization of the expert informal knowledge for road safety inspections based on the geometric parameters obtained.

		Classification (DT process)		
		Low	**Medium**	**High**
Ground truth (Road safety expert)	**Low**	2	0	0
	Medium	1	5	0
	High	0	0	3

Table 8. Risk validation results. LU-722 (Case study 4).

Fig. 7 shows the evaluation risk resulting from the DT and corresponding to each one of the circular elements that define the horizontal alignment. It should be noted that the transition curves were categorized according to the risk level resulting from the circular elements that compose these curves.



523 Fig. 7. Risk mapping for the different geometric elements of the horizontal alignment. Case study 1 LU-
524 722 Stretch 1. Red line: High risk. Orange line: Medium risk. Green line: Low risk.

4. Discussion and Conclusions

The main contribution of this research, related to remote sensing, is to show the application possibilities that the MLS technology can offer to road infrastructure risk assessment by means of: (i) an automatic risk mapping of the road based on geometrical consistency indexes and (ii) accurate stability criteria derived from these indexes. This is done through the integral approach presented, which combines geometry and risk, using 3D MLS point clouds. This approach represents a novel method of evaluating security in roads for the engineering community and/or road managers.

Even though the achieved quality by the alternative approaches to MLS outlined in the introduction (e.g. airborne laser scanner or aerial imagery) is highly promising (around 90% and above), these quality values are not yet comparable with those values required in road projects (centimetric accuracy). Therefore, datasets coming from MLS are the suitable source for obtaining the horizontal road alignment from which to derive the geometric consistency indices.

Ultimately, a high degree of reliability was obtained in the extraction of the geometrical elements of the road horizontal alignment. This is evident considering the total number of circular elements detected, as well as the low discrepancies obtained for the main geometric parameters (R and CCR_i). The minimum differences observed in the geometric consistency indexes and in the stability criteria are also noteworthy.

The methodology presented in this paper for the evaluation of road safety, through the twofold approach (geometry plus risk) and supported by artificial intelligence techniques (inductive process through a decision tree), can suitably complement other safety approaches, such as EuroRAP, which is based on a statistical-qualitative approach, sensitive to the distorting effect of road accidents. In this sense, the methodology developed allows us to objectify the intrinsic risk that the geometric design confers to the road. Furthermore, the scalability of this approach is also a significant and additional advantage, since it could be applied to other road typologies or countries by means of adapting the formulations and the specific models to the road design standards in each country.

On the other hand, the majority of studies based on road safety are focused on external causes to the road geometry, such as probability of crashes, types of drivers, level of use

556 and existing conditions of the pavement surface, among others, or searching patterns that
557 can explain accident causes in a particular section of the road. However, the inherent risk
558 that the geometric parameters themselves provide have received less attention. In this
559 sense, it is worth to highlight that the exposed approach in this paper represents a novelty
560 regarding the manner to assess road safety from MLS data and supported by an inductive
561 reasoning process, based on a decision tree, which provides a potential risk assessment
562 based on geometric parameters exclusively.

563 In future research, an extension of the categorization and evaluation of the road risk by
564 DT would be desirable, incorporating those geometric constraints related to the vertical
565 alignment (vertical agreement, *Kv* parameter and real superelevation) and completing it
566 with the available sight distance. Furthermore, the idea of this approach is to offer a future
567 alternative to those roads where road marks are not available (situation that can exist in
568 Spanish secondary roads) or roads with road marks in bad state or partially removed.

569 Finally, and in the framework of prevention and roadway deficiency detection, the
570 exposed methodology together with the data acquisition from MLS can be an adequate
571 and effective tool for road safety inspections. Through this approach, and taking into
572 account the service life of road infrastructures, it would be possible to have a complete
573 risk evolution according to the changes that may have been necessary to make to the
574 original designs.

575 **Acknowledgements**

576 Authors thank especially, INSITU Ingeniería S.L. for the mobile laser scanner data and
577 the equipment used for this work. This research has been partially supported by the
578 INROAD project (TSI-100505-2016-019) Energy, Tourism and Digital Society Ministry
579 (National projects: Strategical action / Call 2016).

580 **References**

581 Ai, C., Tsai, Y.J., 2015. Critical Assessment of an Enhanced Traffic Sign Detection Method Using
582 Mobile LiDAR and INS Technologies. Journal of Transportation Engineering 141 (5), 04014096.

583 Alexander, C., Tansey, K., Kaduk, J., Holland, D., Tate, N.J., 2010. Backscatter coefficient as an
584 attribute for the classification of full-waveform airborne laser scanning data in urban areas. ISPRS
585 Journal of Photogrammetry and Remote Sensing 65 (5), 423-432.

Al-Turaiki, I., Alshahrani, M., Almutairi, T., 2016. Building predictive models for MERS-CoV infections using data mining techniques. Journal of Infection and Public Health 9 (6), 744-748.

Alshehhi, R., Marpu, P.R., 2017. Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. ISPRS Journal of Photogrammetry and Remote Sensing 126 245-260.

Andrasik, R., Bil, M., 2016. Efficient road geometry identification from digital vector data. Journal of Geographical Systems 18 (3), 249-264.

Arranz Justel, J.J., 2013. Diseño, optimización y análisis de sistemas basados en técnicas láser, para el modelado geométrico, registro y documentación, aplicados a entidades de interés patrimonial. PhD. Thesis. Technical University of Madrid, Madrid, Spain, 499 pp.

Azimi, S.M., Fischer, P., Körner, M., Reinartz, P., 2018. Aerial LaneNet: Lane Marking Semantic Segmentation in Aerial Imagery using Wavelet-Enhanced Cost-sensitive Symmetric Fully Convolutional Neural Networks. arXiv preprint arXiv:1803.06904.

Barton, D.N., Saloranta, T., Moe, S.J., Eggestad, H.O., Kuikka, S., 2008. Bayesian belief networks as a meta-modelling tool in integrated river basin management - Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. Ecological Economics 66 (1), 91-104.

Bitenc, M., Lindenbergh, R., Khoshelham, K., van Waarden, A.P., 2011. Evaluation of a LIDAR Land-Based Mobile Mapping System for Monitoring Sandy Coasts. Remote Sensing 3 (7), 1472-1491.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and Regression trees, Chapman & Hall/CRC. Taylor & Francis Group, Washington, D.C., USA.

Cabo, C., Kukko, A., Garcia-Cortes, S., Kaartinen, H., Hyyppa, J., Ordonez, C., 2016. An Algorithm for Automatic Road Asphalt Edge Delineation from Mobile Laser Scanner Data Using the Line Clouds Concept. Remote Sensing 8 (9), 740.

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

Carletta, J., 1996. Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics 22 (2), 249-254.

Castro, M., Lopez-Cuervo, S., Parens-Gonzalez, M., de Santos-Berbel, C., 2016. LIDAR-based roadway and roadside modelling for sight distance studies. Survey Review 48 (350), 309-315.

Cavendish, J.C., Field, D.A., Frey, W.H., 1985. An apporach to automatic three-dimensional finite element mesh generation. International Journal for Numerical Methods in Engineering 21 (2), 329-347.

Chang, L., Chien, J., 2013. Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree model. Safety Science 51 (1), 17-22.

Chang, L., Wang, H., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Analysis and Prevention 38 (5), 1019-1027.

Chicco, D., 2017. Ten quick tips for machine learning in computational biology. Biodata Mining 10, 35.

Clode, S., Rottensteiner, F., Kootsookos, P.J., 2005. Improving city model determination by using road detection from lidar data, In: Joint Workshop of ISPRS and DAGM-CMRT05, Anonymous ISPRS, pp. 159-164.

Clode, S., Kootsookos, P.J., Rottensteiner, F., 2004. The automatic extraction of roads from LIDAR data, In: The International Society for Photogrammetry and Remote Sensing's Twentieth Annual Congress, Anonymous ISPRS, pp. 231-236.

Cover, T.M., Thomas, J.A., 1991. Chapter 1. Introduction and Preview, In: Elements of Information Theory, John Wiley & Sons, Hoboken, New Jersey, USA.

da Costa, J.O., Prudencio Jacques, M.A., Cunha Soares, F.E., Freitas, E.F., 2016. Integration of geometric consistency contributory factors in three-leg junctions collision prediction models of Portuguese two-lane national highways. Accident Analysis and Prevention 86, 59-67.

de Oña, J., Lopez, G., Abellan, J., 2013. Extracting decision rules from police accident reports through decision trees. Accident Analysis and Prevention 50, 1151-1160.

DGT, 2017. <http://www.dgt.es/es/prensa/notas-de-prensa/2016/20160509-dos-cada-tres-fallecidos-accidente-trafico-producen-carreteras-convencionales.shtml> (Accessed 10/01, 2017).

Diaz-Vilarino, L., Gonzalez-Jorge, H., Bueno, M., Arias, P., Puente, I., 2016. Automatic classification of urban pavements using mobile LiDAR data and roughness descriptors. Construction and Building Materials 102, 208-215.

Domingos, P., 2012. A Few Useful Things to Know About Machine Learning. Communications of the ACM 55 (10), 78-87.

Eftekharzadeh, S.F., Khodabakhshi, A., 2014. Safety evaluation of highway geometric design criteria in horizontal curves at downgrades. International Journal of Civil Engineering 12 (3), 326-332.

EU, 2008. Directive 2008/96/EC of the European Parliament and of the Council of 19 November 2008 on Road Infrastructure Safety Management European Commission.

EuroRAP, 2018. <http://www.eurorap.org/> (Accessed 01/22, 2018).

Fischler, M.A., Bolles, R.C., 1981. Random Sample Consensus - a Paradigm for Model-Fitting with Applications to Image-Analysis and Automated Cartography. Communications of the ACM 24 (6), 381-395.

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. Publication No: 99-171, Federal Highway Administration (FHWA), Georgetown Pike, USA.

Galathiya, A., Ganatra, A., Bhensdadia, C., 2012. Improved decision tree induction algorithm with feature selection, cross validation, model complexity and reduced error pruning. International Journal of Computer Science and Information Technologies 3 (2), 3427-3431.

Garach, L., de Ona, J., Lopez, G., Baena, L., 2016. Development of safety performance functions for Spanish two-lane rural highways on flat terrain. Accident Analysis and Prevention 95, 250-265.

Garcia-Rodenas, R., Lopez-Garcia, M.L., Teresa Sanchez-Rico, M., 2017. An Approach to Dynamical Classification of Daily Traffic Patterns. Computer-Aided Civil and Infrastructure Engineering 32 (3), 191-212.

Gargoum, S., El-Basyouny, K., 2017. Automated Extraction of Road Features using LiDAR Data: A Review of LiDAR applications in transportation (Contributed Paper). In: 4th International Conference on Transportation Information and Safety (ICTIS). 8–10 August 2017 Banff, Alberta, Canada.

Gonzalez-Jorge, H., Puente, I., Riveiro, B., Martinez-Sanchez, J., Arias, P., 2013. Automatic segmentation of road overpasses and detection of mortar efflorescence using mobile LiDAR data. Optics and Laser Technology 54, 353-361.

Hatger, C., Brenner, C., 2003. Extraction of road geometry parameters from laser scanning and existing databases. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 34 (3/W13), 225-230.

Holgado-Barco, A., Gonzalez-Aguilera, D., Arias-Sanchez, P., Martinez-Sanchez, J., 2015. Semiautomatic Extraction of Road Horizontal Alignment from a Mobile LiDAR System. Computer-Aided Civil and Infrastructure Engineering 30 (3), 217-228.

Holgado-Barco, A., Gonzalez-Aguilera, D., Arias-Sanchez, P., Martinez-Sanchez, J., 2014. An automated approach to vertical road characterisation using mobile LiDAR systems: Longitudinal profiles and cross-sections. ISPRS Journal of Photogrammetry and Remote Sensing 96, 28-37.

Isenburg, M., Liu, Y., Shewchuk, J., Snoeyink, J., 2006. Streaming computation of Delaunay triangulations. Acm Transactions on Graphics 25 (3), 1049-1056.

Javanmardi, M., Javanmardi, E., Gu, Y., Kamijo, S., 2017. Towards High-Definition 3D Urban Mapping: Road Feature-Based Registration of Mobile Mapping Systems and Aerial Imagery. Remote Sensing 9 (10), 975.

Jung, S., Qin, X., Oh, C., 2016. Improving strategic policies for pedestrian safety enhancement using classification tree modeling. Transportation Research Part A-Policy and Practice 85, 53-64.

Kang, B., Choi, K., 2000. Automatic Transliteration and Back-transliteration by Decision Tree Learning. Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC), 1135-1141.

Körting, T. S., 2006. C4. 5 algorithm and multivariate decision trees. Image Processing Division, National Institute for Space Research–INPE Sao Jose dos Campos–SP, Brazil.

Kumar, P., McElhinney, C.P., Lewis, P., McCarthy, T., 2014. Automated road markings extraction from mobile laser scanning data. International Journal of Applied Earth Observation and Geoinformation 32, 125-137.

Kumar, P., McElhinney, C.P., Lewis, P., McCarthy, T., 2013. An automated algorithm for extracting road edges from terrestrial mobile LiDAR data. ISPRS Journal of Photogrammetry and Remote Sensing 85, 44-55.

Kwon, O.H., Rhee, W., Yoon, Y., 2015. Application of classification algorithms for analysis of road safety risk factor dependencies. Accident Analysis and Prevention 75, 1-15.

Lamm, R., Wolhuter, K.M., Beck, A., Rusher, T., 2001. Introduction of a new approach to geometric design and road safety (Contributed Paper). In: 20th South African Transport Conference 'Meeting the Transport Challenges in Southern Africa'. 16 – 20 July 2001 South Africa.

Lamm, R., Psarianos, B., Mailaender, T., 1999. Highway design and traffic safety engineering handbook, MacGraw-Hill, New York, USA.

Lamm, R., Guenther, A.K., Choueiri, E.M., 1995. Safety module for highway geometric design. Transportation Research Record 1512 (9), 7-15.

Lamm, R., Choueiri, E., Mailaender, T., 1991. Side friction demand versus side friction assumed for curve design on two-lane rural highways. Transportation Research Record 1303, 11-21.

López, G., de Oña, J., 2017. Extracting crash patterns involving vulnerable users on two-lane rural highways. Securitas Vialis 9 (1), 1-13.

López, G., de Oña, J., Garach, L., Baena, L., 2016. Influence of deficiencies in traffic control devices in crashes on two-lane rural roads. Accident Analysis and Prevention 96, 130-139.

Marinelli, G., Bassani, M., Piras, M., Lingua, A.M., 2017. Mobile mapping systems and spatial data collection strategies assessment in the identification of horizontal alignment of highways. Transportation Research Part C-Emerging Technologies 79, 257-273.

Mc Elhinney, C., Kumar, P., Cahalane, C., McCarthy, T., 2010. Initial Results from European Road Safety Inspection (Eursi) Mobile Mapping Project. Proceedings of the ISPRS Commission V Mid-Term Symposium Close Range Image Measurement Techniques 38, 440-445.

MFOM. 2018. <https://www.fomento.gob.es> (Accessed 01/09, 2018).

MFOM. 2017. Mapa de Tráfico año 2016. Provincia de Lugo. Red de Carreteras del Estado.

MFOM. 2016. Ministerio de Fomento. Instrucción de Carreteras. Norma 3.1-IC.

MFOM. 1987. Ministerio de Fomento. Normas de Carreteras. 8.2-IC. Marca Viales.

Misaghi, P., Hassan, Y., 2005. Modeling operating speed and speed differential on two-lane rural roads. Journal of Transportation Engineering-ASCE 131 (6), 408-417.

Molina, J.-L., Zazo, S., 2018. Assessment of Temporally Conditioned Runoff Fractions in Unregulated Rivers. Journal of Hydrologic Engineering 23 (5), 04018015.

Molina, J.-L., Zazo, S., Rodriguez-Gonzalvez, P., Gonzalez-Aguilera, D., 2016. Innovative Analysis of Runoff Temporal Behavior through Bayesian Networks. Water 8 (11), 484.

Montella, A., Imbriani, L.L., 2015. Safety performance functions incorporating design consistency variables. Accident Analysis and Prevention 74, 133-144.

Morrall, J., Talarico, R., 1994. Side friction demanded and margins of safety on horizontal curves. Transportation Research Record 1435, 145.

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

PCL, 2018. <http://pointclouds.org/> (Accessed 07/15, 2018).

Pearl, J., 1988. Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, San Francisco, USA.

Pérez-Zuriaga, A.M., Camacho-Torregrosa, F.J., Garcia, A., 2013. Tangent-to-Curve Transition on Two-Lane Rural Roads Based on Continuous Speed Profiles. Journal of Transportation Engineering 139 (11), 1048-1057.

Puente, I., Gonzalez-Jorge, H., Martinez-Sanchez, J., Arias, P., 2013a. Review of mobile mapping and surveying technologies. Measurement 46 (7), 2127-2145.

Puente, I., Gonzalez-Jorge, H., Riveiro, B., Arias, P., 2013b. Accuracy verification of the Lynx Mobile Mapper system. Optics and Laser Technology 45, 578-586.

Quinlan, J.R., 1993. C4. 5: Programs for Machine Learning, Volume 1 of Morgan Kaufmann series in Machine Learning, Morgan Kaufmann, San Mateo, USA.

Quinlan, J.R., 1996. Learning decision tree classifiers. ACM Computing Surveys 28 (1), 71-72.

Rasdorf, W., Findley, D.J., Zegeer, C.V., Sundstrom, C.A., Hummer, J.E., 2012. Evaluation of GIS Applications for Horizontal Curve Data Collection. Journal of Computing in Civil Engineering 26 (2), 191-203.

Riveiro, B., Gonzalez-Jorge, H., Martinez-Sanchez, J., Diaz-Vilarino, L., Arias, P., 2015. Automatic detection of zebra crossings from mobile LiDAR data. Optics and Laser Technology 70, 63-70.

Singh, S., Gupta, P., 2014. Comparative study ID3, Cart and C4. 5 decision tree algorithm: a survey. International Journal of Advanced Information Science and Technology 3 (7), 97-103.

Siskind, V., Steinhardt, D., Sheehan, M., O'Connor, T., Hanks, H., 2011. Risk factors for fatal crashes in rural Australia. Accident Analysis and Prevention 43 (3), 1082-1088.

Sitran, A., Delhaye, E., Uccelli, I., 2016. Directive 2008/96/EC on road infrastructure safety management: an ex-post assessment 5 years after its adoption. Transport Research Arena Tra2016. 14, 3312-3321.

Soler Flores, F., 2014. Estimación de sucesos poco probables mediante redes bayesianas. PhD. Thesis. University of Castilla-La Mancha, Ciudad Real, Spain, 271 pp.

769 Sujatha, C., Selvathi, D., 2015. Connected component-based technique for automatic extraction
770 of road centerline in high resolution satellite images. Eurasip Journal on Image and Video
771 Processing 8.

772 Varela-Gonzalez, M., Gonzalez-Jorge, H., Riveiro, B., Arias, P., 2014. Automatic filtering of
773 vehicles from mobile LiDAR datasets. Measurement 53, 215-223.

774 WEKA, 2018. Weka 3: Data Mining Software in Java. Machine learning Group at the University
775 of Waikato. <https://www.cs.waikato.ac.nz/ml/weka/> (Accessed 01/10, 2018).

776 Witten, I.H., Frank, E., 2005. Data Mining. Practical machine learning tools and techniques.
777 Morgan Kaufmann, San Francisco, USA.

778 Xunta de Galicia, 2016. <http://civ.xunta.gal/seccion-
779 organizacion/c/CIV_Axencia_Galega_de_Infraestruturas?content=Direccion_Xeral_Infraestrutu
780 ras/Plan_aforos/seccion.html&std=plan-aforos.html> (Accessed 01/08, 2018).

781 Yang, B., Fang, L., Li, J., 2013. Semi-automated extraction and delineation of 3D roads of street
782 scene from mobile laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote
783 Sensing 79, 80-93.

784 Yan, L., Liu, H., Tan, J., Li, Z., Xie, H., Chen, C., 2016. Scan Line Based Road Marking
785 Extraction from Mobile LiDAR Point Clouds. Sensors 16 (6), 903.

786 You, K., Sun, L., Gu, W., 2012. Reliability-Based Risk Analysis of Roadway Horizontal Curves.
787 Journal of Transportation Engineering-Asce 138 (8), 1071-1081.

788 Zhang, L.V., Wong, S.L., King, O.D., Roth, F.P., 2004. Predicting co-complexed protein pairs
789 using genomic and proteomic data integration. Bmc Bioinformatics 5, 38.

790 Zhou, L., Vosselman, G., 2012. Mapping curbstones in airborne and mobile laser scanning data.
791 International Journal of Applied Earth Observation and Geoinformation 18 293-304.

792

793