**TRANSPORT**
ISSN 1648-4142 / eISSN 1648-3480
2014 Volume 29(4): 419–430
doi:10.3846/16484142.2014.984329

# A FUNCTIONAL ROAD CLASSIFICATION WITH DATA MINING TECHNIQUES

Antonino D'Andrea, Claudio Cappadona, Gianluca La Rosa, Orazio Pellegrino

*Dept of Civil Engineering, University of Messina, Messina, Italy*

Submitted 15 May 2013; resubmitted 27 August 2013; accepted 28 October 2013

**Abstract.** The current international road standards, in order to give organization and safety, promote the classification of roads according to their technical and functional characteristics beyond their administrative membership, but the procedures are yet strongly based on the expertise's judgment. In fact, although this activity has a great importance for the consequences that produces in terms of responsibility and allocation of economic resources, it is solely based on the quantification of some variables without specifying methods or analytical procedures. In this paper, after an instrumental survey of the road environment, we applied data mining techniques that consider the 'vagueness' of the analysed scenario. The type of algorithms used, therefore, permits to quantify a degree of membership (among 0 and 1) of a road to the groupings provided and to prepare any corrective action in order to direct the final result towards a specific class with greater precision. In addition, this method is very flexible and willing to contain new variables or observations at different times with great easiness. Moreover, the geographical location of the individual observations, as it was done also in this research, can be transferred to a GIS system, with a positive impact on maintenance programs.

**Keywords:** fuzzy; infrastructure; traffic safety; uncertainty; road.

## Introduction

The functional classification of the roads consists of grouping them into homogeneous classes depending on the type of service they provide within a defined geographical area and within a road network. It takes a major concern not only with regard to the creation of new infrastructures, but also to promote functional retrieval and/or the rehabilitation of existing routes instead of expensive new designs.

The international road standards, given the importance of the issue, lead analyst towards specific procedures, very similar from country to country. Usually, in fact, it is necessary to identify certain factors as the following:

- the type of movement (transit, distribution, penetration, access);
- the extent of travel;
- the assumed function in the environmental context;
- the allowed components of traffic.

One of the objectives of the road standards is to identify the various functional levels (primary – connectors, main – manifolds, secondary, local) in the network and match them to the roads, divided into several classes. In the Italian case, these are the following six groups:

- A – highways;
- B – main roads;
- C – secondary roads;
- D – urban roads;
- E – urban district roads;
- F – local roads.

In other countries the division into groups is very similar:

- in the United States and Canada – we find Freeways, Arterials, Collector and Local roads;
- in the UK – Motorway; Primary A-road; Non-primary A-road, B-road, C-road; Unclassified roads;
- in the France – *Autoroutes*, *Route Nationale*, *Départmentales Routes*, *Routes Communales*.

The selection of the class (from A to F, according to the Italian standard) derives from a complex process which comprises surveys of the road context and the following analysis of the collected data: this process, however, can also terminate with a judgment of un-classification, with the consequent necessity of adjustments in the road itself and the need to determine safety measures to be applied during the transition pe-

Corresponding author: Orazio Pellegrino
E-mail: *opellegrino@unime.it*

Copyright © 2014 Vilnius Gediminas Technical University (VGTU) Press
http://www.tandfonline.com/TRAN

Taylor & Francis
Taylor & Francis Group

riod. Although there are slight differences among the different countries (AASHTO 2011; Design Manual for Roads… 2002, FHWA 2004; Lamm *et al.* 1999; Italiano Strada Standard 2001; Bosurgi *et al.* 2011), the classes and performed functions always depend on the demand for drivers' mobility, safety conditions, territorial development expected, availability of different transport modes and environmental protection of the examined area (Kaptein, Claessens 1998; Jaarsma 1997; Giummarra 2003; Chen *et al.* 2008). Among the most complex elements to evaluate we have the traffic generating sites (i.e. the presence of administrative, cultural, economic, commercial, industrial centers) to which is assigned a hierarchy usually depending only on the level of experience of the analyst.

The functional classification is, therefore, one of the tasks of the designer and it is essential in the case of existing roads. A misuse of the road, compared with its proper characteristics, in fact, worsen the users' safety (Kockelman 2001; Karlaftis, Golias 2002; Forkenbrock, Foster 1997; Dickerson *et al.* 2000; Nowakowska 2010; Kashani, Mohaymany 2011) and most usually leads to an increase in social costs and possible inefficiencies in the programming of maintenance (Sohn, Lee 2003; Cafiso *et al.* 2011).

This activity is deceptively simple, because a road mutates easily its geometric and functional characteristics along its length, making it necessary to properly investigate the importance of the collected data. In this regard, the modern digital instruments allow us to perform measurements almost continuously, managing in an optimal way the sampling frequency with relatively low computational costs and ensuring a good reliability of the collected data. However, at the end of the on-site survey, it is possible that the equipment gives us a database with redundant or hidden data, so that useful information can be obtained only after appropriate filtering operations (Pellegrino 2011, 2012a, 2012b). It is necessary, therefore, to organize a procedure capable of acquiring knowledge from the detected data base, presumably from different sources (GPS, GPR, Speed Trap, image analysis, GIS, etc.) – in order to take into account the high degree of uncertainty of its variables (often correlated among them) – and reasonably quickly (Kuehnle, Burghout 1998; Mena 2003; Manduchi *et al.* 2005; Kang, Scott 2008; Paclík *et al.* 2000; Praticò, Giunta 2011). Finally, the evaluation of the collected data must be carried out using techniques that limit the analyst's role, so far predominant in the procedures that are suggested by the regulations (D'Andrea, Pellegrino 2012).

We can find the means to resolve these issues by using soft computing or artificial intelligence techniques such as fuzzy logic, neural networks and genetic algorithms (Dorsey, Coovert 2003; Cafiso *et al.* 2004). The soft computing approach, of course, is not always preferable to other methods, but it produces more realistic results when the number of variables involved is considerable and, especially, when their non linear dependence would render other techniques not applicable (Jang 1993; Güler, Übeyli 2005; Mon 2007; Gu, Oyadiji 2008; Tahmasebi, Hezarkhani 2010; Ravi, Zimmermann 2000; Shanahan *et al.* 2000; Dağdeviren *et al.* 2008). Soft Computing techniques have been applied with profit also for functional classification of roads. Especially in the last 10–15 years, researchers have used them to extract information from the road environment through image analysis, previously recorded with usual camcorders (Zhu *et al.* 1998). The concomitant development of the Intelligent Transportation System (ITS) has encouraged the preparation of fuzzy models or based on Genetic Algorithms (GA), able not only to recognize but also to understand the images with uncertainties controllable by analysis probabilistic (Shanahan *et al.* 2000; Lingras 2001). More recently, soft computing techniques have been applied to evaluate details of the functional classification as, for example, the recognition of the road alignment by means of Artificial Neural Networks (ANN) (García Balboa, Ariza López 2008) or to evaluate traffic flows (Toplak *et al.* 2010).

With this paper we propose the application of a data mining technique that, based on the knowledge of a data set previously collected, allows us to quantify the membership degree of a road to a given functional class, identifying the contribution of the structural, geometrical and functional elements involved in the analysis.

The proposed model, furthermore, has the advantage that it can be updated with additional variables without losing accuracy (but, rather, increasing it) in the calculation of final solutions. The achievement of this objective is fundamental because it would allow, both during the design of a new road and during the maintenance of an existing road, to act on well-identified characteristics of the scenario that often are not covered by the applied standard.

In order to quantitatively assess the contribution of the proposal, we present a case study of a country road located in Sicily (Italy).

## 1. Method

The procedure suggested by the majority of the international standards is very general and is based on the determination of certain parameters, useful for the recognition of the most opportune class. Amongst them, for example, there are type of movement, offered services, standards of construction, traffic components and route directions. Furthermore, these properties depend on other variables (flows, operating speeds and design, consistency, visibility, planning tools, etc.), often with interdependency between them, which makes this problem extremely complex to solve. Finally, a further element of difficulty is represented by the fact that some constructive and functional characteristics can vary along the track, especially if this is sufficiently extended.

Therefore, taking into account the conditions here briefly introduced, we proposed a methodology based on the following steps:

- survey of the following 25 variables for the definition of the expected class, recorded at appropriate points of the road axis: function in the network, served movement, transport capacity, design speed, operational speed, difference between de-

sign speed and operational speed, stopping sight distance, available sight distance, degree of conditioning between users, roughness of the pavement, radius of the curve, length of the curve, lane width, width of the traffic island, width of shoulders right and left, vertical and horizontal signs, safety barriers, traffic flows (light and heavy vehicles), attractors of traffic, traffic components, admissions, presence of emergency lanes and parking areas, separation of lane directions;
- the survey was performed with electronic equipment and produced a data base in which the rows represent the variables and the columns represent the observation;
- to the data base so formed a data mining technique has been applied, in order to obtain 'knowledge' and delete information less useful.

In particular, we have thought of applying an algorithm of fuzzy type, in order to detect 'borderline' situations – in which the group assignment is controversial. Moreover, the possibility that all the variables may direct the analyst to a clear decision is essentially remote. Our proposal is based, therefore, on a fuzzy clustering algorithm, which allowed to quantify the degree of membership of the road to every scheduled class.

In Table 1 we reported all the 25 variables that will serve to further analysis and their values that, ideally, should belong to 5 different classes: the first 4 (A, B, C and F) refer to the typology prescribed by Italian legislation for the roads. The fifth class is, instead, a condition of un-classification. Even if the designation of classes varies from country to country, the procedure is, however, of very general use.

Further below we give some details on the determination of certain variables more complicated to determine.

### 1.1. Visibility Distances and Design Speed

The effectively Available Sight Distance (ASD) was obtained using a commercial software (Civil Design® by Digicorp, *http://www.digicorpingegneria.com*), through the preparatory reconstruction of horizontal and vertical alignment of the road, including the identification of all the obstacles in the three dimensions of the space able to influence the vision of the driver (Bosurgi *et al.* 2010). The 3D analysis, as it is well known, is more precise than the traditional 2D analysis, because it allows considerations of the effective trajectory of the vision radius, following the altimetry of the road and the obstruction of the obstacles with their effective height.

Table 1. Variables used to classify the roads and their recommended values for belonging to the A, B, C, F (rural roads only) groups

| | CATEGORY GROUP | | A | B | C | F | NO |
|---|---|---|---|---|---|---|---|
| 1 | FUNCTIONAL NETWORK CLASSIFICATION | | State-wide | Regional | Provincial | Municipal | 0 |
| 2 | MOBILITY FUNCTION | | Connector | Collector | Penetration | Local | 0 |
| 3 | OFFERED SERVICE | Capacity | 7200 | 3600 | 1800 | 1200 | 400 |
| 4 | | *V*(design) | 115 | 95 | 80 | 70 | 30 |
| 5 | | *V*85 | 135 | 115 | 95 | 80 | 60 |
| 6 | | Stopping sight distance (offered–demanded) | 100 | 70 | 50 | 40 | 0 |
| 7 | | Interaction between drivers | 100 | 75 | 50 | 25 | 1 |
| 8 | | Comfort – roughness | 4 | 3 | 2 | 2 | 0 |
| 9 | STANDARD CONSTRUCTION | *V*85–*V*(design) | 10 | 15 | 20 | 25 | 40 |
| 10 | | Minimum radius | 339 | 178 | 118 | 45 | 40 |
| 11 | | Minimum length | 79.86 | 65.97 | 55.56 | 48.61 | 20.00 |
| 12 | | Lane – minimum width | 3.75 | 3.75 | 3.75 | 3.50 | 2.50 |
| 13 | | Traffic island – minimum width | 2.60 | 2.50 | 0.00 | 0.00 | 0.00 |
| 14 | | Left shoulder – minimum width | 0.70 | 0.50 | 0.00 | 0.00 | 0.00 |
| 15 | | Right shoulder – minimum width | 2.50 | 1.75 | 1.50 | 1.00 | 0.00 |
| 16 | | Horizontal sign | 4 | 4 | 3 | 3 | 0 |
| 17 | | Vertical sign | 4 | 4 | 4 | 4 | 0 |
| 18 | | Safety barriers | 4 | 4 | 4 | 4 | 0 |
| 19 | | Light vehicular flow | 1100 | 1000 | 600 | 450 | 100 |
| 20 | | Heavy vehicular flow | 110 | 100 | 60 | 45 | 10 |
| 21 | CATEGORY OF ATTRACTORS | | 1 | 2 | 3 | 4 | 0 |
| 22 | TRAFFIC COMPONENTS | | Restrict | Restrict | All eligible | All eligible | All eligible |
| 23 | ACCESS | | Not allowable | Not allowable | Allowable | Allowable | Allowable |
| 24 | EMERGENCY LANE AND STOPPING AREA | | No | Separate areas | Stopping area | Stopping area | No |
| 25 | SEPARATION OF DIRECTION | | Yes | Yes | No | No | No |

*Notes*: the fifth class (NO) is a condition of un-classification; this table will constitute a term of comparison for all future observations.

The calculation of the Stopping Sight Distance (SSD), as required by Italian Road Standard (Italiano Strada Standard 2001), is carried out by solving the equation below, substantially similar to those contained in many international standard:

$$D_A = D_1 + D_2 = \frac{V_0}{3.6}\cdot\tau - \frac{1}{3.6^2}\times \int_{V_0}^{V_1} \frac{V}{g\cdot\left(f_1(V)\pm\frac{1}{100}\right)+\frac{R_a(V)}{m}+r_0(V)}dV, \quad (1)$$

where: $D_1$ – distance covered over time τ [m]; $D_2$ – braking distance [m]; $V_0$ – design speed of the vehicle at the beginning of braking, equal to the design speed obtained punctually from the design speed diagram [km/h]; $V_1$ – final design speed of the vehicle, in which $V_1 = 0$ in case of stopping [km/h]; $i$ – longitudinal slope of road [%]; τ – total reaction time (perception, reflection, reaction and performance) [sec], Italian Road Standard (Italiano Strada Standard 2001) recommends to put this variable equal to $2.8–0.01\cdot V_0$; $g$ – gravitational acceleration [m/sec$^2$], equal to 9.8 m/sec$^2$; $m$ – mass of the vehicle [kg], Italian Road Standard (Italiano Strada Standard 2001) recommends to put this variable equal to 1250 kg; $R_a$ – aerodynamic resistance [N], Italian Road Standard (Italiano Strada Standard 2001) recommends to put this variable equal to $(2.61\cdot10–5\cdot V_0^2)\cdot m$; $f_l$ – longitudinal coefficient for braking. Italian standard recommends to infer this variable from literature; $r_0$ – unitary resistance to rolling, generally negligible [N/kg].

Knowledge of the geometry of the road has also permitted the valuation of the design speed $V_d$, obtained by following the indications provided by Italian Road Standard (Italiano Strada Standard 2001) – greatly similar to those contained in other countries' standards. In the case we studied, it is not important to report the course of $V_d$ on the whole map, but only on the curves. The speed on these elements can be obtained from the following expression, that links the radius $R$ [m], the inferred design speed $V_d$ [km/h], the coefficient of lateral friction (function of the $V_d$) and the cross-sectional slope of road $q$ [%]:

$$V_d = \sqrt{127\cdot R\cdot(f_t + q)}. \quad (2)$$

The previous formula is complicated by the fact that the variable $f_t$, in turn, depends from $V_d$.

### 1.2. *V*85 Determination

The campaign of survey was carried out during daylight hours, with dry and regular paving and good meteorological conditions. Of course, according to the procedure for this type of measurements (Lamm *et al.* 1999), the conditioning factors due to the traffic were not considered, nor did motorcycles, commercial or heavy vehicles, or cars with time spacing between them less than 5 sec take part in the analysis.

The speed was surveied by using a laser speed gun. This instrument, as it is known, measures the round-trip time for light to reach a vehicle and reflect back and shoots a very short burst of infrared laser light, then waiting for it to reflect off the vehicle.

For every place, the speeds of over 250 vehicles were recorded, obtaining from the analysis a minimal number not beneath 100 isolated vehicles. In total, more than 9000 passages were therefore acquired and all the data collected were processed in order to obtain the 85th percentile of the values of the speeds ($V$85). The analysis, as already stated, is based on the study of 45 cross sections.

For every analysed section, the main parameters of interest were assessed in order to reconstruct the distribution of the speeds and, therefore, determine the $V$85.

The density of relative frequency of the class $f_i$ was calculated with the well-known equation:

$$f_i = \frac{n_i}{n\cdot A}. \quad (3)$$

From the relative frequency, we can then pass to the cumulated frequency through the equation:

$$F_i = A\cdot\sum f_i. \quad (4)$$

In which $F_i$ is the frequency to which the vehicles travel at a lower or equal speed to the considered. In this way, it was possible to reconstruct the course of the distributions of the frequency in relation to the position of each section.

Referring to the Eqs (3) and (4) reported above:

- $n_i$ is the absolute frequency, that is the number of elements pertaining to the class;
- $n$ is the total number of the elements in the series of data;
- $A$ is the amplitude of the class.

The information so collected allowed the reconstruction of the distributions of the frequency in relation to the position of each section and the calculation of the representative value of the 85th percentile.

### 1.3. Classification of Other Characteristics of the Road Scenario

In order to complete the survey of the road environment, further variables (admissions, intersections, horizontal and vertical signs, barriers, roughness of the surface pavement, etc.) have been identified and classified in four categories, based on the analyst's judgment. Although this is a subjective determination, however, it was simply found or not the presence or the degree of effectiveness of the element under consideration and, therefore, we think that the difference with what is perceived by the user is negligible.

### 1.4. Brief Notes about Clustering

Generally, cluster analysis can be used to deal with a very wide range of problems, like classification, optimization, pattern recognition, prediction, decision support, especially when:

- the system is non-linear, time-variant or ill defined;
- the variables are continuous;

– a mathematical model is either too difficult or expansive to organize;
– there are too many or noisy inputs.

The objective of data mining or, more specifically, of cluster analysis in this paper is the classification of objects according to similarities among them, in order to organize data into groups. The main characteristic of these techniques is to detect the underlying structure in data, not only for classification and pattern recognition, but also for model reduction and optimization. In the recent past, some researchers (Jang 1993) have ascertained the convenience to pair the methods previously seen in order to maximize their benefits. For example, one of the most effective procedures, that is the fuzzy clustering approach, consists of a structure that permits objects to belong to several clusters simultaneously, with different degrees of membership. In this way, the analyst should avoid to classify in a too much simplistic way certain phenomena within a single category when they have, instead, common characters to several classes, albeit with different degrees of membership.

Clustering techniques are generally used to classify similar objects and, especially, to organize data in predetermined groups by identifying hidden structures in the source data (Abonyi, Feil 2007). The data submitted for analysis originate from physical observations or surveys and each observation consists of $n$ measured features, grouped into an $n$-dimensional vector:

$$x_k = \left[ x_{k1}, x_{k2}, ..., x_{kn} \right]^T, \; x_k \in R^n. \tag{5}$$

Therefore, a set of $N$ observations can be represented as a matrix $N \times n$:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nn} \end{bmatrix}. \tag{6}$$

A cluster is representative of similar elements with respect to other belonging to another cluster. This characteristic is measured in analytical way as the normal distance between the center of the cluster and the data that belong to it. In the traditional view of hard clustering, an element of the data set belongs only to one cluster without no possibility to belong to another. In the last years, the growth of soft computer techniques, as fuzzy logic, has permitted to propose methods in which an object can belong to a number of clusters c simultaneously, with different membership degrees between 0 and 1. Naturally, the sum of the different membership degrees related to the interested clusters must be equal to one. The structure of the partition matrix $U = [N \times c]$, then, is the following:

$$U = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1c} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{N1} & \mu_{N2} & \cdots & \mu_{Nc} \end{bmatrix}, \tag{7}$$

where: $c$ is the number of fuzzy subsets or of clusters.

This matrix $U=[\mu_{ik}]$ is subject to the following conditions:

$$\mu_{ik} \in \left[0,1\right], \; 1 \le i \le c, \; 1 \le k \le N; \tag{8}$$

$$\sum_{i=1}^{c} \mu_{ik} = 1, \; 1 \le k \le N; \tag{9}$$

$$0 < \sum_{k=1}^{N} \mu_{ik} < N, \; 1 \le i \le c. \tag{10}$$

With these premises, the fuzzy partitioning space for $X$ is represented by the set:

$$M = \left\{ U \in R^{c \times N} \left| \begin{array}{l} \mu_{ik} \in [0,1], \; \forall i, \; k; \\ \sum_{i=1}^{c} \mu_{ik} = 1, \; \forall k; \; 0 < \sum_{k=1}^{N} \mu_{ik} < N, \; \forall i \end{array} \right. \right\}. \tag{11}$$

The $i$th column of $U$ contains the values of the membership function of the $i$th fuzzy subset of $X$. With reference to the equation (5), the sum of each column is 1 and thus the total membership of each $x_k$ in $X$ equals one.

Given the more realistic compliance with the uncertain nature of the variables collected in this paper, we have used the algorithm called Fuzzy C-means instead of other methods referred to hard clustering. This procedure regards the minimization of an objective function, termed C-means functional, defined as:

$$J\left(X;U,V\right) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m \left\| x_k - v_i \right\|_A^2, \tag{12}$$

where: $V$ is the vector containing the centers of the clusters and $m$ is a weighting exponent >1 that influences the fuzziness of the results:

$$V = \left[ v_1, v_2, \ldots, v_c \right], \; v_i \in R^n. \tag{13}$$

The minimization of the C-means functional $J$ is a problem of nonlinear optimization, usually solved through a simple Picard iteration through the first-order conditions for stationary points of $J$ equation.

The stationary points are obtained by applying appropriate constraints (Eq. (5)) to $J$ through the Lagrange multipliers and the setting of the gradients $\left(\overline{J}\right)$ respect to $U$, $V$ and $\lambda$ to zero:

$$\overline{J}\left(X;U,V,\lambda\right) = \sum_{i=1}^{c} \sum_{k=1}^{N} \left(\mu_{ik}\right)^m D_{ikA}^2 + \sum_{k=1}^{N} \lambda_k \left( \sum_{i=1}^{c} \mu_{ik} - 1 \right), \tag{14}$$

where the following expression is the squared inner-product distance norm:

$$D_{ikA}^2 = \left\| x_k - v_i \right\|_A^2 = \left( x_k - v_i \right)^T A \left( x_k - v_i \right). \tag{15}$$

If

$$D_{ikA}^2 > 0, \forall i, k \text{ and } m > 1, \tag{16}$$

then:

$$\left(U, V\right) \in M \times R^{n \times c} \tag{17}$$

can minimize $J$ only if:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{D_{ikA}}{D_{jkA}}\right)^{\frac{2}{m-1}}}, \; 1 \le i \le c; 1 \le k \le N \tag{18}$$

and

$$v_i = \frac{\sum_{k=1}^{N} \mu_{ik}^m \, x_k}{\sum_{k=1}^{N} \mu_{ik}^m}, 1 \le i \le c \, . \tag{19}$$

The above equations can be easily automated within computing environments such as *Matlab*® (*http://www.mathworks.com*) or *Mathematica*® (*http://www.wolfram.com/mathematica*), greatly reducing the computational cost.

### 1.5. Feature Extraction

Successively, we have applied a features extraction technique to summarize the so classified data into a lower dimensional space, to remove second order dependencies and for displaying results in a more comfortable way. In order to map the high dimensional data point into a lower space, we have used the projection technique called Principal Component Analysis (PCA).

In particular, the PCA technique (Ravi *et al.* 2000) transforms a number of potential correlated variables into a predetermined number of uncorrelated variables (the principal components). The first component contains information regarding the variability in the data, while the other components (in general only one) consider the remaining variability as possible. The aim of this analysis is not only to find new significant variables but, furthermore, to reduce the dimensionality of the data set.

The algebraic solution is based on an important property of eigenvector decomposition.

In particular, the first principal component presents the same direction of the associated eigenvector with the largest eigenvalue. The direction of the second principal component is identified by the associated eigenvector with the second largest eigenvalue.

In order to reduce the size of the data set, its covariance matrix is defined in the following way:

$$F = \frac{1}{N-1}\sum_{k=1}^{N}\left(x_k - \bar{x}\right)\cdot\left(x_k - \bar{x}\right)^T, \tag{20}$$

where: $\bar{x}$ is the mean of the data:

$$\bar{x} = \frac{1}{N}\sum_{k=1}^{N} x_k \tag{21}$$

and $N$ is equal to the number of objects in the data set.

The projection of the data onto a hyper-plane is based on the first few $q$ nonzero eigenvalues and the corresponding eigenvectors of the following expression:

$$F \cdot U = U \cdot \Lambda, \tag{22}$$

where: $U$ is a $n \times n$ matrix that presents the unit lengths eigenvectors in its columns; $\Lambda$ is the diagonal matrix with the corresponding eigenvalues $\lambda_i, \ldots, \lambda_n$ along the diagonal.

The variance of the data set is:

$$\sigma^2 = \sum_{i=1}^{n}\lambda_i. \tag{23}$$

If the first 2 greatest eigenvalues are used to visualize the original high dimensional data, the sum of the remainder eigenvalues is lost. The eigenvectors are the principal components and the eigenvalues are the relative variances. Therefore:

$$y_k = U^T \cdot x_k \tag{24}$$

is the representation of the $k$th sample in the new basis and its approximation in the original space is:

$$\hat{x}_k = UU^T x_k. \tag{25}$$

This procedure is useful to decrease the size of the data set but it causes the loss of the connection between the output and the input variables.

## 2. Results

In order to evaluate the proposed procedure we have studied a 22 km long rural road, called 'SS 113' that connects Messina with Trapani (Italy). The infrastructure is old, antecedent to modern road standards and, for this reason, it is composed of a succession of straight stretches and circular curves, without the presence of transition elements.

The alignment is very winding, with short straight stretches interposed between the circular curves. The variability of the radii of the horizontal bends is really elevated and spans from a minimum of 24 m to a maximum of 3300 m. Longitudinal slope is modest for the entire development ($0 \le i \le 5\%$). Moreover, the road is characterized by low volumes of traffic, the absence of intersections and a sufficient geometric consistency, that is such as not to induce abrupt manoeuvres by drivers. The main geometric characteristics were found by the authors through the examination of 3D-digital cartography and with the aid of a differential GPS. In the next phase, after the reconstruction of the road geometry, we have computed the design and the $V85$ speed, the sight distances and the other variables, all in correspondence of the bisector of the circular curves. The final results are summarized in the Table 2.

At the macro level, we divided the procedure into two phases:

- in the first phase we prepared a data set (Table 1) in which the values of the 25 variables have identified with precision 5 clusters representing the four groups A, B, C and F provided by Italian standard, plus a fifth group that acts as a pole of attraction for all those observations that could represent situations of un-classification of the road;
- the second phase has assigned the observations recorded and reported in Table 2 to the 5 clusters previously identified (A, B, C, F and NO).

Table 2. The observations recorded in the center of the bends

*The first 12 observations recorded in the center of the bends*

| | Obs1 | Obs2 | Obs3 | Obs4 | Obs5 | Obs6 | Obs7 | Obs8 | Obs9 | Obs10 | Obs11 | Obs12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 |
| 4 | 73 | 100 | 76 | 51 | 100 | 100 | 92 | 45 | 46 | 58 | 40 | 51 |
| 5 | 67 | 80 | 73 | 53 | 80 | 86 | 72 | 56 | 52 | 72 | 49 | 58 |
| 6 | 0.12 | 53.30 | 8.13 | 0.61 | 0.43 | 0.43 | 1.42 | 0.21 | 2.47 | 0.20 | 0.85 | 0.61 |
| 7 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 |
| 8 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 2 |
| 9 | 7 | 20 | 3 | 2 | 20 | 14 | 20 | 11 | 6 | 14 | 9 | 7 |
| 10 | 200 | 700 | 220 | 80 | 500 | 800 | 350 | 60 | 62 | 110 | 45 | 80 |
| 11 | 89 | 79 | 98 | 134 | 86 | 160 | 162 | 190 | 172 | 126 | 111 | 84 |
| 12 | 4.35 | 3.07 | 3.07 | 3.25 | 3.25 | 4.19 | 3.26 | 3.07 | 3.07 | 3.59 | 3.25 | 3.79 |
| 13 | 0.50 | 0.36 | 0.36 | 0.36 | 0.50 | 0.50 | 0.36 | 0.36 | 0.36 | 0.50 | 0.50 | 0.36 |
| 14 | 0.30 | 0.00 | 0.00 | 0.00 | 0.30 | 0.30 | 0.00 | 0.00 | 0.00 | 0.30 | 0.30 | 0.00 |
| 15 | 1.25 | 1.00 | 1.00 | 1.00 | 1.25 | 1.25 | 1.00 | 1.00 | 1.00 | 1.25 | 1.25 | 1.00 |
| 16 | 3 | 2 | 2 | 2 | 3 | 4 | 3 | 2 | 2 | 3 | 3 | 2 |
| 17 | 1 | 2 | 1 | 2 | 2 | 2 | 3 | 1 | 2 | 2 | 1 | 3 |
| 18 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 3 | 4 | 2 | 3 | 2 |
| 19 | 141 | 141 | 137 | 137 | 137 | 137 | 137 | 137 | 137 | 137 | 137 | 137 |
| 20 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| 21 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*The last 12 observations recorded in the center of the bends*

| | Obs13 | Obs14 | Obs15 | Obs16 | Obs17 | Obs18 | Obs19 | Obs20 | Obs21 | Obs22 | Obs23 | Obs24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 | 1200 |
| 4 | 45 | 37 | 45 | 37 | 44 | 40 | 56 | 62 | 66 | 73 | 100 | 73 |
| 5 | 57 | 48 | 57 | 50 | 48 | 53 | 57 | 68 | 73 | 71 | 94 | 68 |
| 6 | 5.96 | 34.23 | 2.41 | 0.63 | 0.57 | 0.85 | 0.88 | 2.39 | 0.69 | 0.12 | 0.43 | 43.27 |
| 7 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 26 | 29 | 29 | 29 |
| 8 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 9 | 12 | 11 | 12 | 13 | 5 | 13 | 1 | 6 | 7 | 2 | 6 | 5 |
| 10 | 60 | 37 | 60 | 38 | 55 | 45 | 100 | 130 | 150 | 200 | 1000 | 200 |
| 11 | 36 | 69 | 61 | 48 | 53 | 55 | 36 | 104 | 32 | 43 | 62 | 38 |
| 12 | 3.25 | 4.50 | 3.25 | 3.25 | 3.58 | 3.76 | 3.25 | 4.76 | 3.77 | 3.25 | 5.30 | 3.82 |
| 13 | 0.50 | 0.36 | 0.50 | 0.50 | 0.36 | 0.36 | 0.50 | 0.50 | 0.36 | 0.50 | 0.50 | 0.50 |
| 14 | 0.30 | 0.00 | 0.30 | 0.30 | 0.00 | 0.00 | 0.30 | 0.30 | 0.00 | 0.30 | 0.30 | 0.30 |
| 15 | 1.25 | 1.00 | 1.25 | 1.25 | 1.00 | 1.00 | 1.25 | 1.25 | 1.00 | 1.25 | 1.25 | 1.25 |
| 16 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 | 3 | 3 | 3 |
| 17 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 |
| 18 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| 19 | 137 | 137 | 137 | 137 | 137 | 137 | 137 | 137 | 137 | 124 | 124 | 124 |
| 20 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 8 | 8 | 8 |
| 21 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 1 | 1 | 1 |
| 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 23 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

All the observations have a degree of membership (between 0 and 1) with respect to the five groups and the highest value suggests the predominant class. For example, in the case of the Observation No 1, the highest value is 0.53 and is representative of the class F. It Means that the class F can be attributed (with some certainty) to the entire section subtended from that observation. On the contrary, this stretch of road cannot fall into classes A or B, since the values of membership of such clusters are negligible (0.00 and 0.02 respectively).

In details, the application of the fuzzy clustering procedure can be schematized in the following phases:

- Collection of two data sets with $N = 5$ observations (first phase) and $N' = 24$ observations (second phase). Because of the heterogeneity of the units contained in the data sets, it was necessary to perform a normalization procedure not reported here, given its banality;
- Imposition of the number of clusters equal to 5;
- Establishment of the termination tolerance $\varepsilon > 0$ (in this case $\varepsilon = 110^{-6}$);
- Initialization of the partition matrix randomly, such that $U^{(0)} \in M$.
- Repetition for $L = 1; 2; \ldots$ of the following phases:
  - Computation of the five cluster centers:

$$v_i^{(L)} = \frac{\sum_{k=1}^{N} \left(\mu_{ik}^{L-1}\right)^m x_k}{\sum_{k=1}^{N} \left(\mu_{ik}^{L-1}\right)^m}, 1 \le i \le c;$$

  - Computation of the distances:

$$D_{ikA}^2 = \left\|x_k - v_i\right\|_A^2 = \left(x_k - v_i\right)^T A\left(x_k - v_i\right) \text{ with } 1 \le i \le c \text{ and } 1 \le k \le N;$$

  - Updating of the partition matrix:

$$\mu_{ik}^L = \frac{1}{\sum_{j=1}^{c} \left(\frac{D_{ikA}}{D_{jkA}}\right)^{\frac{2}{m-1}}}, 1 \le i \le c; 1 \le k \le N$$

until $\left\|U^{(L)} - U^{(L-1)}\right\| < \varepsilon$.

The shape of the clusters is determined by the choice of the particular A in the distance measure ($A = I$) that induces the standard Euclidean norm:

$$D_{ikA}^2 = \left\|x_k - v_i\right\|_A^2 = \left(x_k - v_i\right)^T \left(x_k - v_i\right).$$

The application of the Fuzzy C-Means algorithm returns the structure matrix representing the partition matrix $U$ ($c \times N$) (Table 3), the distance matrix containing the square distances between data points and cluster centers $D_{ik}^2$ ($c \times N$), the cluster centers $v_i$ ($c \times n$) (Table 4); $N = 5$ is the number of the initial observations; $c = 5$ is the clusters number; $n = 25$ is the number of input variables.

Finally, the insertion of new observations (Table 2) has allowed to evaluate the membership to the clusters without knowing the value of the density and to determine the partition matrix for the evaluated data set $U^*$ ($c \times N'$) and the distance matrix representative of the distances between the evaluated data points and the cluster centers $D_{ik}^{*2}$ ($c \times N'$), with $N' = 24$ number of new observations (Table 5).

Through the Tables 5, therefore, it is possible to identify the preponderant class to which report all the observations and, therefore, the entire road. Fig. 1 shows how the general infrastructure belongs to the class F but also evidences an attraction to the cluster representing situations of un-classification (red dashed line).

Finally, the reduction of the size of the matrix via the PCA procedure, allowed to represent the data in a more compact form and to highlight the closeness to the 5 clusters (Fig. 2).

Table 3. Partition matrix U (it permits to assign the cluster number to a specific class; in this case A – cluster 2; B – cluster 3; C – cluster 1; F – cluster 4; NO – cluster 5)

| | A | B | C | F | NO |
|---|---|---|---|---|---|
| Cluster 1 | 1.1E–11 | 1.8E–11 | **1.0E+00** | 1.8E–10 | 1.6E–11 |
| Cluster 2 | **1.0E+00** | 6.9E–11 | 1.1E–11 | 8.2E–12 | 4.5E–12 |
| Cluster 3 | 6.9E–11 | **1.0E+00** | 1.8E–11 | 1.4E–11 | 6.5E–12 |
| Cluster 4 | 8.2E–12 | 1.4E–11 | 1.8E–10 | **1.0E+00** | 1.9E–11 |
| Cluster 5 | 4.5E–12 | 6.5E–12 | 1.6E–11 | 1.9E–11 | **1.0E+00** |

Table 4. Matrix containing the $v_i$ cluster centers respect to every variable

| | A | B | C | F | NO |
|---|---|---|---|---|---|
| 1 | 0.500 | 1.000 | 0.750 | 0.250 | 0.001 |
| 2 | 0.500 | 1.000 | 0.750 | 0.250 | 0.001 |
| 3 | 0.210 | 1.000 | 0.470 | 0.120 | 0.001 |
| 4 | 0.590 | 1.000 | 0.760 | 0.470 | 0.001 |
| 5 | 0.620 | 1.000 | 0.810 | 0.480 | 0.001 |
| 6 | 0.500 | 1.000 | 0.700 | 0.400 | 0.001 |
| 7 | 0.490 | 1.000 | 0.750 | 0.240 | 0.001 |
| 8 | 0.500 | 1.000 | 0.750 | 0.500 | 0.001 |
| 9 | 0.500 | 0.000 | 0.380 | 0.630 | 1.000 |
| 10 | 0.260 | 1.000 | 0.460 | 0.020 | 0.001 |
| 11 | 0.210 | 0.350 | 0.270 | 0.170 | 0.001 |
| 12 | 0.450 | 0.450 | 0.450 | 0.360 | 0.001 |
| 13 | 0.001 | 1.000 | 0.960 | 0.001 | 0.001 |
| 14 | 0.000 | 1.000 | 0.710 | 0.001 | 0.001 |
| 15 | 0.600 | 1.000 | 0.700 | 0.400 | 0.001 |
| 16 | 0.750 | 1.000 | 1.000 | 0.750 | 0.001 |
| 17 | 0.750 | 1.000 | 1.000 | 0.750 | 0.001 |
| 18 | 1.000 | 1.000 | 1.000 | 1.000 | 0.001 |
| 19 | 0.500 | 1.000 | 0.900 | 0.350 | 0.001 |
| 20 | 0.550 | 1.000 | 0.910 | 0.410 | 0.001 |
| 21 | 0.600 | 0.200 | 0.400 | 1.000 | 0.001 |
| 22 | 0.000 | 1.000 | 1.000 | 0.000 | 0.001 |
| 23 | 1.000 | 0.000 | 0.000 | 1.000 | 1.000 |
| 24 | 0.500 | 1.000 | 0.750 | 0.500 | 0.001 |
| 25 | 0.001 | 1.000 | 1.000 | 0.001 | 0.001 |

Table 5. Partition matrix (the sum of the values in every column is equal to 1)

| *The membership grade compared to the 5 clusters of the first 12 observations recorded in the center of the bend* | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obs1 | Obs2 | Obs3 | Obs4 | Obs5 | Obs6 | Obs7 | Obs8 | Obs9 | Obs10 | Obs11 | Obs12 |
| A | 0.00 | 0.01 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| B | 0.02 | 0.04 | 0.02 | 0.02 | 0.03 | 0.05 | 0.03 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| C | 0.18 | 0.24 | 0.19 | 0.18 | 0.22 | 0.24 | 0.20 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 |
| F | 0.53 | 0.43 | 0.53 | 0.55 | 0.47 | 0.41 | 0.50 | 0.54 | 0.54 | 0.54 | 0.55 | 0.55 |
| NO | 0.25 | 0.29 | 0.26 | 0.25 | 0.28 | 0.29 | 0.27 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 |
| *The membership grade compared to the 5 clusters of the last 12 observations recorded in the center of the bend* | | | | | | | | | | | | |
| | Obs13 | Obs14 | Obs15 | Obs16 | Obs17 | Obs18 | Obs19 | Obs20 | Obs21 | Obs22 | Obs23 | Obs24 |
| A | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| B | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.06 | 0.02 |
| C | 0.18 | 0.18 | 0.18 | 0.17 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.18 | 0.25 | 0.18 |
| F | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.55 | 0.54 | 0.54 | 0.53 | 0.38 | 0.53 |
| NO | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.25 | 0.26 | 0.30 | 0.26 |

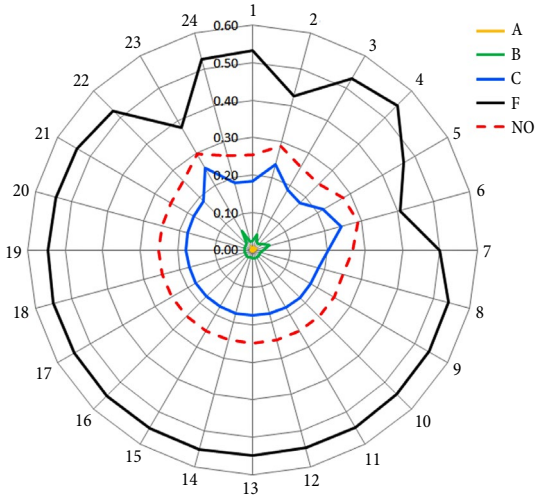


Fig. 1. Graphical representation of the membership grade in every section. It is clear the predominance of the F class except in 2, 6 and 23 sections

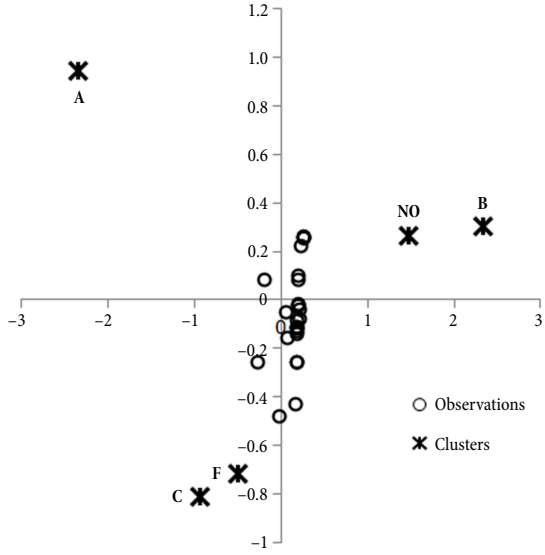


Fig. 2. Reduction of the n-dimensional matrix in 2-dimensional matrix and visualization of the two clusters centres and the observations. The separation in two cluster was also evident from the examination of the results in *n*-dimensions (Tables 4 and 5)

## 3. Discussion

An examination of the tables and figures included in the Results section allows us to highlight some advantages of the proposed procedure. As it is well known, the classification of a road is hardly ever an easy operation. This eventuality represents a rare and desirable case, but actually, the analyst has to manage a multiplicity of parameters whose values are often contradictory to each other. That is the great weakness of all the manual procedures: none legislation, in fact, specifies the way in which we can carry out the analytical process. For example, a road may have geometrical and constructive characteristics adequate to be classified as C, but, at the same time, the traffic flows would address it towards the group F. Situations like this are highly conflictive and can compromise the quality of the analysis. In fact, the analyst would need to know in quantitative terms the conformity of the road against all groups, in order to take the subsequent decision with due prudence and knowledge. Furthermore, the full membership to an expected class would permit to predict and organize the most correct maintenance operations in order to improve the functional characteristics of the infrastructure up to the desired level.

The uncertainty and the extreme subjectivity of the traditional methodologies regarding these issues are exceeded by the procedure proposed here. In particular, we have quantified the degree of membership to a predetermined class (cluster) of each road section. This information is already sufficient for the stakeholder of the infrastructure: for example, if the degree of membership to the group C is limited by an insufficient available sight distance, he can act on the obstacles (barriers, walls, trees, etc.), or on the other parameters involved, to bring the section close to the desired cluster.

The examination of the Fig. 1 allows to evaluate synthetically the content of the Table 5. In particular, the graph shows the sections 2, 6 and 23 in which there is a strong inconsistency compared to adjacent portions of the road. In fact, while all the other sections have a high prevalence of the group F (black solid line around the value 0.53), these three ones have a relatively low degree of membership to it (black solid line around the value 0.40), though still prevalent than other classes. The same graph shows that, at the same time, there is a certain tendency of these sections to acquire characteristics of the group C and, therefore, better functional characteristics. To inspect the reasons for these results it is sufficient to examine the Table 2, where we can see how the three sections have abnormal values of the following variables:

- *design speed* very high (100 km/h) and, therefore, very large radiuses.
- *V85 speed* high enough (80, 86 and 94 km/h respectively for sections 2, 6 and 23).
- remarkable *sight stopping distance* (but only in section 2).

Since the outcome of the analysis cannot be seen as positive (it is, in fact, an obvious lack of homogeneity), the stakeholder should introduce solutions to mitigate the speed, in order to make the route more uniform.

The response of the model to the fuzzy clustering analysis also highlights a certain affinity (between 0.25 and 0.30) towards the NO cluster. This time the examination of the Table 2 allows us to see how some variables do not respect the nominal limits of the four standard classes (A, B, C and F): among the most evident there are the *V*85, the *stopping sight distance* and the *available sight distance*, the presence of emergency lanes and parking areas. It is clear, therefore, that this procedure does not diminish the role of the analyst, but rather makes him capable to take more rational decisions.

At the end, the Fig. 2 summarizes a procedure (PCA) to decrease the size of the sample. In this case, it has been used to reduce it to a two-dimensional matrix and appreciate the disposition of the observations compared to the 5 clusters. It would be interesting to perform the analysis only with the more important and delete the others, in order to make the subsequent survey less expansive. However, in this phase of the research this additional step has been avoided to focus on the actual validity of the presented procedure.

We solved our problem with the clustering technique called Fuzzy C-Means. Only later, the results have been simplified in two dimensions through the PCA procedure, solely for arrange them in a more comfortable way. Although there are many other techniques applicable to this type of problem, we believe that the FCM is the most suitable one for its simplicity and because the analytical structure returns the final result accompanying it to a degree of uncertainty.

## Conclusions

The functional classification of a road is an obligatory step for the designer or for the stakeholder of a new or an existing road. In fact, this operation establishes the geometrical standard, orientates future maintenance operations (usually characterized by limited budgets), identifies the right constructive elements to ensure or improve drivers' safety, follows the development of the territory as required by the planning instruments. In general, however, the analyst accomplishes this task by determining only few synthetic indices, often of a qualitative nature, without any analytical detail and so providing a result strongly based on his own experience.

This approach causes not only an approximation of the final outcome, but also a real incapability to identify the contribution of each variable involved in the analysed scenario.

In this paper, we wanted to propose a procedure based on a fuzzy clustering technique in order to make some improvements to the analysis required by the road rules.

First, the choice of a clustering technique of this type takes into account the vagueness of the survey, even in the final results. If, instead, we had applied a hard type technique (K-means clustering, K-medoids, etc.) we would have had a very sharp classification, so inducing the analyst's thought towards a false sense of accuracy of the final result.

Secondly, the vagueness of the outcome of the analysis, when it occurs, allows the analyst to operate through two different ways, both desirable:

- acceptance of a condition quite critical but in the awareness of the distance with the theoretical class of the road (in essence, the difference between 1 and the calculated degree of membership);
- improvement of the so determined scenario with appropriate measures on the variables identified as the most vulnerable. In both cases, the analyst always has a very significant role, but may base his choice on the support of a sufficiently evolved analysis.

This procedure can be applied with profit by the manager of the road and can be further developed to refine some problems of little scientific interest, but of considerable practical importance. It is possible, for example, to increase significantly both the number of variables and the observations, achieving improved reliability of the final results. It is also desirable to implement this technique in a Geographic Information System in order to use the data typically found in these instruments and offer an additional instrument for the knowledge of the road context.

From the scientific point of view, a sufficiently simple development to perform is to study the variables that most affect the studied phenomenon with feature selection techniques. In this way, it is possible to limit the cost of the survey and, above all, the analyst turns his attention only to the parameters that really count for the quantification of the final result. As mentioned, at this stage this step has not been deepened, also because it requires database with a very significant number of observations with respect to the number of the input variables.

## References


AASHTO. 2011. *A Policy on Geometric Design of Highways and Streets*. 6th edition, American Association of State Highway and Transportation Officials (AASHTO), Washington, DC. 912 p.

Abonyi, J.; Feil, B. 2007. *Cluster Analysis for Data Mining and System Identification*. Birkhäuser. 306 p.

Bosurgi, G.; D'Andrea, A.; Pellegrino, O. 2010. Could drivers' visual behaviour influence road design?, *Advances in Transportation Studies* 22: 17–30.

Bosurgi, G.; D'Andrea, A.; Pellegrino, O. 2011. Context sensitive solutions using interval analysis, *Transport* 26(2): 171–177. http://dx.doi.org/10.3846/16484142.2011.589425

*Design Manual for Roads and Bridges (DMRB)*. 2002. Volume 6: Road Geometry. Section 1: Links. TD 9/93 Highway Link Design. UK Department for Transport. Available from Internet: http://www.dft.gov.uk/ha/standards/dmrb

Cafiso, S.; La Cava, G.; Montella, A. 2011. Safety inspections as supporting tool for safety management of low-volume roads, *Transportation Research Record* 2203: 116–125. http://dx.doi.org/10.3141/2203-15

Cafiso, S.; Lamm, R.; La Cava, G. 2004. Fuzzy model for safety evaluation process of new and old roads, *Transportation Research Record* 1881: 54–62. http://dx.doi.org/10.3141/1881-07

Chen, H.; Namdeo, A.; Bell, M. 2008. Classification of road traffic and roadside pollution concentrations for assessment of personal exposure, *Environmental Modelling & Software* 23(3): 282–287. http://dx.doi.org/10.1016/j.envsoft.2007.04.006

Dağdeviren, M.; Yüksel, İ.; Kurt, M. 2008. A fuzzy analytic network process (ANP) model to identify faulty behavior risk (FBR) in work system, *Safety Science* 46(5): 771–783. http://dx.doi.org/10.1016/j.ssci.2007.02.002

D'Andrea, A.; Pellegrino, O. 2012. Application of fuzzy techniques for determining the operating speed based on road geometry, *Promet – Traffic & Transportation* 24(3): 203–214. http://dx.doi.org/10.7307/ptt.v24i3.313

Dickerson, A.; Peirson, J.; Vickerman, R. 2000. Road accidents and traffic flows: an econometric investigation, *Economica* 67(265): 101–121. http://dx.doi.org/10.1111/1468-0335.00198

Dorsey, D. W.; Coovert, M. D. 2003. Mathematical modeling of decision making: a soft and fuzzy approach to capturing hard decisions, *Human Factors* 45(1): 117–135. http://dx.doi.org/10.1518/hfes.45.1.117.27228

FHWA. 2004. *Flexibility in Highway Design*. US Department of Transportation, Federal Highway Administration (FHWA). 205 p. Available from Internet: http://www.fhwa.dot.gov/environment/publications/flexibility

Forkenbrock, D. J.; Foster, N. S. J. 1997. Accident cost saving and highway attributes, *Transportation* 24(1): 79–100. http://dx.doi.org/10.1023/A:1017978227554

García Balboa, J. L.; Ariza López, F. J. 2008. Generalization-oriented road line classification by means of an artificial neural network, *GeoInformatica* 12(3): 289–312. http://dx.doi.org/10.1007/s10707-007-0026-z

Giummarra, G. J. 2003. Establishment of a road classification system and geometric design and maintenance standards for low-volume roads, *Transportation Research Record* 1819: 132–140. http://dx.doi.org/10.3141/1819a-20

Gu, Z. Q.; Oyadiji, S. O. 2008. Application of MR damper in structural control using ANFIS method, *Computers & Structures* 86(3–5): 427–436. http://dx.doi.org/10.1016/j.compstruc.2007.02.024

Güler, İ.; Übeyli, E. D. 2005. Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, *Journal of Neuroscience Methods* 148(2): 113–121. http://dx.doi.org/10.1016/j.jneumeth.2005.04.013

Italiano Strada Standard. 2001. *Norme funzionali e geometriche per la costruzione delle intersezioni stradali*. (in Italian).

Jaarsma, C. F. 1997. Approaches for the planning of rural road networks according to sustainable land use planning, *Landscape and Urban Planning* 39(1): 47–54. http://dx.doi.org/10.1016/S0169-2046(97)00067-4

Jang, J.-S. R. 1993. ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man and Cybernetics* 23(3): 665–685. http://dx.doi.org/10.1109/21.256541

Kang, H.; Scott, D. M. 2008. An integrated spatio-temporal GIS toolkit for exploring intra-household interactions, *Transportation* 35(2): 253–268. http://dx.doi.org/10.1007/s11116-007-9146-4

Kaptein, N. A.; Claessens, F. M. M. 1998. *Effects of Cognitive Road Classification on Driving Behaviour: a Driving Simulator Study*. TNO Human Factors Research Institute, Soesterberg, The Netherlands. 43 p.

Karlaftis, M. G.; Golias, I. 2002. Effects of road geometry and traffic volumes on rural roadway accident rates, *Accident Analysis & Prevention* 34(3): 357–365. http://dx.doi.org/10.1016/S0001-4575(01)00033-1

Kashani, A. T.; Mohaymany, A. S. 2011. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models, *Safety Science* 49(10): 1314–1320. http://dx.doi.org/10.1016/j.ssci.2011.04.019

Kockelman, K. M. 2001. Modeling traffic's flow-density relation: Accommodation of multiple flow regimes and traveler types, *Transportation* 28(4): 363–374. http://dx.doi.org/10.1023/A:1011815913359

Kuehnle, A.; Burghout, W. 1998. Winter road condition recognition using video image classification, *Transportation Research Record* 1627: 29–33. http://dx.doi.org/10.3141/1627-05

Lamm, R.; Psarianos, B.; Mailaender, T. 1999. *Highway Design and Traffic Safety Engineering Handbook*. McGraw-Hill Professional Publishing. 1088 p.

Lingras, P. 2001. Unsupervised rough set classification using GAs, *Journal of intelligent information systems* 16(3): 215–228. http://dx.doi.org/10.1023/A:1011219918340

Manduchi, R.; Castano, A.; Talukder, A.; Matthies, L. 2005. Obstacle detection and terrain classification for autonomous off-road navigation, *Autonomous Robots* 18(1): 81–102. http://dx.doi.org/10.1023/B:AURO.0000047286.62481.1d

Mena, J. B. 2003. State of the art on automatic road extraction for GIS update: a novel classification, *Pattern Recognition Letters* 24(16): 3037–3058. http://dx.doi.org/10.1016/S0167-8655(03)00164-8

Mon, Y.-J. 2007. Airbag controller designed by adaptive-network-based fuzzy inference system (ANFIS), *Fuzzy Sets and Systems* 158(24): 2706–2714. http://dx.doi.org/10.1016/j.fss.2007.06.005

Nowakowska, M. 2010. Logistic models in crash severity classification based on road characteristics, *Transportation Research Record* 2148: 16–26. http://dx.doi.org/10.3141/2148-03

Paclík, P.; Novovičová, J.; Pudil, P.; Somol, P. 2000. Road sign classification using Laplace kernel classifier, *Pattern Recognition Letters* 21(13–14): 1165–1173. http://dx.doi.org/10.1016/S0167-8655(00)00078-7

Pellegrino, O. 2011. Road context evaluated by means of fuzzy interval, *Cognition, Technology & Work* 13(1): 67–79. http://dx.doi.org/10.1007/s10111-010-0155-2

Pellegrino, O. 2012a. Prediction of driver's workload by means of fuzzy techniques, *The Baltic Journal of Road And Bridge Engineering* 7(2): 120–128. http://dx.doi.org/10.3846/bjrbe.2012.17

Pellegrino, O. 2012b. The application of drivers' cognitive reliability using fuzzy techniques, *Applied Artificial Intelligence* 26(6): 517–540. http://dx.doi.org/10.1080/08839514.2012.670973

Praticò, F. G.; Giunta, M. 2011. Speed distribution on low-volume roads in Italy: from inferences to rehabilitation design criteria, *Transportation Research Record* 2203: 79–84. http://dx.doi.org/10.3141/2203-10

Ravi, V.; Reddy, P. J.; Zimmermann, H.-J. 2000. Pattern classification with principal component analysis and fuzzy rule bases, *European Journal of Operational Research* 126(3): 526–533. http://dx.doi.org/10.1016/S0377-2217(99)00307-0

Ravi, V.; Zimmermann, H.-J. 2000. Fuzzy rule based classification with *FeatureSelector* and modified threshold accepting, *European Journal of Operational Research* 123(1): 16–28. http://dx.doi.org/10.1016/S0377-2217(99)00090-9

Shanahan, J.; Thomas, B.; Mirmehdi, M.; Martin, T.; Campbell, N.; Baldwin, J. 2000. A soft computing approach to road classification, *Journal of Intelligent and Robotic Systems* 29(4): 349–387. http://dx.doi.org/10.1023/A:1008158907779

Sohn, S. Y.; Lee, S. H. 2003. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, *Safety Science* 41(1): 1–14. http://dx.doi.org/10.1016/S0925-7535(01)00032-7

Tahmasebi, P.; Hezarkhani, A. 2010. Application of adaptive neuro-fuzzy inference system for grade estimation; case study, Sarcheshmeh porphyry copper deposit, Kerman, Iran, *Australian Journal of Basic and Applied Sciences* 4(3): 408–420.

Toplak, W.; Koller, H.; Dragaschnig, M.; Bauer, D.; Asamer, J. 2010. Novel road classifications for large scale traffic networks, in *2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 19–22 September 2010, Funchal, Portugal, 1264–1270. http://dx.doi.org/10.1109/ITSC.2010.5625182

Zhu, Z.; Yang, S.; Xu, G.; Lin, X.; Shi, D. 1998. Fast road classification and orientation estimation using omni-view images and neural networks, *IEEE Transactions on Image Processing* 7(8): 1182–1197. http://dx.doi.org/10.1109/83.704310