inter.noise 27-30 AUGUST HONG KONG 2017

Error Analysis of Real-time Acoustic Maps for Dynamap

Roberto BENOCCI[1]; H. Eduardo ROMAN[2]; Maura SMIRAGLIA[1]; Giovanni ZAMBON[1]

[1] Dipartimento di Scienze dell'Ambiente e del Territorio e di Scienze della Terra (DISAT),

Università di Milano-Bicocca, Piazza della Scienza 1, 20126 Milano, Italy

[2] Dipartimento di Fisica, Università di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy

ABSTRACT

In this work we evaluate the mean statistical errors incurred within the real-time noise mapping of Dynamap for the equivalent noise levels of each acoustic map. In order to estimate the total error we rely on the Central Limit Theorem which asserts that for the sum of *N* uncorrelated random processes the total variance of the sum, σ_N^2, is given by the sum of the individual variances. In the present case, we have identified four processes whose fluctuations contribute to the total variance (error), which denote as σ_T^2, yielding the result $\sigma_T^2 = \sigma_{pred}^2 + \sigma_{stat}^2 + \sigma_{comp}^2 + \sigma_{sample}^2$, corresponding to the intrinsic prediction error of the method, σ_{pred}^2, the statistical variance of the equivalent noise levels measured by the monitoring stations, σ_{stat}^2, the different cluster compositions for different time intervals, σ_{comp}^2, and the variance due to stratified sampling, σ_{sample}^2.The overall mean error is expected to be bounded by about (1-3) dB.

Keywords: Noise, Traffic, Acoustic Maps I-INCE Classification of Subjects Number(s): 76.

1. INTRODUCTION

Dynamap is a LIFE+ project aimed at developing a dynamic noise mapping system able to detect and represent in real time the acoustic impact of road infrastructures through the use of a limited number of noise monitoring stations. It is based on the idea of finding a suitable set of roads which display similar traffic noise behavior (temporal noise profile over a whole day) so that one can group them together into a single noise map [1- 6]. Each map thus represents a group of road stretches whose traffic noise will be updated periodically, typically every five minutes during daily hours and every hour during night. The information regarding traffic noise will be taken continuously from (24) monitoring stations distributed appropriately over the urban zone of interest. This project will be, therefore, an effective tool for local administrations to assess traffic strategies and transport policies.

To achieve this goal, we have performed a detailed analysis of traffic noise data, recorded every second from 93 monitoring stations randomly distributed over the whole urban area of the City of Milan. From the analysis, we develop a model for predicting the traffic noise of an arbitrary road stretch within the same area. Our final results are presented for a restricted area, the urban Zone 9 of Milan. We have separated the whole set of (about 2000) stretches into six groups, each one represented by a noise map, and give a prescription for the locations of the future 24 monitoring stations. From our analysis, it is estimated that the mean overall error for each group of stretches (noise map), averaged over the 24 hours, is about 2 dB.

[1] roberto.benocci@unimib.it
[2] eduardo.roman@mib.infn.it
[1] m.smiraglia@campus.unimib.it
[1] giovanni.zambon@unimib.it

2. THE TWO-CLUSTER NOISE DISTRIBUTION FUNCTIONS

2.1 The equivalent noise level clustering

The analysis of hourly traffic noise from the 93 monitoring stations have been performed using standard clustering techniques (see [1,2]), yielding two clusters, here denoted as Cluster 1 and 2, which display similar time noise dependences. The results are shown in Figure 1, for the mean normalized acoustic equivalent levels $\bar{\delta}_{ik}$ for each cluster k=1,2 as a function of the hour i of the day.

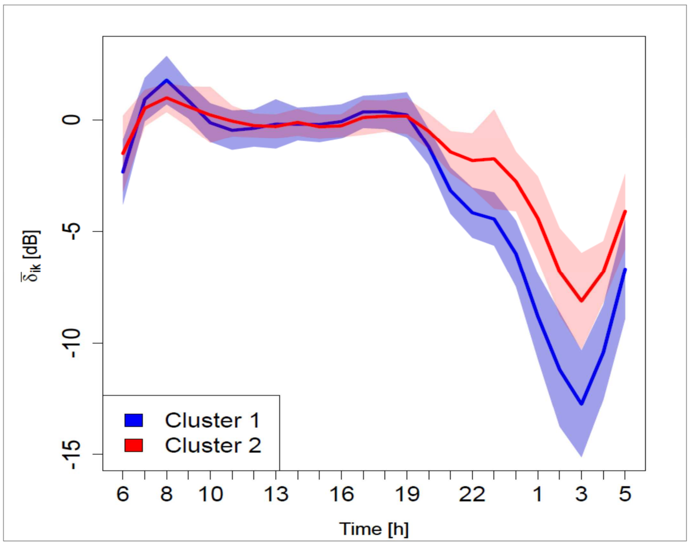

Figure 1 – Mean cluster normalized equivalent level profiles (continuous lines), $\bar{\delta}_{ik}$ [dB], for the two clusters, k=(1,2), as a function of hour i of the day (also indicated with the letter h). For illustration, we show the corresponding dispersion of the hourly values within one standard deviation (colored bands). We will use also the notation $\delta_{C1}(h)$ and $\delta_{C2}(h)$ for the mean equivalent noise levels of Cluster 1 and 2 as a function of hour h. It was found that Cluster 1 has 56 elements and Cluster 2 has 37 components.

These two clusters summarize the typical behavior roads can be divided into. In particular, Cluster 1 represents roads characterized by strong variations of the equivalent level between day and night periods and therefore of the corresponding traffic flow in these two periods. On the contrary, Cluster2 presents less variations which can be associated with rather high traffic flow rate both during the day and night period.

2.2 The non-acoustic parameter x and its distribution function P(x)

In order to predict traffic noise for a given road stretch, when a direct measurement is not practicable, we need to define a non-acoustic parameter which we denote generically as x and is related to model calculations of the traffic flow on that road stretch. The hourly behavior of the traffic noise for a given road stretch n, characterized by a value x_n, can be described in terms of the distribution functions of the variable x obtained from each one of the noise Clusters 1 and 2. The corresponding distribution functions, $P_1(x)$ and $P_2(x)$, are shown in Figure 2, for the choice of x given by the logarithm of the total daily traffic flow rate x = Log(T_T). Other choices of x lead to similar qualitative behavior. We concentrate here on Log(T_T) which has been taken as our working choice.

As one can see from Figure 2, there is a conspicuous overlap between the two distributions suggesting that a sharp separation into two clusters is not possible in general. We conclude that a given value of x has components in both Clusters, meaning that the temporal evolution of the noise for that given road stretch is partly due to Cluster 1 and partly due to Cluster 2. The idea of the method is to evaluate the probability β_1 that x belongs to cluster 1, and β_2 = 1- β_1 that it belongs to Cluster 2. The corresponding values of β are given by the relations,

$$\beta_1(x) = P_1(x)/(P_1(x)+ P_2(x)) \quad \text{and} \quad \beta_2(x) = P_2(x)/(P_1(x)+ P_2(x)). \qquad (1)$$

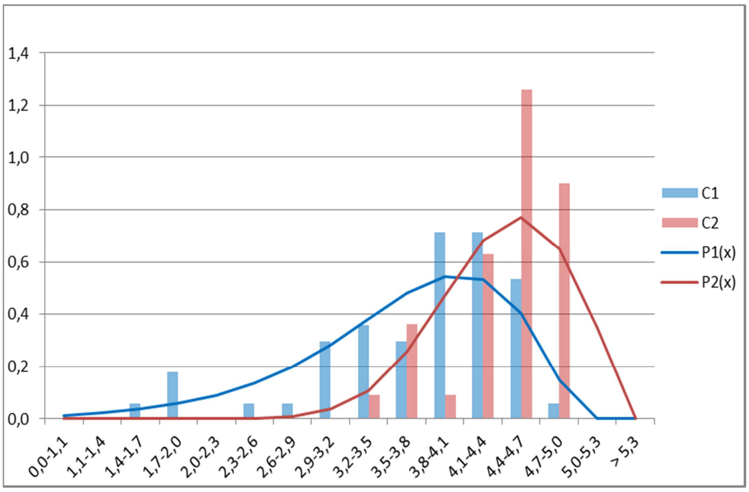

Figure 2 - Distribution functions $P_1(x)$ and $P_2(x)$ for $x = \log(T_T)$, on noise Clusters 1 and 2, respectively. The continuous lines represent logarithmic fits to the actual histograms using the cumulative distribution functions, and are used here to determine $\beta_{1,2}$ using Eq. (1) accurately.

2.3 The prediction of the equivalent noise levels from the Clusters 1 and 2

The paper Using the values of $\beta_{1,2}$ we can predict the hourly variations $\delta_x(h)$ for a given value of x according to,

$$\delta_x(h) = \beta_1(x)\, \delta_{C1}(h) + \beta_2(x)\, \delta_{C2}(h) \tag{2}$$

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

$$\varepsilon^2 = \frac{1}{24} \sum_{h=1}^{24} [\delta_x(h) - \delta_{meas}(h)]^2. \tag{3}$$

2.4 The distribution function of x for Zone 9 of Milan and choice of monitoring stations

One important requirement for the location of monitoring stations is that the corresponding road stretches, where they are positioned, have values of x covering essentially all the relevant values of the urban zone under consideration. To check this for the 93 stations we have used, the distribution function of x for the zone 9 (Z9) of Milan has been determined and compared with the P(x) obtained from the stretches where the monitoring stations were positioned. The results are shown in Figure 3, suggesting that the choice of station locations was appropriate. To be noted is that the later has been fitted analytically from the 93 x-values, which are more or less scattered over the same range of x as the one for Z9.

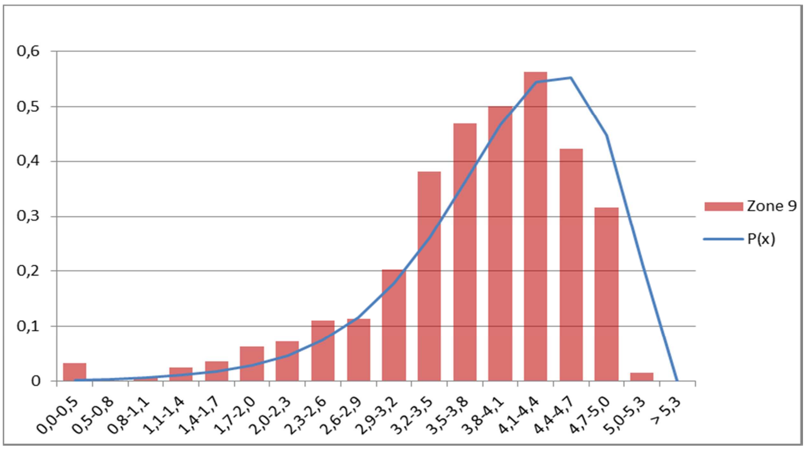

Figure 3 – Distribution function P(x) of $x = \text{Log}(T_T)$ for the whole zone 9 of Milan City (Histogram). The continuous line represents the same fitting procedure as in Figure 2, from the 93 road stretches of the monitoring stations.

2.5 The acoustic maps and the associated groups of road stretches

For the effective implementation of Dynamap, we have only 24 stations at our disposal. In this case, the station locations should not be chosen purely at random, but distributed in such a way as to approximate as much as possible the empirical distribution of x from Z9 (see Figure 3). For this reason, we have divided the whole range of x values from Z9 into (arbitrarily) 6 pieces, called groups, so that each one contains approximately the same number of road stretches, thus mimicking a uniform distribution of locations. The decision of taking six groups is dictated by practical reasons about the number of acoustic maps that will be handled in the actual implementation of Dynamap. To be noted is that, once the number of groups has been chosen, six in this case, the arbitrariness in the determination of the groups is removed.

Once we have determined the six groups of x values (Table 1), the acoustic map of each group will be obtained from Eq. (2) by using the mean values $\bar{\beta}_1$ and $\bar{\beta}_2$ within each group. They read,

$$\bar{\beta}_1(\bar{x}_g) = P_1(\bar{x}_g)/(P_1(\bar{x}_g)+ P_2(\bar{x}_g)) \quad \text{and} \quad \bar{\beta}_2(\bar{x}_g) = P_2(\bar{x}_g)/(P_1(\bar{x}_g)+ P_2(\bar{x}_g)), \tag{4}$$

where $\bar{x}_g$ is the mean value of x within group g. The values of β for each group are reported in Table 1.

Table 1 - Mean values of $\bar{\beta}_1$ and $\bar{\beta}_2$ for the six groups of x = log(T_T) within Z9.

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
$\bar{\beta}_1$	0,99	0,81	0,63	0,50	0,41	0,16
$\bar{\beta}_2$	0,01	0,19	0,37	0,50	0,59	0,84

3. ESTIMATION OF THE ERRORS IN PREDICTING THE ACOUSTIC MAPS

In this Section we evaluate the mean statistical errors incurred within the present method for predicting the hourly equivalent noise levels of each acoustic map, i.e. for each group of road stretches. For convenience the six groups will be denoted by capital letters: g=A,B,C,D,E,F.

In order to estimate the total error we rely on the Central Limit Theorem (CLT) which asserts that for the sum of N uncorrelated random processes the total variance of the sum, σ_N^2,is given by the sum of the individual variances. In our case, we have identified five processes whose fluctuations contribute to the total error, σ_T^2, yielding the result,

$$\sigma_T^2 = \sigma_{pred}^2 + \sigma_{stat}^2 + \sigma_{comp}^2 + \sigma_{sample}^2, \tag{5}$$

corresponding to the intrinsic prediction error of the method, σ_{pred}^2, the statistical variance of the equivalent noise levels measured by the monitoring stations, σ_{stat}^2, the different cluster compositions for different time intervals, σ_{comp}^2, and the variance due to stratified sampling, σ_{sample}^2. In what follows, we discuss each term in Eq. (5) separately.

3.1 The error of the prediction of the equivalent noise level: σ_{pred}^2

This error has been discussed above in Eq. (3). In order to estimate it for each group g, we calculate the mean errors, $\bar{\varepsilon}_1^2$ and $\bar{\varepsilon}_2^2$, for the predictions of the monitoring station noise levels for Cluster 1 and Cluster 2, using Eq. (3). Once these mean errors are known we estimate the mean prediction error inside each group according to the relation,

$$\sigma_{pred}^2(\bar{x}_g) = \bar{\beta}_1(\bar{x}_g)\ \bar{\varepsilon}_1^2 + \bar{\beta}_2(\bar{x}_g)\ \bar{\varepsilon}_2^2, \tag{6}$$

where $\bar{\beta}_1(\bar{x}_g)$ and $\bar{\beta}_2(\bar{x}_g)$ are given in Table 1.

3.2 The statistical variance of monitoring stations from Cluster 1 and 2: σ_{stat}^2

The variance has been calculated by following the three time zones of the day, daily, evening and night, within which we have used (5,15,60) mins intervals, respectively. In the calculations, we have considered the mean measured noise levels from each station in the hourly Cluster 1 and Cluster 2, $\bar{\sigma}_1^2$ and $\bar{\sigma}_2^2$, respectively. Therefore, for each group g we have the following relation,

$$\sigma_{stat}^2(\bar{x}_g) = \bar{\beta}_1(\bar{x}_g)\ \bar{\sigma}_1^2 + \bar{\beta}_2(\bar{x}_g)\ \bar{\sigma}_2^2. \tag{7}$$

3.3 Cluster composition Error on time intervals smaller than the hour: σ^2_{comp}

Our clustering calculations are based on hourly data of equivalent traffic noise. In the practical applications, we will deal with time intervals of 5 mins, 15 mins and an hour. Therefore the question arises whether the composition (set of road stretches) belonging to each Cluster remains the same, and if does not, how to quantify the differences. We have therefore repeated the clustering process for the 5 mins and 15 mins data and compared them with the hourly results, in the following way. For each time interval τ = (5, 15) mins, we have obtained the mean values $\delta_{meas,C1,2(\tau)}(\tau)$ every τ mins, averaged over the roads belonging to Cluster 1 and 2 as given by the clustering analysis for time interval τ. Then, we have calculated the same mean quantities but using the cluster composition obtained from the hourly Clusters, $\delta_{meas,C1,2(h)}(\tau)$, and obtain the mean-square differences,

$$\sigma^2_{comp,C1,2}(\tau) = \frac{1}{N(\tau)} \sum_{t=1}^{N(\tau)} [\delta_{meas,C1,2(\tau)}(t) - \delta_{meas,C1,2(h)}(t)]^2, \quad (8)$$

where N(τ)= 24 (60/τ), is the number of data points within the 24 hours. In order to calculate the error associated to each group, we multiply each Cluster variance in Eq. (8) by the corresponding mean $\bar{\beta}_1(\bar{x}_g)$ and $\bar{\beta}_2(\bar{x}_g)$, for Cluster 1 and 2, respectively,

$$\sigma^2_{comp}(\bar{x}_g) = \bar{\beta}_1(\bar{x}_g)\ \sigma^2_{comp,C1}(\tau) + \bar{\beta}_2(\bar{x}_g)\ \sigma^2_{comp,C2}(\tau). \quad (9)$$

3.4 Error associated with the Stratified Sampling method: σ^2_{sample}

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 [7]. 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_{pop}, the mean $\bar{x}$ of all samples taken from the same population will be approximately equal to the mean of the population (expected value). Furthermore, all of the samples will follow an approximate normal distribution pattern, with all variances σ^2_x being approximately equal to the variance of the population divided by each sample's size n, that is $\sigma^2_x = \sigma^2_{pop}/n$. The maximum error σ_{sample}, 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 $Z_\alpha\sigma_x$ with Z_α = 1.96) is, $\sigma_{sample} = Z_\alpha\sigma_x$. The minimum number of elements of a sample n_{min} for a correct estimation of the mean of the population within an accuracy ± σ_{sample} is:

$$n_{min} = Z^2_\alpha\ \sigma^2_{pop}\ /\ \sigma^2_{sample}. \quad (11)$$

In general, σ^2_{pop} is unknown, therefore, we used in its place the sample variance σ^2 , i.e.

$$n_{min} = Z^2_\alpha\ \sigma^2\ /\ \sigma^2_{sample}. \quad (12)$$

In our case, the overall sample size consists of 192 measurements to be divided into each cluster. According to Eq. (12), the actual expected deviation σ_{sample} for the two clusters with a number of measurements, n_{c1}, n_{c2} , respectively, reads,

$$\sigma_{sample_i} = Z_\alpha\sigma_i\ /\ \sqrt{n_{ci}} \quad (i=1,2). \quad (13)$$

Figures (4-6) show the expected deviation σ_{sample} calculated for a temporal discretization τ = (5, 15, 60) min.

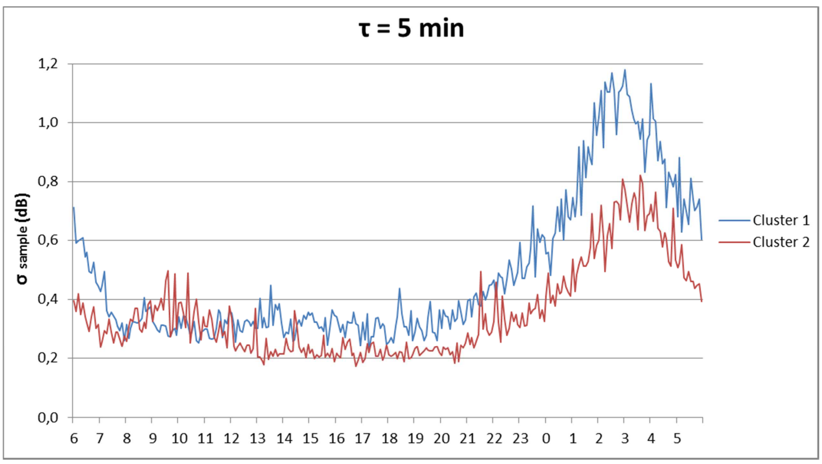

Figure 4 - Expected deviation σ_{sample} of the sample mean from the population mean calculated

for a temporal discretization τ=5 min.

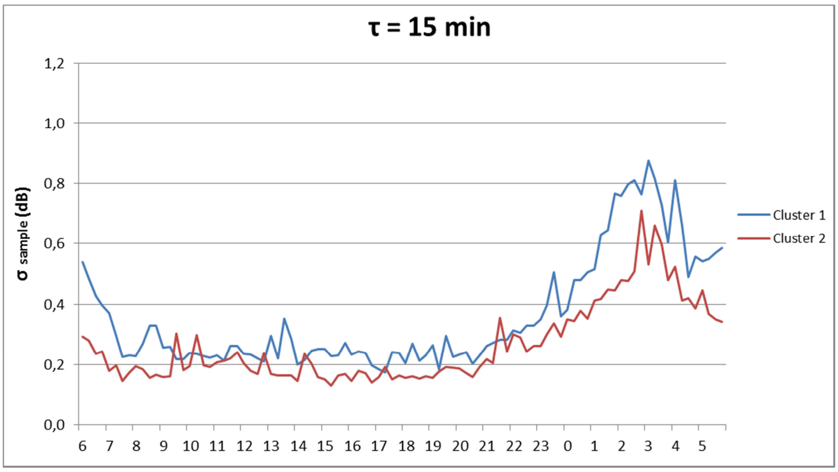

Figure 5 - Expected deviation σ_{sample} of the sample mean from the population mean calculated for a temporal discretization τ=15 min.

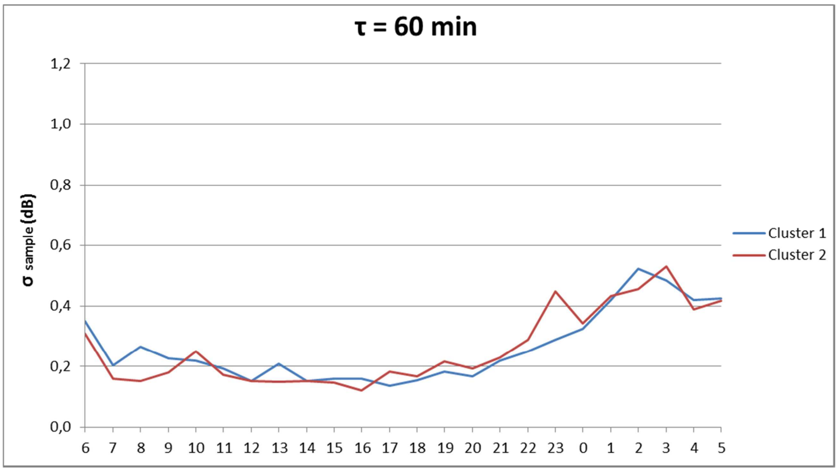

Figure 6 - Expected deviation σ_{sample} of the sample mean from the population mean calculated for a temporal discretization τ=60 min.

In Table 2, the mean expected deviation $\overline{\sigma_{sample}}$ calculated within the three time intervals T_h defined as: T_h=(07:00 - 21:00),(21:00 - 01:00),(01:00 - 07:00), for the three discretization times τ=(5, 15, 60) min. and the two clusters is reported. The time interval subdivision allows us to estimate an overall mean $\overline{\sigma_{sample_1}}$=0.37 dB and $\overline{\sigma_{sample_2}}$ =0.33 dB for cluster 1 and 2, respectively. As is apparent, the sampling error can be neglected as compared to those reported in Table. 3.

Table 2: Mean expected deviation $\overline{\sigma_{sample}}$ calculated within the three time intervals used in this work: T_h= (07:00 - 21:00), (21:00 - 01:00), (01:00 - 07:00), for the three discretization times τ= (5, 15, 60) min, for the two clusters. The average daily values are therefore: 0.37 dB for C_1 and 0.33 dB for C_2.

Interval	**τ (min)**	**Cluster 1**		**Cluster 2**		**Tot. n° of measurements**
		n°measurements	$\overline{\sigma_{sample_1}}$ (dB)	n°measurements	$\overline{\sigma_{sample_2}}$ (dB)	
07:00-21:00	5	110	0.32	82	0.26	192
21:00-01:00	15	106	0.36	86	0.30	
01:00-07:00	60	125	0.44	67	0.42	

Since the stratified sampling method represents an intrinsic error due to the finite number of measurements, we take the same value $\sigma^2_{sample}(\bar{x}_g)$ =0.35 dB for each group g.

3.5 Summary of error components for each group g: σ^2_T

We report in Table 3 the values obtained for the different error contributions in Eq. (5) to the total error σ_T for each group g.

Table 3 – Mean daily values of the error contributions to σ_T[dB] for each group (acoustic map) g. The different symbols correspond to: The intrinsic prediction error of the method, σ_{pred}, the statistical variance of the equivalent noise levels measured by the monitoring stations, σ_{stat}, the different cluster compositions for different time intervals, σ_{comp}, and the error due to stratified sampling, σ_{sample}.

Error/Group	A	B	C	D	E	F
σ_{pred}	1.47	1.44	1.40	1.37	1.36	1.31
σ_{stat}	1.53	1.47	1.40	1.35	1.32	1.22
σ_{comp}	0.14	0.15	0.17	0.18	0.18	0.20
σ_{sample}	0.35	0.35	0.35	0.35	0.35	0.35
σ_T	2.16	2.09	2.02	1.96	1.94	1.83

4. CONCLUSIONS

At the present intermediate stage, Dynamap project has delivered important results that can be summarized as follows:

- Measurement of a statistical relevant road traffic noise sample.
- Statistical analysis to determine characteristic traffic noise behavior (sect. 2.1).
- Assignment of non-monitored road to one of the two clusters by means of a non-acoustic parameter (e. g. traffic flow rate) (sect. 2.2).
- Prediction of traffic noise of a generic road through the "mixing" of two mean noise time profiles (sect. 2.3).
    - Definition a restricted area, the urban Zone 9 of Milan, where to install the monitoring stations.
    - Separation of the whole set of (about 2000) stretches into six groups, each one represented by a noise map.
    - Method for the locations of the future 24 monitoring stations.

In particular, in this paper, we focused on the evaluation of the different uncertainty contributions associated with the dynamic noise mapping representation. The total S.D., reported in Table 3, represents the error in the prediction of the noise level for Dynamap for each group of road stretches (within zone 9 of Milan) averaged over 24 hours by taking into account the three time intervals T_h=(07:00 - 21:00), (21:00 - 01:00), (01:00 - 07:00), associated with the three discretization times τ= (5, 15, 60) min, respectively. The error remains confined within about 2 dB. This value is consistent with the reference bound reported in the GPG for Strategic Noise Mapping of the European Commission Working Group [8].

ACKNOWLEDGEMENTS

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

REFERENCES

1. Zambon G., Benocci R., Brambilla G. Statistical Road Classification Applied to Stratified Spatial Sampling of Road Traffic Noise in Urban Areas. Int. J. Environ. Res. 2016; 10(3):411-420.
2. Zambon G., Benocci R., Brambilla G. Cluster categorization of urban roads to optimize their noise Monitoring. Environ. Monitoring Assessment 2016; 188(26): DOI 10.1007/s10661-015-4994-4.
3. Zambon, G., Benocci, R., Bisceglie A., 2015. Development of optimized algorithms for the classification of networks of road stretches into homogeneous clusters in urban areas. In: The 22nd ICSV, Florence, Italy, 2015.
4. Zambon, G., Benocci, R., Angelini, F., Brambilla, G., Gallo, V., 2014. Statistics-based functional classification of roads in the urban area of Milan. In: The 7th Forum Acusticum, Krakow, Poland, 2014.
5. Zambon, G., Benocci, R., Bisceglie A., Roman, H. E., 2016. Milan dynamic noise mapping from few monitoring stations: Statistical analysis on road network. In: The 45th INTERNOISE, Hamburg, Germany, 2016.
6. Zambon G. et Al. Traffic noise monitoring in the city of Milan: construction of a representative statistical collection of acoustic trends with different time resolutions, Proceedings of the 23rd International Congress on Sound and Vibration, Florence, Italy, 12–16 July, (2015).
7. Kaufman L., Rousseeuw P. Finding Groups in Data, Wiley Series in Probability and Mathematical Statistics; 1990.
8. European Commission working Group. Assessment of Exposure to noise (WG-AEN). Good Practice Guide for Strategic Noise Mapping and the Production of Associated Data on Noise Exposure. Version 2. August 2007.