MADRID
inter.noise 2019
June 16 - 19
NOISE CONTROL FOR A BETTER ENVIRONMENT

Comparison of different clustering techniques for traffic noise analysis in the city of Milan

Benocci Roberto[1], Roman Hector Eduardo[2]
Dipartimento di Scienze dell'Ambiente e della Terra (DISAT)
Università degli Studi di Milano-Bicocca
Piazza della Scienza 1
20126 Milano, Italy

ABSTRACT

In this paper, we present the results of a comparative analysis on the traffic noise in the city of Milan within the framework of a co- financed project by the European Commission through the Life+ 2013 program called Dynamap. Dynamap is based on the idea of finding a suitable set of roads that display similar traffic noise behaviour (24 hours temporal noise profile). The dataset is made of the traffic noise recorded in 93 sites distributed over the entire city. Different unsupervised clustering algorithms have been studied: the so-called hard and soft clustering. In order to apply efficiently the Dynamap method, it is important that we have a continuity or soft passage between clusters as the non-acoustic parameter, which has been introduced to describe non-monitored roads, changes. The hard clustering coupled to a density distribution of the non-acoustic parameter in the two obtained road- cluster behaviours, revealed more efficient than the soft one though its inherent smooth approach to clustering.

Keywords: Noise mapping, Environment, Annoyance
I-INCE Classification of Subject Number: 30

1. INTRODUCTION

Dynamic noise mapping is a challenge for all big cities because it provides a direct means to intervene when noise levels are overcoming their limits. A European Life project termed Dynamap rests on the concept that a limited number of real-time noise measurements can be used to build up a noise map representative of a large urban area [1]. The underlying idea considers the traffic noise source as dependent on the urban context. The statistical analysis showed that both the hourly noise level profiles of roads and vehicle flow rates may be divided into two general behaviours (clusters) [2, 3].

[1] roberto.benocci@.unimib.it
[2] eduardo.roman@mib.infn.it

Such result suggests that the traffic noise can be employed as a parameter to attribute to a non-monitored road a noise behavior. Previous works showed that a representative noise parameter is provided by the available information on the total daily vehicle flow as obtained by a traffic flow model developed by the Municipality of Milan [4]. The traffic noise in the city of Milan has been characterized by a long monitoring campaign which involved 93 sites distributed over the entire city. In order to obtain a statistically significant sample, we adopted a stratification sampling approach based on roads showing similar noise trend profiles. The use of clustering techniques proved to provide a better efficiency and a robust sample [5]. The latest upgrade of Dynamap project can be found in [6, 7, 8]. Here, we present a comparative analysis of different clustering techniques whose results suggest that the adopted method, consisting in using a binary classifier coupled to the distribution of the non-acoustic parameter in the obtained groups, gives more useful information than those ones provided by soft clustering techniques.

2. CUSTERING ALGORITHMS

Unsupervised clustering algorithms are commonly employed to find similarities among data and group them together according to different schemes. Algorithms such as hierarchical agglomeration [9], k-means algorithm [10], partitioning around medoids (PAM) [11] are usually considered to this purpose. In general, the number of clusters is chosen in such a way as to obtain a reasonable compromise between satisfactory discrimination between data but keeping the number of groups to a minimum. As for statistical computing analysis and graphics, we used the statistical software R [12]. The results of the analysis of the acquired noise profiles have been deeply investigated in previous works [13, 14]. In synthesis, the use a R package "clValid" [15] allowed to rank each algorithm using an index based on its performance [16] and providing a two-cluster hierarchical agglomeration at the first place, followed also by a two-cluster groups by k-means and PAM methods. In such way, each monitored road has been assimilated, in term of noise profile over 24 h, to one of the two found groups. In order to extend such properties to non-monitored roads, a known independent parameter must be used to allow such procedure. In-depth studies [4], revealed that the logarithm of total vehicle flow could be considered to this purpose.

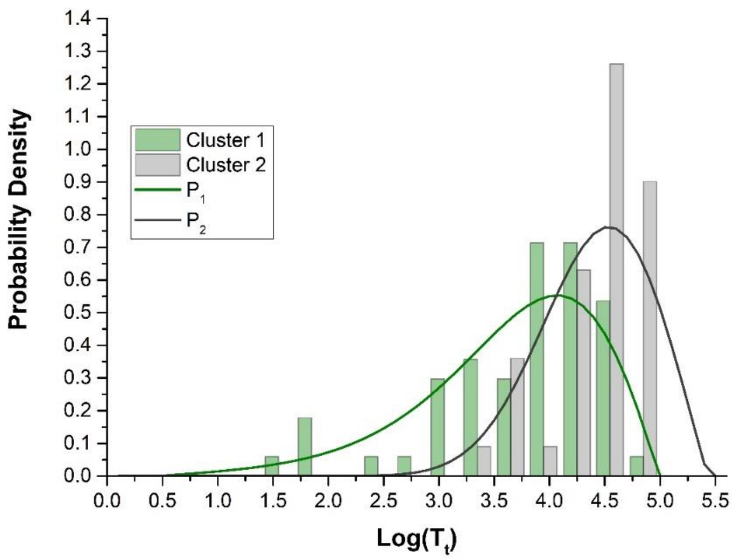

Figure 1: *Histogram and probability distribution P_1 and P_2 of the non-acoustic parameter $Log(T_t)$ in the two clusters.*

In figure 1, we can observe how such parameter is distributed in the two obtained clusters. The large overlapping interval suggests that each road noise temporal profile, characterized by a given non-acoustic parameter, might be regarded as a combination of the two main mean cluster noise profiles. Figures 2 and 3 give the number of events and the cumulative probability for Cluster 1 and 2 as a function of the non-acoustic parameter $Log(T_t)$. As alternative method, figure 1 suggested also the idea to consider as a clustering technique not a binary classifier but rather a "soft clustering" approach. In general, in non-fuzzy clustering (also known as hard clustering), data is divided into distinct clusters, where each data point can only belong to exactly one cluster. In fuzzy clustering, data points can potentially belong to multiple clusters. Membership grades are assigned to each of the data and membership grades indicate the degree to which a street arch belong to each cluster. Therefore, streets on the edge of a cluster might be part of a cluster to a weaker degree than streets in the center of cluster. Therefore, as alternative to traditional clustering methods, such as hierarchical clustering and k-means clustering, we can opt for fuzzy clustering algorithms of which one of the most widely used is the Fuzzy C-means clustering (FCM) algorithm [17, 18] as well as the model-based clustering [19, 20]. Both methods use a soft assignment, where each data point has a probability of belonging to each cluster.

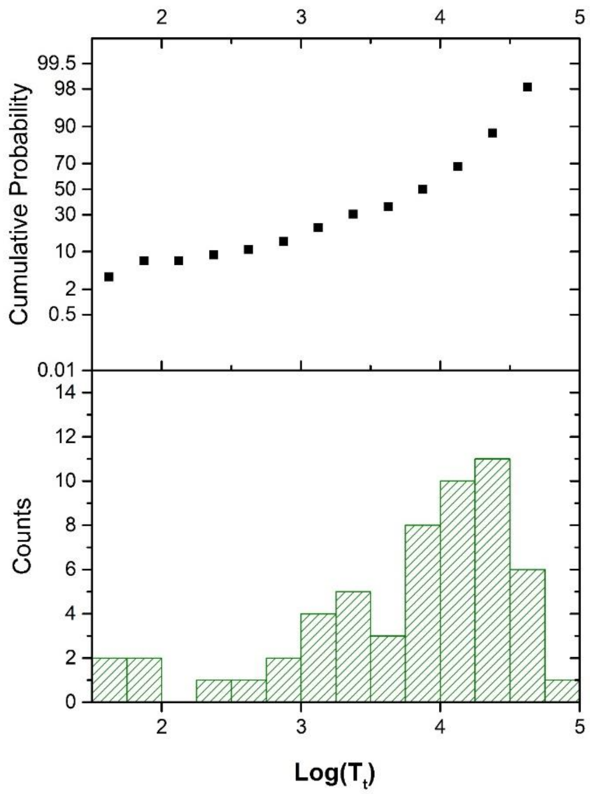

Figure 2: *Number of events and cumulative probability for Clusters 1 as a function of the non-acoustic parameter $Log(T_t)$. Bin size is 0.25.*

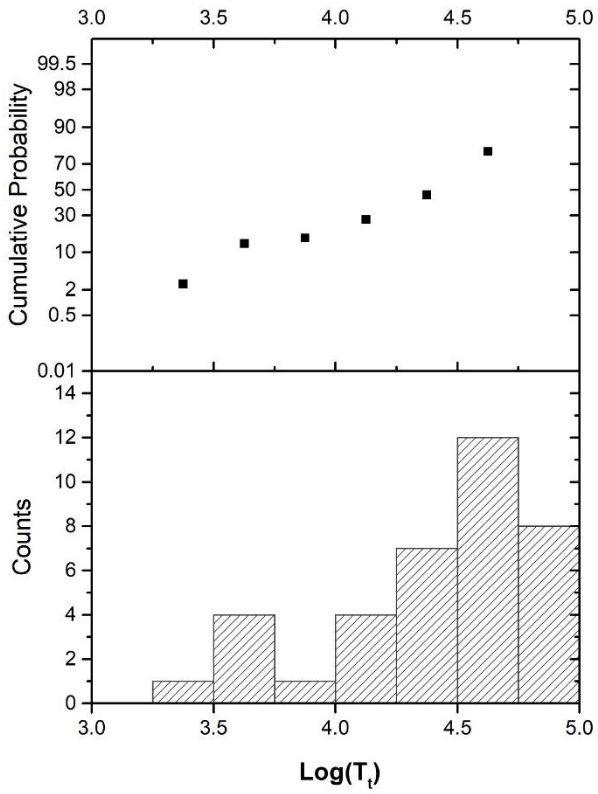

Figure 3: *Number of events and cumulative probability for Clusters 2 as a function of the non-acoustic parameter Log(T_t). Bin size is 0.25.*

3. RESULTS

The results of the mean cluster normalized equivalent level profiles, $\overline{\delta_{\iota k}}$ *(dB),* for the two clusters, k=(1, 2), as a function of hour of the day and for the different clustering algorithms are illustrated in figures 4 and 5. Both figures give similar results confirming the robustness of the clustering procedure. Details on the measurements and the standard analysis related to this result are reported in [1, 5] as well as other examples of traffic related clustering analysis [21, 22].

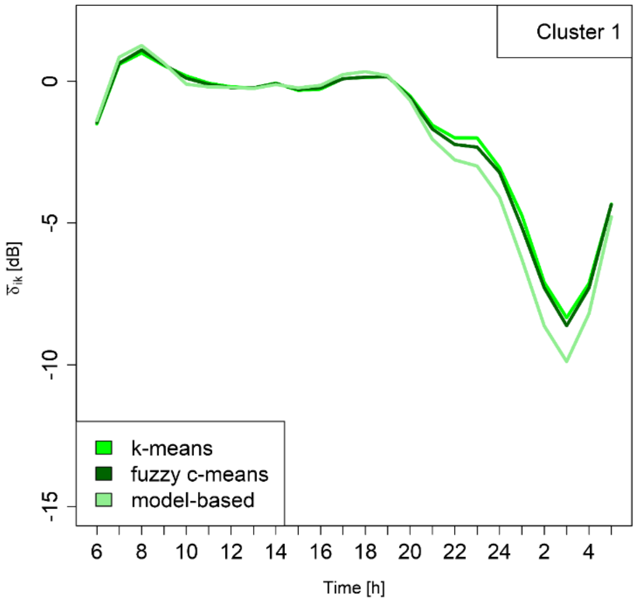

Figure 4: *Mean cluster normalized equivalent level profiles, $\overline{\delta_{\iota k}}$ (dB), obtained for the three clustering methods for Cluster 1 as a function of the hour of the day.*

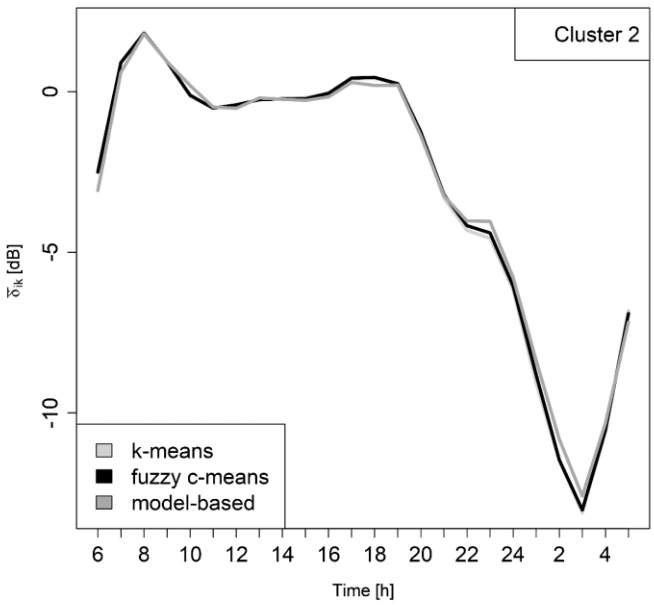

Figure 5*: Mean cluster normalized equivalent level profiles, $\overline{\delta_{\iota k}}$ (dB), obtained for the three clustering methods for Cluster 2 as a function of the hour of the day.*

In figures 6, 7 and 8, we show the scatter plot obtained from the Multi- Dimensional Scaling (MDS) results for the three clustering algorithms. The MDS provides a visual representation of the pattern of proximities among the data. Each point represents a road shown with its code number. In all figures, data have the same distribution over the two-dimensional representation; what differs is in the concentration ellipse, which provides

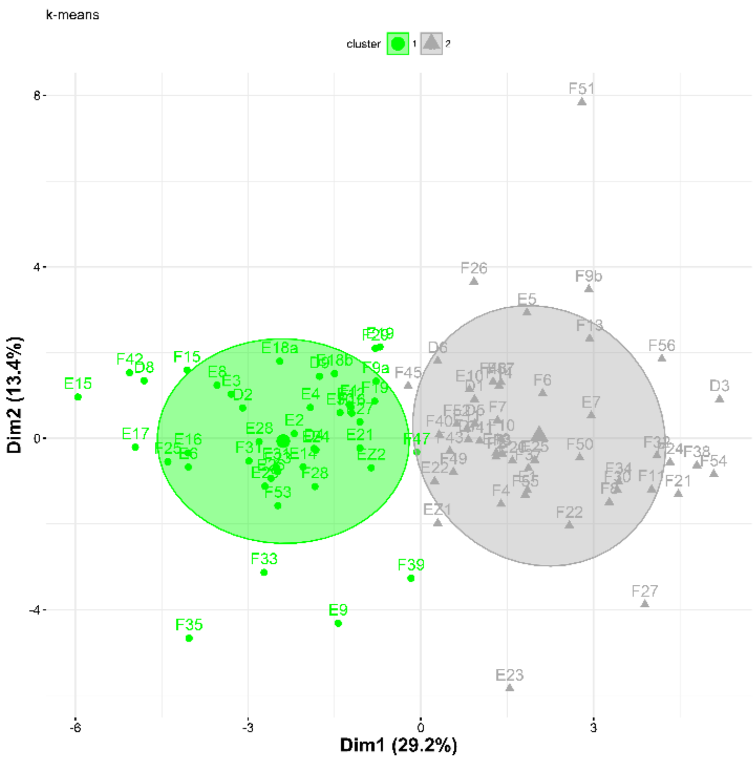

Figure 6*: Clustering results from Multi-Dimensional Scaling for k-means algorithm. The ellipse represents the confidence level at 0.68. The two clusters are marked in different colors.*

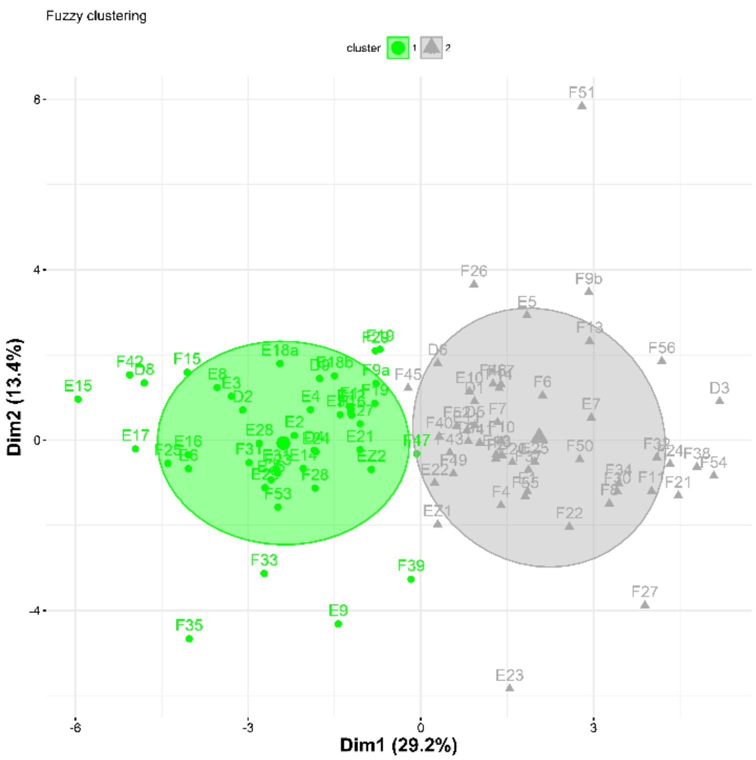

Figure 7: *Clustering results from Multi-Dimensional Scaling for Fuzzy algorithm. The ellipse represents the confidence level at 0.68. The two clusters are marked in different colors.*

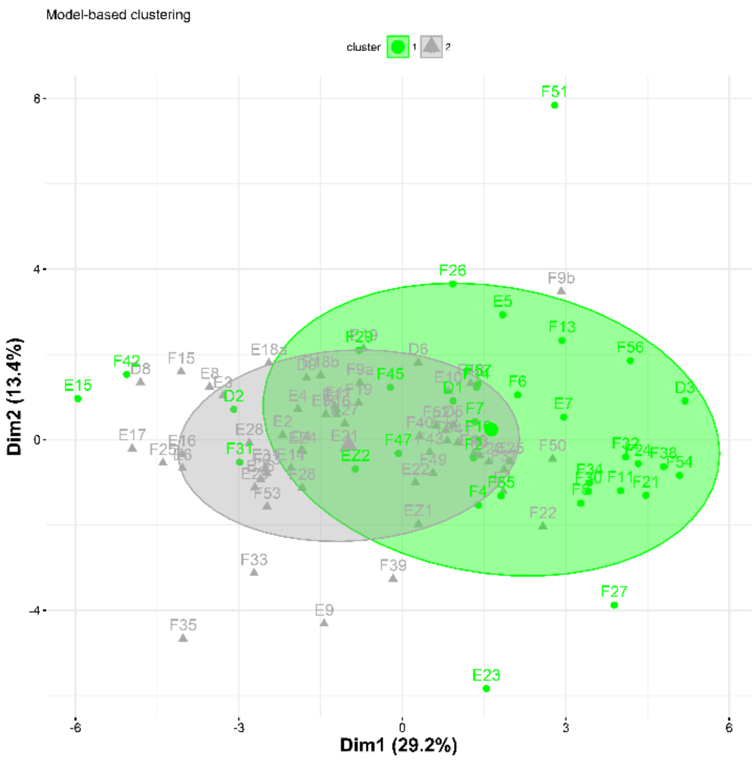

Figure 8: Clustering results from Multi-Dimensional Scaling for model-based algorithm. The ellipse represents the confidence level at 0.68. The two clusters are marked in different colours.

the confidence level for a normal distribution. In figures 6-8 the ellipse provides the 68% confidence level for the population mean. k-means and fuzzy clustering algorithms give the same probability distribution (see figures 6 and 7), whereas the model-based shows an overlapping of the two clusters for the same confidence level (figure 8). For this reason, we concentrated on the first two algorithms.

4. DISCUSSION

As anticipated above, each road noise temporal profile, may be regarded as a combination of the two main mean cluster noise profiles of figures 4 and 5. This means that for a road characterized by a non-acoustic parameter, which we denote as *x*, the corresponding noise temporal profile will have components in both clusters that is partly due to Cluster 1 and partly due to Cluster 2. The idea of the method is to evaluate the probability β_1 that *x* belongs to Cluster 1 and the probability $\beta_2 = 1-\beta_1$ that it belongs to Cluster 2. The corresponding values of β are given by the following relations:

$$\beta_1(x) = \frac{P_1(x)}{P_1(x)+P_2(x)} \tag{1}$$

and

$$\beta_2(x) = \frac{P_2(x)}{P_1(x)+P_2(x)} \tag{2}$$

where P_1 and P_2 represent the probability distribution shown in figure 1. Using the values of $\beta_{1,2}$, we can predict the hourly variations $\delta_x(h)$ for a given value of *x* according to:

$$\delta_x(x) = \beta_1(x) \cdot \delta_{C1}(h) + \beta_2(x) \cdot \delta_{C2}(h) \tag{3}$$

In principle, in order to exploit efficiently our method, it is important that we have a continuity or a soft passage between Cluster 1 and Cluster 2 as the non-acoustic parameter changes. This requirements is translated into β values changing with continuity as a function of $Log(T_t)$. In figure 9, we can observe the behaviour of β_1 as a function of $Log(T_t)$ for the two cluster algorithms: k-means and Fuzzy.

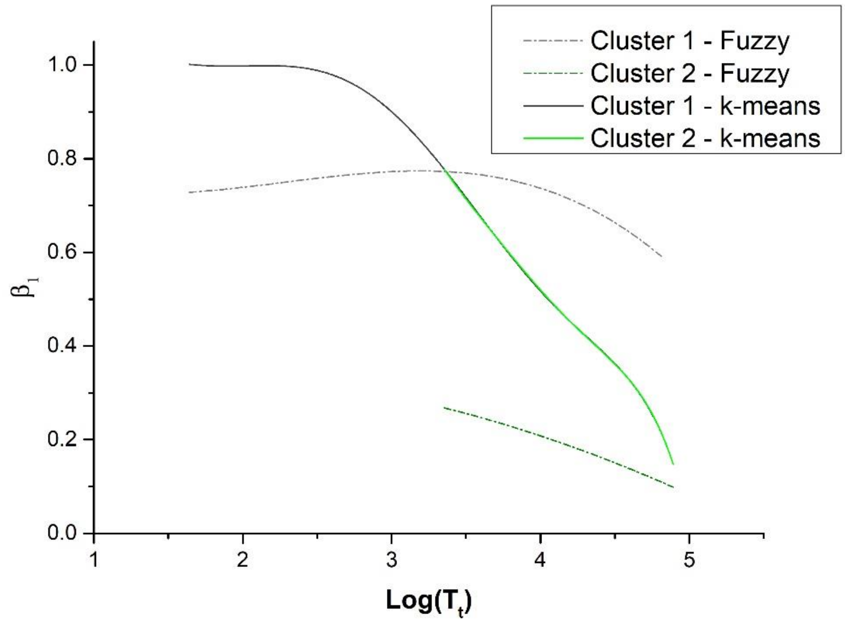

Figure 9: *Probability P_1 that a given road with non-acoustic parameter $Log(T_t)$ belongs to Cluster 1. Continuous lines refer to k-means, dashed lines to Fuzzy clustering. All the curves are the result of a fitting.*

In particular, for k-means algorithm, β_1 has been obtained from the density distribution of figure 1. The result is a smooth passage between Cluster 1 and 2, though the clustering algorithm used is a "hard" one. On the contrary, the Fuzzy algorithm is not able to provide such requirement though its soft approach to clustering. We omitted to draw Model-based clustering algorithm because its efficiency to separate the two clusters has been judged poor as it can be clearly seen in figure 8.

5. CONCLUSIONS

In this paper, we presented the results of a comparative analysis on the traffic noise in the city of Milan, which involved 93 sites distributed over the entire city. Different clustering algorithms have been studied: the so-called hard and soft clustering. This latter provides, naturally, roads behavior which are in-between the two mean noise clusters provided by the hard clustering. However, the binary classifier coupled to a non-acoustic parameter is able to provide a soft passage between Cluster 1 and Cluster 2 as the non-acoustic parameter changes. The Fuzzy algorithm does not meet this requirement though its inherent soft approach to clustering.

6. ACKNOWLEDGEMENTS

This research has been partially funded by the European Commission 247 under project LIFE13 ENV/IT/001254 Dynamap.

7. REFERENCES

1. Zambon, G.; Benocci, R.; Bisceglie, A.; Roman, H.E.; Bellucci, P. The LIFE DYNAMAP project: *Towards a procedure for dynamic noise mapping in urban areas*. Appl. Acoust. 2017, 124, 52-60.

2. Smiraglia, M.; Benocci, R.; Zambon, G.; Roman, H.E. Predicting Hourly Traffic Noise from Traffic Flow Rate Model: *Underlying Concepts for the DYNAMAP Project*. Noise Mapp. 2016, 3, 130-139.

3. Zambon, G.; Benocci, R.; Angelini, F.; Brambilla, G.; Gallo, V. *Statistics-based functional classification of roads in the urban area of Milan*. In Proceedings of the 7th Forum Acusticum, Krakow, Poland, 7-12 September 2014.

4. Zambon, G.; Benocci, R.; Bisceglie, A.; Roman, H.E. *Milan dynamic noise mapping from few monitoring stations: Statistical analysis on road network*. In Proceedings of the 45th INTERNOISE, Hamburg, Germany, 21-24 August 2016.

5. Zambon, G.; Benocci, R.; Brambilla, G. *Statistical Road Classification Applied to Stratified Spatial Sampling of Road Traffic Noise in Urban Areas*. Int. J. Environ. Res. 2016, 10, 411-420.

6. Zambon, G. ,Roman, H.E. ,Smiraglia, M., Benocci, R. "*Monitoring and prediction of traffic noise in large urban areas,* Applied Sciences (Switzerland), 8 (2) 2018.

7. R. Benocci, F. Angelini, M. Cambiaghi, A. Bisceglie, H. E. Roman, G. Zambon, R. M. Alsina-Pagès, J. C. Socorò, F. Alìas, F. Orga, "*Preliminary Results of Dynamap Noise Mapping Operations.*" Proceed. of 47th Internoise Conference 2018, 26-29 August 2018 Chicago (USA).

8. G. Zambon, F. Angelini, M. Cambiaghi, H. E. Roman, R. Benocci, "*Initial verification measurements of Dynamap predictive model"*, Proceed. of Euronoise Conference, Heraklion, Crete (Gr), 27-31 May 2018

9. Ward J. H., "*Hierarchical Grouping to Optimize an Objective*." Journal of the American Statistical Association, 58, 236-244, 1963.

10. Hartigan J. A., Wong M. A., "*k-means clustering*", Applied Statistics, 28, 100-108, 1979.

11. Kaufman L., Rousseeuw P. J., "*Finding Groups in Data."* Wiley Series in Probability and Mathematical Statistics, Hoboken (NJ), 1990.

12. http://www.r-project.org/

13. Zambon G., Angelini F., Benocci R., Bisceglie A., Radaelli S., Coppi P., Bellucci P., Giovannetti A., Grecco R., "*DYNAMAP: a new approach to real-time noise mapping"*, Proceed. of Euronoise Conference, Maastricht, June 1-3, 2015.

14. Zambon G., Benocci R., Brambilla G, "*Cluster categorization of urban roads to optimize their noise monitoring"*, Environ Monit. Assess (2016) 188-26, DOI 10.1007/s10661-015-4994-4

15. Brock G., Pihur P., Datta S., Datta S., "*clValid: An R Package for Cluster."* Journal of Statistical Software, 25, 1-22, 2008.

16. Pihur V., Datta S., Datta S., "*Weighted Rank Aggregation of Cluster Validation Measures: a Monte Carlo Cross-entropy Approach*." Bioinformatics, 23(13), 1607-1615, 2007.

17. Dunn J. C, "*A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters.*" Journal of Cybernetics. 3 (3): 32-57, 197. doi:10.1080/01969727308546046. ISSN 0022-0280.
18. Bezdek J. C., "*Pattern Recognition with Fuzzy Objective Function Algorithms*". ISBN 0-306-40671-3, 1881.
19. Fraley C., Raftery A. E., "*Model-Based Clustering, Discriminant Analysis, and Density Estimation*", Journal of the American Statistical Association 97 (458): 611-31, 2002.
20. Fraley C., Raftery A. E., Murphy T. B., Scrucca L., "*Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification", and Density Estimation*. Technical Report No. 597, 2012, Department of Statistics, University of Washington. https://www.stat.washington.edu/research/reports/2012/tr597.pdf.
21. Zambon G., Roman H. E., Benocci R., "*Vehicle Speed Recognition from Noise Spectral Patterns*", International Journal of Environmental Research, 11 (4), 449-459, 2017, DOI 10.1007/s41742-017-0040-4
22. Zambon G, Roman HE, Benocci R (2017) "*Scaling model for a speed- dependent vehicle noise spectrum.*" J Traffic Transp. Eng. (English Edition) 4(3):230-239. doi:10.1016/j.jtte.2017.05.001 11