Milan dynamic noise mapping from few monitoring stations: statistical analysis on road network

Giovanni ZAMBON[1]; Roberto BENOCCI[2]; Alessandro BISCEGLIE[3]; H. Eduardo ROMAN[4]

[1-2-3] Dipartimento di Scienze dell'Ambiente e del Territorio e di Scienze della Terra, University of Milano - Bicocca, Italy

[4] Dipartimento di Fisica "G. Occhialini", Università degli Studi di Milano–Bicocca, Milan, Italy

ABSTRACT

One of the main objectives of "Dynamap" project is to realize a dynamic noise map within a given area of the city of Milan, containing about 2000 road arches, by using approximately 20 continuous measuring stations. This map is scheduled to be updated dynamically within a period of time varying from 5 minutes to 1 hour. The ability to obtain reliable maps within a complex urban context is linked to the fact that the vehicle flow patterns are very regular. In the first phase of the project, we monitored nearly 100 roads for a period of at least 24 hours. Then, these roads have been aggregated by means of a cluster analysis, depending on their noise level profiles. The analysis of the distribution of a non-acoustical parameter in the clusters, such as the vehicular flow rate, allowed to attribute a specific noise profile to a road not present in the database. It was also possible to select the most representative arches, among the 2000 arches to be mapped, where the monitoring stations will be installed. It was finally estimated the error associated with the noise levels of the dynamic maps as a function of the updating time interval.

Keywords: urban noise mapping; dynamic map; clustering algorithm.
I-INCE Classification of Subjects Number(s): 52.3, 76.9

1. INTRODUCTION

The DYNamic Acoustic MAPping (DYNAMAP) project, co-financed in the framework of the LIFE 2013 program, has the aim of developing a dynamic approach for noise mapping that is capable of updating environmental noise levels through a direct link with a limited number of noise monitoring stations. A dedicated system will be implemented for noise related to road infrastructures in two pilot areas: a part of the agglomeration of Milan and a section of the Motorway A 90 around Rome. One of the main goals of the project is to create a low-cost sensors network, properly sized in respect of the whole road 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. Noise maps are updated by scaling the noise levels of pre-calculated noise maps as functions of the difference observed between measured and calculated original grid data (at specific points). This operation should be performed for each road in the mapping area. The total map is updated by energetic summation of single source levels from updated basic noise maps. In this paper, we provide a general method to identify and optimize the number and the location of sites to be monitored for the maps updating in urban area, where the road network is large and quite complex. Generally, roads sharing the same characteristics for some parameters, such as vehicle flow capacity and number of lanes, are grouped together. Such parameters are usually included in the functional classification of roads and linked to their role played in urban mobility. However, this classification generally doesn't reflect the actual use of roads and the actual noise emission. For a better description of the real behaviour of noise in complex network, we approached the problem considering an aggregation method based upon similarities among the 24-h continuous acoustic monitoring of the hourly equivalent LAeqh levels. For this purpose, a statistical

[1] giovanni.zambon@unimib.it
[2] roberto.benocci@unimib.it
[3] alessandro.bisceglie@unimib.it
[4] eduardo.roman@mib.infn.it

cluster analysis has been conducted as a tool for grouping together road segments sharing the same noise behavior, starting from an acoustic measurements dataset performed in Milan. 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). Within the DYNAMAP project, 24 roads representative of the clusters will be identified in the urban pilot area and continuously monitored to provide noise levels for noise map updating. The maps that will be obtained using this method will be associated with an error that will depend on the time of representation chosen.

2. STATISTICAL ANALYSIS

2.1 Acoustic Database

The dataset considered in the present work refers to the road traffic noise of the city of Milan, Italy, and is made of 218 patterns of 24-h continuous monitoring of the hourly equivalent levels, L_{Aeqh}, in 93 different sites 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 [1]. 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), we referred each i^{th} hourly L_{Aeqhij} level of the j^{th} temporal pattern to its corresponding daytime level, L_{Aeqdj}:

$$\delta_{ij} = L_{Aeqh_{ij}} - L_{Aeqdj} \qquad (i = 1,\dots, 24\ h;\ j = 1,\dots, 93) \tag{1}$$

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.

2.2 Cluster Analysis

The Italian legislative classification of roads is mainly based on the geometry of the road and, therefore, not always corresponds to its actual use by the urban traffic. This is the reason why we decided to approach the sampling issue not inside each functional group but 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 [2-6] based upon cluster categorization. The clustering method applied is widely described in [7]. Here we report the main results.

The analysis 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 [8].

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 and of roads of class E for cluster 2. 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 1 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.

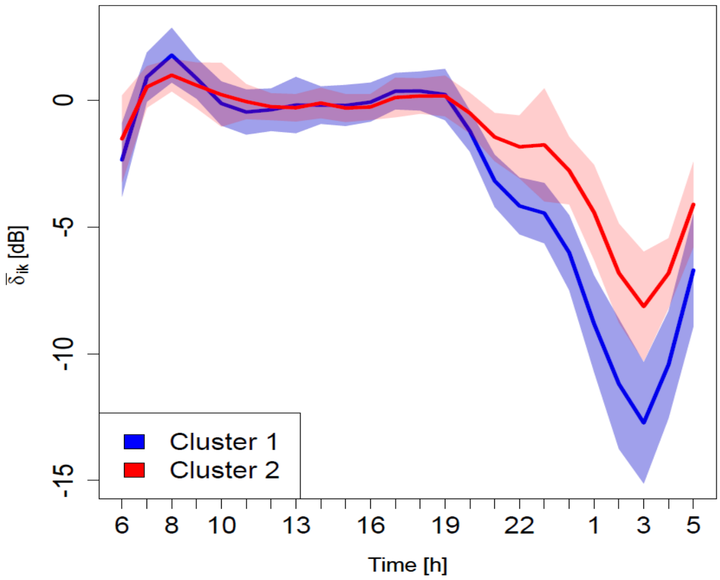


Figure 1 – Mean cluster profiles, $\bar{\delta}_{ik}$, and the corresponding ± standard deviation

2.3 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, $\bar{\delta}_{ik}$, (i= 1,….,24*(60/Integr.Time); k= 1, 2) for cluster 1 and 2 are shown in the Figures. 2-3.

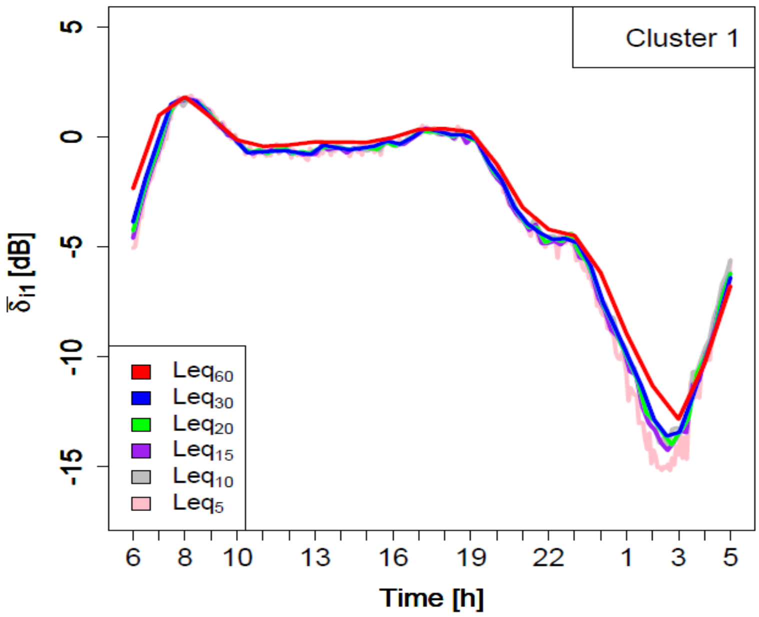


Figure 2 – Comparison between the mean profiles $\bar{\delta}_{i1}$ for cluster 1 with different temporal discretization as a function of daily time

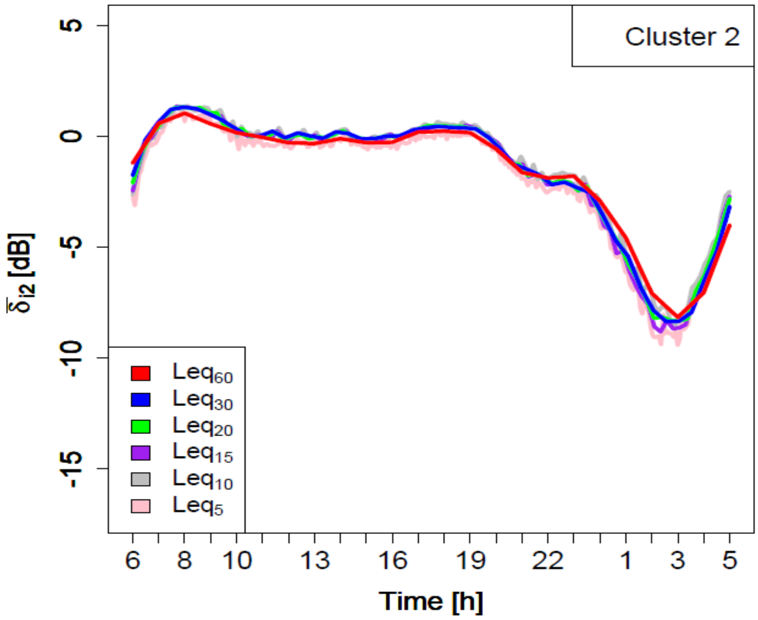


Figure 3 – Comparison between the mean profiles $\bar{\delta}_{i2}$ for cluster 2 with different temporal discretization as a function of daily time

The mean profiles with different temporal discretization present similar trends for both clusters. Higher oscillations are observed for smaller temporal intervals especially in the night period. The composition of the clusters remains essentially similar to the hourly data.

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.

2.4 Error Analysis at Different Temporal Discretization

In Figure 4, we report the standard deviations [dB] as a function of daily time for the different time intervals for all arches considered. According to Figure 4, we can derive a few suggestions on how to choose a proper temporal representation of the dynamic map maintaining a reasonable error.

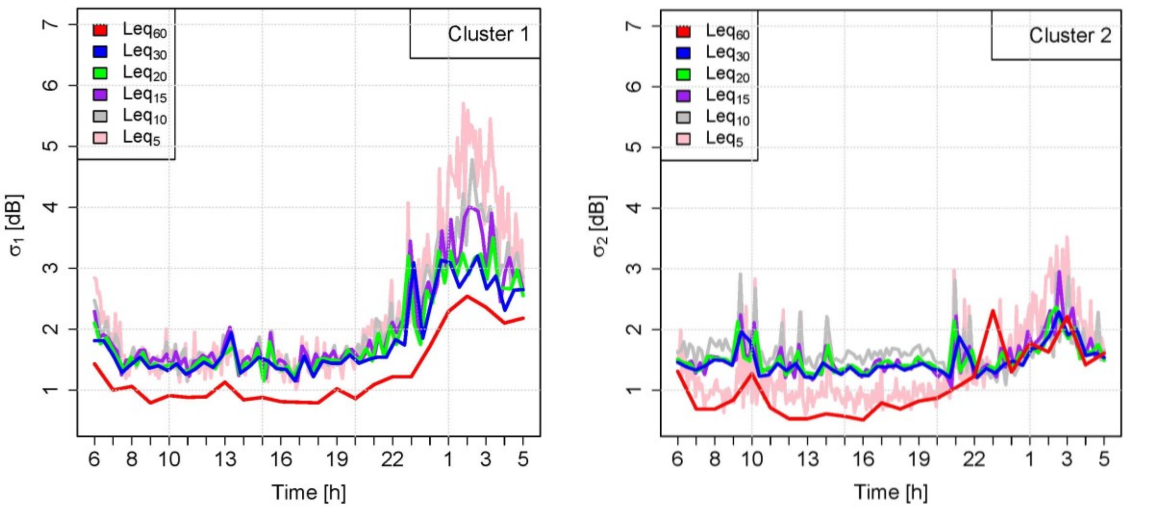
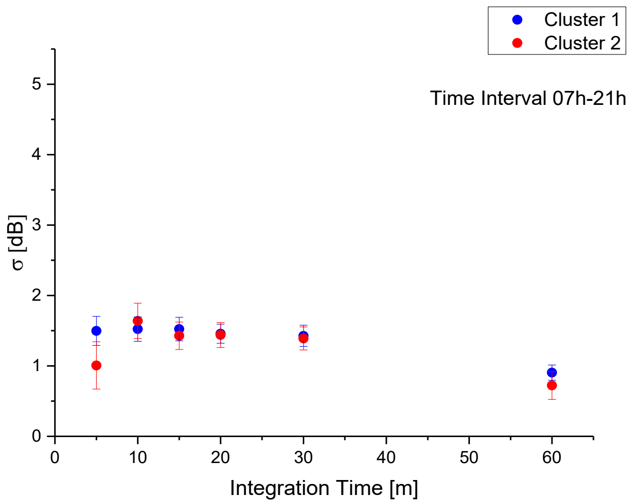


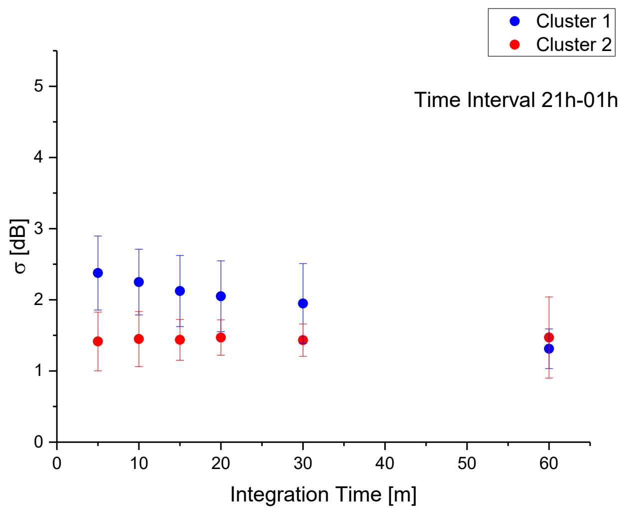
Figure 4 – Standard deviations [dB] as a function of daily time for the different time intervals for all arches considered for cluster 1 and 2

For this reason, we opted for the following division of daily time:

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

In Figure 5, we present the mean standard deviations as a function of the integration time and for the three 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.





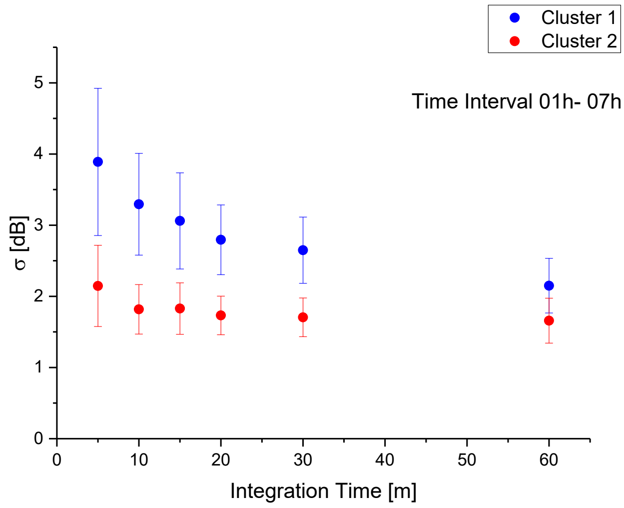


Figure 5 – Mean standard deviations, σ, [dB] as a function of integration time for the different time intervals.

3. NON-ACOUSTIC PARAMETER

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".

The non-acoustic parameter can be chosen by considering two (or more) hours during the day. 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,

$$x = \sqrt{(\mathrm{Log}F_i)^2 + (\mathrm{Log}F_j)^2} \tag{2}$$

with F_i and F_j being the flows at the i^{th} and j^{th} 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. One can add more hours to the definition of x in Eq.(2) 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 for the noise measurements. For each cluster, we have calculated the values of x according to Eq.(2) and studied their distribution functions P(x). We are interested in finding an analytical representation for both P(x) and, therefore, we 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 Figure 6.

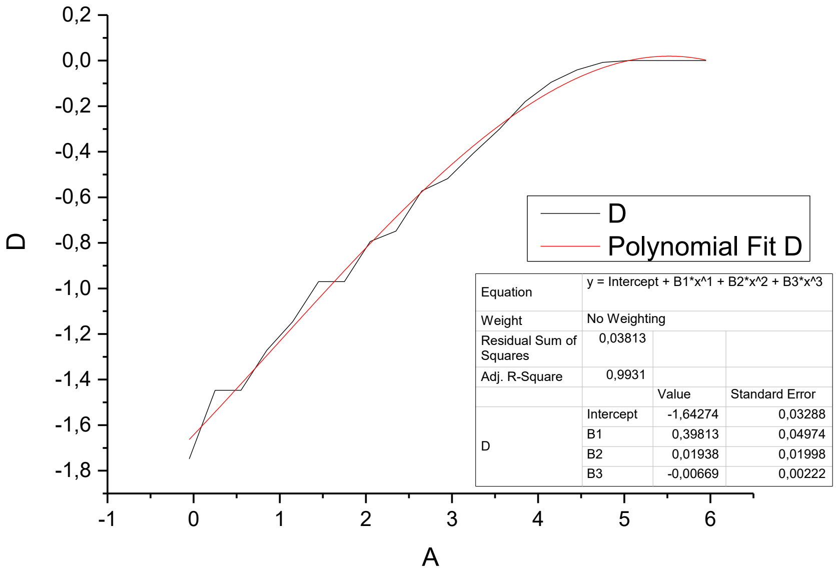


Figure 6 – The cumulative distribution I(x) is fitted using the analytical expression I(x) = $10^{f(x)}$, where f(x) is a polynomial of third degree. In the plot we have used the notation A=x and D=I(x).

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 Figure 7 we compare the histograms of x with the analytical fit. We can see from Figure 7 that both distribution functions 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.

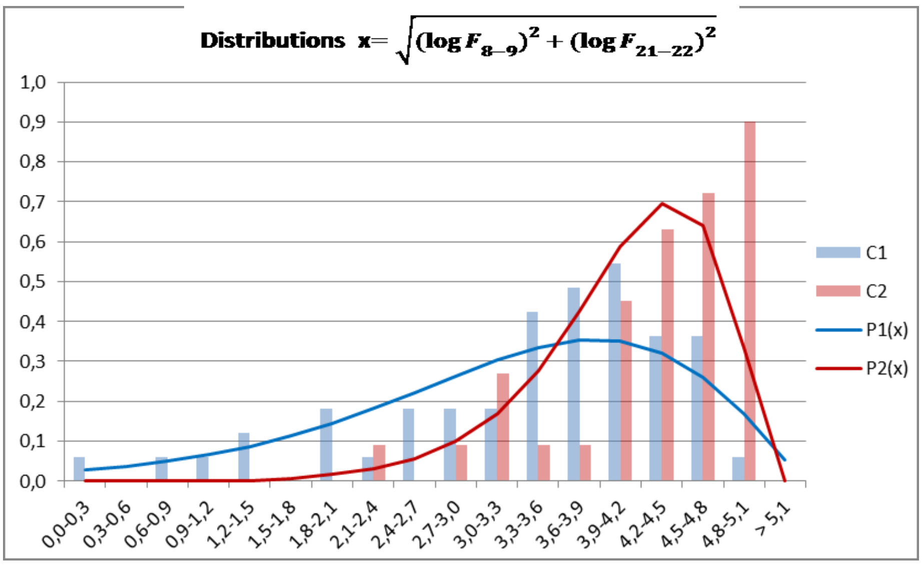


Figure 7 – Distribution Function P(x) vs. non-acoustic parameter x for Cluster 1 and 2. The histograms of x are represented by the vertical bars, while the analytical fits by the continuous lines obtained using the procedure explained in Figure 6

Due to the large superposition of the two cluster distributions $P_1(x)$ and $P_2(x)$ shown in Figure 7, 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 (α1, α2) of the linear combination can be obtained for each value of x using the relations: $\alpha_1=P_1(x)$ and $\alpha_2=P_2(x)$. 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, 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:

$$\beta_1 = \frac{\alpha 1}{\alpha 1+\alpha 2} \quad \text{and} \quad \beta_2 = \frac{\alpha 2}{\alpha 1+\alpha 2} \tag{3}$$

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

$$\delta_{pred}(h) = \beta_1 * \delta_{C1}(h) + \beta_2 * \delta_{C2}(h) \tag{4}$$

with $\delta_{C1}(h)$ and $\delta_{C2}(h)$ representing the mean values of the noise (Fig. 1) for each Cluster 1 and 2, respectively. The error made in using Eq. (4) can be estimated by calculating the standard deviation ε of the prediction $\delta_{pred}(h)$ from the measured values $\delta_{meas}(h)$, that is:

$$\varepsilon^2 = \frac{1}{24}\sum_{h=1}^{24}\left[\delta_{meas}(h) - \delta_{pred}(h)\right]^2 \tag{5}$$

3.1 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 1).

The mean errors for the corresponding predictions (Eq. 5) 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 1, 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 Figure. 8, 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.

Table 1 - 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)
$X = \sqrt{[(\text{Log}F_{8\text{-}9})^2 + (\text{Log}F_{21\text{-}22})^2]}$	1.34	0.57	0.37	2.98
$X = \text{Log}(T_D)$	1.36	0.57	0.36	2.86
$X = \text{Log}(T_N)$	1.37	0.59	0.43	2.84
$X = \sqrt{[(\text{Log}T_D)^2 + (\text{Log}T_N)^2]}$	1.31	0.56	0.29	2.80
$X = \text{Log}(T_T)$	1.33	0.57	0.31	2.90

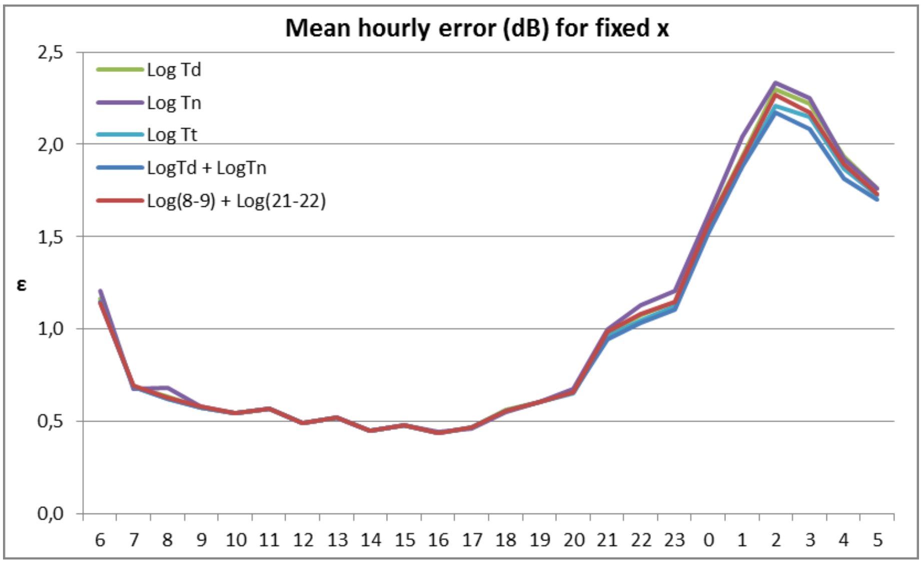


Figure 8 – Mean hourly error (dB) for the five non-acoustic parameter x shown in Table 1

4. 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. The number of maps 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. 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. 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 on finding those hours which better represent the behavior of the flow inside each one of the groups we are looking for.

4.1 Discretization of the parameter x for the case x= $LogT_T$

As an exemple we present the results for x=$LogT_T$. It includes the parametrizations of the distribution functions $P_{1,2}(x)$, total distribution function for x within the urban zone of interest, its 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. (3) and (4). Here, we report the results for the case x= $LogT_T$. Analytical fit functions for $P_{1,2}(x)$ are the following:

$f_1(x)$ = -1.56586-0.24459x+0.28834x2-0.03526x3

$f_2(x)$ = -15.2673+7.01263x -1.02922x2+0.04708x3

where f_1 and f_2 refer to cluster 1 and 2 respectively.

The comparison with the empirical data yields Figure 9. The six groups are now found according to the diagram shown below (Fig. 10).

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). 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. 9). 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.

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 α1 and α2 from the analytical formulas for P(x).

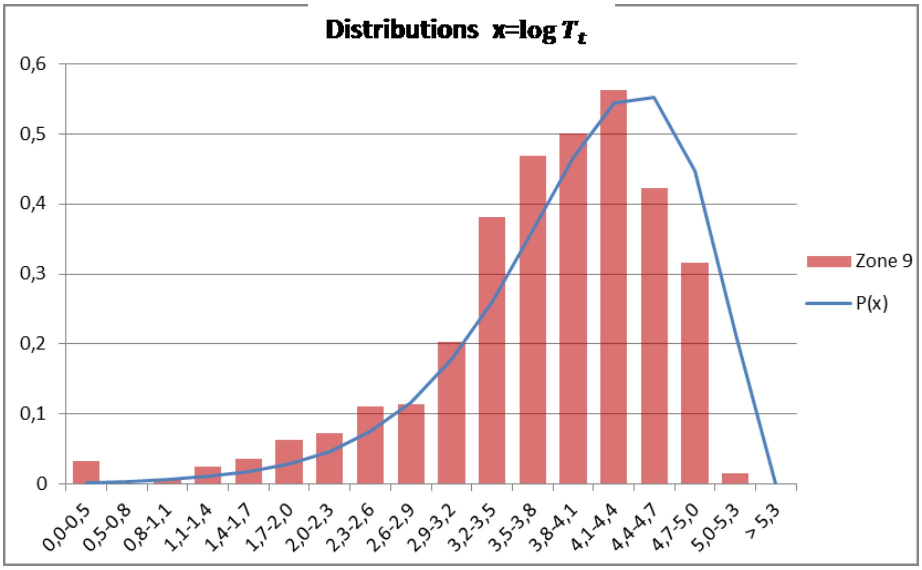


Figure 9 – Distribution of x= LogT_T for the whole Zone 9 of Milano(Histogram).The continuous line (P(x)) represents the corresponding distribution of x for the 93 arches of the recording stations

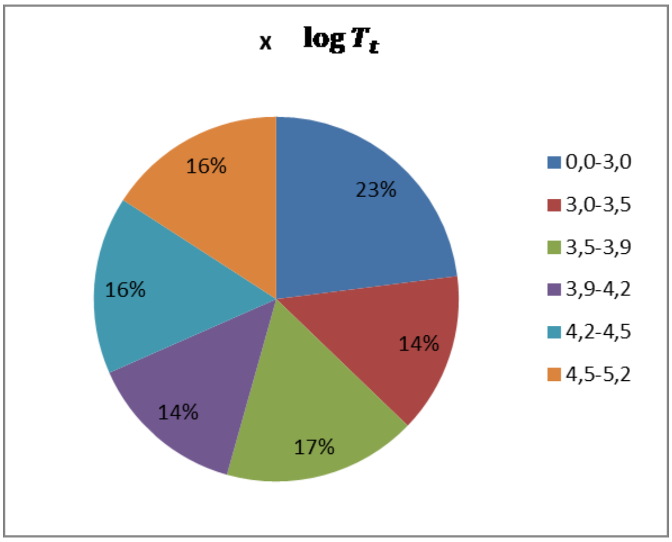


Figure 10 - The six groups of the parameter x = log T_T

4.2 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.

Here, we report the results for the choice x=Log T_T. The top 6 stretches are displayed in Table 2.

Table 2 - 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

The mean group flows are as displayed in Figure 11.

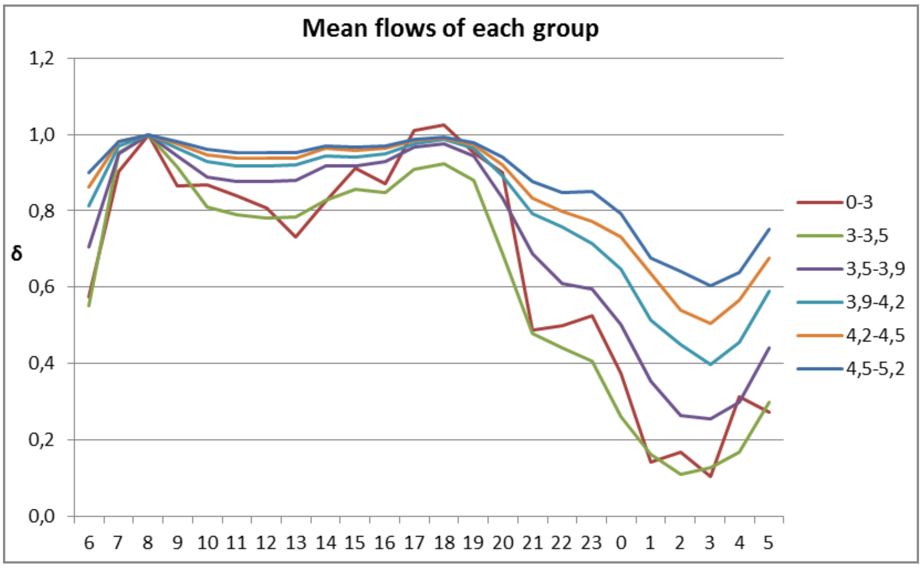


Figure 11 - Mean flows for roads inside each one of the 6 groups shown in Figure 10. This graph shows the mean hourly variations of the flow for each of the 6 groups by normalizing them using the rush hour values (08-09).

5. CONCLUSIONS

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, $x = \mathrm{Log}T_T$, is a suitable choice. The location of the recording stations, in number of 24, have been suggested by analyzing the traffic flows inside each group. 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, one can locate 4 stations, for a total of 24.

ACKNOWLEDGEMENTS

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

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

REFERENCES

1. "Tecniche di rilevamento e di misurazione dell'inquinamento acustico" Italian decree D.M. Ambiente 16/3/1998.
2. S. Fidel; Nationwide urban noise survey, J. Acoust. Soc. Am., 64, 198-206, 1978.
3. A. L. Brown, K. C. Lam, Urban noise surveys, Appl. Acoust., 20, 23-39, 1987.
4. K. Kumar, V. K. Jain, A study of noise in various modes of transports in Delhi, Appl. Acoust., 43, 57-56, 1994.
5. H. M. E. Miedema, H. Vos, Exposure-response relationships for transportation noise, J. Acoust. Soc. Am. 104, 3432-3445, 1998.
6. A. Garcia, L. J. Faus, Statistical analysis of noise levels in urban areas, Appl. Acoust., 34, 227-247, 1991.
7. G. Zambon, R. Benocci, G. Brambilla, Cluster categorization of urban roads to optimize their noise monitoring. Environ Monit Asses (2016) 188:26. DOI 10.1007/s10661-015-4994-4.
8. G. Brambilla, G. B. Cannelli, S. Santoboni, "Valutazione e Previsione dell'Inquinamento Acustico Urbano - Mappa del Rumore di Roma", CNR, Istituto di Acustica "O.M. Corbino", Scientific Report N. 56 (1984)