LIFE13 ENV/IT/001254

Sizing the monitoring network

Technical Report

LIFE – DYNAMAP
Dynamic Acoustic Mapping – Development of low cost sensors networks for real time noise mapping

Deliverable Number and Title:	B1 – Sizing the monitoring network - Technical Report
Action Number – Title:	B1 – Sizing the monitoring network
Dissemination Level:	R (Restricted to Beneficiaries)
Status:	Final Version
Release Date:	22/02/2016
Author(s):	Roberto Benocci, Alessandro Bisceglie, Eduardo Roman, Maura Smiraglia, Giovanni Zambon
Reviewer(s):	Giovanni Zambon, Giovanni Brambilla, Sandro La Monica
Document code:	LIFE-DYNAMAP_BICOCCA_B1-SIZING THE MONITORING NETWORK_V01 - 30/10/2015
Contact person:	Giovanni Zambon
Postal address:	Dipartimento di Scienze dell'Ambiente e del Territorio e di Scienze della Terra Piazza della Scienza 1, 20126 Milano ITALY
Telephone:	+39 02 64482744
Fax:	+39 02 64482794
E-mail:	giovanni.zambon@unimib.it
Project Website:	www.life-dynamap.eu

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF KEYWORDS

Traffic noise emission; urban noise mapping; dynamic map; street categorization; clustering algorithm; continuous monitoring; statistical analysis.

EXECUTIVE SUMMARY

In this report is described the B1 Action: “Sizing the monitoring network”. Given the large number of roads present in Milan city, obtaining the dynamics acoustics map of this city requires application of a statistical approach where the roads having similar flow conditions—and thus similar noise trends—are grouped (clustered) together. In order to obtain these groups (clusters), an extensive measurement campaign was executed. In our analysis, we have taken traffic noise data from 93 stations distributed in the city. Of these 93 stations, about 20 fall inside the Milan pilot area, of interest here for developing the acoustic map. The data are recorded continuously, and different data are aggregated for selected time intervals, such as τ=(5, 10, 15, 20, 30, 60) mins. From the hourly data of noise levels, we apply a cluster analysis, and classify the 93 stations into two main clusters, denoted here as cluster 1 and 2. The hourly noise levels are very similar during the morning and afternoon hours, while more conspicuous differences are seen during the night hours. Cluster 1 contains those streets for which the noise falls down strongly during the night, corresponding to streets carrying a low traffic flow. Cluster 2 refers to those highly traffic streets, for which the level of noise remains relatively high even during the night hours. For each street in each cluster we calculate a non-acoustic parameter x. The latter has been chosen from many different possibilities. In summary, we found that the logarithm of the total daily flow, that is the log of the sum of the flow over the 24 hours, $x=LogT_T$, is a suitable choice. We then calculate the distribution of x for each cluster, 1 and 2, denoted as P1(x) and P2(x). For a given street i, we obtain its value of xi, and calculate the probability that such a stretch belongs to each one of the two clusters, i.e. β1= P1(xi) / (P1(x1)+P2(xi)), β2= P2(xi) / (P1(x1)+P2(xi)), respectively. Using these two probabilities, we estimate the noise level at stretch i, δi(h), according to the sum, δi(h)= β1 δ1(h)+ β2 δ2(h), where δ1,2(h) are the normalized mean noise level values from each cluster 1 and 2, recorded by the noise stations at hour h. In order to construct the noise maps, we study the distribution of the non-acoustic parameter P(x) for each stretch in the zone 9 (the pilot area selected). We check that it overlaps with the corresponding distribution of x corresponding to the stretches where the recording stations are located. This to ensure that the data recording contains essentially all the information over the whole range of values of x from the zone 9 of interest. Then, we divide the total interval of x values into six intervals or groups, each one containing a similar number of streets. Inside each group n, we determine the mean value xn and obtain the corresponding values of β1(n) and β2(n). From these values of β, we apply the above relation for δi(h) which yields the same noise level for each stretch i inside the group n. That is, to all the stretches belonging to group n we associate the same noise level δn (τ), obtained at time interval τ as measured by all the stations from clusters 1 and 2. As one can see, all recording stations together determine the six acoustic maps for each group, each one characterized its own mean values of β1(n) and β2(n). The location of the recording stations, in number of 24, have been suggested by analyzing the traffic flows inside each group n. We have created a list of stretches which have an hourly flow similar to the mean traffic flow of the group. The list covers a set of the first best choices, from which the final decision about the location of the recording stations can be taken. Inside each group n, one can locate 4 stations, for a total of 24. Once the 24 stations have been chosen and implemented, one needs to classify them into the two clusters, 1 and 2. We have found that 5 mins is a good compromise for updating the acoustic maps during daily hours, while during the night a more conservative hourly update is recommended. This is based on our analysis of the errors taken in the predictions, ranging up to about 2 dB during the day, and increasing up to 5-6 dB during night hours. We have checked that mean values of noise levels calculated over 5 mins are essentially the same as those calculated over 1 hour. The cluster composition remains essentially the same as those of the clusters obtained from the hourly data. In conclusion, we found that the maps can be updated every 5 mins in the time interval from 7-21 hours, every 15 mins in the time interval 21- 01 hours, while for the night 21-7 hours the update can be made every hour. Errors estimates in the action project described in this report (B1) will be checked and if necessary corrected during the action project B7 (System testing and fault analysis).

ACKNOWLEDGEMENTS

Special thanks to AMAT and Milan Municipality for providing their traffic model database and other Milan road network information.

SIZING THE MONITORING NETWORK
Technical report

1. INTRODUCTION

The DYNAMAP project aims at developing a prototype system in an effort to achieve the real-time noise mapping of road infrastructures in two pilot areas: a part of the agglomeration of Milan and a section of the Motorway A 90 in Rome. One of the main goals of the project is to create a low-cost sensor network able to make the five-year update of the strategic noise maps required by the European Directive 2002/49/EC more efficient and less expensive.
In this action a general method to identify and optimize the number of sites to be monitored for maps updating has been identified. To that end, application of techniques of spatial sampling was used to identify a limited number of arches, representative of the entire road network, where sound measurements will be performed.
In urban environment, such techniques require the identification of different types of road segments (arches), representative of different clusters. The definition of clusters was based on homogeneous noise levels temporal trends. Each road stretch of the network were then assigned to one predefined cluster.
The preliminary segmentation of the road network was defined by the presence of intersection nodes. As it is impossible to monitor all roads in the pilot area, a database of daily road traffic noise levels has been carried out.
The database information led to draw, by means of a statistical cluster analysis, a limited number of road types, with homogeneous noise trend levels over the day and night periods to be used as a reference. Each cluster, obtained by this method, has been statistically described also by non-acoustic information to be used to assign the non-monitored road segments to the different clusters (e.g. rush hour traffic flow or average daily traffic volume).
In order to assign road arches to the clusters, each segment of the entire road network needs to be classified on the basis of non-acoustic parameters. The choice of number and type of parameters to be used for the allocation of arches to clusters has been chosen among the ones that best represent the previously division into the clusters obtained by means of the sound levels trends.
Monitoring sites will then be identified for each cluster of roads.

2. DESCRIPTION OF THE URBAN NETWORK

The whole road network of Milan is composed of roughly 20.000 arches

The large number of sections which the network is divided into is given by several intersections and junctions, typical of a highly urbanized context. Furthermore, a road is often splitted in two or more roadways (carriageways). Also the volumes of circulating traffic are obviously determined by the structure of the network.
According to the new *Italian Road Code*, the roads are classified with respect to their structural, technical and functional features in the following typologies:

- A: Highways
- B: Main suburban roads
- C: Secondary suburban roads
- D: High flow urban roads
- E: Arterial urban roads
- F: Local Roads.

As represented in Fig. 1, Milan road network is mainly composed of urban and local roads.

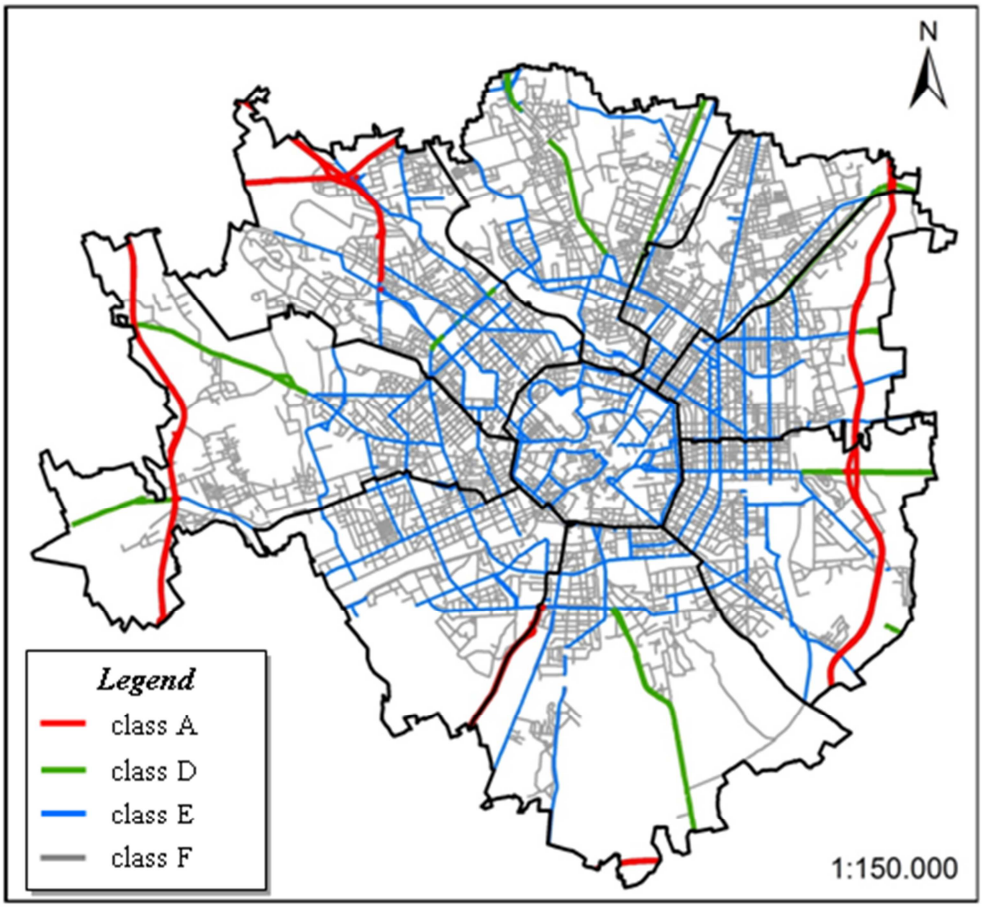


Fig. 1 - The road categories of Milan.

Implementation of the Dynamap system in urban context will be realized on an Pilot Area (ref. Action A1), representative of the conditions of traffic and acoustic climate of the entire city.

The study of noise propagation in the urban area of the city of Milan will be obtained by implementing a tridimensional model of the pilot area inside an acoustic calculation software. Acoustic computation will be run on the basis of the European standard algorithm for the vehicular traffic noise calculation (NMPB) [15]. Information regarding road network layout, land orography and the built up area shape, such as all the static features of the model, will be taken from geographical information system database of Milan municipality.
Instead, the definition of acoustic characteristics of sources in the calculation model is based on data obtained from a model of traffic distribution, developed by AMAT, one of the DYNAMAP project partners.

3. DATA COLLECTION

This chapter describes the activities carried out to create the collection of noise trends required for the statistical analysis.
Nowadays noise measurements are used mainly to validate results obtained from computational models and they are usually not directly involved into the noise mapping process.

The construction of a database, containing the noise emission of road infrastructure obtained in previous measurements and in new measurements in the city of Milan, is necessary to characterize the "acoustic behavior" of homogeneous groups of roads.
Therefore one of the first actions of the project is the realization of a large-scale noise monitoring survey specifically made to improve the previous noise data acquired with new ones.
To this end, a wide traffic noise monitoring campaign was planned on the entire area of the city of Milan, considering the space-time variability of the noise emissions generated by traffic sources. For each measurement point, different series of equivalent noise levels with different time resolutions were obtained for the period of 24 hours.

3.1. Acoustic data

The first activity involved the archiving of previous noise measurements. From all the historical data available; only those specifically related to road sources and with a duration of 24 hours were selected.
The Department of Environment and Territory and Earth Sciences has carried out since 2007 several measures of acoustic monitoring in the city of Milan.
The collected data have different origins. Some sound level measurements were performed for research purposes, others for collaborative activities carried out for institutions such as the Lombardy Region, the Municipality of Milan and the Milan Territory Environment Agency Mobility AMAT.
The dataset of previous continuous noise monitoring consists in 49 sites, related to 8 different road categories.
The increase of the historical database with noise trends coming from other institutions or research organizations is one of the next targets of the research team group.

The second phase of the study involved the planning and execution of a new campaign of acoustic monitoring, closely related to the purposes of the project DYNAMAP.
In order to create a representative statistical sample of the entire road network of the city, the following general criteria were adopted to identify the measuring sites:

- homogeneous distribution on the entire metropolitan area and between the nine city's districts of Milan;
- uniform distribution between the different road categories (A, D, E, F) and road subclasses (E1, E2, F0, F1; F2, F3);

- various urban scenarios (urban canyons with different conformation, open sound field, etc.);
- different road surface type;
- no influence of other roads on the monitored road stretch;
- absence of other noise sources (tram lines, railways, airports, etc).

In the monitoring campaign, three different types of monitoring units have been used. These units are:

- fixed monitoring stations (Fig. 2 A);
- semi-permanent monitoring stations (Fig. 2 B-C);
- monitoring stations placed on cart or on mobile laboratory (Fig. 2 D-E).

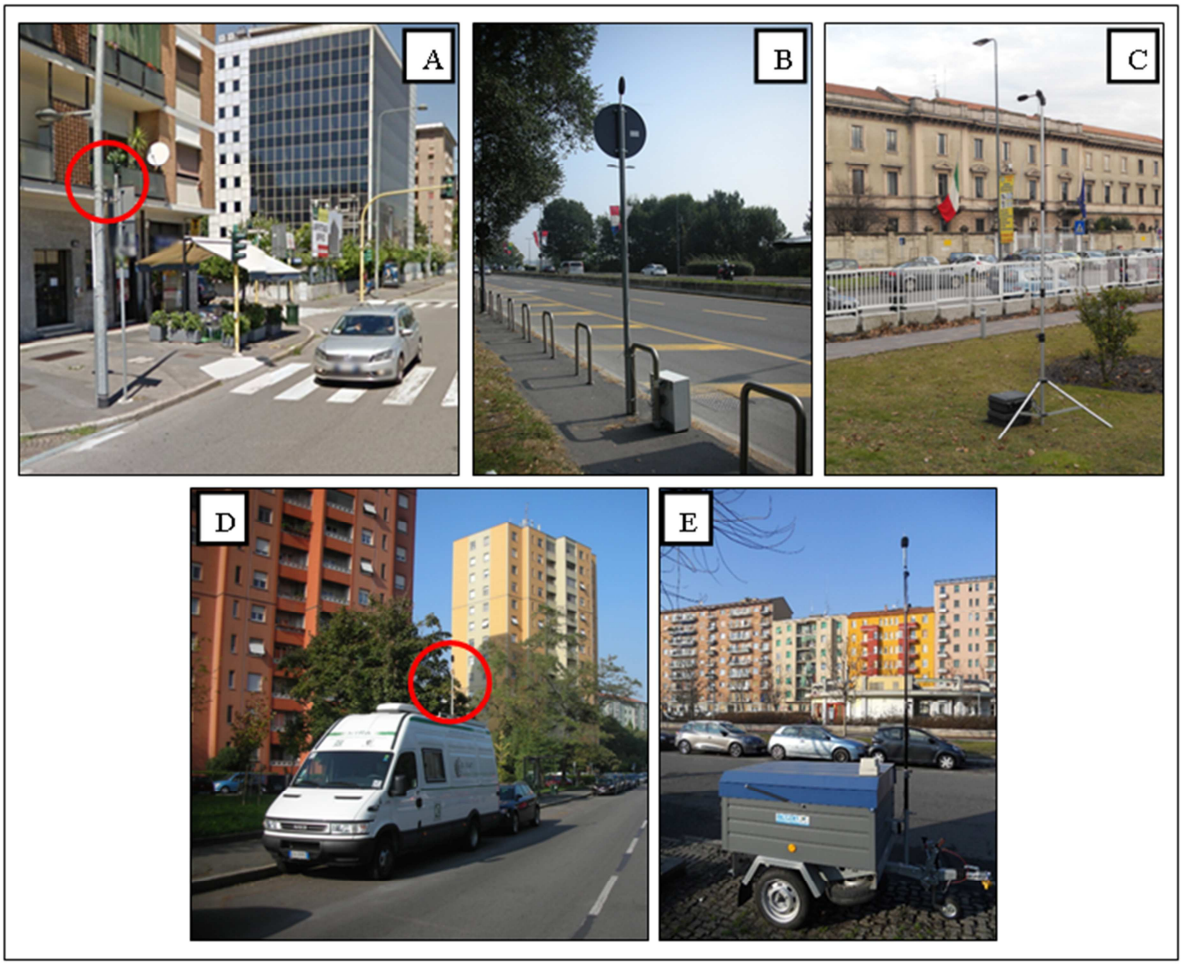


Fig. 2 - The monitoring stations used during the DYNAMAP noise monitoring campaign.

All monitoring stations are equipped with a class 1 sound level meter able to obtain the main noise indexes and the spectral trends in third-octave bands with a temporal resolution of 1 second.
The monitoring activity was sized on a minimum measurement time of 24 hours, starting at 6 a.m. and eventually protracted for several days.

3.2. Processing of the acoustic measurements

Subsequently to the execution of the noise measurements, all acquired data was elaborated in a three steps procedure.
The first operation involved the exclusion of noise records of public holidays and anomalous days (such as days of school closures) in order to extrapolate only the acoustic data relative to a typical weekday.

The second operation involved the correction of the acoustic data with weather data. The units of acoustic monitoring in fact are not equipped with weather stations, therefore, it was necessary to associate every single noise level with weather data of rainfall and wind speed measured in the different weather stations of ARPA (Regional Agency for Environmental Protection) located on Municipality of Milan (Fig. 3).
To get weather data related to the real weather conditions detectable near each measurement site, each site was selected and compared to the reference weather station closest to the point of measurement itself.
The effect on the acoustic data of rainfall events was evaluated considering a method proposed by ARPA. In detail, the criteria used are:

- precipitation < 2 mm: irrelevant;
- precipitation > 2 mm: influential on individual hourly data;
- precipitation > 4 mm: influential for the time immediately following the event weather.

To evaluate the influence of atmospheric turbulence on the acoustic data, the criterion used is indicated by Italian normative. The acoustic data were discarded when:

- wind speed > 5 m / sec

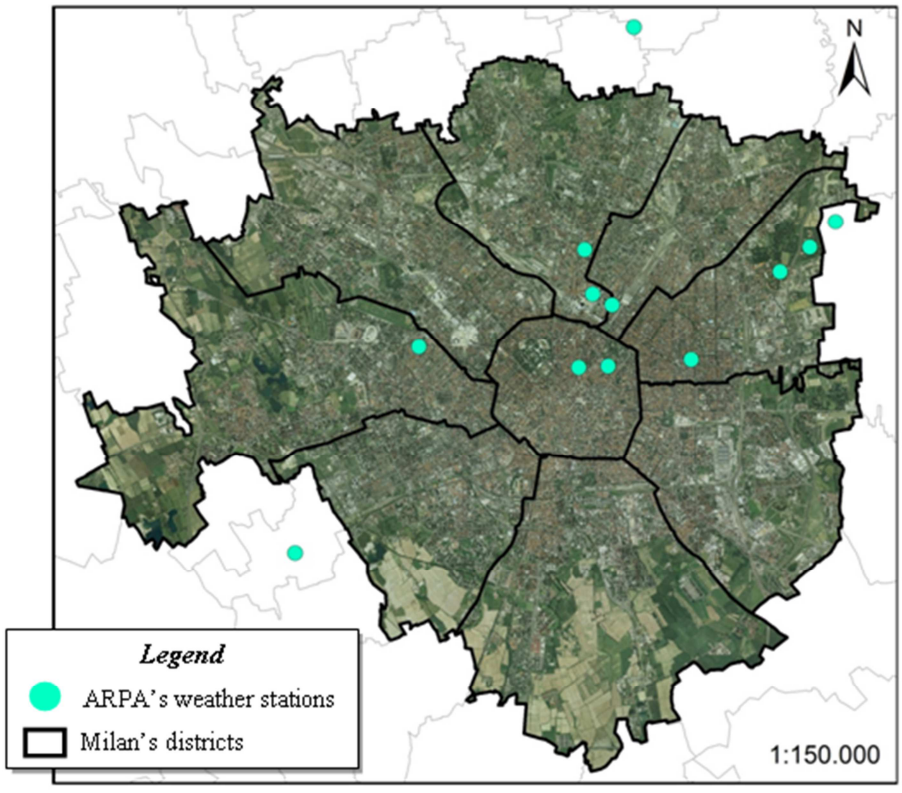


Fig. 3 - The dislocation of ARPA's weather stations in the territory of Milan.

The third corrective operation of the original acoustic data involved the elimination of extraordinary or abnormal events (such as sirens, horns, airplane transits, noisy human activities, technical facilities, etc.), since their presence can affect and alter the equivalent noise level.
This operation is important because it allows to achieve an acoustic data closely related to the single traffic source.

The identification of the extraordinary events was based on the comparison and analysis of the sonograms. Sonograms are graphic representations of a sequence of spectra in time, where the sound pressure level, in a chromatic scale, is expressed in function of time and frequency (Fig. 4).

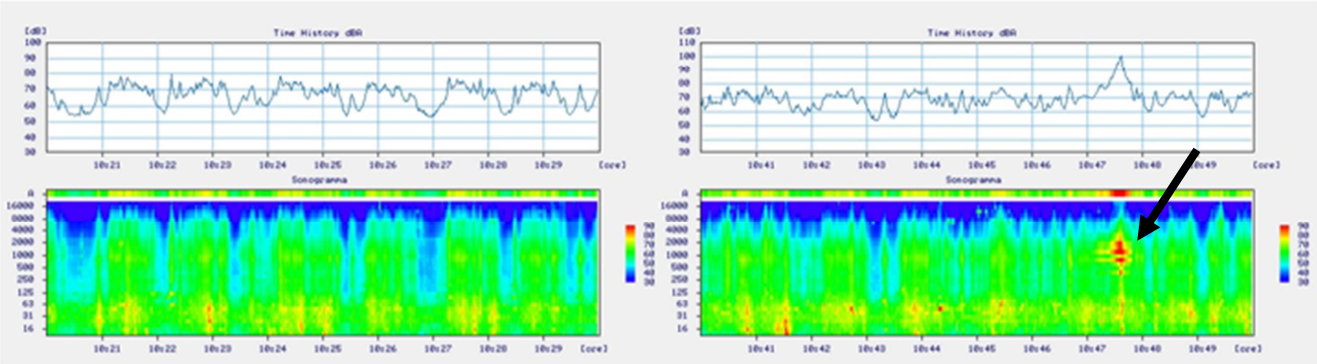


Fig. 4 - Sonograms related to road traffic in normal conditions and in the presence of an anomalous event (ambulance siren).

After the corrections of the acquired data, the sound pressure levels were calculated by integrating the acoustic data on different time intervals, respectively 5, 10, 15, 20, 30 and 60 minutes.
The succession of equivalent noise levels on 24 hours represents the trend noise, which constitutes the basis for subsequent statistical analysis. Each noise dataset consists of 6 distinct temporal profiles obtained with different sampling of data per second.
Overall, a total amount of 231 noise trends was obtained: 123 coming from previous noise monitoring activities and 108 coming from the DYNAMAP noise monitoring campaign (Fig. 5). These 231 trends well describe the noise emissions of 102 roads of Milan (Table 1). Road categories A and D have been almost entirely described (see Fig. 1), while the E and F road classes, that show an high internal variability, will be the object of future noise surveys.

Table 1 - The results of the noise monitoring campaigns.

Functional class of road	Number of roads monitored	Number of daily noise trends collected
A	4	13
D	9	20
E1	21	61
E2	10	21
F0	18	34
F1	10	15
F2	15	32
F3	15	35
Total	**102**	**231**

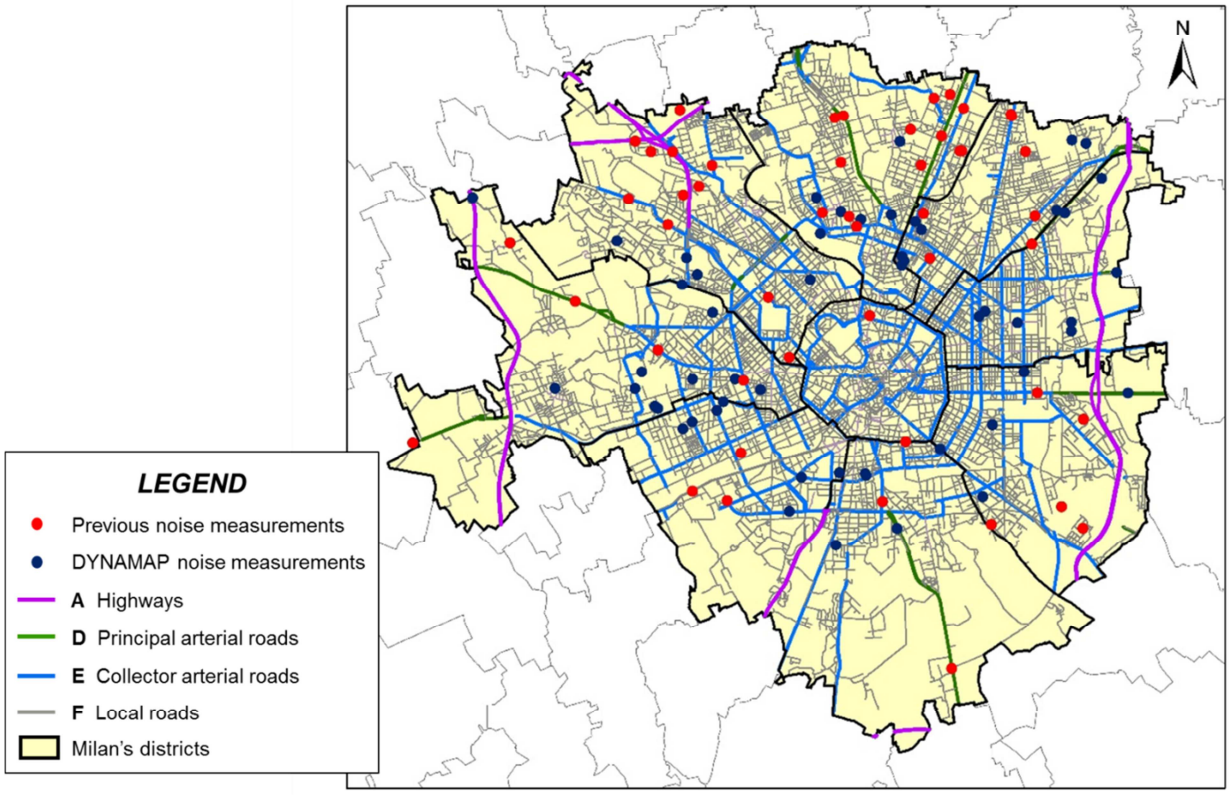


Fig. 5 - The location of noise measurements positions.

3.3. Traffic data

The opportunity of applying a statistical method (as provided by the project DYNAMAP) for the description of the entire urban road network requires that, in addition to the acoustic information relating to a road stretch, a "non-acoustic" parameter has to be found to complete the description of all stretches of the road network.
The parameter most easily obtainable and directly comparable with the acoustic data is the daily vehicular traffic, expressed in number of vehicles per hour and referred to each stretch of the road network.
Traffic flows can be obtained using two approaches:

- *network modeling*: extended to all road stretches; the data extrapolated can be affected by the confidence limits connected to the modeling approach;
- *experimental measurements*: consisting in on-site traffic flows counting (through radar instruments, portable traffic analyzers or magnetic loops).

The simulation model of the road network developed by AMAT is defined as a "macro-scale traffic static allocation equilibrium model". It consists in assigning to the arches of a graph which schematically represents the road network of Milan and hinterland, the traffic flows defined by means of appropriate “matrices origin/destination”; this matrix defines the amount of movements for each possible relationship between the zones in which the geographical area is divided, for different types of vehicle/user. Matrices origin/destination used were obtained from a specific survey campaign on the mobility of people in the Milan area.

For each road section, generally defined between two intersections or nodes, an arch in the graph corresponds, split in the two directions of flow.
The characteristics of outflow of the arch, that is the change in driving times on the road section, are defined by appropriate parameters depending on the geometric features, functional and traffic conditions existing in the simulation reference period.
The allocation model is a multi-class typology, namely three different matrices, corresponding respectively to passenger cars, commercial vehicles and motorcycles, are defined and assigned, in order to take into account the behavioral differences between different users categories. All types of vehicle contribute, with appropriate equivalence coefficients, to the whole occupancy of road capacity.
Overall, the main information available from the traffic model for each arch are:

- hourly traffic flow for 24 hours
- traffic composition by three vehicles typologies
- maximum capacity of each arch (static information)

This represents the non-acoustic available dataset related to the whole road network analysed.

The adopted calculation method of traffic volumes involves some empty fields, i.e. incorrect hourly values or not entirely reliable values.
This occurs mainly in low flows local roads, which could be not considered by the system of origins and destinations. As a result, the analysis of data available from traffic dataset could not completely cover the network over the whole time interval of 24 hours, in particular with respect to the night period.
In order to link the acoustic monitoring data to non-acoustic road data (traffic data) of the whole network, different arches (linear features physically distinct in the model) related to the same road have been merged together. The combined topographic element has the sum of the relative single traffic values in the associated attribute database.

In Table 2 a summary of the characteristics of the road network related to the portion of land identified as pilot area is reported.

Table 2 - Summary of road data, Milan Pilot Area.

item	Pilot Area
N° original arches	2652
N° elaborated arches	2075
Total daily traffic, Average value	16721
Total daily traffic, Median value	7872
Total daily traffic, St. Dev.	21152

3.4. Production of a Geodatabase

All data presented were collected on a GIS platform and inserted into a relational geodatabase (DBMS) purposely created. This tool, containing the acoustic data, the data of traffic flows and the georeferenced position of the measuring points, also collects and integrates a large series of information that may be useful for further consideration and elaboration. In detail, these informations are specifically:

- the identification code of the measurement;

- the type of monitoring units used;
- the date and the duration of the measure;
- characteristics of the measurement stations (microphone height, reflective surfaces and obstacles, the distance between the noise source and the roadway edge);
- information about the road monitored such as identification code, name and functional class of the road;
- photographic documentation;
- presence of structures (schools, hospitals, rest homes, etc.) in the neighborhood.

4. STATISTICAL ANALYSIS

4.1. Acoustic Database

The dataset considered in the present work refers to road traffic noise in the city of Milan, Italy, and is formed by 218 patterns of 24-h continuous monitoring of the hourly equivalent levels, L_{Aeqh}, in 93 different sites (this number derives from the validation of all the sites investigated in the monitoring campaign and related to traffic data) representing 8 functional road classes, named A, D, E and F, the last two divided into two and four sub-groups respectively. In the present study the sub-groups in each class were merged. Data were recorded on weekdays and in absence of rain, as required by the Italian decree D.M. Ambiente 16/3/1998 [19]. Because of the non-homogeneity of L_{Aeqh} level dataset, due to different monitoring conditions, such as different distances from the road but also to the condition of the street itself (its geometry, the presence of reflecting surfaces and obstacles along sound propagation and types of paving), we referred each i^{th} hourly L_{Aeqhij} level of the j^{th} temporal pattern to its corresponding daytime level, L_{Aeqdj}:

$$\delta_{ij} = L_{Aeqh_{ij}} - L_{Aeqdj} \text{ [dB]} \qquad (i = 1, \ldots\ldots, 24\ h;\ j = 1, \ldots\ldots, 93) \qquad (1)$$

The daytime level, L_{Aeqd}, was chosen as reference for the hourly L_{Aeqh} because this descriptor is more often available than the nighttime value, L_{Aeqn}. For all the 93 sites the vehicle flow rate at rush-hour (time interval 7:30 a.m. - 8:30 a.m.) was available too. In 47 sites, where the monitoring extended over more days, the median of the δ_{ij} hourly values was considered, as this index is less influenced by the presence of outliers. Figure 6 shows the 24-hour mean pattern $\overline{\delta}_{im}$ (green line, with m = D, E, F) and the corresponding ± the standard deviation for each road functional class (light green area). We did not take into account roads of class A due to the poor statistical relevance.

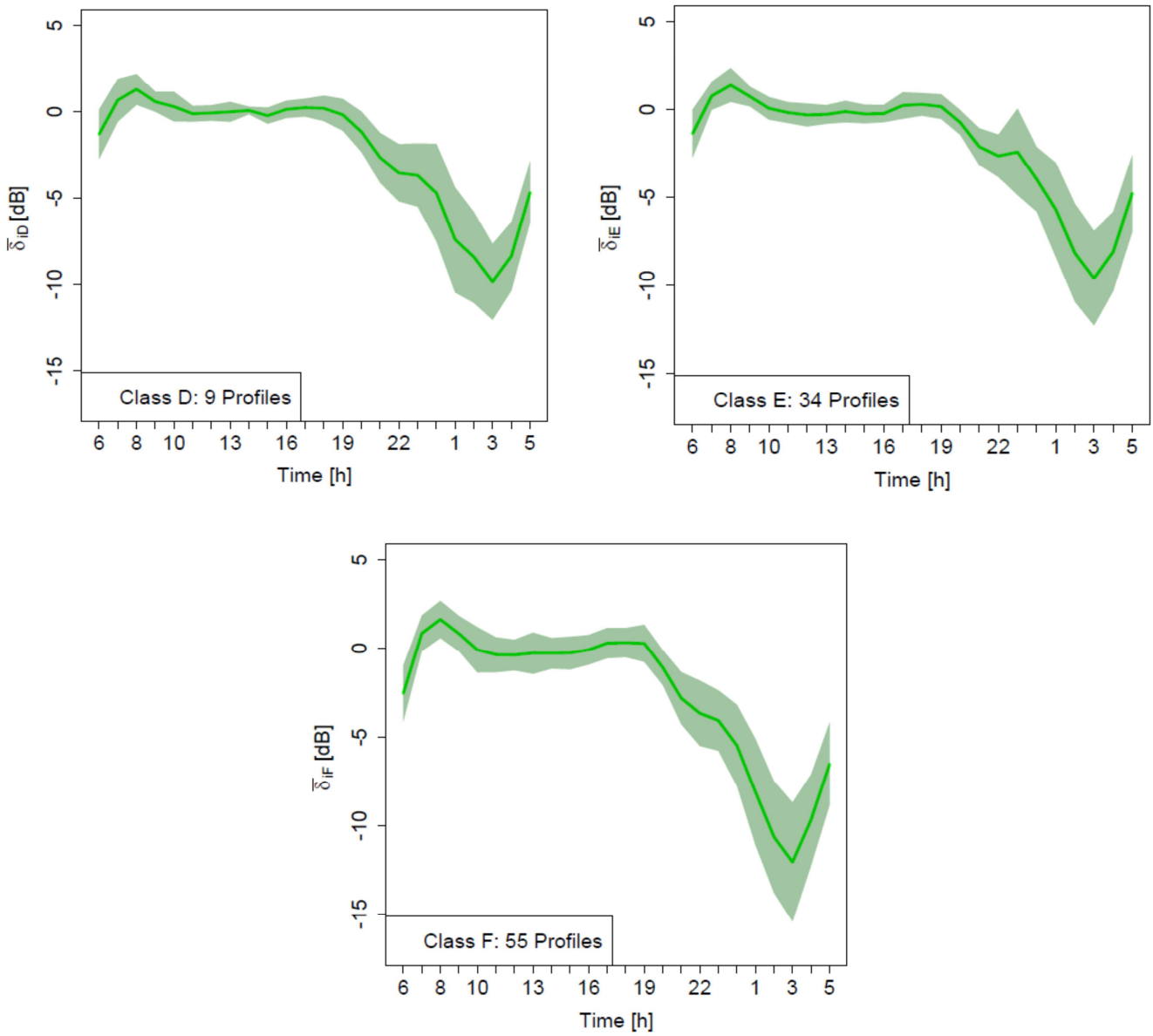


Fig. 6 - 24-hour mean patterns $\overline{\delta}_{im}$ (green line with m = D, E, F) and the corresponding ± standard deviation for each road functional class (light green area).

4.2. Cluster Analysis

Plotting together all the 24-hour mean patterns, $\overline{\delta}_i$, as it can be observed in Figure 7, we notice a strong overlapping of the mean profiles, $\overline{\delta}$, and their standard deviation bands, over all the day and night period. This fact suggests that roads, which share similar noise behavior, fall in different functional classes. This is the reason why we decided to approach the sampling issue not inside each functional group (stratified sampling) by firstly statistically analyzing the temporal profiles series, δ_{ij}, in order to obtain homogeneous groups of roads with similar noise behavior. This approach has the advantage of providing a better description of the real behavior of the complex road network of the city of Milan and, therefore, of improving the sampling efficiency [1-5] based upon cluster categorization.

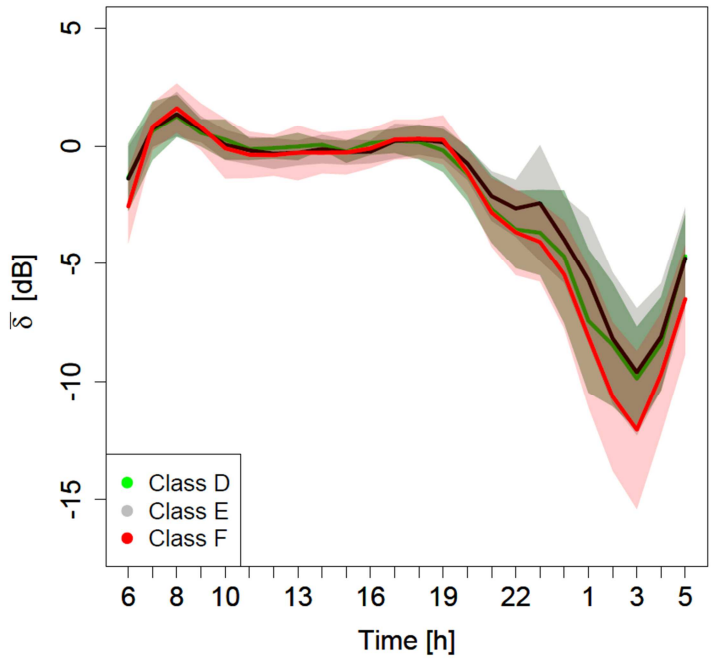


Fig. 7 - 24-h mean normalized noise level profiles, $\overline{\delta}$, of the 93 monitored sites with the corresponding standard deviation band grouped by functional classes. Roads of class A have been excluded because of poor statistical relevance.

Therefore, in addition to the functional road classification, we chose to classify roads according to statistic criteria based on the noise emission of the road. Generally such emission depends on the road activity, its use in the urban context, its width, the presence of reflecting surfaces, obstacles, type of paving, etc. These parameters are usually included in the functional classification of roads and linked to the role played in the urban mobility. However, this classification generally does not reflect the actual use of roads. For a better description of the real behavior of noise in complex scenarios such as the mobility network of the city of Milan, we approached the problem considering an agglomeration method based upon similarities among the 24-h continuous monitoring of the hourly equivalent L_{Aeqh} levels. After computing the δ_{ij} values, the corresponding time pattern provides a tool to group together roads following the same vehicular dynamics, therefore allowing a more real description of such road networks. For this reason, unsupervised clustering algorithms were applied to group together patterns found to be “similar” to one another. Various algorithms (hierarchical agglomeration using Ward algorithm [6], K-means algorithm [7], Partitioning Around Medoids [8], Expectation Maximization algorithm implemented by the “mclust” module [9]) were considered, and their results compared. The number of clusters was chosen as a compromise between satisfactory discrimination and the need to limit the number of groups. The range of solutions for clustering was set from three groups (for a straightforward comparison with the functional road classes considered) to two (corresponding to the minimal discrimination). Euclidean distance was chosen as the metric of the distance among observations. The statistical software R, an open-source software environment for statistical computing and graphics, was applied for the clustering [10]. A detailed discussion on the statistical analysis is given in [11] where a validating test of the clustering results was carried out by the “clValid” package [12] in order to assess the quality of the clustering and assign a score to the different clustering algorithms. The results obtained by the best performing algorithms (hierarchical with two clusters) were further tested to check for their independence, confirming the robustness of the obtained clusters.
The obtained clusters were formed of roads belonging to different classes, as reported in Table 3. This confirms that the road traffic is primarily linked to the effective urban mobility rather than its

functional classification, as shown also by the outcomes of previous studies such as in [13], where the monitoring data of road traffic noise collected in 244 sites distributed in 37 Italian cities were analyzed by K-means clustering.

Table 3 - Composition of clusters for different number of groups compared to the road classes

Hourly cluster	Road category			Total
	D	E	F	
1	5 (9%)	12 (21%)	39 (70%)	56
2	4 (24%)	20 (54%)	13 (35%)	37
Total	9	32	52	93

Regarding the clustering process results, the two-cluster solution represents a satisfying balance between an adequate differentiation among time patterns and the need to get a simple practical solution. The two clusters appeared to be formed primarily of the contributions from different temporal profiles belonging to roads of class F for cluster 1 (made up of 39 temporal profiles corresponding to 70% of total) and of roads of class E for cluster 2 (made up of 20 temporal profiles corresponding to 54% of total). Class D roads are almost equally distributed over the two clusters. This result confirms that the noise time patterns are not directly linked to the road classification. Figure 8, which accounts for the two-cluster solution reported in Table 3, shows the profiles of mean cluster values, $\overline{\delta}_{ik}$ with (k = 1, 2), and the corresponding ± the standard deviation for each cluster. Cluster 1 (blue line) presents one peak at the hourly interval 8-9. It remains close to L_{Aeqd} until 19 h, afterwards it goes down in the night period till 3 h and then it starts raising again. Cluster 2 (red line) shows one lower peak at 8-9 h and higher values at nighttime. In the remaining time period, it shows a similar behavior of cluster 1. In order to visualize the level of similarity of each cluster, a Multidimensional Scaling (MDS) is presented in Fig. 9, which shows a satisfying division between the two groups.

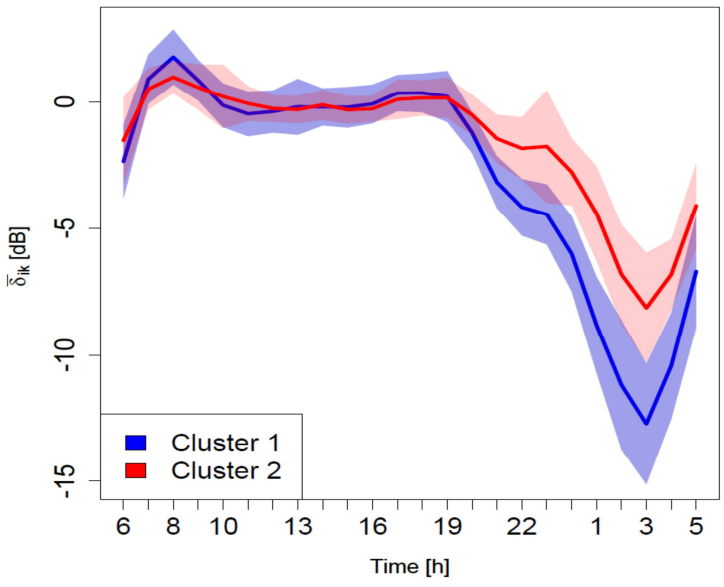


Fig. 8 - Mean cluster profiles, $\overline{\delta}_{ik}$, and the corresponding ± standard deviation.

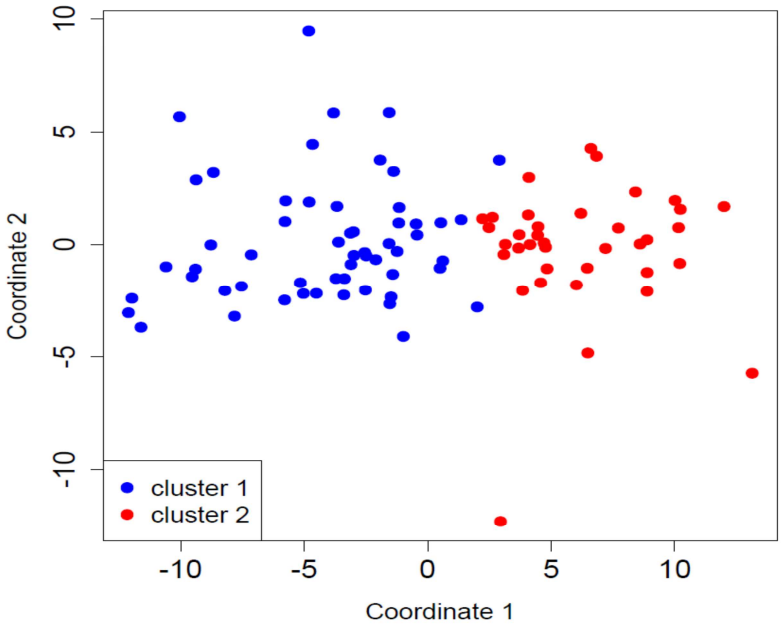


Fig. 9 - MDS of the two cluster results.

The hourly data were not normally distributed over the whole day period as shown by the Shapiro-Wilk's test results in Figure 10 (the time intervals of δ_{ij} not satisfying the normal distribution for Cluster 1 and 2 at α=0.05 significance level are placed below the red line).

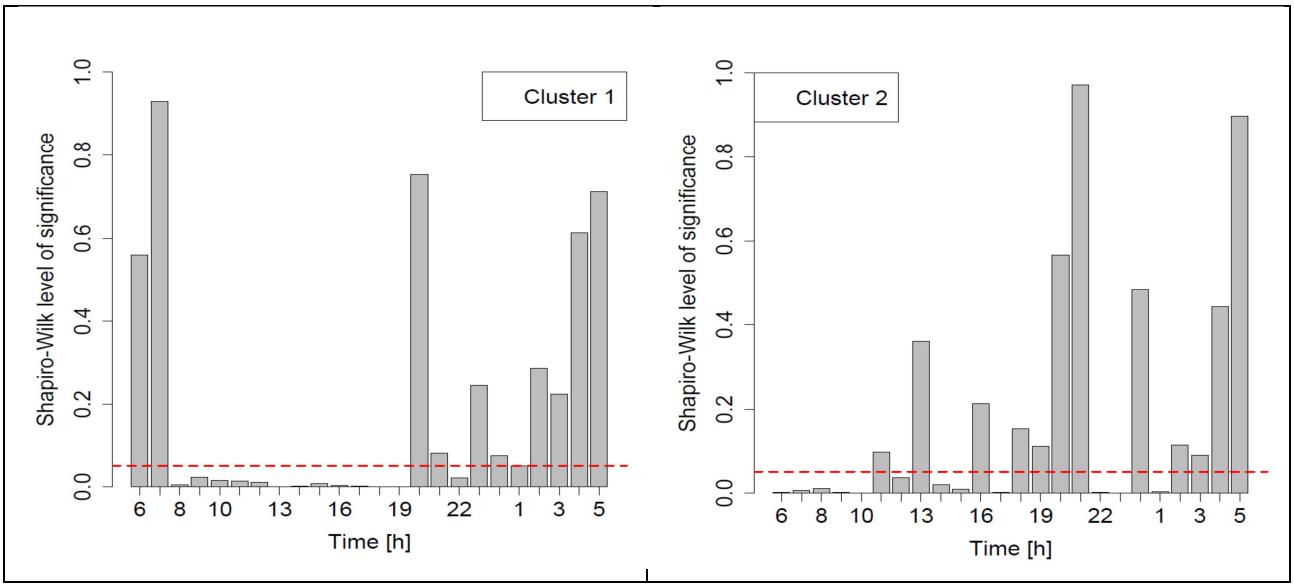


Fig. 10 - Shapiro-Wilk results for the normal distribution test of δ_{ij} . The time intervals of δ_{ij} not satisfying the normal distribution for Cluster 1 and 2 at α=0.05 significance level are placed below the red line

Thus, to check the independence of the two average cluster profiles, both the Student's *t* and Mann–Whitney–Wilcoxon tests were performed on each couple of profiles. The two-tailed *p*-values for both tests were much lower than α = 0.05 significance level for most of the morning, evening and night

hours as it is also clear from Figure 11 and 12. Therefore, the hypothesis of dependence was rejected (with the exception of the hours above the reference red line).

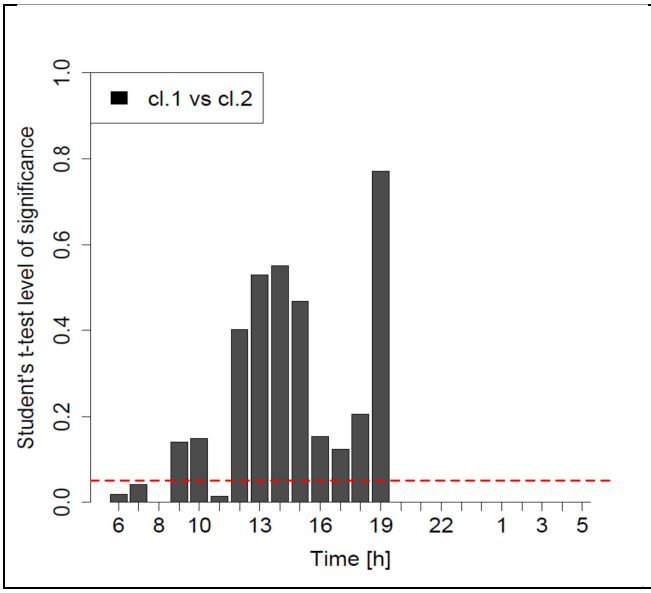


Fig. 11 - Student's *t*-test results for the two clusters.

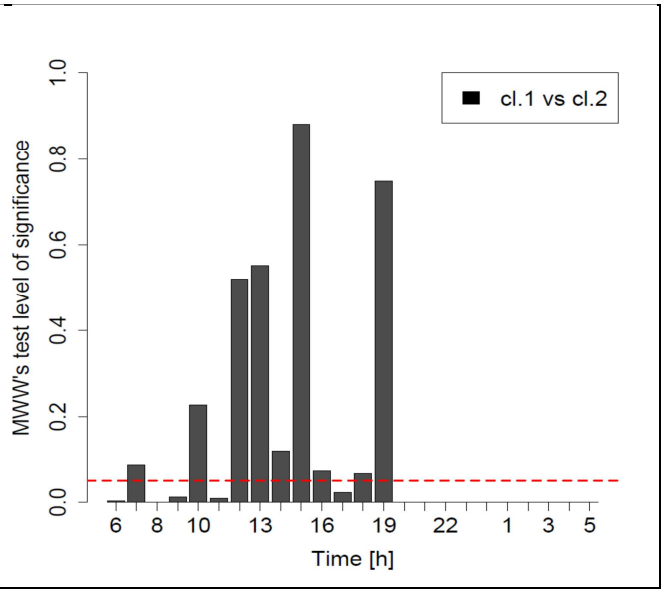


Fig. 12 - Mann–Whitney–Wilcoxon's test results for the two clusters.

We think that a new criterion should be developed for a more realistic classification of streets from the point of view of their acoustic features. In fact, the present legislation on traffic noise limitations are based on the standard acoustic classification of streets which is based on their geometric and physical properties, rather than on the effective traffic flow. The possibility that this new scenario could be actually applied in future legislations is a by-product of the present investigation.

4.3. Stratified spatial sampling

In stratified spatial sampling the sample is split up into strata (sub-samples) in order to decrease variances of sample estimates, to use partly non-random methods applied to sub-groups or clusters or to study strata individually [14].

The central limit theorem states that, given a sufficiently large sample size *n* from a population of size *N* with a finite level of variance σ^2, the mean of all samples $\bar{x}$ from the same population will be approximately equal to the mean of the population. Furthermore, all of the samples will follow an approximate normal distribution pattern, with all variances σ_x^2 being approximately equal to the variance of the population divided by each sample's size: $\sigma_x^2 = \frac{\sigma^2}{n}$. The maximum error *E*, that is the largest expected deviation of the sample mean from the population mean with the stated confidence level *1-α* (for *1-α* = 95%, the amplitude of the Gaussian distribution is $z_\alpha \cdot \sigma_x$ with $z_\alpha = 1.96$) is:

$$\varepsilon = z_\alpha \cdot \sigma_x \qquad (2)$$

The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ± *ε* is:

$$n_{\min} = \frac{z_\alpha^2 \cdot s^2}{\varepsilon^2} \qquad (3)$$

where *s* is the sample standard deviation.
In our case, the overall sample size consists of 203 measurements (to be divided into each class: functional or cluster). We fixed the maximum error ε = 1 dB and we calculated the minimum number of elements n_{min}. The results are reported in Figs. (13-18). As we can clearly observe, the computed minimum number of sample elements n_{min} for a correct estimation of the mean of the population within an accuracy ± ε*= 1 dB* is very well respected for the cluster categorization. In fact, it remains well below the number of performed measurements in the case of 5, 15 and 60 minutes (with just the exception of the 5 min. discretization in the night hours between 2 and 4 am). The most important thing is that the cluster approach to road categorization is much more efficient than the functional one as, in the latter case, the minimum number of elements n_{min} overcomes the number of performed measurements over a large portion of the day. Shown in each figure by dashed lines is the number of performed measurements.

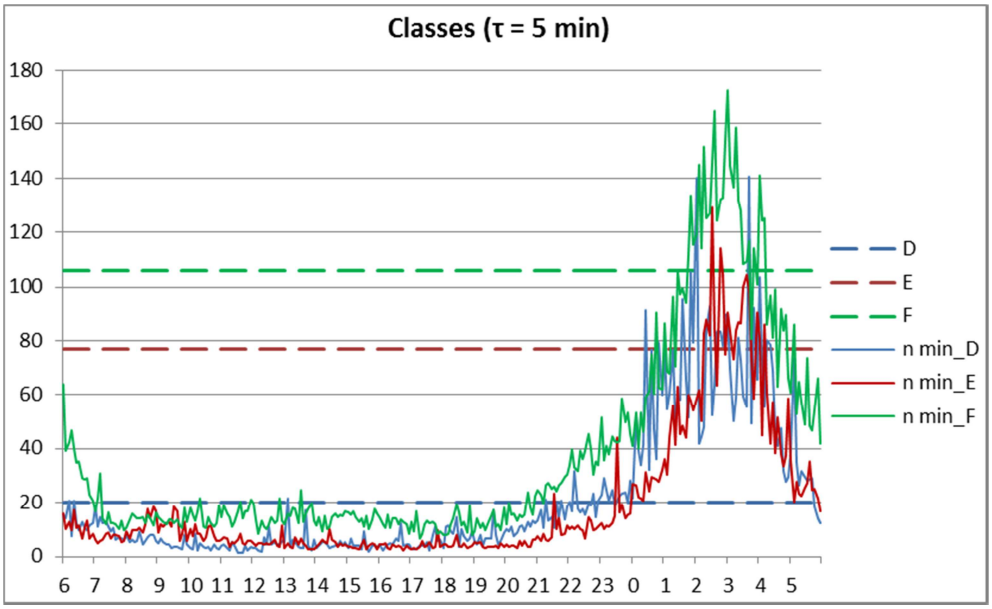


Fig. 13 - The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ε *=1 dB,* for the case of functional categorization of roads and with a temporal discretization of 5 minutes. The number of performed measurements in each class is shown by dashed lines.

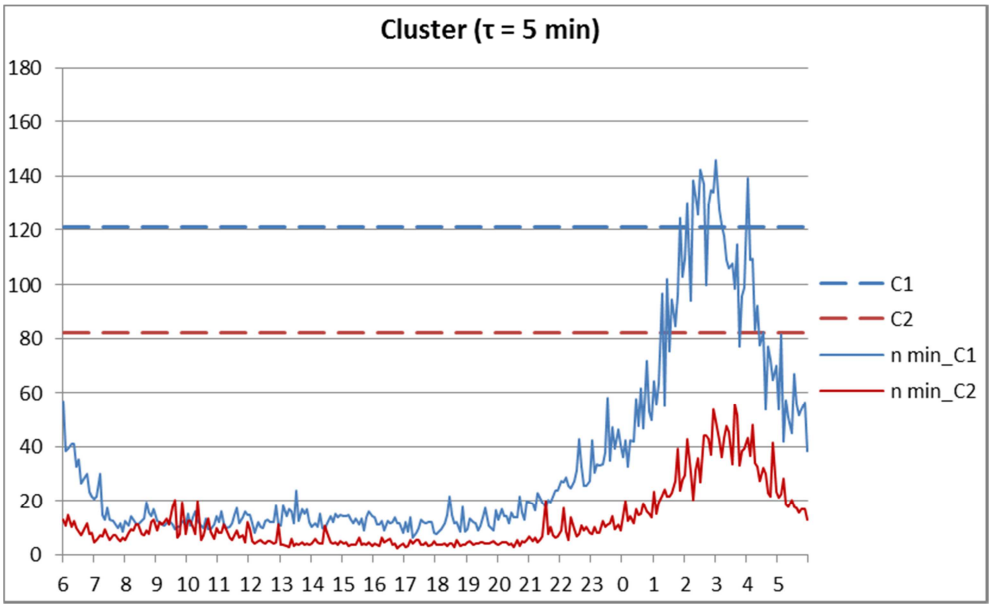


Fig. 14 - The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ε *=1 dB,* for the case of cluster categorization of roads and with a temporal discretization of 5 minutes. The number of performed measurements in each cluster is shown by dashed lines.

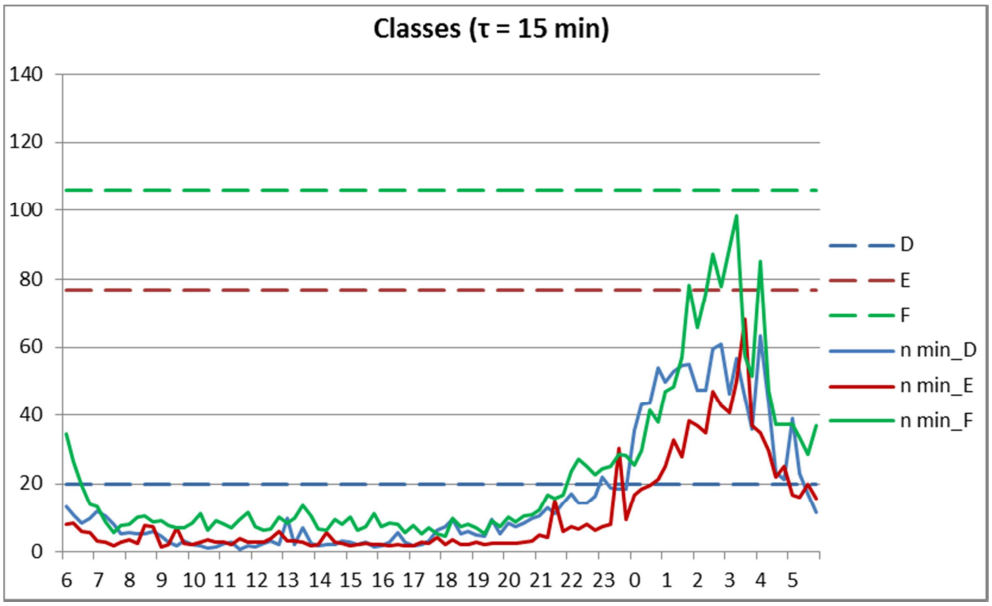


Fig. 15 - The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ε *=1 dB,* for the case of functional categorization of roads and with a temporal discretization of 15 minutes. The number of performed measurements in each class is shown by dashed lines.

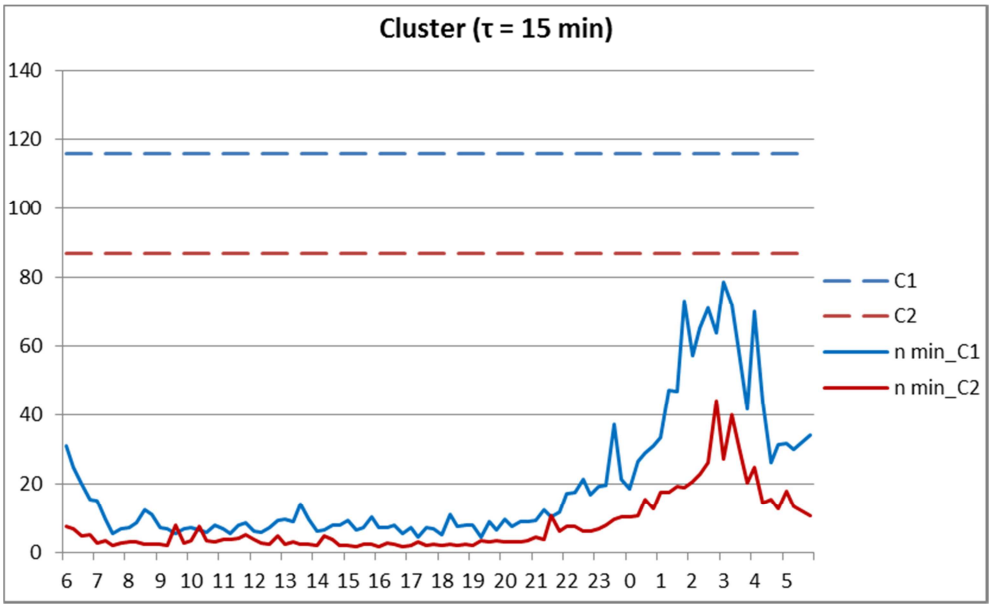


Fig. 16 - The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ε *=1 dB,* for the case of cluster categorization of roads and with a temporal discretization of 15 minutes. The number of performed measurements in each cluster is shown by dashed lines.

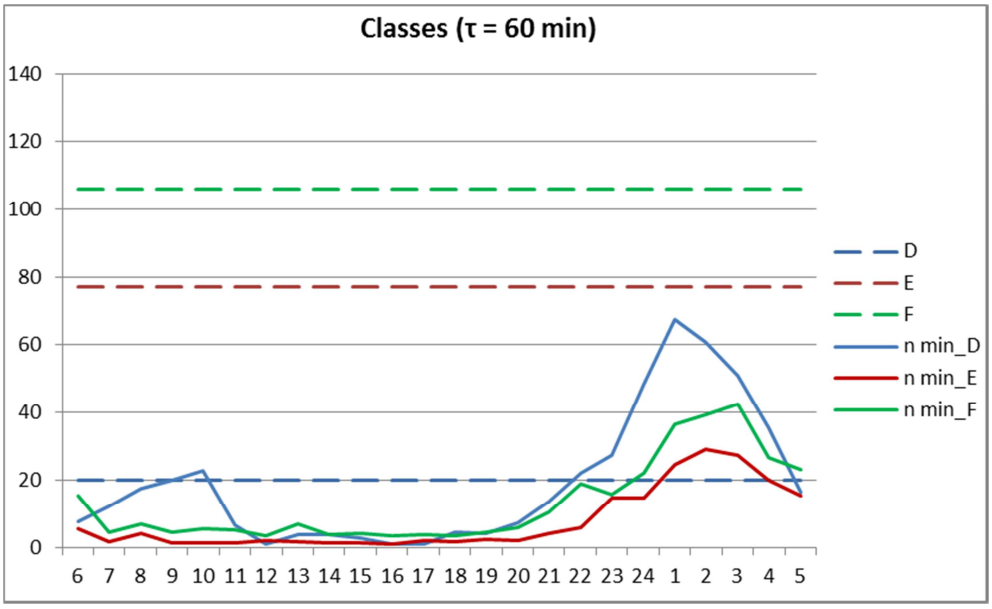


Fig. 17 - The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ε =1 dB, for the case of functional categorization of roads and with a temporal discretization of 60 minutes. The number of performed measurements in each class is shown by dashed lines.

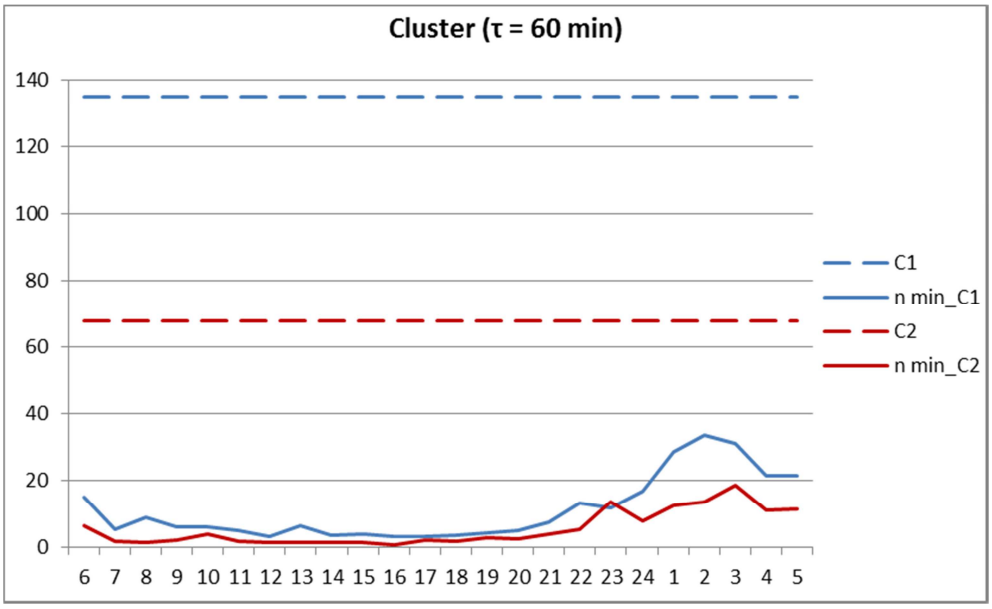


Fig. 18 - The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ε =1 dB, for the case of cluster categorization of roads and with a temporal discretization of 60 minutes. The number of performed measurements in each cluster is shown by dashed lines.

In order to build up a dynamic map, it is necessary that each road belonging to the pilot area could be attributed to one of the two clusters

In fact, unlike the functional classification of roads, the two obtained cluster profiles cannot be applied straightforward without any indication linking them to a specific feature. Such limitation can be overcome by associating each mean cluster profile with a corresponding "non-acoustic parameter". As a first application, traffic flow rate at rush hour has been considered as a non-

acoustic parameter for each of the 93 roads. In fact, the rush hour represents one of the hours for which the two mean cluster profiles results independent. The traffic data were provided by the AMAT agency, in charge of the traffic mobility management of the city of Milan[1]. Figure 19 shows the box plots of the vehicle flow rate at rush hour for the two-mean-cluster profiles. In particular, their interquartile range does not overlap. Thus, a vehicular flow rate at rush hour of 1500 vehicles/hour can be considered as threshold between the two profiles, that is roads featuring lower values (< 1500 vehicles/hour) can be associated with cluster 1, whereas higher flow rates (> 1500 vehicles/hour) can be allocated to cluster 2.

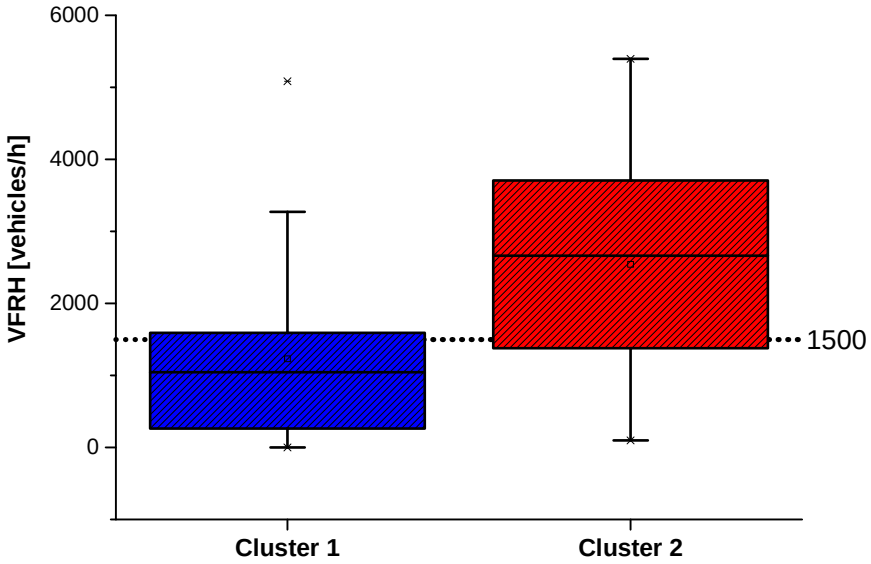


Fig. 19 - Box plots of the vehicle flow at rush hour (VFRH) for the two mean cluster profiles. Roads with traffic flow rate at rush hour VFRH < 1500 vehicles/hour can be associated with Cluster 1; roads with traffic flow rate at rush hour VFRH > 1500 vehicles/hour to Cluster 2.

In order to determine a possible indication on how to search for one a more efficient non-acoustical parameter or a combination of them, we proceeded calculating the correlation coefficient between different traffic flows and the corresponding measured traffic noise as a function of the daily hours each for cluster, 1 and 2 (Fig. 20). We considered for this analysis the following non-acoustic parameters:

- Normal Traffic Flow (total number of vehicles per hour)
- Equivalent Traffic Flow (one heavy vehicle is equivalent to 8 light vehicles in terms of produced noise)
- Log Equivalent Traffic Flow
- Log Normalized Equivalent Traffic Flow (the equivalent traffic flow is normalized with its maximum value)

[1] *http://www.amat-mi.it/it/mobilita/dati-strumenti-tecnologie/

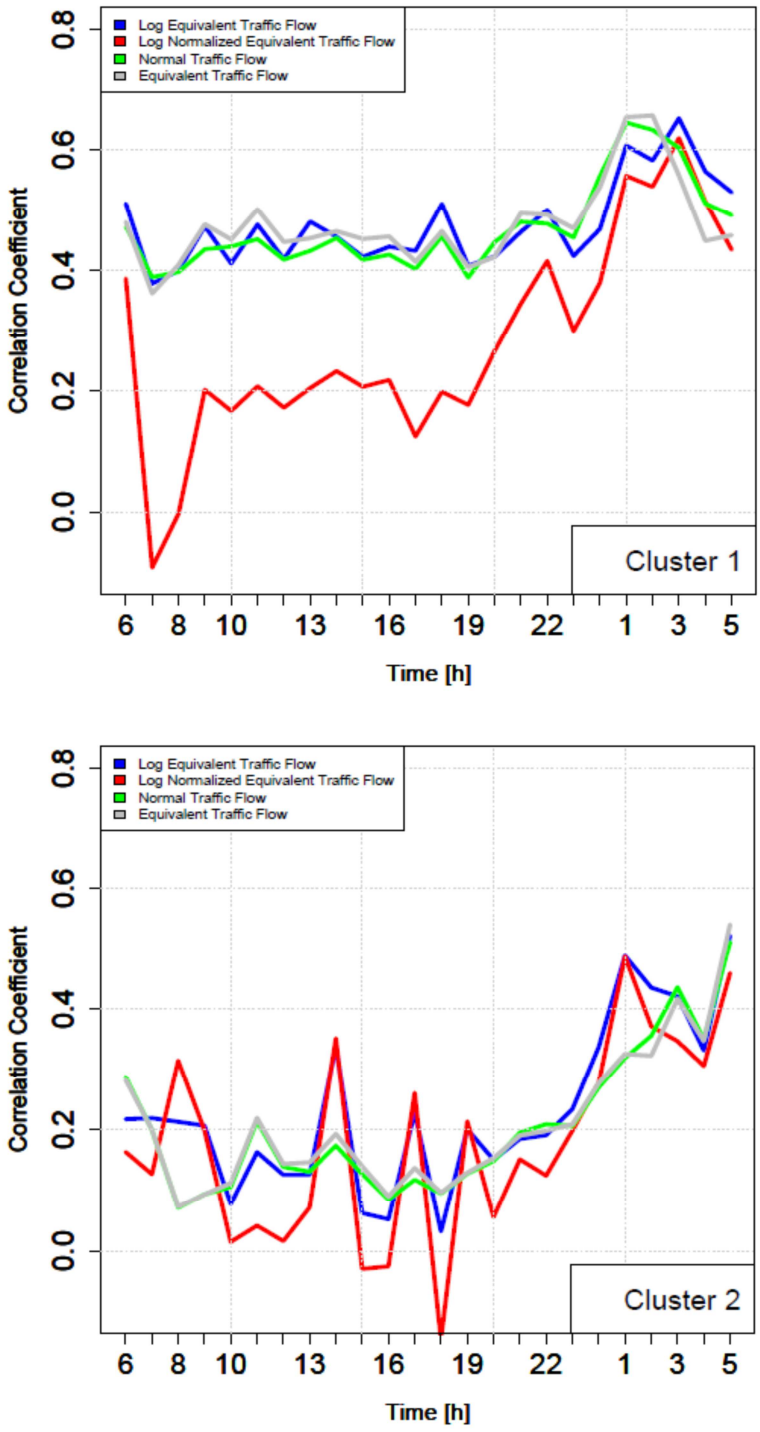


Fig. 20 - Correlation coefficients between traffic flows and noise for the clusters, 1 and 2, as a function of daily hours. The flows considered are, from top to bottom: Log Equivalent Traffic Flow (blue color), Log Normalized Equivalent Traffic Flow (red color), Normal Traffic Flow (green color), and Equivalent Traffic Flow (grey color).

Cluster 1 presents higher correlation coefficients for the Normal Traffic Flow , the Equivalent Traffic Flow and the Log Equivalent Traffic Flow. Cluster 2 presents quite low correlation coefficients for all the non-acoustic parameters. However, also in this case, the Log Normalized Equivalent Traffic Flow presents the worse performance.

Afterward, we calculated the difference between the results obtained for cluster 1 and 2, choosing as non-acoustic parameters those showing the higher difference between the two clusters (Fig. 21). In this case, as our purpose is to obtain the best separation between the two clusters, we assumed as" best performant" parameters those presenting a "high" correlation with one cluster and a "low" correlation with the other. As is clear in Fig. 21, from these preliminary results we can definitely discard the parameter Log Normalized Equivalent Traffic Flow. As for the choice of the "optimal" non-acoustic refer to paragraph 5.

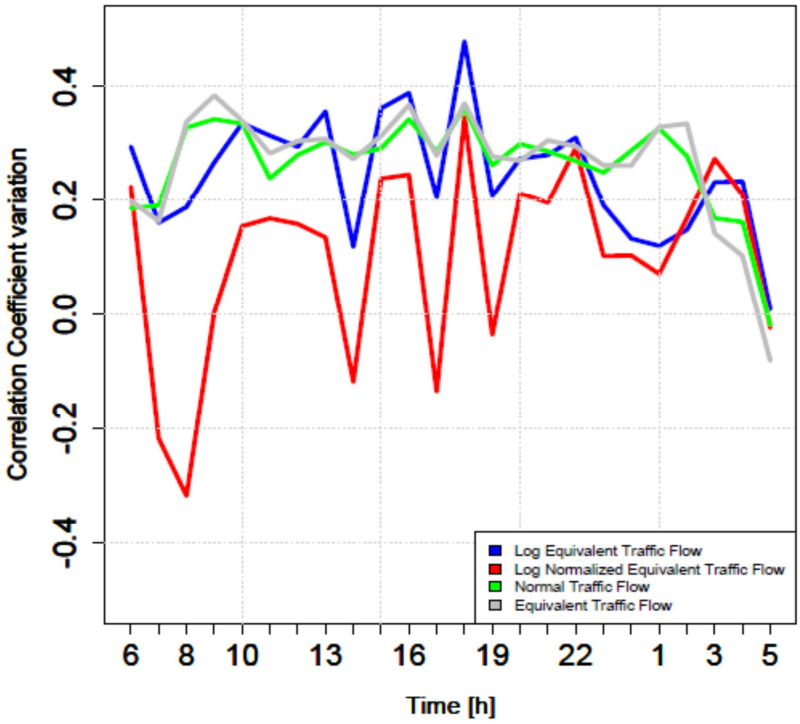


Fig. 21 - Correlation coefficient variation between the clusters, 1 and 2, as a function of daily hours. The flows considered are, from top to bottom: Log Equivalent Traffic Flow (blue color), Log Normalized Equivalent Traffic Flow (red color), Normal Traffic Flow (green color), and Equivalent Traffic Flow (grey color).

4.4 Comparative analysis among profiles of different temporal discretization

Another necessary issue related to noise mapping regards the smallest time interval a noise map can be updated without losing significant information from the original data (hourly levels). To this purpose, we extracted five new level profiles with temporal resolution of 30, 20, 15, 10, 5 minutes. For this analysis we considered 95 hourly noise profiles.

Each level profile with different temporal discretization was statistically analyzed and the results of the mean values, $\overline{\delta}_{ik}$, (i= 1,….,24*(60/Integr.Time); k= 1, 2) the for cluster 1 and 2 are shown in the Figs. (22-23) below.

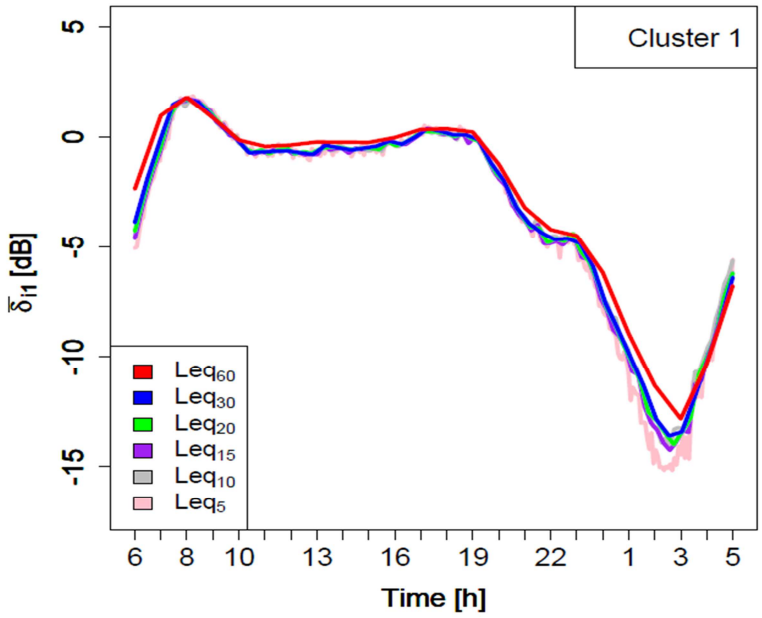


Fig. 22 – Representation of the mean profiles $\overline{\delta}_{i1}$ for cluster 1 with different temporal discretization as a function of daily time.

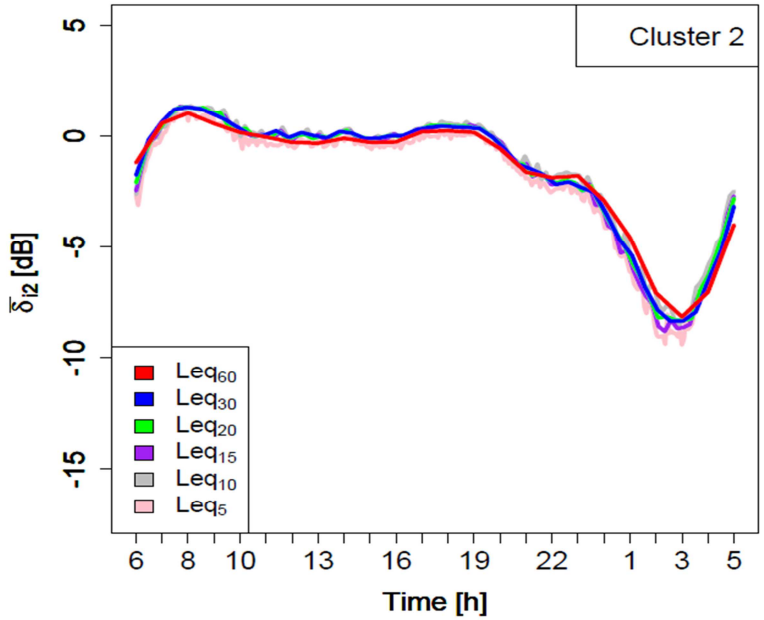


Fig. 23 – Representation of the mean profiles $\overline{\delta}_{i2}$ for cluster 2 with different temporal discretization as a function of daily time.

The mean profiles with different temporal discretization present similar trends for both clusters. Higher oscillations are observed for smaller temporal intervals especially in night period. The composition of the clusters remain essentially similar to the hourly data, as reported in Table 4. Here, we report the temporal discretization for time intervals: (60, 30, 20, 15, 10, 5) min. The second column refers to the two clusters, 1 and 2, while the central ones display the cluster compositions in terms of the road category; finally, the last column to the right reports the total number of elements in each cluster.

This result suggests that our description in terms of noise clusters (1 and 2) is robust and appropriate for classifying roads in large urban areas also with a shorter integration time.

Table 4 - Composition of clusters for different temporal discretization.

TEMPORAL DISCRETIZATION [min]	CLUSTER	ROAD CATEGORY			TOTAL
		D	E	F	
60	1	5 (8,0%)	12 (20,0%)	43 (72,0%)	60
	2	4 (11,0%)	18 (51,0%)	13 (37,0%)	35
30	1	5 (9,3%)	10 (18,5%)	39 (72,2%)	54
	2	4 (9,8%)	20 (48,8%)	17 (41,5%)	41
20	1	5 (8,9%)	11 (19,6%)	40 (71,4%)	56
	2	4 (10,3%)	19 (48,7%)	16 (41,0%)	39
15	1	5 (9,1%)	10 (18,2%)	40 (72,7%)	55
	2	4 (10,0%)	20 (50,0%)	16 (40,0%)	40
10	1	5 (7,6%)	15 (22,7%)	46 (69,7%)	66
	2	4 (13,8%)	15 (51,7%)	10 (34,5%)	29
5	1	5 (8,5%)	11 (18,6%)	43 (72,9%)	59
	2	4 (11,1%)	19 (52,8%)	13 (36,1%)	36

4.5 Error Analysis at Different Temporal Discretization

In Figure 24 and 25 we report the standard deviations [dB] as a function of daily time for the different time intervals for all arches considered, and the mean standard deviation [dB] for each temporal discretization, respectively. The highest values are found during the night time and the “high frequency” temporal discretization (up to 6 dB).

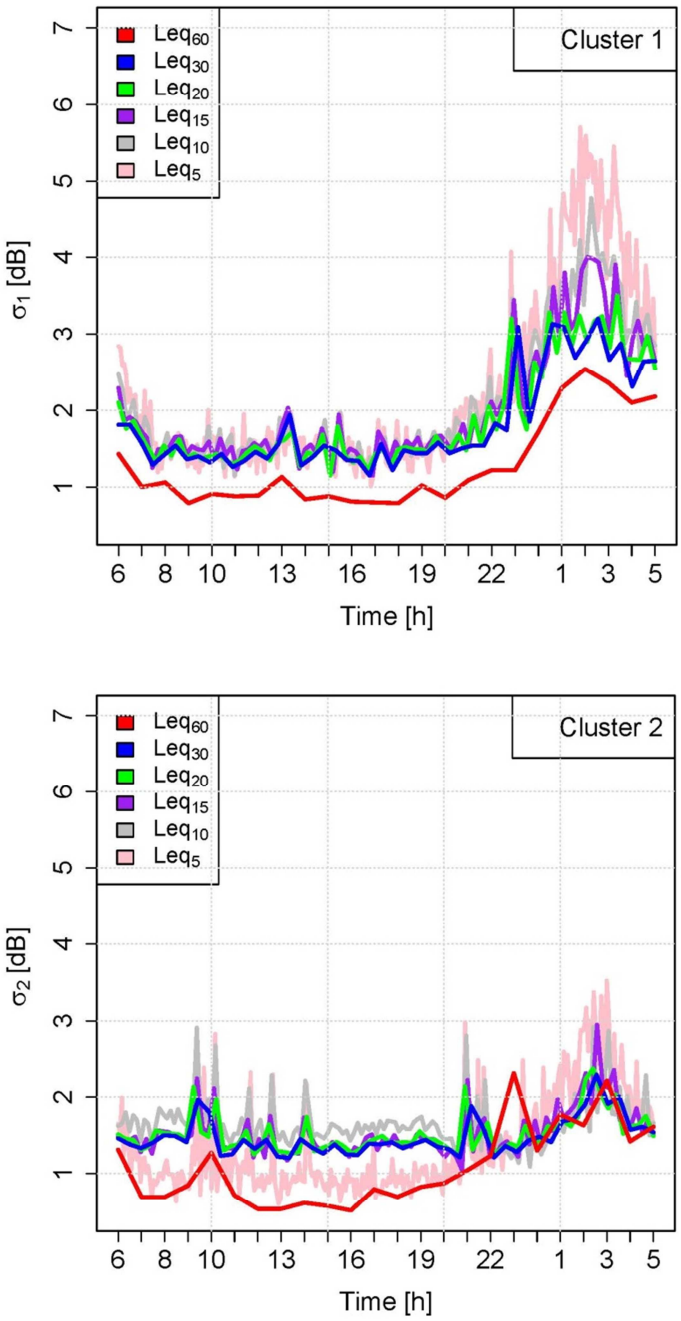


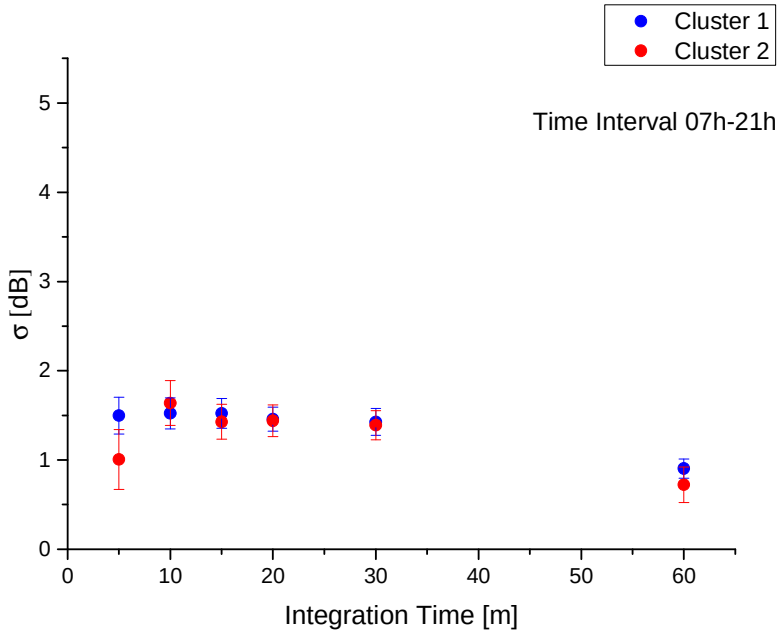
Fig. 24- Standard deviations [dB] as a function of daily time for the different time intervals for all arches considered for cluster 1 and 2 .

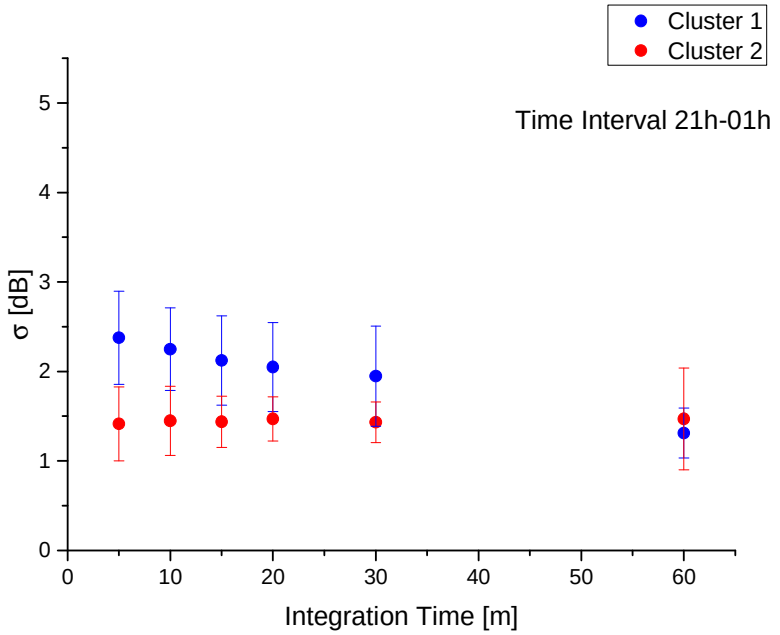
According to Fig. 24, we can derive a few suggestions on how to choose a proper temporal representation of the dynamic map maintaining a reasonable error. For this reason we opted for the following division of daily time:

1. 07h-21h
2. 21h-01h
3. 01h-07h

In Fig. 25 and Table 5 we present the mean standard deviations as a function of the integration time and for the different time intervals. We can clearly recognize that for the time period 07h-21h we can

choose as updating time the 5 minutes discretization. It presents a standard deviation below 1.5 dB comparable with the 60 minutes integration time. For the time interval 21h-01h an updating time of 15 minutes seems to represent a good compromise between a reasonable integration time and accuracy. For the nighttime interval we can increase the integration time to one hour. This choice has the advantage to keep the uncertainty quite constant (below σ=1.5 dB with "variability" Δσ=0.3 dB for daytime, σ=2.1 dB and Δσ=0.5 dB for evening time and σ=2.1 dB and Δσ=0.4 dB for nighttime) during the night, evening and day period acting on a proper integration time.





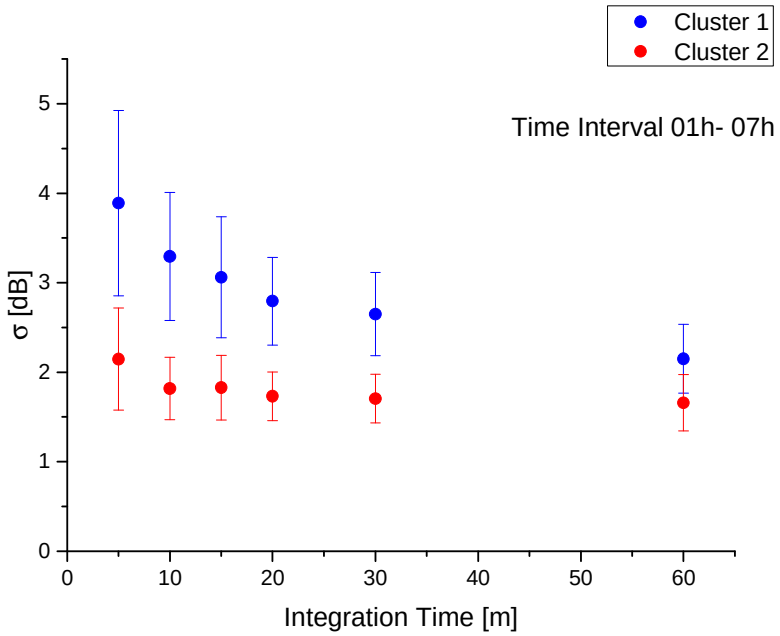


Fig. 25 - Mean standard deviations, σ, [dB] as a function of integration time for the different time intervals.

Table 5 – Mean standard deviations, σ, (dB), and its variability, Δσ, (dB) for different period of the day and for different temporal discretization.

	Integration Time	σ[07h-21h]	Δσ	σ [21h-01h]	Δσ	σ[01h-07h]	Δσ
CLUSTER 1	5	1,497	0,206	2,377	0,520	3,889	1,035
	10	1,522	0,172	2,249	0,462	3,294	0,715
	15	1,523	0,168	2,123	0,500	3,060	0,675
	20	1,457	0,136	2,049	0,497	2,794	0,490
	30	1,427	0,150	1,947	0,560	2,648	0,465
	60	0,904	0,108	1,312	0,278	2,150	0,384
CLUSTER 2	5	1,007	0,336	1,415	0,412	2,147	0,570
	10	1,640	0,250	1,448	0,387	1,818	0,348
	15	1,428	0,195	1,438	0,287	1,828	0,361
	20	1,439	0,177	1,470	0,248	1,732	0,271
	30	1,390	0,163	1,432	0,227	1,705	0,272
	60	0,723	0,198	1,470	0,571	1,658	0,316

4.6 Error Analysis for different cluster composition.

As one can see from Table 4, clusters at different time intervals do not have the same composition as for the clusters obtained at the time interval of one hour. In the following, we quantify the difference between the mean noise measurements obtained from the hourly Clusters, $\delta_{\tau,i}(C_n)$, and from those obtained by their specific clusters, $\delta_{\tau,i}(C'_n)$, at the considered time interval τ (integration time). This is the maximum error you would obtain in case of erroneous attribution of an arch in a cluster, with

respect to the allocation that would occur with the hourly cluster. These are shown in Figs. (26-30) for τ = (30, 15, 5) min. We estimate the mean error by evaluating the mean square differences at each time i, according to:

$$\varepsilon_\tau^2 = \frac{1}{N}\sum_{h=1}^{N}[\,\delta_{\tau,h}(C_n) - \delta_{\tau,h}(C'_n)]^2. \qquad \text{Eq. (4)}$$

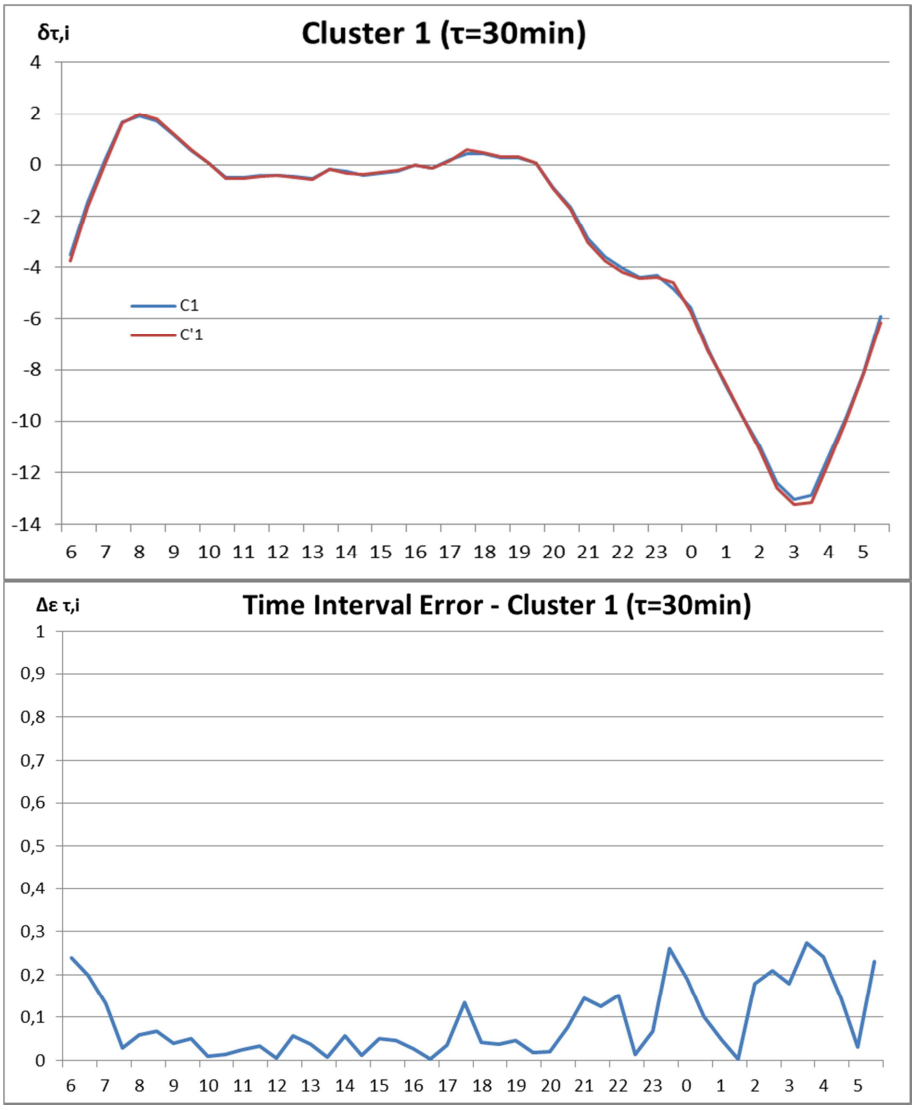


Fig. 26 - Time variation of $\delta_{\tau,i}(C_1)$ and $\delta_{\tau,i}(C'_1)$ for τ=30 min as a function of time i (upper panel). As one can see from the lower panel the maximum deviations are of the order of 0.3 dB. The mean error is ε_1=0.118.

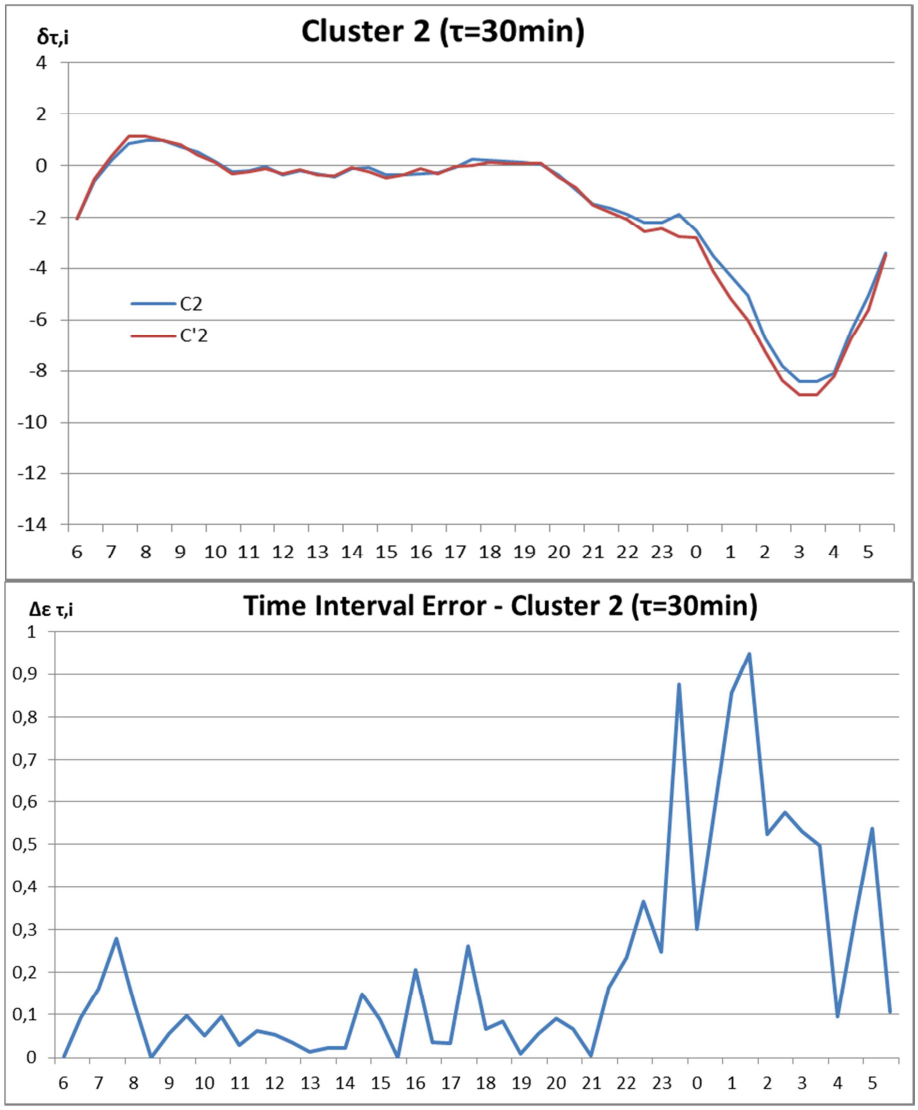


Fig. 27 - Time variation of $\delta_{\tau,i}(C_2)$ and $\delta_{\tau,i}(C'_2)$ for τ=30 min as a function of time i (upper panel). As one can see from the lower panel the maximum deviations are of the order of 0.7 dB. The mean error is ε_2=0.323.

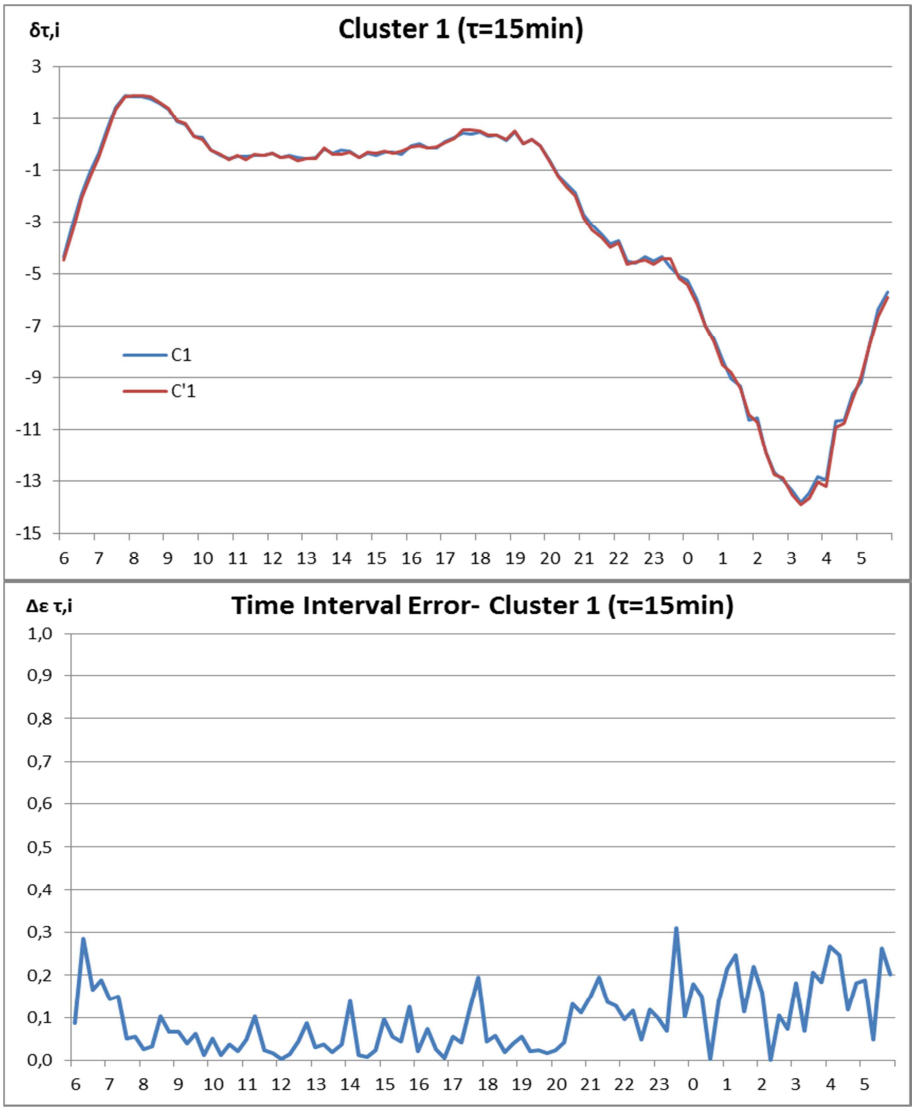


Fig. 28 - Time variation of $\delta_{\tau,i}(C_1)$ and $\delta_{\tau,i}(C'_1)$ for τ=15 min as a function of time i (upper panel). As one can see from the lower panel the maximum deviations are of the order of 0.4 dB. The mean error is ε_1=0,121.

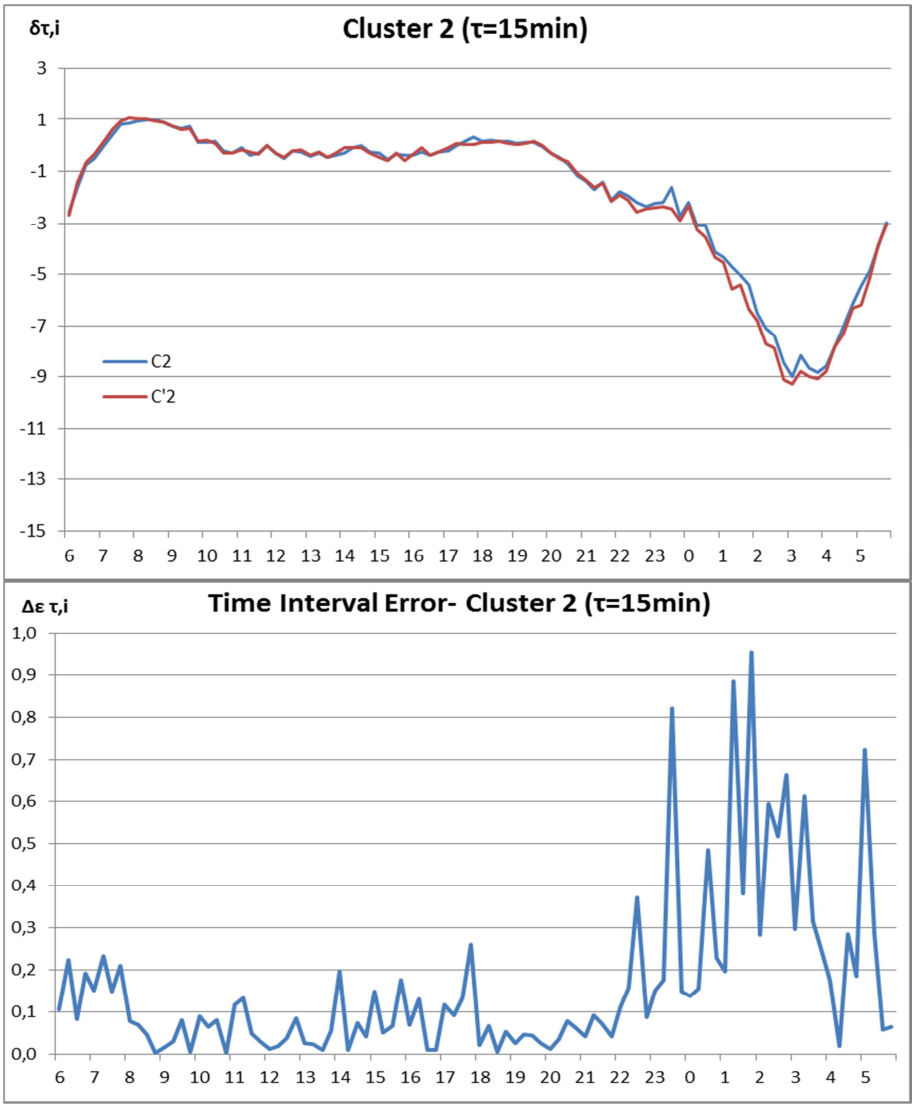


Fig. 29 -Time variation of $\delta_{\tau,i}(C_2)$ and $\delta_{\tau,i}(C'_2)$ for τ=15 min as a function of time i (upper panel). As one can see from the lower panel the maximum deviations are of the order of 0.8 dB. The mean error is ε_2=0,256.

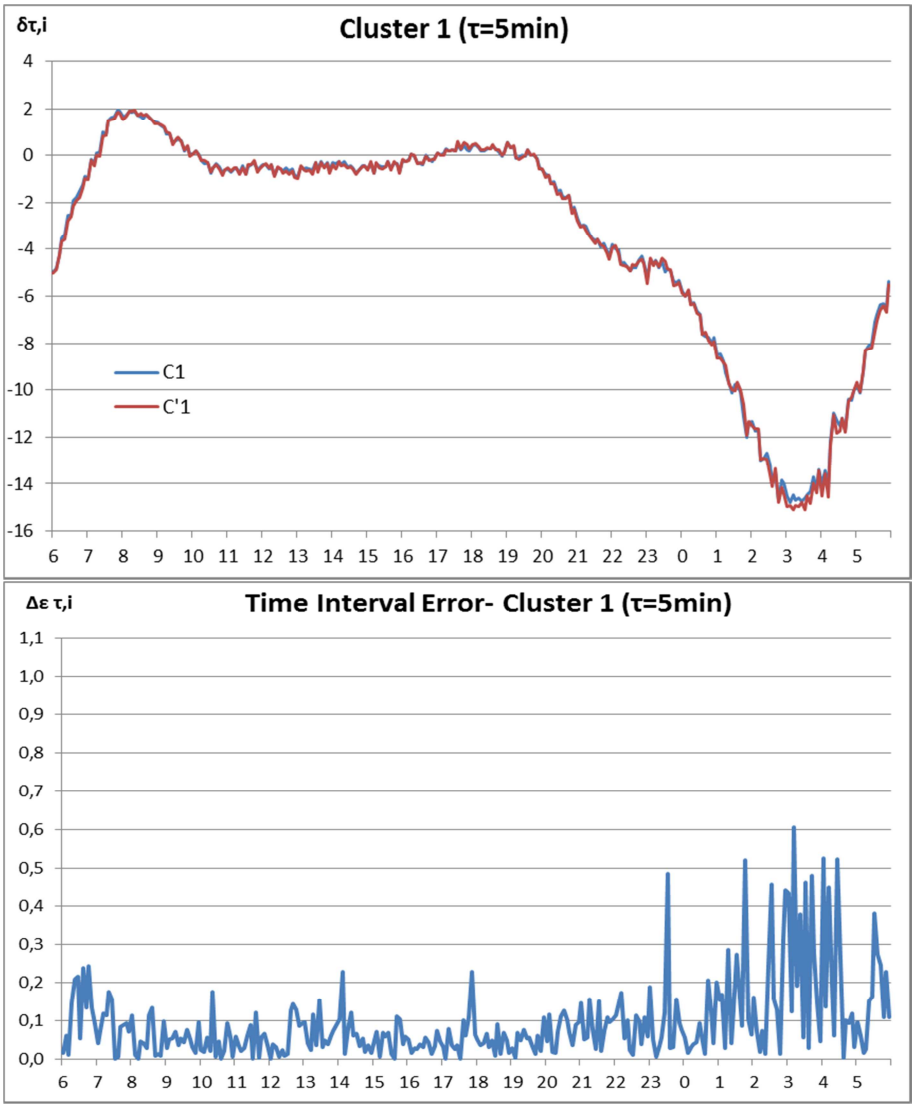


Fig. 30 - Time variation of $\delta_{\tau,i}(C_1)$ and $\delta_{\tau,i}(C'_1)$ for τ=5 min as a function of time i (upper panel). As one can see from the lower panel the maximum deviations are of the order of 0.7 dB. The mean error is ε_1=0.140.

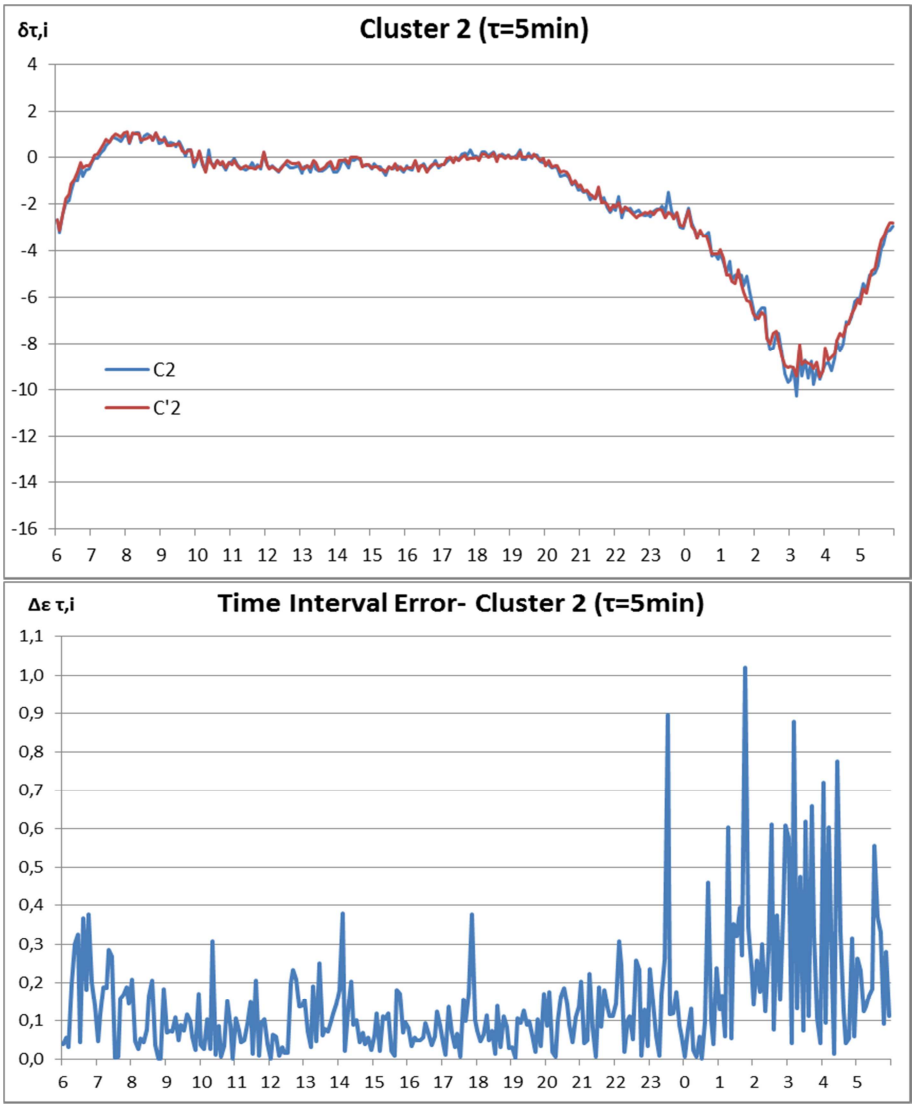


Fig. 31 -Time variation of $\delta_{\tau,i}(C_2)$ and $\delta_{\tau,i}(C'_2)$ for τ=5 min as a function of time i (upper panel). As one can see from the lower panel the maximum deviations are of the order of 0.9 dB. The mean error is ε_2=0.214.

These results suggest that despite the different cluster compositions at different time intervals the hourly cluster structure is quite robust and can be used for predicting the traffic noise behavior at shorter time intervals such as 30, 15 and 5 mins. In Table 6 we report the mean errors $\bar{\varepsilon}$ and the corresponding standard deviations over the 24 hours, while in Table 7 the errors are splitted into two time zones, the (07h-21h) and (21h-01h) for comparison.

Table 6 - The mean error, $\bar{\varepsilon}$, and its deviation, σ, calculated considering the 24 hours, for different time interval τ , as a result of the different cluster compositions.

Cluster	$\bar{\varepsilon}$ (τ=5min)	σ	$\bar{\varepsilon}$ (τ=15min)	σ	$\bar{\varepsilon}$ (τ=30min)	σ
1	0,140	0,104	0,121	0,075	0,118	0,080
2	0,214	0,157	0,256	0,199	0,323	0,247

Table 7 - The mean error, $\bar{\varepsilon}$, and its deviation, σ, calculated for the two hourly periods [07h-21h] and [21h-01h] for the time intervals τ=[5,15]min, as a result of the different cluster compositions.

Cluster	$\bar{\varepsilon}$ (τ=5min) [07h-21h]	σ	$\bar{\varepsilon}$ (τ=15min) [21h-01h]	σ
1	0,072	0,043	0,144	0,068
2	0,117	0,071	0,283	0,202

5. AGGREGATION OF THE ROAD STRETCHES

The main object of aggregation is to find a suitable set of roads which display similar traffic noise behavior so that one can group them together into a single noise map. All the road stretches inside a group will be described by the same noise map variations. Typically, one expects to deal with few such maps, which can be estimated in a number between 4 and 6 maps. The previous analysis of noise suggests that noise can be clustered quite suitably into just two main groups. Of course, this means that in a first approximation, we could deal with just two groups of stretches. This appears to be quite a drastic approximation and therefore we wish to find a method which can generalize this result but still keeping this simple description valid as the starting point. By this we mean that noise measurements will be taken on few stretches inside the zone of interest, and clustered together according to the scheme discussed in Part 4. Once we have obtained the mean behavior of noise inside each one of the two clusters, we can use this result in order to 'interpolate' the behavior of noise between the two regimes. The interpolation scheme will be discussed in detail in this section. In order to be able to predict the behavior of noise for those stretches for which the traffic noise is not known, we look at the corresponding values of hourly traffic flows, which are known for all stretches in the urban zone. Thus the problem is to find a map between traffic flow and noise. This implies the need to introduce a suitable 'non-acoustic' parameter which should accurately represent the traffic flow and be sufficiently correlated with the corresponding traffic noise level. The choice of the non-acoustic parameter relies in finding those hours which better represent the behavior of the flow inside each one of the groups we are looking for.
As we have seen in Part 4, values of the flow at the rush hour (08:09) can be considered as the first choice for a non-acoustic parameter, also representing a good discriminant for the different behaviors of noise. We have seen that the corresponding mean flows inside each one of the two reference clusters (the noise clusters 1 and 2) yield box plots which essentially do not overlap and have a quite definite threshold (i.e. 1500 vehicles/hour). In order to estimate the robustness of this choice, we have studied several definitions of non-acoustic parameters, hereafter denoted by the letter x. Thus, the rush hour flows is just one possible choice for x.

5.1 The non-acoustic parameter and the interpolation method between two mean noise behaviors

The non-acoustic parameter can be chosen by considering two (or more) hours during the day, and not just one such as the rush hour. For illustration of the method, we consider the following definition of the non-acoustic parameter, x, which combines values of the logarithm of the traffic flows, $F_{i,j}$, at different hours (i,j), added according to the formula,

$$x = \sqrt{(LogF_i)^2 + (LogF_j)^2} \qquad \text{Eq. (5)}$$

with F_i and F_j being the flows at the ith and jth hours. The use of the logarithm of the flow is dictated by the need to deal with values which may differ appreciably between each other, such as for instance the morning and night traffic flows. The choice above for the sum of squares resembles the standard procedure of considering an Euclidean metric, and we have found it suitable for our clustering problems. In other words, the (x=Log F) acts as a coordinate in flow space, and by considering two coordinates implies the use of a two-dimensional space for representing each stretch, which can be easily visualized on a two dimensional plot. One can add more hours to the definition of x in Eq.(5) and several possibilities emerge. We have studied several combination of pairs of hours, and triplets of hours too. We summarize the results later below.
In the following, we report results obtained by combining the rush hour (8:00-9:00) with the night hours (21:00-22:00). We have considered the two clusters obtained in Part 4 for the noise measurements. For each cluster, we have calculated the values of x according to Eq.(5) and studied their distribution functions P(x). We are interested in finding an analytical representation for both P(x) and therefore have studied the corresponding cumulative distributions of x, which have been fitted using an analytical expression $I(x) = 10^{f(x)}$, where f(x) is a polynomial of third degree. The results of I(x) for Cluster 1 are reported in Fig. 32.

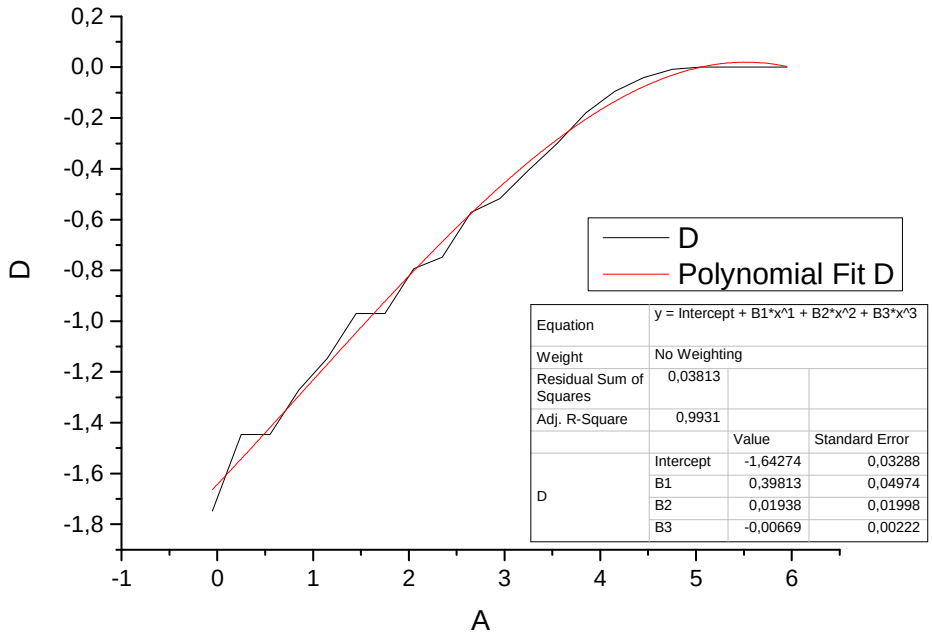


Fig. 32 - The cumulative distribution I(x) is fitted using the analytical expression $I(x) = 10^{f(x)}$, where f(x) is a polynomial of third degree. In the plot we have used the notation A=x and D=I(x).

The probability distribution P(x) can be obtained from the analytical fit of the cumulative distribution I(x) according to the relation: P(x) = (Ln10)*f'(x)*I(x), where f'(x) is the derivative of f(x). In Fig. 33 we compare the histograms of x with the analytical fit. We can see from Fig. 33 that both distribution function overlap significantly. This behavior is typical of any chosen non-acoustic parameter, and therefore it is an intrinsic property of x for the present problem, even for the case in which we consider the rush hour, x=F(08:09). The conclusion is that one cannot separate the stretches in an obvious way by just looking at their distribution functions. However, we can find an approximation scheme which may interpolate between the two distributions. The idea behind this interpolation scheme is that we do not associate each stretch to just one of the two noise clusters, 1 or 2, rather we determine the probability that a single stretch belongs to each one of them. Thus, as a result we

find an interpolation scheme which provides us with a prediction of the temporal behavior of the noise for an arbitrary road stretch.

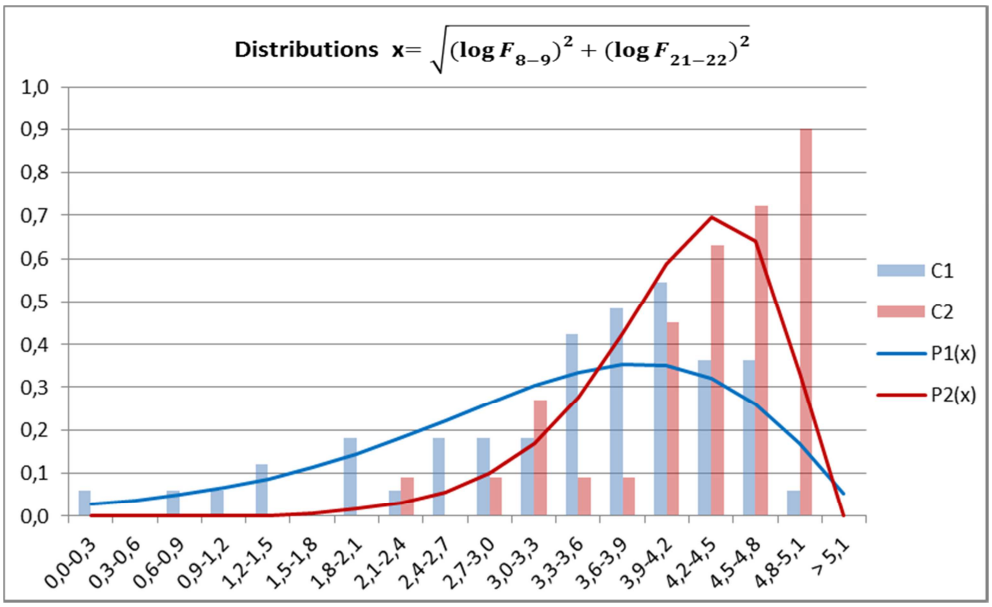


Fig. 33 - Distribution functions P(x) versus non-acoustic parameter x for Cluster 1 and 2. The histograms of x are represented by the vertical bars, while the analytical fits by the continuous lines obtained using the procedure explained in Fig. 32.

The calculus scheme is therefore:
$P_1(x)$ = number of events for which x falls within the interval (x_k, x_{k+} Δ) in cluster 1
$P_2(x)$ = number of events for which x falls within the interval (x_k, x_{k+} Δ) in cluster 2

Due to the large superposition of the two cluster distributions $P_1(x)$ and $P_2(x)$ shown in Fig. 33, we consider a linear combination between the two time dependencies of the traffic noise from each cluster, for the given value of x. The weights (α_1, α_2) of the linear combination can be obtained for each value of x using the relations: $\alpha_1=P_1(x)$ and $\alpha_2=P_2(x)$. That is, for a given value of x we find the 'components' α_1 and α_2 from the analytical expressions of P(x) discussed above. The values of $\alpha_{1,2}$ represent the 'probability' that a given road characterized by its own value of x belongs to the corresponding Cluster, 1 and 2. As one can see, we do not consider a sharp threshold for x, based for instance on the VFRH (vehicular flow rush hour), but the resulting hourly behavior h of the noise for that road is a linear combination of the mean noises measured for Cluster 1 and 2, denoted respectively as $\delta_{C1}(h)$ and $\delta_{C2}(h)$. The predicted traffic noise behavior, $\delta_{pred}(h)$, of a given value of x is then obtained by normalizing the values of $\alpha_{1,2}$ denoted as β:

$$\beta_1 = \frac{\alpha 1}{\alpha 1+\alpha 2} \quad \text{and} \quad \beta_2 = \frac{\alpha 2}{\alpha 1+\alpha 2}. \qquad \text{Eq. (6)}$$

Using the values of β we can predict the hourly variations δ(h) for a given value of x according to,

$$\delta_{pred}(h) = \beta_1 * \delta_{C1}(h) + \beta_2 * \delta_{C2}(h) \qquad \text{Eq. (7)}$$

with $\delta_{C1}(h)$ and $\delta_{C2}(h)$ representing the mean values of the noise (Fig. 8) for each Cluster 1 and 2, respectively. The error made in using Eq. (7) can be estimated by calculating the standard deviation ε of the prediction $\delta_{pred}(h)$ from the measured values $\delta_{meas}(h)$, that is

$$\varepsilon^2 = \frac{1}{24} \sum_{h=1}^{24} [\, \delta_{meas}(h) - \delta_{pred}(h)]^2 \qquad \text{Eq. (8)}$$

A weighted average is applied using the two mean cluster results $\delta_{1,2}(h)$ (Eq. 7), and in this way the best possible approximation for the hourly variation of the stretch having non-acoustic parameter x is obtained. In other words, instead of attributing a single cluster hourly behavior to the road stretch considered, we take a weighted average of the two cluster mean values $\delta_{1,2}(h)$, where the weights represent the probability that the value of x belongs to each one of the clusters. They provide us with a sort of 'projection' of the x value onto each noise cluster (1,2). This is not equivalent to averaging over all stretches, of course for general weights. Only in the particular case in which both weights are equal to ½ the weighted average coincides with the average over all noise measurements, that is without clustering.

In Fig. 34 we display two typical results, one characterized by a small error ε (Road D5) and a second one having a large error value (Road D3). We have repeated the calculation by excluding the roads data, D5 and D3, respectively, in the calculation of the mean noise values for Cluster 1 and 2, in order to test whether the results are biased by their inclusion in the clusters. We did not find any sensitive variation with respect to the results shown here.

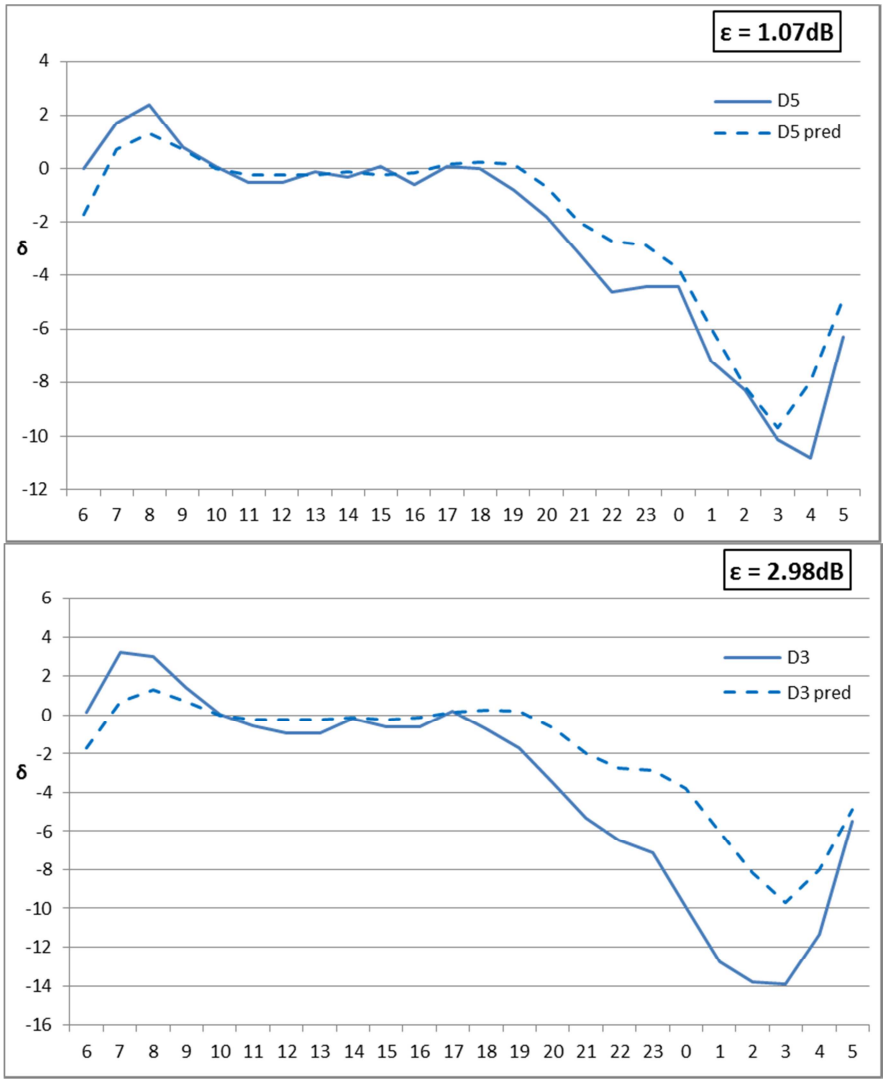


Fig. 34 - Comparison of the noise prediction (dashed line) according to Eq. (6) with the measured values (continuous line) of the traffic noise versus daily hours for roads D5 and D3. The y-axis is in dB and the mean error ε are: ε =1.07 dB (D5) and ε =2.98 dB (D3).

5.2 Different choices for the non-acoustic parameter

Next we consider alternative choices for the non-acoustic parameter, such as $x=T_D$ referring to the daily traffic value (within 06-22), $x=T_N$ the night traffic values (22-06), their logarithmic counterparts, their combinations and the logarithm of the total daily flow T_T (see Table 8).

The mean errors for the corresponding predictions according to Eq. (8) were calculated for all roads belonging to each Cluster, 1 and 2, for different choices of the non-acoustic parameter, and are reported in Table 8, together with the corresponding standard deviations of the errors, and the minimum and maximum errors in each case. The corresponding mean hourly errors are shown in Fig. 35, suggesting that essentially there are no differences in the predictions for the different non-acoustic parameters. The latter display slightly variations during the night hours. In the next section we discuss a quantitative test for further discriminating between the non-acoustic parameters from another point of view.

Table 8 - Different choices of the non-acoustic parameter and the corresponding total errors of the predictions. Here, T_D refers to the daily mean flow value (within 06-22), T_N the night mean values (22-06), and their logarithmic counterparts. The last row reports the logarithm of the total daily flow T_T. Values of mean error ε determined from the 93 measured traffic noises.

Non-acoustic parameter X	Mean error (dB)	σ (dB)	Min error (dB)	Max error (dB)
$x = \sqrt{[(LogF_{8-9})^2 + (LogF_{21-22})^2]}$	1,34	0,57	0,37	2,98
$x = Log\ (T_D)$	1,36	0,57	0,36	2,86
$x = Log\ (T_N)$	1,37	0,59	0,43	2,84
$x = \sqrt{[(LogT_D)^2 + (LogT_N)^2]}$	1,31	0,56	0,29	2,80
$x = Log(T_T)$	1,33	0,57	0,31	2,90

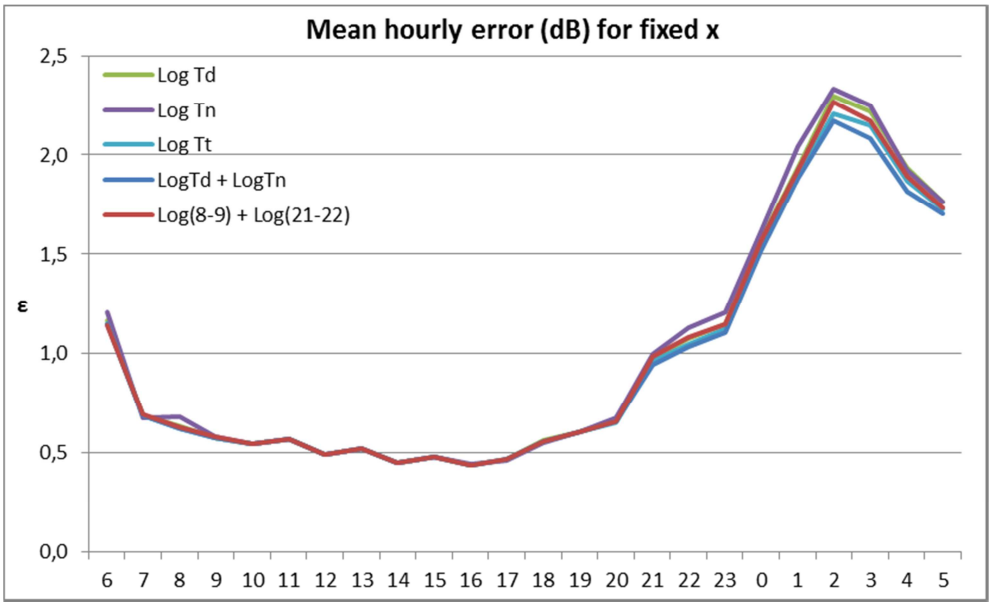


Fig. 35 - Mean hourly error (dB) for the five non-acoustic parameter x shown in Table 8.

5.3 Additional criterion for determining the non-acoustic parameter x

5.3.1 The Box plots

In the following we discuss a general criterion to support the choice of the non-acoustic parameter in addition to the results shown above (Sect. 5.2) for the mean error ε. The criterion is based essentially on the study of the box plots, suggesting a simple way to classify the parameter x according to its accuracy in separating the flows according to the Clusters 1 and 2 obtained from the noise measurements, and its availability from the model of traffic employed. This criterion must be considered as complimentary to the results shown in Sect. 5.2.

The results are shown in Fig. 36 for different choices of x, suggesting that ($x = \sqrt{[(\mathrm{Log}T_D)^2 + (\mathrm{Log}T_N)^2]}$), i.e. the sum of the logs (T_D,T_N) and log T_T are suitable non-acoustic parameters from the point of view of the separation of the stretches into two clusters.

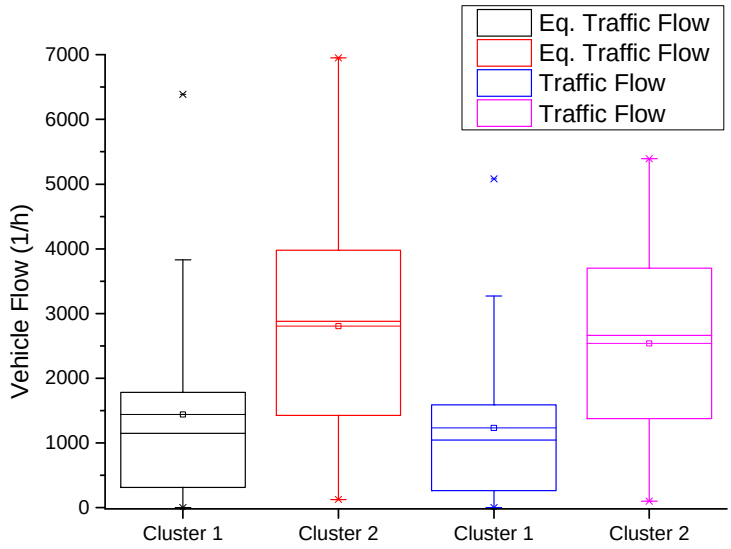


Fig. 36 (a) - Equivalent traffic flows and normal Traffic flows for the Rush hour (08-09).

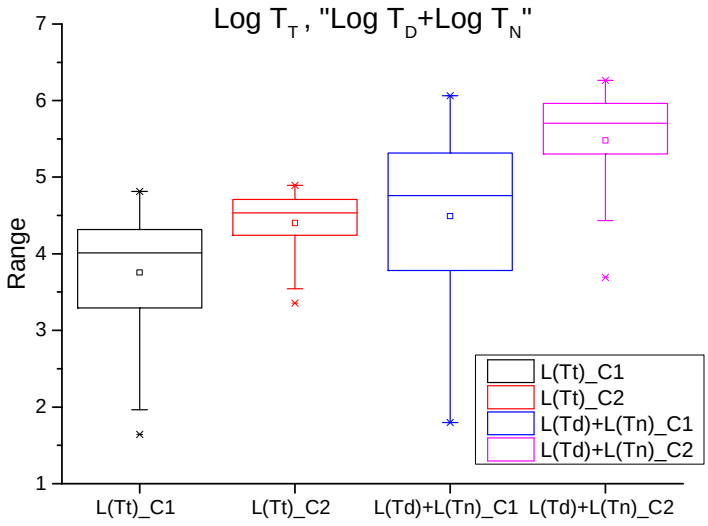


Fig. 36 (b) - Logarithm of the T_T [L(Tt)] and sum of log T_D and log T_N [L(Td)+L(Tn)] ($x = \sqrt{[(\mathrm{Log}T_D)^2 + (\mathrm{Log}T_N)^2]}$).

5.3.2 Receiver Operating Characteristic (ROC) analysis

In general, the receiver operating characteristic (ROC) [25], or ROC curve, is a graphical method to evaluate the performance of a binary classifier. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The index related to the Area Under the Curve (AUC) is equivalent to the probability that the result of the test on a group of roads with non-acoustic parameter over the threshold belongs to the proper cluster. In table 9, the threshold values and the correspondent AUC are reported for each non-acoustic parameter considered.
The best result is obtained for log Tn and its combinations. However, due to the uncertainty of the night flow rate calculated by the traffic model, we opted for log Tt, which provides an AUC of 79.6 %. As a benchmark, we also report the rush hour vehicle flow rate , F_{8-9}, which yields an AUC of 76.2 % and a threshold value of 2007 vehicles/hour. The rush hour vehicle flow rate was initially employed as discriminating parameter between the two clusters.
In Fig. 37 the ROC curve for log Tt. is displayed together the calculated AUC (79.6%) and its 95% confidence level interval.

Table 9: Threshold value of the binary classifier (non- acoustic parameter) and the correspondent AUC.

Non-Acoustic Parameter	Threshold Value	Area Under the Curve (AUC) (%)
log Tt	4.45	79.6
log Td	4.42	79.1
log Tn	3.21	81.3
$\sqrt{(\log F_{8-9})^2 + (\log F_{21-22})^2}$	4.16	78.9
$\sqrt{(\log F_{8-9})^2 + (\log F_{18-19})^2}$	4.69	77.5
$\sqrt{(\log F_{8-9})^2 + (\log F_{3-4})^2}$	3.65	79.2
$\sqrt{(\log F_{8-9})^2 + (\log F_{21-22})^2 + (\log T_t)^2}$	6.20	79.1
$\sqrt{(\log F_{8-9})^2 + (\log T_n)^2}$	4.47	80.1
$\sqrt{(\log T_d)^2 + (\log T_n)^2}$	5.30	80.8
$\sqrt{(\log T_d)^2 - (\log T_n)^2}$	1.02	72.1
F_{8-9}	2007	76.2

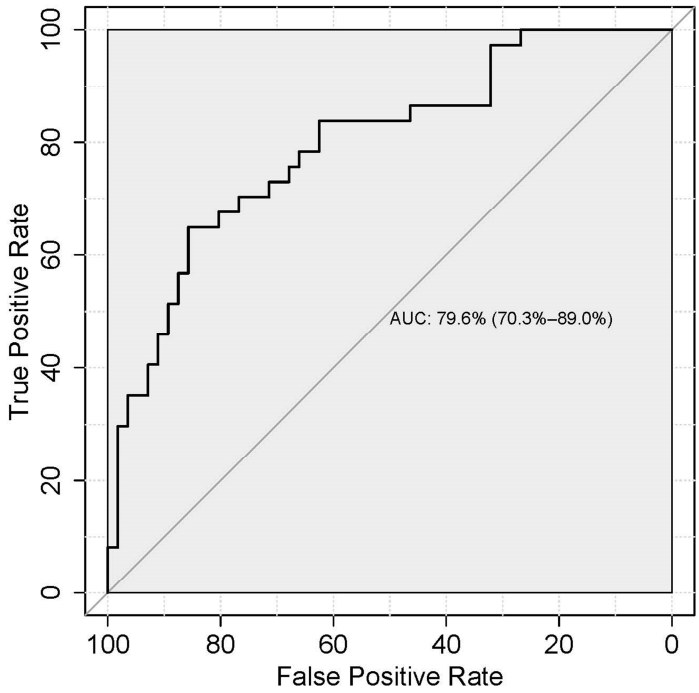


Figure 37: ROC curve for the non-acoustic parameter log Tt. The AUC of 79.6% is reported together its 95% confidence level interval.

As a comparison, we plotted in figure 38 the results obtained from the ROC analysis, AUC, against those obtained considering the ranking provided by the corresponding probability counterpart calculated according to Eq. (8).
The fair agreement between the two methods, with an R2=0.62, suggests that both represent a proper evaluation procedure to judge the goodness of x.

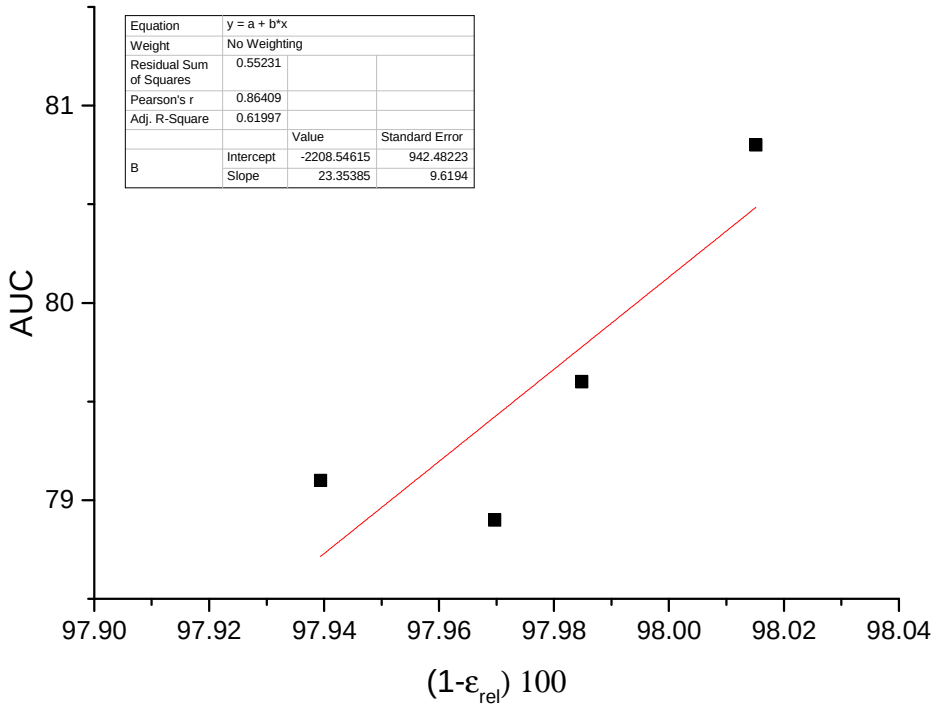


Figure 38: Linear regression of the ROC analysis AUC values against the corresponding probability counterpart calculated according to Eq. (8). In the graph, we report the most significant x values: logTt, logTd, $\sqrt{(\log F_{8-9})^2 + (\log F_{21-22})^2}$, $\sqrt{(\log T_d)^2 + (\log T_n)^2}$.

5.4 Discretization of the parameter x for two cases

Since the above results suggest that both $x= \sqrt{[(LogT_D)^2 + (LogT_N)^2]}$ and $x=Log\ T_T$ are suitable choices for the non-acoustic parameter, we wish to present results for these two cases separately. They include the parametrizations of the distribution functions $P_{1,2}(x)$, total distribution function for x within the urban zone of interest, their separation in (six) groups with the determination of the stretches inside each group, and the corresponding mean values of β to be used according to Eq. (6) and (7). In addition to this we provide with a result to evaluate the error in the prediction of the hourly noise behaviour of two stretches by using the mean values of β instead of its own. The latter result to be compared with the owns shown in Fig. 34.

a) Results for $x= \sqrt{[(LogT_D)^2 + (LogT_N)^2]}$

The corresponding distributions $P_{1,2}(x)$ are reported in Fig. 39, where the polynomial fits for each f(x) are:

$f_1(x) = -2{,}03436 + 0{,}2357x + 0{,}08372x^2 - 0{,}01106x^3$

$f_2(x) = -9{,}21266 + 3{,}22974x - 0{,}35699x^2 + 0{,}01199x^3$

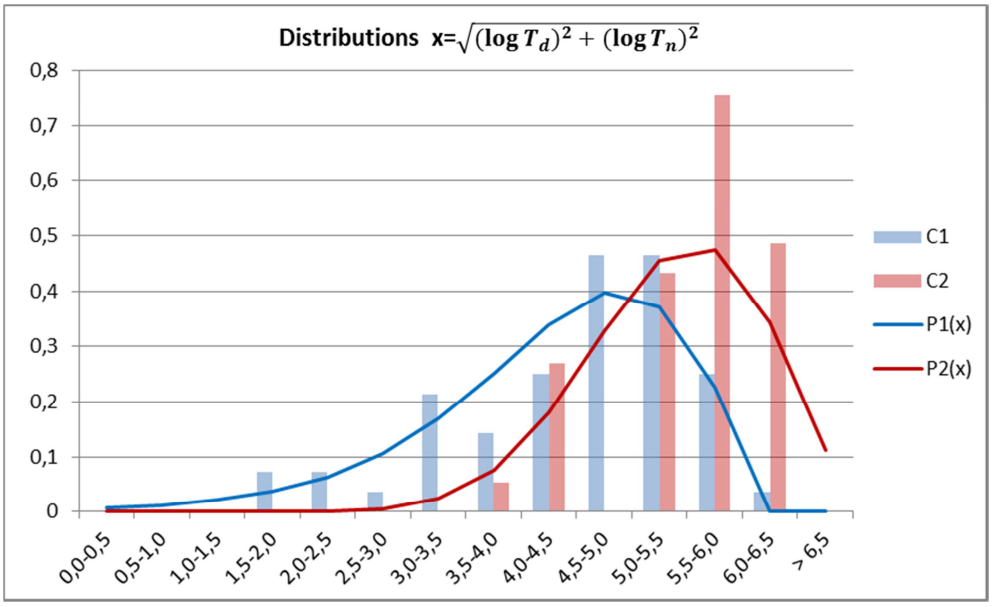


Fig. 39 - Probability distribution functions P(x) for Cluster 1 and 2, for the parameter $x=\sqrt{[(LogT_D)^2 + (LogT_N)^2]}$.

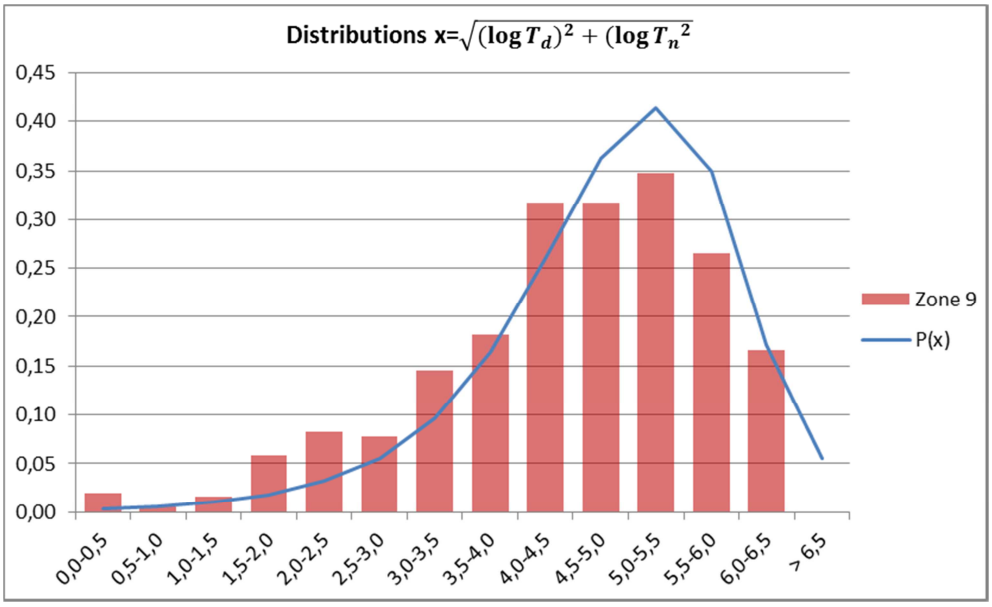


Fig. 40 - Distribution of $x=\sqrt{[(LogT_D)^2 + (LogT_N)^2]}$ for the whole Zone 9 of Milano (Histogram).The continuous line (P(x)) represents the corresponding distribution of x for the 93 arches of the recording stations.

We have calculated x for each of the 1900 roads of Zone 9 of Milan (from a total of 2075 roads, of which we have not considered those which present vanishing flow over the whole day).
The distribution of the parameter x over the whole set of streets in the required zone has been obtained and compared with the resulting distribution of x over the 93 road stretches used in our measurements. From the theoretical model used for this scope, some stretches (specifically 175 cases) have a flow rate of less than one vehicle per day. For these stretches, the parameter x would take the value zero, yielding an artificial peak at x=0. Since the model is expected to fail in those circumstances, that is of extremely low traffic flows, we have decided to take them out of the

database. Also because such roads are characterized by very low levels of noise and have not been monitored, so that their inclusion into the analysis would have altered the final result and conclusions.

Their distribution is reported in Fig. 40. As one can see, the distribution of x obtained from the measurements at the 93 noise recording stations is consistent with the whole distribution of x from the zone 9. This is a very important requirement for the chosen stations so that their x values cover essentially all the range of x values from the zone to be predicted the behavior of the traffic noise.
From this distribution we have chosen 6 intervals of x so that they result to be almost equally populated (cake graphic), which include however those roads with zero flow contained in the first group (Fig. 41).

Even if the levels of traffic noise for road stretches with very low traffic flows yield extremely low values of noise, we have kept them and included them in the group having the lowest values of x. In the actual implementation, the corresponding road stretches will follow the time variations of the lowest x group. We expect that due to their intrinsically low noise values the error made in this case will be of no significance.

Each stretch belongs to a single group, and all stretches within the group behave the same, they have the same temporal variations. The larger the number of groups, the better will be their time resolution. However, this will result in a larger number of acoustic maps to be implemented. Given the fact the total number of monitoring noise stations is rather small (24 in total), we have to reach a compromise in order to have a reasonable number of stations inside each group. We took this number to be 24/6=4.

To each group of x values there will be associated a noise map which will be updated in real time by measuring the noise in few selected locations for the roads inside each group. We discuss the way in which the noise station locations can be obtained below.

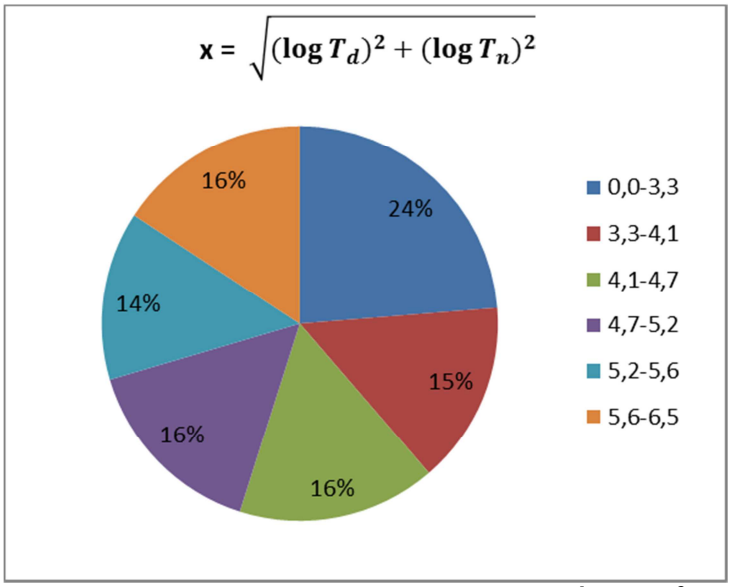


Fig. 41 - The six groups of x values for the parameter $x=\sqrt{[(LogT_D)^2 + (LogT_N)^2]}$.

Inside each interval or group of x, we have calculated the mean values $\bar{\beta}_1$ ($\bar{\beta}_2$ = 1-$\bar{\beta}_1$) by dividing it into 100 smaller parts to obtain the values of α_1 and α_2 from the analytical formulas for P(x). From those values we obtain the mean results reported in the Table 10.

Table 10 - Mean values of β for the six groups of $x=\sqrt{[(LogT_D)^2 + (LogT_N)^2]}$.

Range of x	0,0 – 3,3	3,3 – 4,1	4,1 – 4,7	4,7 – 5,2	5,2 – 5,6	5,6 – 6,5
β_1	0,98	0,78	0,62	0,51	0,42	0,15
β_2	0,02	0,22	0,38	0,49	0,58	0,85

These mean values of β can be used according to the scheme described by Eq. (7) in predicting the traffic noise associated to roads belonging to a given group x. The predictions for the traffic noise by using the mean values of β instead of the actual ones for each road is shown in Fig. 42 in the two examples discussed above, D5 and D3.

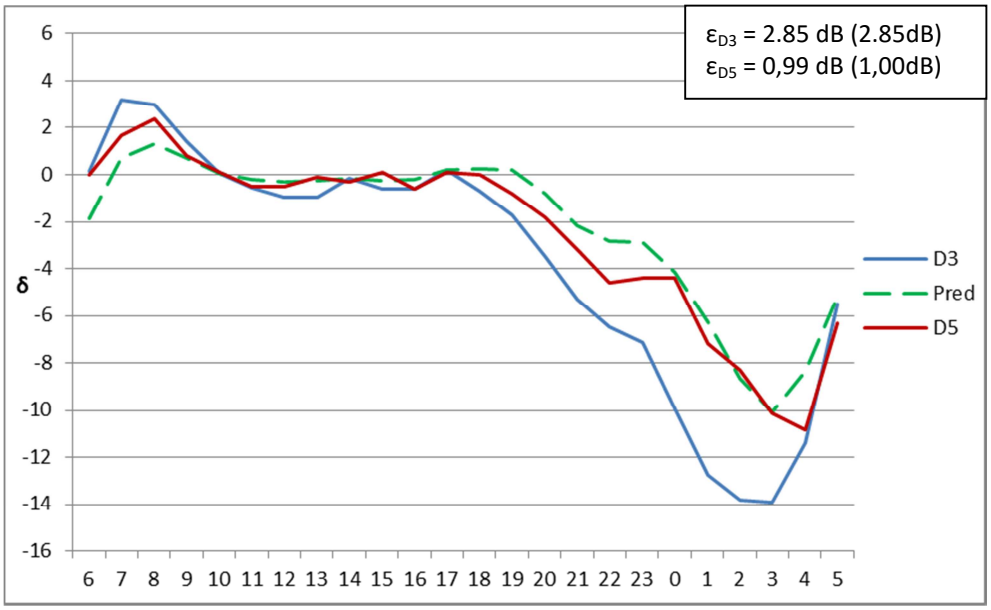


Fig. 42 - Comparison of predicted traffic noise with the measured one for D5 and D3 using the mean values of β from the corresponding group x. The errors are reported in the inset to be compared with similar results from Fig. 34.

The mean error obtained using the mean group value $\bar{\beta}$ instead of the single ones is now 1.32 dB, essentially the same value obtained in Table 8.

b) Results for x= $LogT_T$

Here we report the results of this choice of x. Analytical fits functions for P(x)'s are:

$f_1(x) = -1{,}56586-0{,}24459x+0{,}28834x^2-0{,}03526x^3$

$f_2(x) = -15{,}2673+7{,}01263x -1{,}02922x^2+0{,}04708x^3$

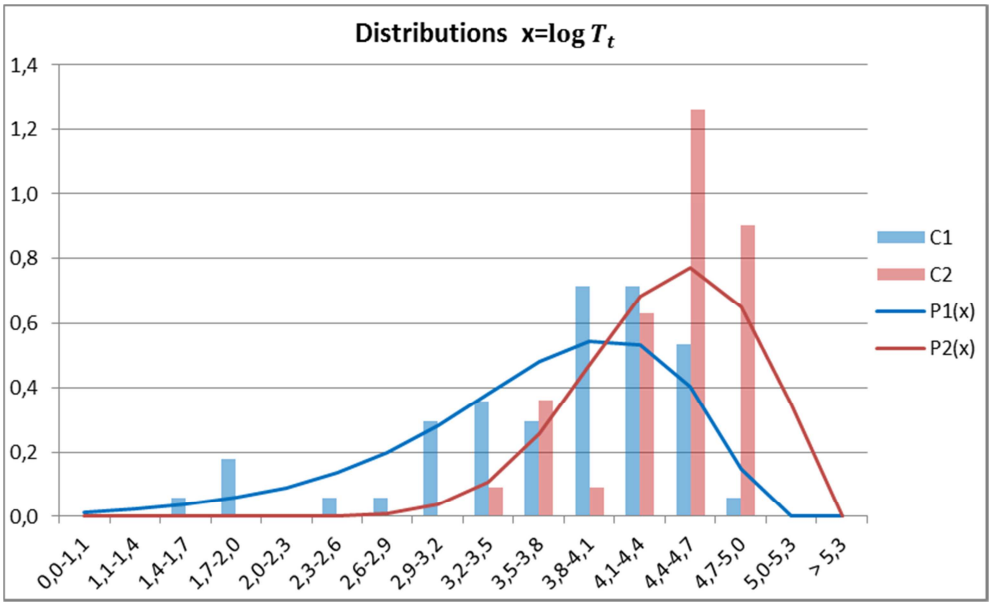


Fig. 43 - Distribution functions $P_1(x)$ and $P_2(x)$ for x = log T_T.

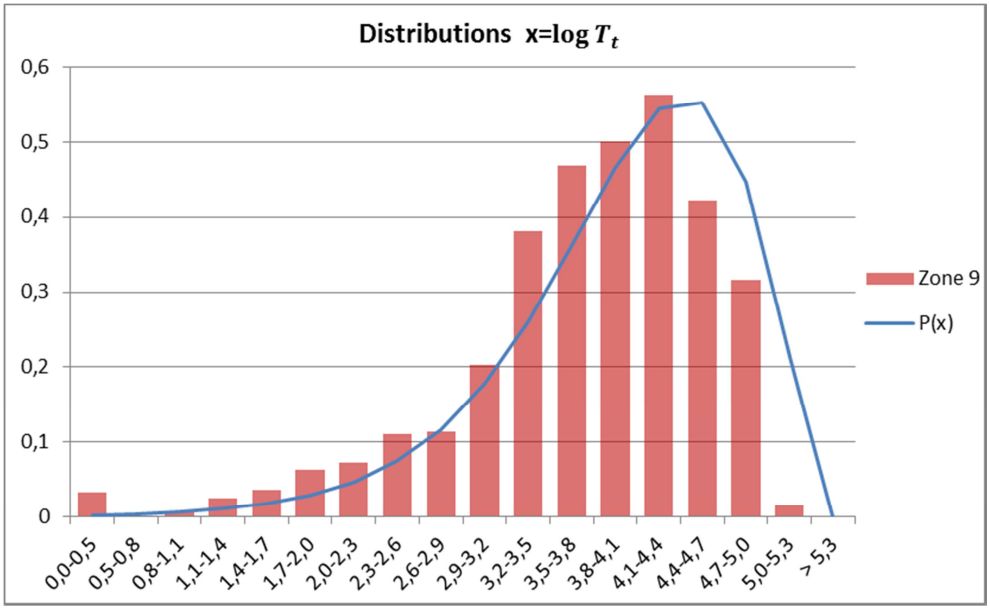


Fig. 44 - Distribution of x= LogT_T for the whole Zone 9 of Milano(Histogram).The continuous line (P(x)) represents the corresponding distribution of x for the 93 arches of the recording stations.

And the comparison with the empirical data yields Fig. 44. The six groups are now found according to the diagram shown below (Fig. 45).

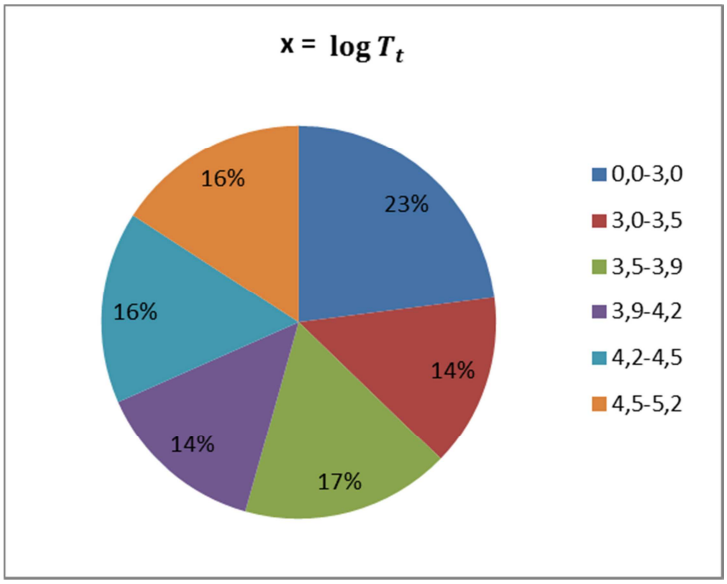


Fig. 45 - The six groups of the parameter x = log T_T.

And the mean values of β are given by,

Table 11 - Mean values of β for the six groups of x = log T_T.

Range of x	0,0 – 3,3	3,3 - 3,5	3,5 - 3,9	3,9 - 4,2	4,2 - 4,5	4,5- 5,2
β_1	0,99	0,81	0,63	0,50	0,41	0,16
β_2	0,01	0,19	0,37	0,50	0,59	0,84

5.5 Identification of monitoring network sites inside Zone 9

In what follows, we discuss a procedure to select the locations of the noise recording stations inside each one of the six groups of x. The choice is based on the use of the hourly flow taken from a model of traffic. To do that, we calculate the hourly mean traffic flow inside each group.

a) Results for $x= \sqrt{[(LogT_D)^2 + (LogT_N)^2]}$

From each group, we calculate the mean standard deviation of the road flow with respect to its mean δ, in the same spirit as in Eq. (8). The list of roads is sorted from the closest distance to the mean (top road) till the 20th distance. From this Table one can decide therefore where to locate the noise recording stations, and the 20 roads found give some room for finding the most appropriate streets according to the different local urban constraints, such as suitability of the street, surrounding building structure and local sources of additional noise which should be considered when choosing the best places for the station.

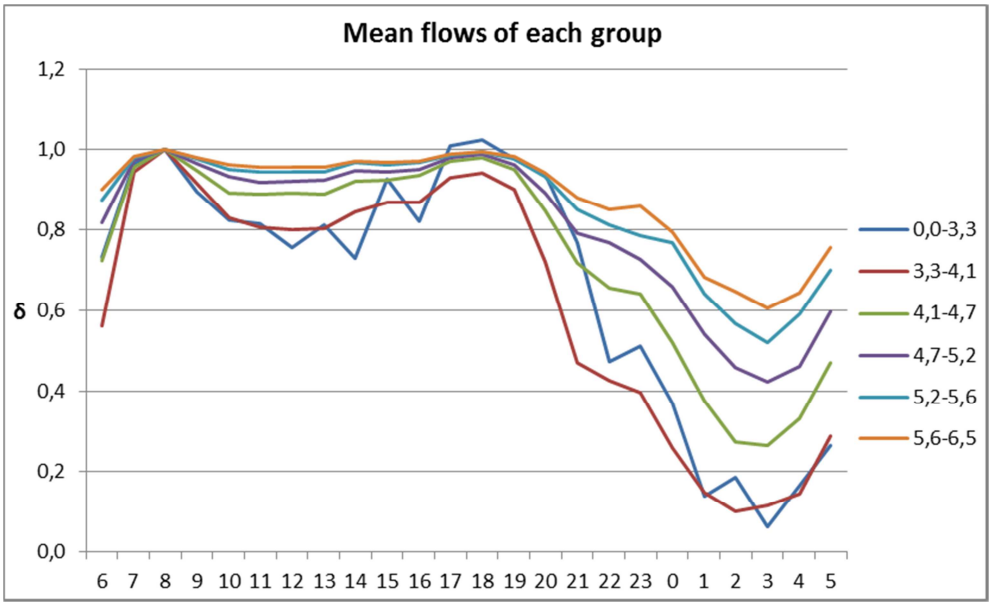
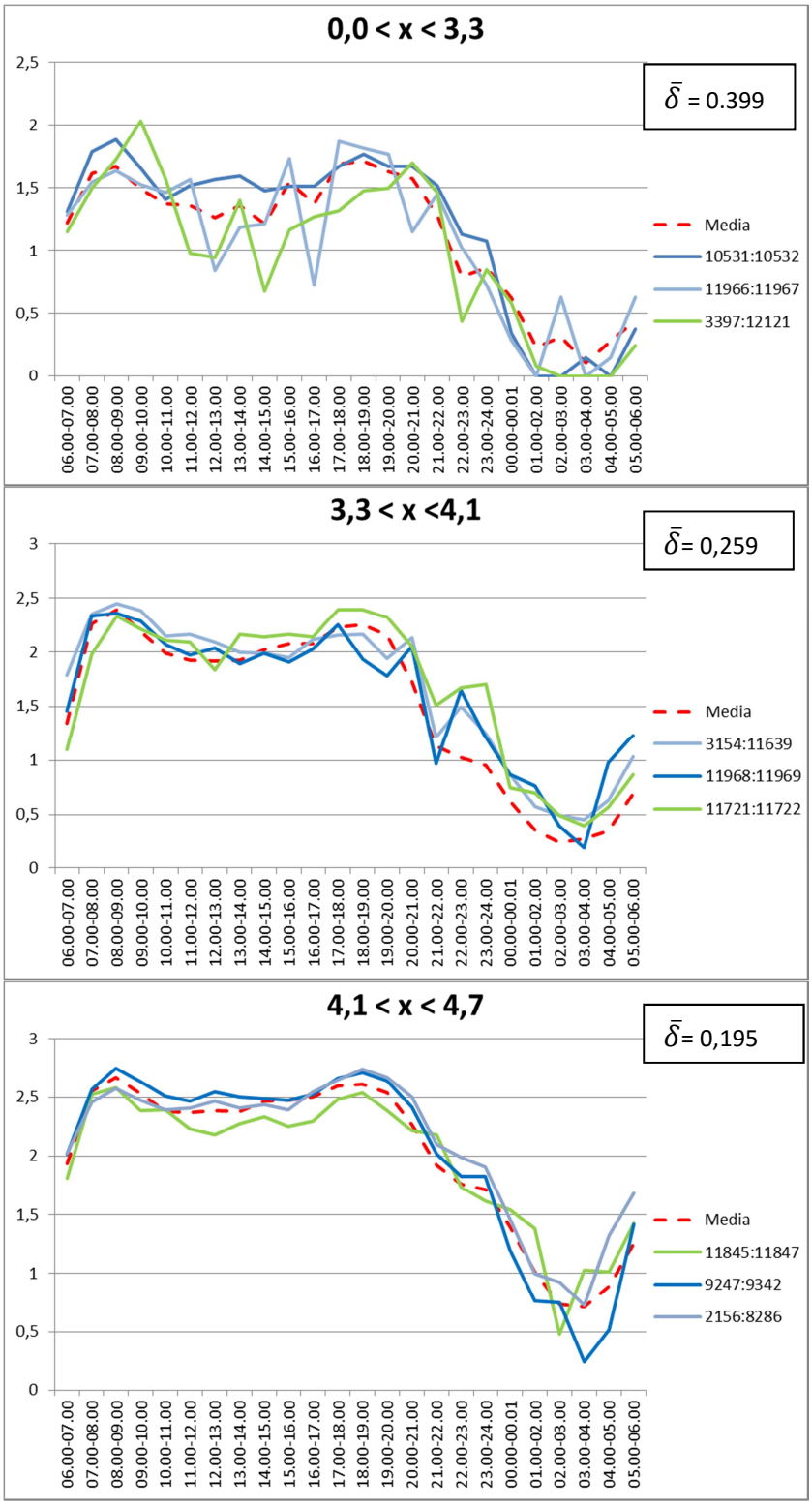


Fig. 46 - Mean flows for roads inside each one of the 6 groups shown in Fig. 45. This graph shows the mean hourly variations of the flow for each of the 6 groups by normalizing them using the rush hour values (08-09).

Table 12 - In the Table we report the first 20 roads for each group of x in which their logarithm of the flow has the lower distance from the mean group value calculated over the 24 hours.

Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
0,0-3,3	δ	3,3-4,1	δ	4,1-4,7	δ	4,7-5,2	δ	5,2-5,6	δ	5,6-6,5	δ
10531:10532	0,20	2953:3124	0,23	9247:9342	0,16	11927:12023	0,10	12039:16188	0,07	11985:11986	0,06
11966:11967	0,24	2953:11972	0,23	2156:8286	0,17	3132:11965	0,10	3121:16188	0,07	Murat_A	0,06
11966:33166	0,24	3154:11639	0,23	11845:11847	0,17	3103:11927	0,10	6974:12039	0,07	2160:11984	0,06
3397:12121	0,26	3934:13721	0,24	3939:13730	0,17	3103:8808	0,11	6974:12008	0,07	11984:11985	0,06
4196:12121	0,26	3935:13721	0,24	2116:9353	0,18	Pirelli_C	0,12	3168:12008	0,07	Stelvio_H	0,06
5049:5051	0,30	2168:10517	0,25	2083:2162	0,19	Pirelli_D	0,12	3064:12024	0,07	Zara_K	0,07
8763:11750	0,32	23024:23028	0,25	12035:16179	0,20	3094:12023	0,12	12026:16189	0,08	Zara_L	0,07
8763:8764	0,32	23028:23033	0,25	16178:16179	0,20	Fermi_AD	0,12	3121:16189	0,08	Zara_D	0,07
3382:12017	0,33	23035:23036	0,25	2103:8286	0,20	3095:3096	0,12	3064:8774	0,08	Zara_C	0,07
3029:3042	0,41	3668:23035	0,25	12157:12194	0,20	3095:3102	0,12	8774:12026	0,08	Zara_E	0,08
3157:11651	0,57	2158:13703	0,25	12161:12194	0,20	8808:11926	0,13	Da_Seregno_M	0,09	Zara_B	0,08
6742:12026	0,59	2527:23161	0,27	12161:12193	0,20	33215:33216	0,14	2095:12118	0,10	Gioia_M	0,08
31400:31401	0,37	2527:2554	0,27	12167:12193	0,20	11882:11884	0,14	Maciachini_E	0,10	Zara_M	0,08
10297:10301	0,39	8812:11938	0,28	5045:17300	0,20	10313:10314	0,14	Maciachini_F	0,10	Istria_A	0,08
3938:16008	0,41	2769:12083	0,28	5041:5045	0,20	Emanueli_D	0,14	13719:13720	0,10	Sauro_A	0,08
10298:10301	0,42	3165:11995	0,28	3067:3069	0,21	Emanueli_E	0,14	3395:12118	0,10	Fermi_AA	0,08
8790:8791	0,50	11968:11969	0,28	11887:31396	0,21	3172:3173	0,14	3395:3929	0,10	Fermi_AB	0,08
5423:5424	0,50	2538:23161	0,28	13771:26857	0,21	3173:13822	0,14	2108:27223	0,10	Fermi_Z	0,08
5427:5428	0,65	11721:11722	0,28	13771:31422	0,21	Cozzi_G	0,14	2141:27223	0,10	Zara_N	0,08
13741:13742	0,71	16019:23024	0,28	27227:31381	0,21	11926:11965	0,14	3396:6759	0,10	Zara_O	0,08

In the following plots (Fig. 47) we report the results for each group, giving the top three roads with lowest distances from the mean, just to get an idea of what is the overall behavior of δ inside each group.



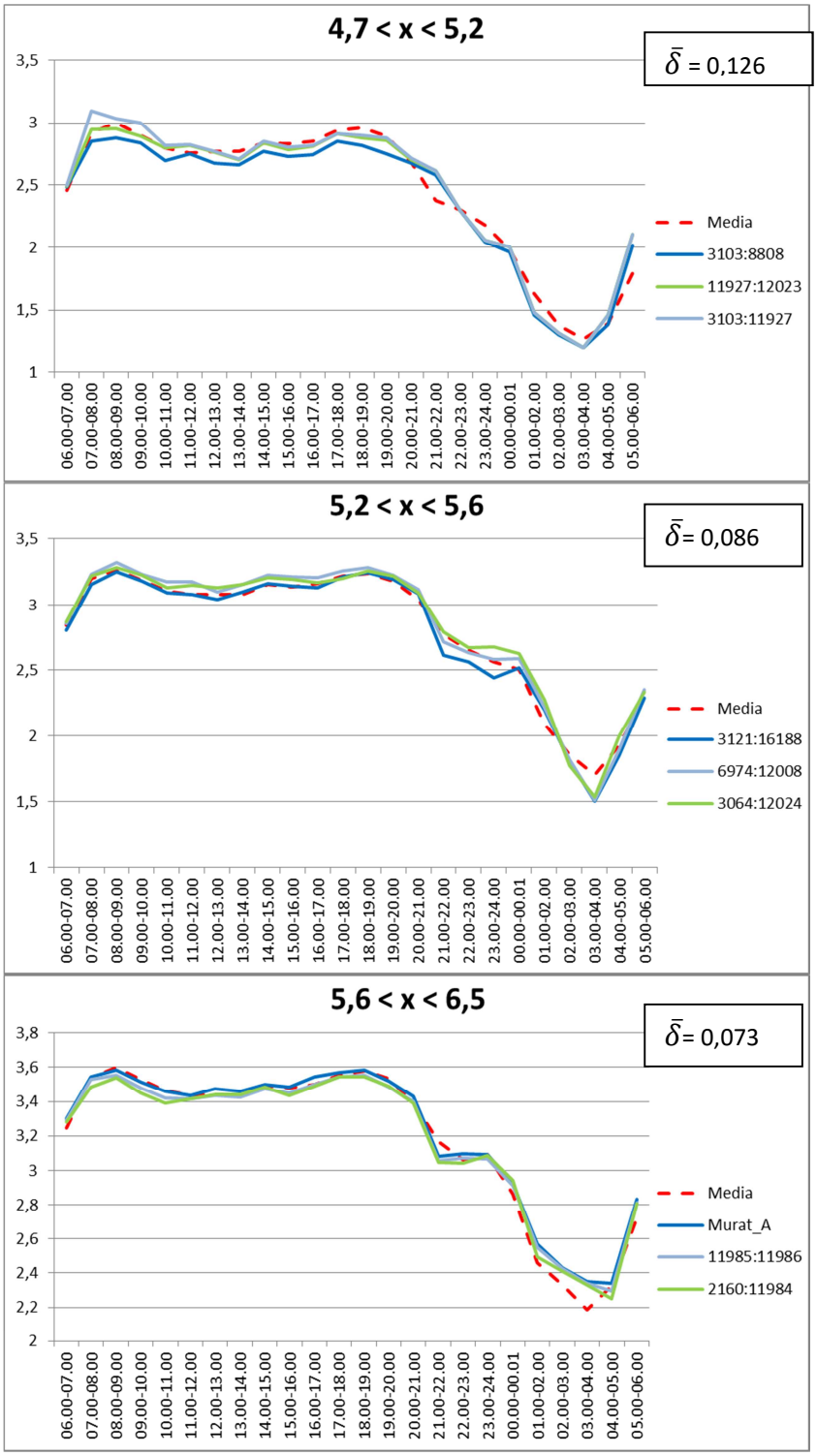


Fig. 47 - Top three roads inside each group of x values for the non-acoustic parameter $x=\sqrt{[(LogT_D)^2 + (LogT_N)^2]}$. The mean values of the flows are given by the dashed lines, and the mean values of δ are reported in the inset.

b) Results for x= LogT_T

Here, we report the results for the choice x=Log T_T. The top 20 stretches are displayed in Table 13. A closer look shows that indeed the stretches are quite similar to the ones obtained for the choice 5.6a (Table 12).

Table 13 - In the Table we report the first 20 roads for each group of x in which their logarithm of the flow has the lower distance from the mean group value calculated over the 24 hours.

Group 1		Group 2		Group 3		Group 4		Group 5		Group 6	
0,0-3,0	δ	3,0-3,5	δ	3,5-3,9	δ	3,9-4,2	δ	4,2-4,5	δ	4,5-5,2	δ
10531:10532	0,24	2168:10517	0,23	9247:9342	0,17	3103:8808	0,09	12039:16188	0,06	Stelvio_H	0,06
12093:13727	0,25	2953:11972	0,23	11845:11847	0,19	11927:12023	0,10	3121:16188	0,06	11985:11986	0,06
11966:11967	0,30	2953:3124	0,23	12161:12193	0,19	8808:11926	0,11	6974:12039	0,06	Murat_A	0,06
11966:33166	0,30	2158:13703	0,24	12167:12193	0,19	Fermi_AD	0,11	3064:8774	0,09	2160:11984	0,06
12092:13727	0,32	3154:11639	0,25	12157:12194	0,19	3103:11927	0,11	8774:12026	0,09	11984:11985	0,07
3029:3042	0,33	3165:11995	0,26	12161:12194	0,19	3132:11965	0,12	6974:12008	0,09	Zara_K	0,07
20540:31375	0,33	23024:23028	0,27	3939:13730	0,20	11926:11965	0,12	12994:13767	0,09	Zara_L	0,07
3965:3968	0,33	23028:23033	0,27	2479:31157	0,20	Pirelli_C	0,12	3168:12008	0,09	Zara_M	0,07
2172:27405	0,34	23035:23036	0,27	2103:8286	0,20	Pirelli_D	0,12	3031:3907	0,09	Sauro_A	0,07
27405:31375	0,34	3668:23035	0,27	2116:9353	0,21	Emanueli_D	0,12	2108:27223	0,09	Zara_D	0,07
3397:12121	0,34	2769:12083	0,27	12035:16179	0,21	Emanueli_E	0,12	2141:27223	0,09	Zara_C	0,07
4196:12121	0,34	2527:23161	0,27	16178:16179	0,21	3931:10314	0,12	12026:16189	0,10	Zara_E	0,07
5049:5051	0,34	2527:2554	0,27	2156:8286	0,21	3094:12023	0,13	3121:16189	0,10	Zara_N	0,07
34391:34392	0,35	11968:11969	0,27	3921:4065	0,21	10313:10314	0,13	2093:31418	0,10	Zara_O	0,07
34392:34393	0,35	3934:13721	0,28	27227:31381	0,21	3095:3096	0,14	3064:12024	0,10	Zara_B	0,08
3032:3041	0,36	3935:13721	0,28	27227:31383	0,21	3095:3102	0,14	12145:12146	0,10	Istria_A	0,08
3032:6031	0,36	2538:23161	0,28	2083:2162	0,21	3172:3173	0,14	12146:12147	0,10	Gioia_M	0,08
3042:6032	0,36	5049:13710	0,28	12103:16003	0,23	3173:13822	0,14	Pasta_A	0,10	Fermi_AA	0,08
6031:6032	0,36	16019:23024	0,28	12104:16003	0,23	Cozzi_G	0,14	Girardengo_B	0,10	Fermi_AB	0,08
3382:12017	0,36	5047:16019	0,28	11887:31396	0,23	10322:16000	0,15	Girardengo_C	0,10	Fermi_Z	0,08

The mean group flows are as follows:

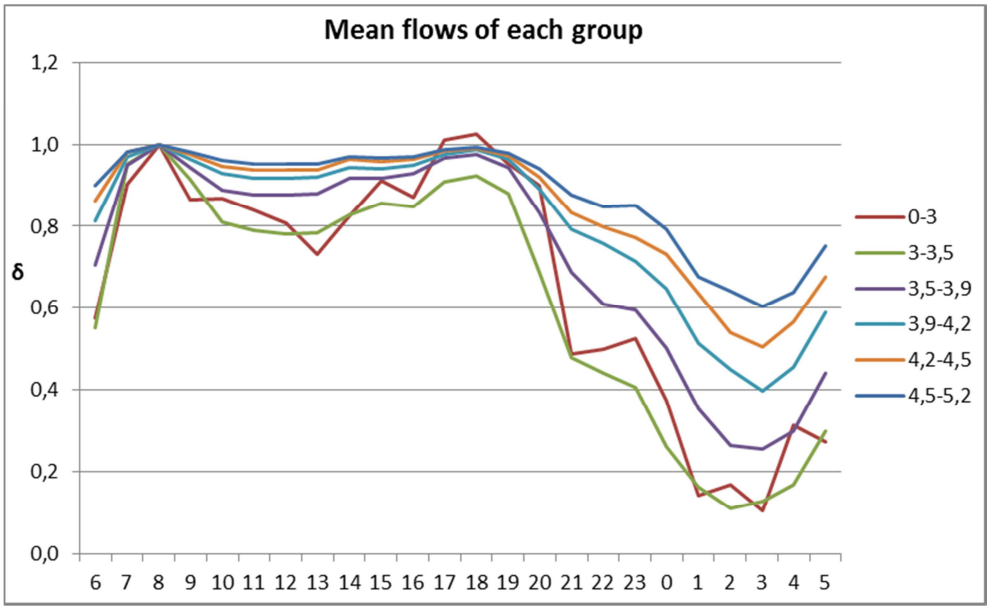
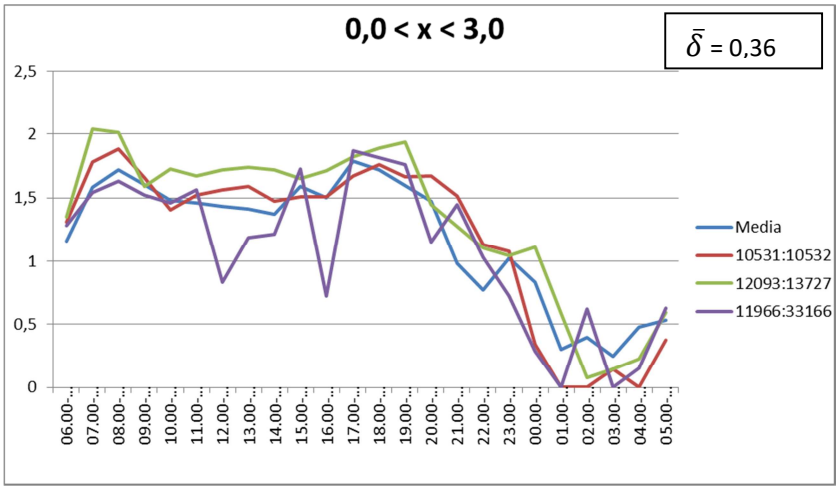
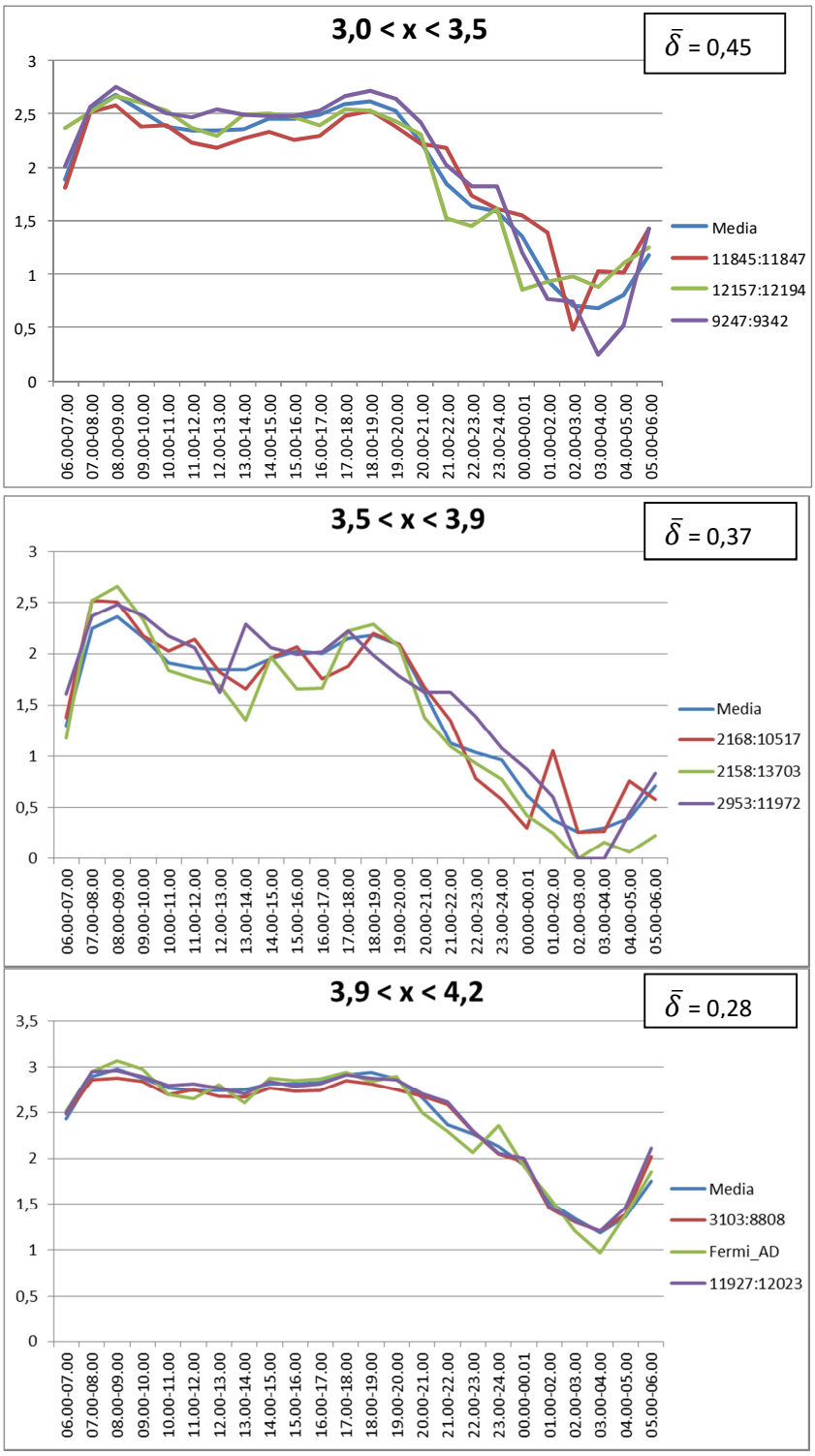


Fig. 48 - Mean flows for roads inside each one of the 6 groups shown in Fig. 43. This graph shows the mean hourly variations of the flow for each of the 6 groups by normalizing them using the rush hour values (08-09).

In the following plots (Fig. 49) we report the results for each group, giving the top three roads with lowest distances from the mean, just to get an idea of what is the overall behavior of δ inside each group.




3,0 < x < 3,5
$\bar{\delta}$ = 0,45
Media
11845:11847
12157:12194
9247:9342
3,5 < x < 3,9
$\bar{\delta}$ = 0,37
Media
2168:10517
2158:13703
2953:11972
3,9 < x < 4,2
$\bar{\delta}$ = 0,28
Media
3103:8808
Fermi_AD
11927:12023
06.00-07.00
07.00-08.00
08.00-09.00
09.00-10.00
10.00-11.00
11.00-12.00
12.00-13.00
13.00-14.00
14.00-15.00
15.00-16.00
16.00-17.00
17.00-18.00
18.00-19.00
19.00-20.00
20.00-21.00
21.00-22.00
22.00-23.00
23.00-24.00
00.00-00.01
01.00-02.00
02.00-03.00
03.00-04.00
04.00-05.00
05.00-06.00

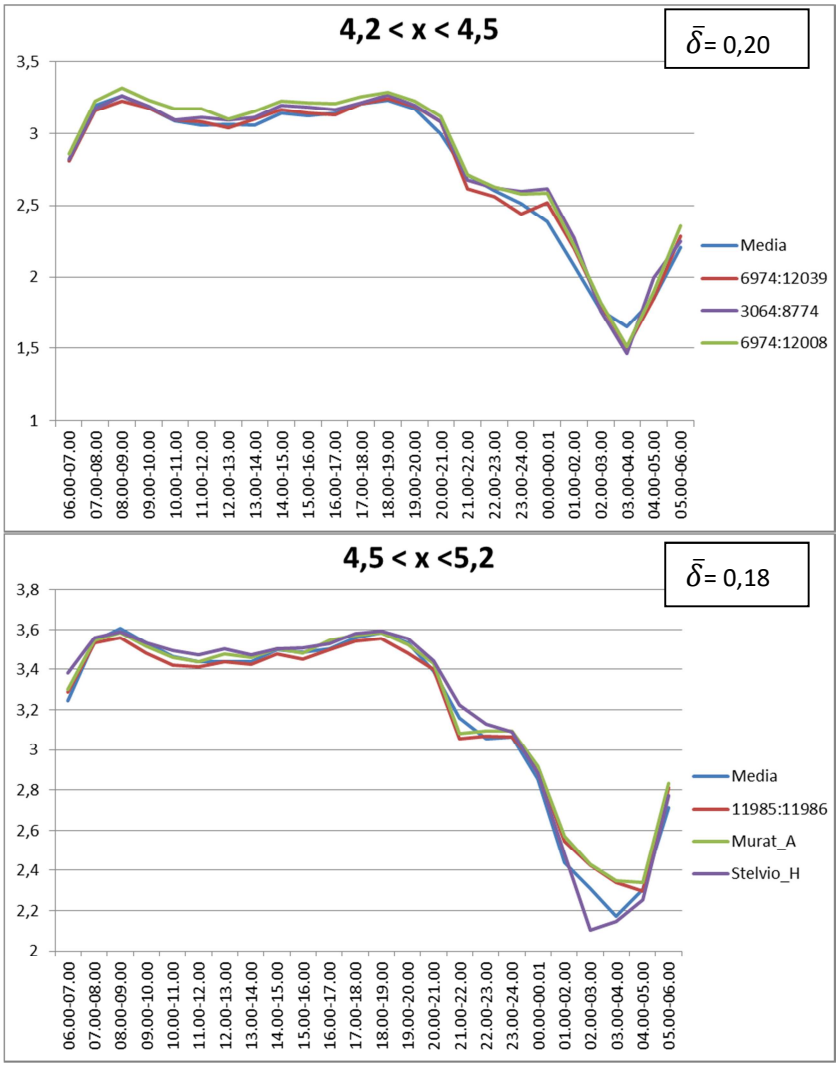


Fig. 49 - Top three roads inside each group of x values for the non-acoustic parameter x = log T_T. The mean values of the flows are given by the dashed lines, and the mean values of δ are reported in the inset.

6. CONCLUSIONS

We have described a method to predict the traffic noise on any street in a city, by recording the traffic noise from few recording stations appropriately located inside the urban zone of interest. We summarize in the following the procedure and the way in which it can be applied in the practice.
In our analysis, we have taken traffic noise data from 93 stations distributed more or less at random in the city of Milan. Of these 93 stations, about 20 fall inside the urban zone 9, of interest here for developing the acoustic map. The data are recorded continuously, and different data are aggregated for selected time intervals, such as τ=(5, 10, 15, 20, 30, 60) mins. We aim at estimate the traffic noise level of any street by relying on data from the corresponding traffic flow. A model is used to obtain such non-acoustic parameter.

Method: From the hourly data of noise levels, we apply a cluster analysis, and classify the 93 stations into two main clusters, denoted here as cluster 1 and 2,. The hourly noise levels are very similar during the morning and afternoon hours, while more conspicuous differences are seen during the

night hours. Cluster 1 contains those streets for which the noise falls down strongly during the night, corresponding to streets carrying a low traffic flow. Cluster 2 refers to those highly traffic streets, for which the level of noise remains relatively high even during the night hours. For each street in each cluster we calculate a non-acoustic parameter x. The latter has been chosen from many different possibilities. In summary, we found that the logarithm of the total daily flow, that is the log of the sum of the flow over the 24 hours, $x=LogT_T$, is a suitable choice. This also because it is a quantity known for most of the stretches of interest. We then calculate the distribution of x for each cluster, 1 and 2, denoted as $P_1(x)$ and $P_2(x)$. For a given street i, we obtain its value of x_i, and calculate the probability that such a stretch belongs to each one of the two clusters, i.e. $\beta_1= P1(x_i) / (P_1(x_1)+P_2(x_i))$, $\beta_2= P_2(x_i) / (P_1(x_1)+P_2(x_i))$, respectively. Using these two probabilities, we estimate the noise level at stretch i, $\delta_i(h)$, according to the sum, $\delta_i(h)= \beta_1 \delta_1(h)+ \beta_2 \delta_2(h)$, where $\delta_{1,2}(h)$ are the normalized mean noise level values from each cluster 1 and 2, recorded by the noise stations at hour h.

Acoustic maps: In order to construct the noise maps, we study the distribution of the non-acoustic parameter for each stretch in the zone 9, P(x). We check that it overlaps with the corresponding distribution of x corresponding to the stretches where the recording stations are located. This to ensure that the data recording contains essentially all the information over the whole range of values of x from the zone 9 of interest. Then, we divide the total interval of x values, for which P(x)>0, into six intervals or groups, each one containing a similar number of streets. Inside each group n, we determine the mean value x_n and obtain the corresponding values of $\beta_1(n)$ and $\beta_2(n)$. From these values of β, we apply the above relation for $\delta_i(h)$ which yields the same noise level for each stretch i inside the group n. That is, to all the stretches belonging to group n we associate the same noise level $\delta_n(\tau)$, obtained at time interval τ as measured by all the stations from clusters 1 and 2. As one can see, all recording stations together determine the six acoustic maps for each group, each one characterized its own mean values of $\beta_1(n)$ and $\beta_2(n)$. This procedure is useful in those cases in which one or few stations fail to record the data, or are under anomalous traffic conditions.

Location of the recording stations: The location of the recording stations, in number of 24, have been suggested by analyzing the traffic flows inside each group n. We have created a list of stretches which have an hourly flow similar to the mean traffic flow of the group. The list covers a set of the first best choices, from which the final decision about the location of the recording stations can be taken. Inside each group n, one can locate 4 stations, for a total of 24.

Practical implementation: Once the 24 stations have been chosen and implemented, one needs to classify them into the two clusters, 1 and 2. This can be done while the system is already working or by measuring the noise levels from each station before running the acoustic maps. The system is able to start working even if the clustering of the stations has not been determined. In this case, one can use the mean values measured by the 4 stations in each group to determine the noise levels of the stretches inside each group. Even before the clustering has been completed, one can measure the traffic flows for each station and determine the corresponding values of x. These measured values can be used then to establish the distributions $P_1(x)$ and $P_2(x)$, when the clusters 1 and 2 are known. A good statistics for x can be obtained by measuring x over different days. This will ensure a more realistic approach to the problem of the non-acoustic parameter, to determine the mean values $\beta_{1,2}(n)$.

Time intervals for updating the acoustic maps: We have found that 5 mins is a good compromise for updating the acoustic maps during daily hours, while during the night a more conservative hourly update is recommended. This is based on our analysis of the errors taken in the predictions, ranging up to about 2 dB during the day, and increasing up to 5-6 dB during night hours. We have checked that mean values of noise levels calculated over 5 mins are essentially the same as those calculated over 1 hour. The cluster composition remains essentially the same as those of the clusters obtained

from the hourly data. In conclusion, we found that the maps can be updated every 5 mins in the time interval from 7-21 hours, every 15 mins in the time interval 21-01 hours, while for the night 21-7 hours the update can be made every hour. Errors estimates the action project described in this report (B1) will be checked and if necessary corrected during the action project B7 (System testing and fault analysis).

7. REFERENCES

[1] S. Fidel; Nationwide urban noise survey, J. Acoust. Soc. Am., 64, 198-206, 1978.

[2] A. L. Brown, K. C. Lam, Urban noise surveys, Appl. Acoust., 20, 23-39, 1987.

[3] K. Kumar, V. K. Jain, A study of noise in various modes of transports in Delhi, Appl. Acoust., 43, 57-56, 1994.

[4] H. M. E. Miedema, H. Vos, Exposure-response relationships for transportation noise, J. Acoust. Soc. Am. 104, 3432-3445, 1998.

[5] A. Garcia, L. J. Faus, Statistical analysis of noise levels in urban areas, Appl. Acoust., 34, 227-247, 1991.

[6] J. M. Fields, Effect of personal and situational variables on noise annoyance in residential ar-eas, J. Acoust. Soc. Am., **93**, 2753-2763, 1993.

[7] Carmona del Río F.J., Escobar V.G., Carmona J.T., Vílchez-Gómez R., Méndez Sierra J.A., Rey Gozalo G., Barrigón Morillas J.M. (2011), A Street Categorization Method to Study Urban Noise: The Valladolid (Spain) Study, Environmental Engineering Science, 28(11), 811-817.

[8] Rey Gozalo G., Barrigón Morillas J.M., Gajardo C.P. (2015), Urban noise functional stratification for estimating average annual sound level, J. Acoust. Soc. Am., 137(6), 3198-3208

[9] Barrigon J.M., Gomez V., Mendez J., Vilchez R., and Trujillo J. (2005). A categorization method applied to the study of urban road traffic noise. J. Acoust. Soc. Am. 117, 2844-2852.

[10] Brambilla G., and Gallo V. (2010). Andamenti dei livelli LAeq orari nelle 24 ore del rumore urbano e indicazioni per il campionamento spaziale stratificato (Patterns of 24 h hourly LAeq of urban noise and indications for stratified spatial sampling). Presented at AIA 2010, Siracusa, Italy, May 26-28.

[11] Angelini F., Bisceglie A., Brambilla G., Gallo V., and Zambon G. (2012). Campionamento spaziale stratificato per il rumore da traffico stradale: un'applicazione alla rete viaria di Milano (Stratified spatial sampling of road traffic noise: a case study of the road network in Milan), Presented at AIA 2012, Roma, Italy, July 4-6.

[12] J. H. Ward, *Hierarchical Grouping to Optimize an Objective Function*, Journal of the American Statistical Association, **58**, 236–244, 1963.

[13] J.A. Hartigan, M.A. Wong, *A K-means clustering algorithm*, Applied Statistics, **28**, 100–108, 1979.

[14] L. Kaufman, P. Rousseeuw, *Finding Groups in Data*, Wiley Series in Probability and Mathematical Statistics, 1990.

[15] C. Fraley et al., "*mclust" version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation*, http://cran.r-project.org/web/packages/mclust/index.html.

[16] R Core Team (2015), R: *A language and environment for statistical computing, R Foundation for Statistical Computing*. Vienna, Austria. Available at: www.R-project.org/

[17] R Core Team (2015), R: *A language and environment for statistical computing, R Foundation for Statistical Computing*. Vienna, Austria. Available at: www.R-project.org/

[18] http://www.r-project.org/

[19] G. Zambon , R. Benocci, G. Brambilla, V. Gallo, *Statistics-based functional classification of roads in the urban area of Milan*, Proceedings of Forum Acusticum, 7-12 September, Krakow, 2014.

[20] G. Brock, V. Pihur, S. Datta, S. Datta, *clValid: An R Package for Cluster Validation*, Journal of Statistical Software, **25**, 1-22, 2008.

[21] G. Brambilla, V. Gallo, *Andamenti dei livelli L_{Aeq} orari nelle 24 ore del rumore urbano e indicazioni per il campionamento spaziale stratificato*, Atti AIA 2010, Siracusa, 2010 (in Italian).

[22] Kish L. (1965), *Survey Sampling,* John Wiley & Sons, NY.

[23] Package "*clValid*" version 0.6-4, 2013.

[24] CERTU/SETRA/LCPC/CSTB (1997). NMPB-Routes-96, Bruit des infrastructures *routières, méthode de calcul incluant les effets météorologiques*. CERTU/SETRA/LCPC/CSTB.

[25] Fawcett T, An introduction to ROC analysis, *Pattern Recogn. Lett.* 27, 861–874 (2006).