Urban noise functional stratification for estimating average annual sound level

Guillermo Rey Gozalo[a]
Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, 5 Poniente 1670, 3460000 Talca, Chile

Juan Miguel Barrigón Morillas and Carlos Prieto Gajardo
Departamento de Física Aplicada, Escuela Politécnica, Universidad de Extremadura, Avda. de la Universidad s/n, 10003 Cáceres, Spain

(Received 3 October 2014; revised 19 April 2015; accepted 2 May 2015)

Road traffic noise causes many health problems and the deterioration of the quality of urban life; thus, adequate spatial noise and temporal assessment methods are required. Different methods have been proposed for the spatial evaluation of noise in cities, including the categorization method. Until now, this method has only been applied for the study of spatial variability with measurements taken over a week. In this work, continuous measurements of 1 year carried out in 21 different locations in Madrid (Spain), which has more than three million inhabitants, were analyzed. The annual average sound levels and the temporal variability were studied in the proposed categories. The results show that the three proposed categories highlight the spatial noise stratification of the studied city in each period of the day (day, evening, and night) and in the overall indicators (L_{Adn}, L_{Aden}, and L_{A24}). Also, significant differences between the diurnal and nocturnal sound levels show functional stratification in these categories. Therefore, this functional stratification offers advantages from both spatial and temporal perspectives by reducing the sampling points and the measurement time. © *2015 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4921283]

[GB] Pages: 3198–3208

I. INTRODUCTION

Noise pollution is a major environmental problem which can affect cities of any population size[1,2] and represents a risk to people's health and quality of life.[3] Recent publications relate road traffic noise to health problems that affect a large part of the world's population. The latter are a clear priority in the action plans of healthcare systems: cardiovascular diseases,[4] diabetes,[5] etc. Such health problems and the quality of life as a measure of mental and physical health are associated with noise exposure or noise annoyance, and express degrees of dissatisfaction and disturbance with regard to noise exposure.[6] Therefore, as a first step, precise determination of noise exposure is required.

Most countries today conduct demographic censuses which allow us to estimate with sufficient precision the population residing in the different buildings of a city. The Spanish Statistical Office, UK National Statistics, etc., are the responsible bodies. Nevertheless, these censuses do not facilitate determination of spatial and temporal variability in sound levels, so computational methods are usually recommended by different standards and legislations.[7,8] These methods need a comprehensive spatial and temporal registry of the vehicular traffic flow of a city for adequate characterization of the sound source. However, most cities do not possess vehicle traffic counters or such counters are generally available for main roads only.[9] Therefore, the recommended computation methods for road traffic noise need different kinds of *in situ* measurements. Noise measurements to calibrate the model and to check the precision of the estimated noise values are also necessary.[10]

In this context, our research group has been working for some years on the development of a low-cost sampling method for *in situ* noise measurements. We term this method: categorization method. On the basis of the concept of street functionality, each stratum defined by the categorization method presents a sound level variability lower than the total sound spatial variability in a city. This has shown significant improvements in the reduction of sample points and in the estimation of noise levels in unsampled streets.[1,11] Recently, its applicability has been studied for urban centers whose populations range from 2000 inhabitants to 700 000.[2] It is a firm candidate to substitute the grid method, spatial sampling strategy collected by ISO 1996-2 standard in both the old[12] and the revised.[8] Recent studies show the advantages of the categorization method compared with the grid method.[1,11] Moreover, the definition of categories allows for a simple update when there are changes in the organization of road traffic. Consequently, this method could be applied to urban planning. These methodological innovations have not gone unnoticed in the scientific community, and several authors have adopted them with some variations.[13–16]

The second variable of importance in the sampling strategy is temporal variability. Noise indicators must be determined over the period of a year according to the European Directive[7] and some international standards.[8] This assessment can easily be performed with modern acoustical instrumentation, but noise-monitoring stations are quite expensive and installation (administrative permissions) and measurements can be time-consuming. For this reason, noise-monitoring stations can only be used at very specific locations. Thus, as an alternative, the

---

[a] Electronic mail: guille@unex.es

0001-4966/2015/137(6)/3198/11/$30.00

best temporal strategy is to register measurements for less than a year and extrapolate from them to obtain data for a year. The duration of most of these measurements is from minutes to hours.[13,15,17,18] Unfortunately, methods for assessing the L_{Aden} (day-evening-night sound level) from short-term measurements can produce significant inaccuracies if the measurement period is not representative of the period of a year, as in the case of singular noise events.[19] This error in short-term measurement has a strong relation with sound level variability.[20,21] It is evident that the lower the variability, the less error estimation will occur, and thus, a lower duration and a lower measurement number will be necessary to estimate the sound period evaluated.

The scientific literature describes statistical models used to evaluate the evolution of noise levels, as well as an attempt to predict the levels that would be achieved with a given probability.[22,23] However, many of these studies are restricted to a particular station,[24] and the resultant mathematical models are not applicable to other temporal series.

One important step similar to the functional stratification of spatial sound variability is the functional stratification of temporal sound variability. Thus, the precision of measurements should increase and categories requiring less measurements or a minor duration of short-term measurements should be identified.

In view of the above, the main hypothesis of this study is as follows: the variability of sound levels is related to the functionality of urban streets and this variability presents a statistically significant stratification. This hypothesis considers two perspectives: the spatial one and the temporal one. From the spatial perspective, recent research[25] has demonstrated the existence of significant functional stratification in average sound values of the different time periods (L_{Ad}, L_{Ae}, L_{An}, and L_{Aden}) measured over a week in the town of Cáceres (Spain). In the present study, the functional stratification of the average values of indicators L_{Ad}, L_{Ae}, L_{An}, L_{Adn}, L_{Aden}, and L_{A24}, in measurements over a year in Madrid, which has more than three million inhabitants, was analyzed. Besides, the hypothesis was resolved from a temporal perspective which had not been studied previously. To that end, the distributions of sound levels which were registered over a year in 21 measurement stations located on different kinds of urban roads were analyzed. Last, the relation between temporal variability of sound levels and the success probability in average annual levels was studied.

Section II describes the method and the city where the measurements were carried out. Section III presents and discusses the results. Finally, Sec. IV presents the most relevant conclusions.

II. METHODS

A. Characterization and location of measurement stations

In this study, sonorous values registered over a year in 21 measurement stations in Madrid were analyzed. The measurement stations were those in which road traffic was the main noise source. Madrid is the capital of Spain, and it is strategically located in the geographic center of the Iberian Peninsula. Its population is approximately 3 255 944 inhabitants and its urban area is 605.8 km^2.[26] Madrid is the major business center of Spain and the tertiary sector, the service sector, is the main economic sector. The principal Madrid highways have a radial shape (A1, A2, A3, A4, A5, and A6), and there are also ring roads (M30, M40, M45, and M50). Madrid is 655 m above sea level, because it is on a plateau, and the surrounding mountains account for the weather, which is characterized by hot summers and relatively cold winters. The mean annual temperature and rainfall are 15.0 °C and 400 mm, respectively.[27]

The measurement stations were equipped with the 4435 Brüel & Kjær (Nærum, Denmark) analyzers and the 4184 Brüel & Kjær microphones (compliant with both IEC 61672-1[41] type 1 and ANSI S1.4[42] type 1). The microphones were used with a windscreen and windscreen holder (to protect them from adverse weather conditions) and installed on a mast 4.0 m above ground level. The parameter registered by analyzers was the continuous equivalent A-weighted level integrated every hour ($L_{\mathrm{Aeq,1h}}$) over the years from 2006 to 2011.

The measurement stations were located on different kinds of urban roads and were classified in reference to their functionality according to the proposed definitions of the categorization method:[28]

(1) Category 1 includes those preferred streets whose function is to form connections with other Spanish towns and to interconnect those streets. In general, these streets are indicated by a system of road signs.
(2) Category 2 includes those streets that provide access to the major distribution nodes of the town. For the purpose of this study, a distribution node is considered to exist when at least four major streets meet. This definition does not include any possible nodes of preferred streets as defined in category 1, above. This category also includes streets normally used as alternatives to category 1 streets in the case of traffic saturation.
(3) Category 3 includes streets that lead to regional roads, streets that provide access from streets of category 1 and 2 to centers of interest in the town (hospitals, shopping malls, etc.), and streets that clearly allow communication between streets of category 1 and 2.
(4) Category 4 includes all other streets that clearly allow communication between the three previously defined categories of street, as well as the principal streets of the different districts of the town that were not included in the previously defined categories.
(5) Category 5 comprises the rest of the streets of the town except pedestrian-only streets.
(6) Category 6 comprises all the pedestrian-only streets.

Figure 1 shows the category in which the measurement stations are located: nine in category 1 (sampling points 1 to 9), four in category 2 (sampling points 10 to 13), four in category 3 (sampling points 14 to 17), two in category 4 (sampling points 18 and 19), and two in category 5 (sampling points 20 and 21).

B. Statistical analysis

The continuous equivalent sound level integrated every hour ($L_{\mathrm{Aeq,1h}}$) was chosen for the different statistical tests used to analyze the results and evaluate the quality of the category classification.

First, the spatial variability of average sound levels (L_{Ad}, L_{Ae}, L_{An}, L_{Adn}, L_{Aden}, and L_{A24}) which were registered in the different measurement stations was analyzed. In this analysis of average values, the sound levels from the measurement stations which were 6.0 m further from the curb (where most of the measurement stations were located) were normalized. For corrections, the methods described in some ISO standards were considered.[8,29] Next, it was studied if sound average levels had a significant stratification according to the category where measurement stations were located. This hypothesis was resolved with the nonparametric Kruskal-Wallis and Mann-Whitney U tests.[30,31] These nonparametric tests were used because of the small number of samples, which made a normal distribution unlikely. Although the number of stations is high for these types of continuous measurements, in some categories only a few stations were available for inferential analysis. This is why, from the perspective of the categorization method basis, adjacent categories with a smaller number of measurements were grouped in a new category for the different inferential analyses. These new categories were as follows: category A comprised the measurement stations located in category 1 (nine measurement stations); category B comprised the measurement stations located in category 2 (four measurement stations) and category 3 (four measurement stations); and category C comprised the measurement stations located in category 4 (two measurement stations) and category 5 (two measurement stations).

The Kruskal-Wallis test was used to compare all the categories to identify any significant differences. When such differences were found, Mann-Whitney U tests were used to compare pairs of categories. The Mann-Whitney U test is a nonparametric test for assessing whether two independent samples or observations come from the same distribution. This test was used to compare pairs of separate categories within the same population.

In contrast to previous statistical tests, the receiver operating characteristics analysis (ROC)[32,33] was used to evaluate the discriminative capacity of the categorization method, in other words, its ability to differentiate the sound values of the sampling points between pairs of categories (category *i* versus category *j*).

Originally the categorization method, without knowing the sound values at different sampling points, classified them in different categories. After, once sound levels are recorded and from these, the ROC analysis generates a predictive ROC classification in which sound levels have statistically significant differences. Through the comparison of categories established by both methods, the categorization method's ability to discriminate sonorous values was evaluated. For this, the functional stratification carried out by the categorization method was taken as reference and in the strata proposed by ROC classification the sensitivity (capacity to include previously

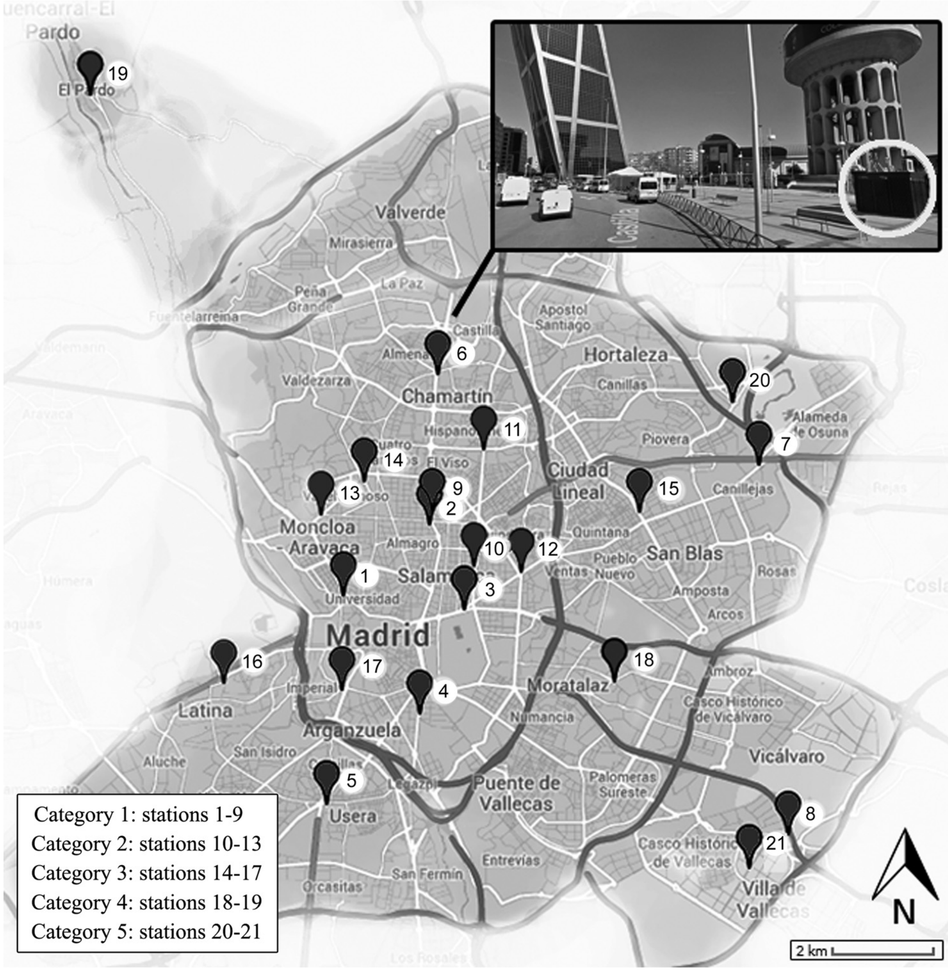

FIG. 1. Location of stations in Madrid city. Category 1: measurement points 1–9; category 2: measurement points 10–13; category 3: measurement points 14–17; category 4: measurement points 18–19; and category 5: measurement points 20–21.

assigned sampling points in the category), non-specificity (proportion of sampling points that were not initially assigned to a certain category but that the ROC classification indicated belonged to that category), and predictive values (proportion of the sampling points that the ROC classification assigned to a category that matched the categories to which they were initially assigned, relative to the total number of sampling points that the ROC classification determined for the category) were calculated with the following equations:

$$\text{sensitivity} = \frac{\text{True } i}{\text{True } i + \text{False } j}, \tag{1}$$

$$\text{non-specificity} = \frac{\text{False } i}{\text{True } j + \text{False } i}, \tag{2}$$

$$\text{predictive value} = \frac{\text{True } i}{\text{True } i + \text{False } i}, \tag{3}$$

where [see Fig. 2(a)]

(1) True i: number of sampling points assigned correctly to category i by ROC classification.
(2) False i: number of sampling points assigned incorrectly to category i by ROC classification.
(3) True j: number of sampling points assigned correctly to category j by ROC classification.
(4) False j: number of sampling points assigned incorrectly to category j by ROC classification.

From the depiction of sensitivity and non-specificity of each of the sampling point is obtained a ROC curve [see Fig. 2(b)]. The optimal cut-off point is the last sampling point belonging to category i (ROC classification) and it has the highest sensitivity and specificity (1 − non-specificity) jointly in the ROC curve. This point is obtained when the distance from the point (0,1) is the lowest. This distance was calculated with the following equation:

$$\text{distance} = \sqrt{(\text{non-specificity})^2 + (1 - \text{sensitivity})^2}. \tag{4}$$

Therefore, the optimal cut-off value is the average of sound values registered in the optimal cut-off point and in the previous sampling point with a lower value.

A ROC curve is a two-dimensional depiction of classifier performance, but a common method to reduce ROC performance to a single scalar value representing expected performance is to calculate the area under the ROC curve (AUC).[34,35] The formal definition is

$$\text{AUC} = \int_0^1 \text{ROC}(v)\,dv, \tag{5}$$

where v is the value of sensitivity-non-specificity sampling points.

AUC value will always be between 0 and 1.0. Therefore, values closer to 1.0 have better discriminatory power. However, because random guessing produces the diagonal line between (0,0) and (1,1), which has an area of 0.5. Values between 0.5 and 0.7 indicate low precision, values between 0.7 and 0.9 are considered useful and values greater than 0.9 indicate high precision.[36] The AUC has an important

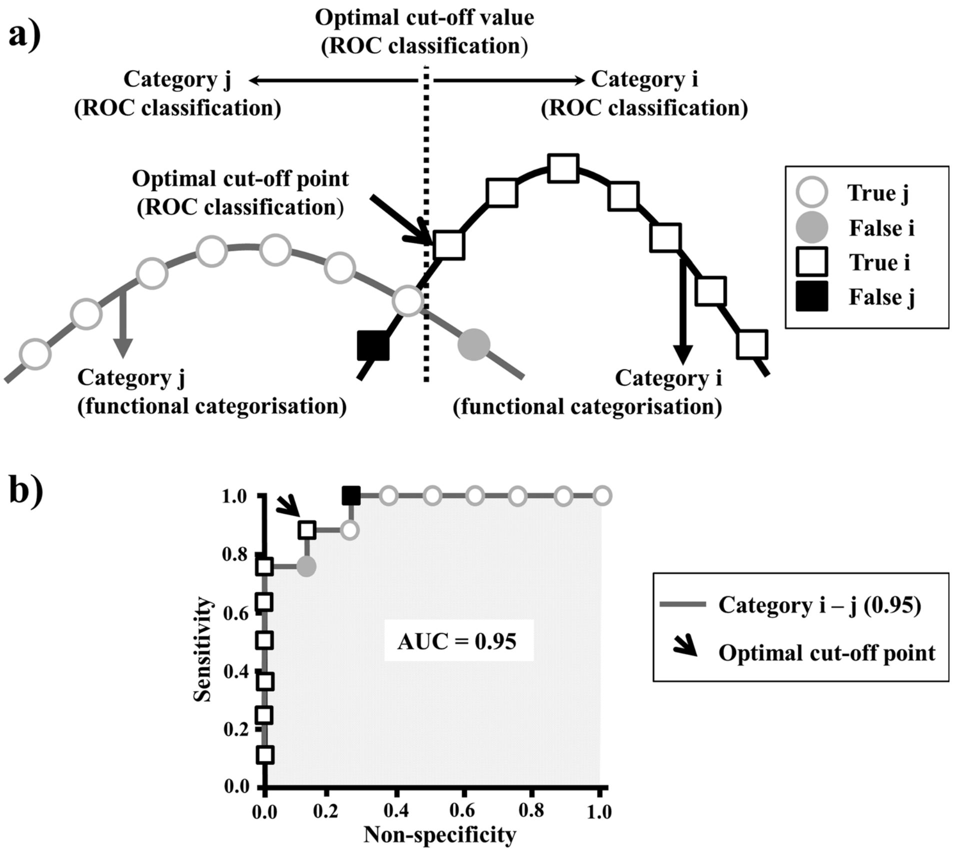

FIG. 2. ROC classification examples: the evaluation of (a) the discrimination ability in the categorization method and (b) the representation of a ROC curve.

statistical property: the AUC of a classifier is equivalent to the probability that the classifier will rank a randomly chosen category *i* higher than a randomly chosen category *j*.[33]

Regarding the predictive value [Eq. (3)], known by Fawcett[33] as precision, is a ratio that indicates the relation between the functional stratification and the ROC classification of sound levels. The closer to 100%, the better the prediction of the classification of sound levels by functional categorization.

After studying the spatial variability of average sound levels, our aim was to analyze if there was a spatial stratification of temporal variability of $L_{\mathrm{Aeq,1h}}$ levels registered in the different measurement stations. To do this, the distribution of sound levels registered over the year was analyzed. The first hypothesis concerned whether the distributions presented significant differences from normal distribution ($\text{p-value} \leq 0.001$). This hypothesis was resolved by the Kolmogorov test and all sound distributions had significant differences in respect of normal distribution ($\text{p-value} \leq 0.001$), similar to information found in road traffic sound level distribution studies.[37] Therefore, in this study parameters such as mean, standard deviation, variance, skewness, kurtosis, etc., were not used. The parameters of median, percentile, and different types of range were analyzed: R_{50} range or interquartile range (percentile P_{75} − percentile P_{25}), R_{80} range ($P_{90} - P_{10}$), R_{90} range ($P_{95} - P_5$), R_{95} range ($P_{97.5} - P_{2.5}$), and R_{99} ($P_{99.5} - P_{0.5}$). The types of range and the differences between median and percentiles gave information about the form of distribution and thus about the variability of sound levels. Moreover, recent studies show the importance of analyzing the percentiles because of their relation with the soundscape perception.[38,39] Average values of these parameters were compared among different categories to look for significant stratification. For this reason, the Kruskal-Wallis test and Mann-Whitney U test were used.

Finally, it was analyzed if the temporal variability registered in the different measurement stations had a significant relation with the success probability of the annual average sound level. The success probability was obtained by the percentage of values $L_{\mathrm{Aeq,1h}}$ which were included in the interval $L_{\mathrm{A24}} \pm \varepsilon$ ($\varepsilon = 0.5$, 1, 2, and 3 dB). The hypothesis was resolved by Spearman's rho.

III. RESULTS AND DISCUSSION

A. Analysis of average sound level variability

Table I shows the mean values of the different sonorous indicators: L_{Ad}, L_{Ae}, L_{An}, L_{Adn}, L_{Aden}, and L_{A24}. In all the sub-day periods studied [day (from 7.0 a.m to 7.0 p.m.) (L_{Ad}), evening (from 7.0 p.m. to 11.0 p.m.) (L_{Ae}), night (from 11.0 p.m. to 7.0 a.m.) (L_{An}), and over the whole day (L_{Adn}, L_{Aden}, and L_{A24})], there is a clear tendency of noise levels to decrease as the category number increases.

TABLE I. Average values and standard deviation of L_{Ax} values of the sonorous indicators. The results are shown separately for each category.

	$\overline{L_{Ax}} \pm \sigma$ (dB)				
Indicator	Category 1	Category 2	Category 3	Category 4	Category 5
L_{Ad}	70.5 ± 1.3	68.5 ± 1.2	65.1 ± 1.9	63.6 ± 0.8	62.5 ± 0.9
L_{Ae}	70.1 ± 1.0	67.8 ± 1.1	64.4 ± 1.6	63.1 ± 1.1	61.4 ± 0.3
L_{An}	66.8 ± 0.9	62.5 ± 1.3	60.4 ± 1.7	58.0 ± 0.6	55.8 ± 0.7
L_{Adn}	73.8 ± 1.6	69.6 ± 1.2	67.6 ± 1.2	65.8 ± 0.2	63.4 ± 1.1
L_{Aden}	74.6 ± 0.9	71.4 ± 1.3	68.5 ± 2.0	66.7 ± 1.2	65.2 ± 0.2
L_{A24}	68.6 ± 1.0	65.6 ± 1.3	62.4 ± 1.5	60.4 ± 1.0	59.1 ± 0.0

Then, it was analyzed if the differences in average values of sonorous indicators among different categories were statistically significant. Before resolving this hypothesis, as mentioned previously, because of the number of data by categories, the categories were grouped into three new categories: category A (category 1), category B (categories 2 and 3), category C (categories 4 and 5). Throughout this study and in the posterior analysis, only these three categories were used.

The hypothesis was resolved first by the Kruskal-Wallis test. This test indicated significant differences ($\text{p-value} \leq 0.001$) for all the sonorous indicators studied. Thus, the Mann-Whitney U test was then applied to analyze the differences among category pairs (Table II). As shown in Table II, the Mann-Whitney U test found significant differences ($\text{p-value} \leq 0.05$) among all pairs of categories studied for all sound indicators analyzed. This finding indicates that the functional stratification of noise levels observed in previous weekly measurement studies is also found for annual measurements and is equally present in all the studied temporal periods. Thus, the categorization method is a very powerful method of spatial noise assessment, allowing the noise values of cities to be characterized by using a reduced number of sampling points.

Finally, to corroborate the quality of the previous results and to obtain more information about the categorization method, the classification capacity of this method was studied via ROC analysis. The results of this analysis are shown in Fig. 3. From the results shown in these graphs, the following can be noted:

(1) Regarding the ROC curve (sensitivity and non-specificity), the AUC indicator (capacity of the method to discriminate correctly the sound levels for two different categories) present values better than 0.94 for all pairs of categories of all sound indicators [see Figs. 3(a)–3(f)]. Thus, the values indicate high precision. These high

TABLE II. Results of the Mann-Whitney U test applied to pairs of categories.

	Category	A	B
L_{Ad}	B	9.9×10^{-4} [a]	—
	C	2.8×10^{-3} [b]	2.8×10^{-2} [c]
L_{Ae}	B	3.3×10^{-4} [a]	—
	C	2.8×10^{-3} [b]	8.1×10^{-3} [b]
L_{An}	B	8.2×10^{-5} [a]	—
	C	2.8×10^{-3} [b]	8.1×10^{-3} [b]
L_{Adn}	B	3.3×10^{-4} [a]	—
	C	2.8×10^{-3} [b]	1.1×10^{-2} [c]
L_{Aden}	B	8.2×10^{-5} [a]	—
	C	2.8×10^{-3} [b]	8.1×10^{-3} [b]
L_{A24}	B	8.2×10^{-5} [a]	—
	C	2.8×10^{-3} [b]	8.1×10^{-3} [b]

[a] Significant at $p \leq 0.001$.
[b] Significant at $p \leq 0.01$.
[c] Significant at $p \leq 0.05$.

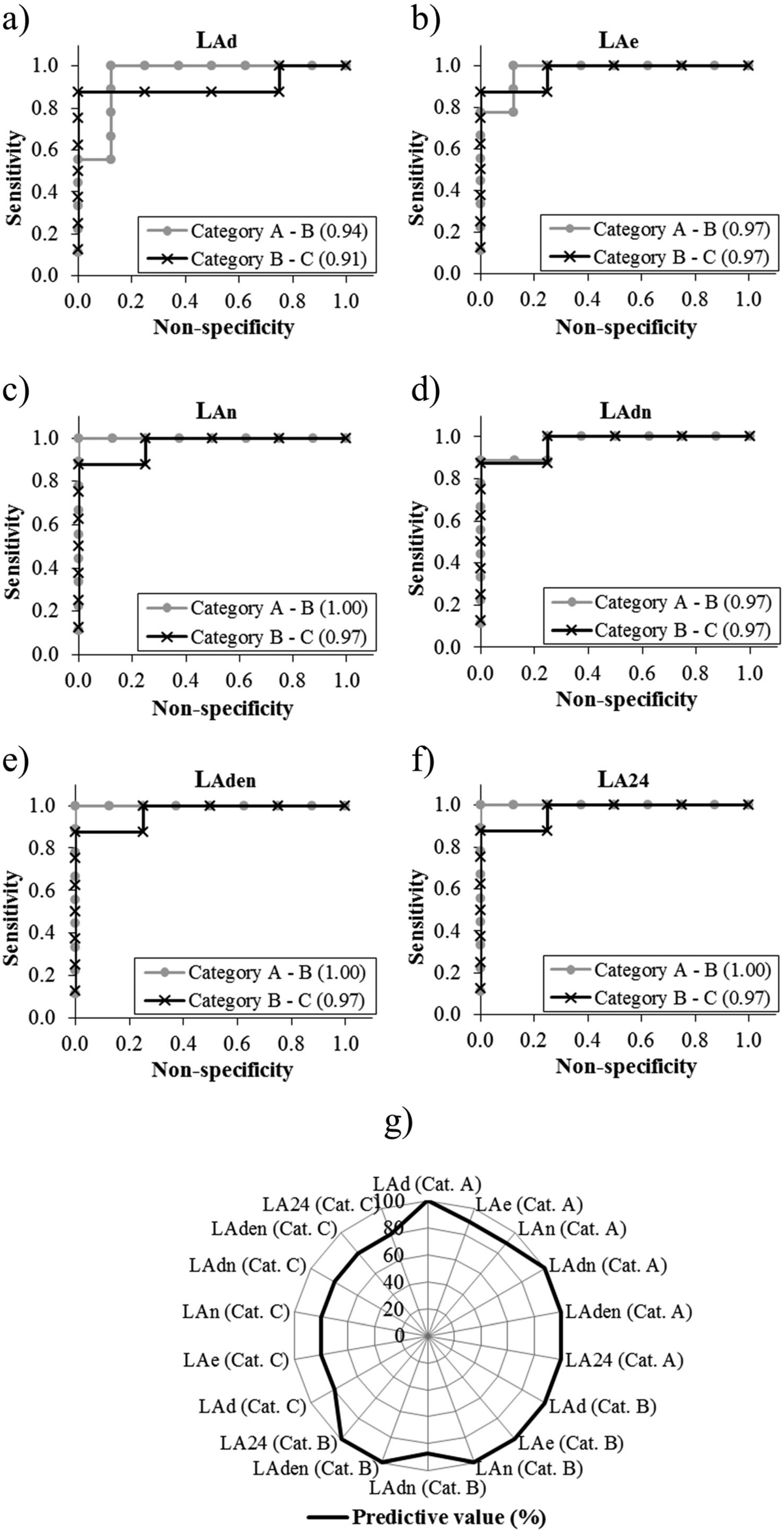

FIG. 3. Results of ROC analysis of sound indicators: (a) L_{Ad}, (b) L_{Ae}, (c) L_{An}, (d) L_{Adn}, (e) L_{Aden}, (f) L_{A24}, and (g) predictive value.

TABLE III. Upper and lower limit obtained from ROC classification for the L_{Ax} (dB) values in the category A, B, and C.

Indicator (dB)	Limit	Category A	Category B	Category C
L_{Ad}	Upper	72.9	68.9	64.4
	Lower	68.9	64.4	61.8
L_{Ae}	Upper	72.0	68.3	64.0
	Lower	68.3	64.0	61.2
L_{An}	Upper	68.7	65.0	59.2
	Lower	65.0	59.2	55.3
L_{Adn}	Upper	76.3	71.9	66.8
	Lower	71.9	66.8	62.6
L_{Aden}	Upper	76.3	73.3	67.7
	Lower	73.3	67.7	65.1
L_{A24}	Upper	70.5	67.4	61.6
	Lower	67.4	61.6	59.1

AUC values [Eq. (4)] indicate, in turn, higher values of sensitivity [Eq. (1)], close to 100%, and very low values of non-specificity [Eq. (2)], close to 0%. The optimal cut-off values of categories A, B, and C are determined from ROC curves, and the results are showed in Table III. These values show the upper and lower limit of sound values registered in different categories according to ROC classification.

(2) Finally, the predictive values of the different strata [Eq. (3)] are very good [see Fig. 3(g)]: categories A and B present values of 100% [except L_{Ae} (category A), L_{An} (category A), and L_{Adn} (category B) which present values of 90%] and category C presents values of 80% for the different sonorous index.

Therefore, for each of the three periods analyzed and for the overall indicators (L_{Adn}, L_{Aden}, and L_{A24}), the results showed the method had high discrimination and predictive capacity. These results suggest a great advance in the validity of the categorization method because of its application to an agglomeration with more than three million inhabitants and sound measurements taken over a year.

Thus, because of its high discrimination and prediction capacity, this procedure seems to be very suitable for further applications such as noise prediction and the design of environmental policy.

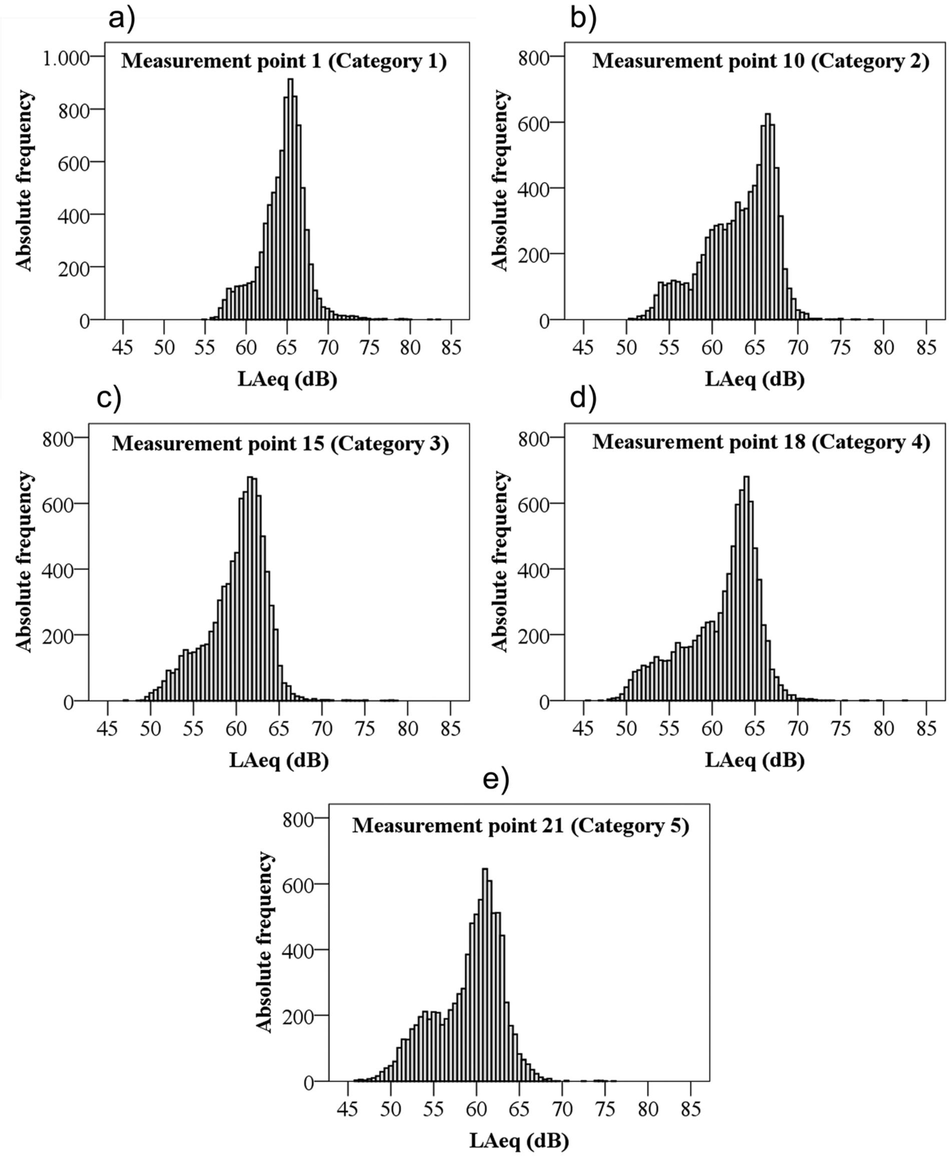

FIG. 4. Histogram of $L_{\mathrm{Aeq,1h}}$ in (a) category 1, (b) category 2, (c) category 3, (d) category 4, and (e) category 5.

TABLE IV. Average values and standard deviation of the different ranges (R_{50}, R_{80}, R_{90}, R_{95}, and R_{99}) of $L_{\mathrm{Aeq,1h}}$ values registered in the measurement stations. The results are shown separately for each category. Range R_{50} describes the difference between percentiles P_{75} and P_{25} (also named interquartile range), R_{80} range is the difference between P_{90} and P_{10}, R_{90} range is the difference between P_{95} and P_5, R_{95} range is the difference between $P_{97.5}$ and $P_{2.5}$ and R_{99} range is the difference between $P_{99.5}$ and $P_{0.5}$.

	Range (dB)				
Category	R_{50}	R_{80}	R_{90}	R_{95}	R_{99}
A	3.5 ± 0.5	7.2 ± 0.6	9.5 ± 0.6	11.2 ± 0.5	14.8 ± 1.6
B	4.9 ± 0.7	9.5 ± 1.0	12.1 ± 1.1	14.2 ± 1.0	17.9 ± 1.0
C	5.6 ± 0.7	11.2 ± 0.9	14.0 ± 1.2	16.2 ± 1.6	21.0 ± 2.7

TABLE V. Results of the Mann-Whitney U test applied to pairs of categories.

	Category	A	B
R_{50}	B	1.8×10^{-3} [a]	—
	C	2.8×10^{-3} [a]	4.9×10^{-2} [b]
R_{80}	B	1.3×10^{-3} [a]	—
	C	6.9×10^{-3} [a]	2.8×10^{-2} [b]
R_{90}	B	1.6×10^{-4} [c]	—
	C	2.8×10^{-3} [a]	2.8×10^{-2} [b]
R_{95}	B	8.2×10^{-5} [c]	—
	C	2.8×10^{-3} [a]	2.8×10^{-2} [b]
R_{99}	B	2.5×10^{-3} [a]	—
	C	2.8×10^{-3} [a]	2.8×10^{-2} [b]

[a] Significant at $p \le 0.01$.

[b] Significant at $p \le 0.05$.

[c] Significant at $p \le 0.001$.

B. Analysis of temporal sound variability

First, before descriptive and inferential analysis of the different statistic parameters related to the variability of sound levels, the distribution of values $L_{\mathrm{Aeq,1h}}$ was analyzed over the year. Figure 4 shows different histogram models obtained for measurement stations located in different categories. *A priori*, it can be observed that the distributions differ among categories and are also different from normal distribution. Figure 4(a) (category 1) has an approximately symmetrical distribution, albeit leptokurtic (slender). The remaining histograms have a noticeably negative skew (left-skewed) which increases when the number of category increases as well. In contrast to categories 2 and 3 [Figs. 4(b) and 4(c)], categories 4 and 5 [Figs. 4(d) and 4(e)] do not show a progressive decrease of sound levels from average values to low values.

Second, all hypotheses were resolved with the aid of statistic inference. The Kolmogorov test verified that all distributions had significant differences from normal distribution ($p\text{-value} \le 0.001$). Thus, different types of range were taken as a measure of the sound variability: range R_{50} derives from the difference of percentiles P_{75} and P_{25} (also named interquartile range); R_{80} range is the difference between P_{90} and P_{10}; R_{90} range is the difference between P_{95} and P_5; R_{95} range is the difference between $P_{97.5}$ and $P_{2.5}$; and R_{99} range is the difference between $P_{99.5}$ and $P_{0.5}$. These types of range give the information about distribution and therefore about the sound variability regarding distance from the median. Table IV shows the average values and the standard deviation of ranges in different categories. A decreasing tendency is observed from category A to category C.

These differences in range were analyzed with Kruskal-Wallis and Mann-Whitney U tests. The Kruskal-Wallis test showed significant differences ($p\text{-value} \le 0.001$) for all the ranges studied. Thus, the Mann-Whitney U test was then applied to analyze the differences among category pairs (Table V). As shown in Table V, the Mann-Whitney U test found significant differences ($p\text{-value} \le 0.05$) among all pairs of categories for all ranges analyzed. Consequently, the functionality of roads allows significant stratification of the variability of sound levels registered over the year. This is very important from the perspective of temporal strategy as it allows estimation of the annual average sound level because it permits reduction in the number or the time of measurements.

The following objective was to look for the differences among the sound levels which meant that the types of range were significantly different among the three analyzed categories. To do this, the distances on both sides of the median, whose sum is the range: percentile P_x − median (M_e) and median (M_e) − percentile P_x were analyzed. Thus, significant stratification caused by differences between average and low sound values (typical difference between diurnal and nocturnal sound values) or between average and high sound values (typical difference between diurnal values) could be detected. The averages of these differences between percentiles and medians are shown in Table VI. It can be seen first that the $P_x - M_e$ value is quite superior to the $M_e - P_x$ value in the different types of range. This difference was foreseeable because the different distributions have a noticeable negative skew (Fig. 4). Second, there is a higher decrease in the $P_x - M_e$ value from category A to category C than in the $M_e - P_x$ value. These differences were analyzed through the Mann-Whitney U test and the results are shown in Table VII. The results show that differences between

TABLE VI. Average values of differences among percentiles (P_x) and median (M_e) and vice versa for different types of range (R_{50}, R_{80}, R_{90}, R_{95}, and R_{99}). The results are shown separately for each category.

	R_{50} (dB)		R_{80} (dB)		R_{90} (dB)		R_{95} (dB)		R_{99} (dB)	
Category	$P_{75} - M_e$	$M_e - P_{25}$	$P_{90} - M_e$	$M_e - P_{10}$	$P_{95} - M_e$	$M_e - P_5$	$P_{97.5} - M_e$	$M_e - P_{2.5}$	$P_{99.5} - M_e$	$M_e - P_{0.5}$
A	2.1	1.4	4.8	2.4	6.3	3.1	7.2	4.0	8.4	6.4
B	3.0	2.0	6.2	3.3	7.9	4.2	8.9	5.2	10.4	7.5
C	3.6	2.0	7.9	3.4	9.6	4.3	10.9	5.4	13.0	8.0

TABLE VII. Results of the Mann-Whitney U test applied to pairs of categories.

	Category	A	B
$P_{75} - M_e$	B	5.2×10^{-3}[a]	—
	C	1.1×10^{-2}[b]	4.9×10^{-2}[b]
$P_{90} - M_e$	B	3.3×10^{-3}[a]	—
	C	6.9×10^{-3}[a]	1.6×10^{-2}[b]
$P_{95} - M_e$	B	2.4×10^{-3}[a]	—
	C	6.9×10^{-3}[a]	4.9×10^{-2}[b]
$P_{97.5} - M_e$	B	1.2×10^{-3}[a]	—
	C	6.9×10^{-3}[a]	6.1×10^{-2}[c]
$P_{99.5} - M_e$	B	3.3×10^{-3}[a]	—
	C	6.9×10^{-3}[a]	2.7×10^{-2}[b]
M_e -P_{25}	B	5.9×10^{-3}[a]	—
	C	1.6×10^{-3}[a]	7.3×10^{-1}[c]
$M_e - P_{10}$	B	2.8×10^{-3}[a]	—
	C	6.8×10^{-3}[a]	9.3×10^{-1}[c]
$M_e - P_5$	B	3.4×10^{-3}[a]	—
	C	6.9×10^{-3}[a]	1.0[c]
$M_e - P_{2.5}$	B	6.1×10^{-3}[a]	—
	C	1.1×10^{-2}[b]	6.1×10^{-1}[c]
$M_e - P_{0.5}$	B	1.0×10^{-1}[c]	—
	C	7.6×10^{-2}[c]	5.7×10^{-1}[c]

[a] Significant at $p \leq 0.01$.
[b] Significant at $p \leq 0.05$.
[c] Non-significant difference ($p > 0.05$).

$P_x - M_e$ values, as occurs in the different types of range, have significant differences (p-value ≤ 0.05). However, the differences between $M_e - P_x$ values between categories B and C in all cases have a p-value > 0.05 (not significant). Therefore, this result corroborates the hypothesis that the stratification of variability among the three analyzed categories is largely the result of the difference between average sound values and low sound values.

Finally, the differences between sound values registered in the diurnal period (L_{Ad}) and nocturnal period (L_{An}) and between the diurnal period (L_{Ad}) and evening period (L_{Ae}) in the different categories were analyzed. The average values of these differences are shown in Table VIII. The results show that differences are more noticeable between different categories in $L_{Ad} - L_{An}$. Then, the averages of these differences were analyzed through the Mann-Whitney U test and the results are shown in Table IX. The results show that differences between nocturnal and diurnal levels are significant among the three categories analyzed. This result differs from results published in previous works, where their categories were reduced to two significantly distinguishable categories.[40] As regards differences between the diurnal and evening level, as was expected from descriptive analyses there were no significant differences (p-value > 0.05).

TABLE VIII. Average values of differences among sound indicators L_{Ad}, L_{An}, and L_{Ae} for each category.

Category	$L_{Ad} - L_{An}$ (dB)	$L_{Ad} - L_{Ae}$ (dB)
A	4.1	0.3
B	5.7	0.3
C	7.2	0.5

TABLE IX. Results of the Mann-Whitney U test applied to pairs of categories.

	Category	A	B
$L_{Ad} - L_{An}$	B	2.5×10^{-3}[a]	—
	C	2.8×10^{-3}[a]	2.8×10^{-2}[b]
$L_{Ad} - L_{Ae}$	B	5.4×10^{-1}[c]	—
	C	7.1×10^{-1}[c]	9.3×10^{-1}[c]

[a] Significant at $p \leq 0.01$.
[b] Significant at $p \leq 0.05$.
[c] Non-significant difference ($p > 0.05$).

The two last analyses resolved the hypothesis that the significant stratification of temporal sound variability among different categories was mainly owed to differences between diurnal and nocturnal sound values.

C. Relation between temporal sound variability and probability of success

In Sec. III B, it was demonstrated that average temporal sound variability through the ranges R_{80}, R_{90}, R_{95}, and R_{99} had a significant functional stratification in the studied categories. This is important from the perspective of estimating the average annual sound value because it could determine those roads which need less time or fewer measurements.

This hypothesis of a relation between the different types of range and the probability of success was analyzed through Spearman's rho. The success probability was obtained from the percentage of values $L_{Aeq,1h}$ which were included in the interval $L_{A24} \pm \varepsilon$ ($\varepsilon = 0.5$, 1, 2, and 3 dB). The Spearman's rho results are shown in Table X. The correlation coefficients are very near to unity and with a p-value ≤ 0.01 indicate a highly significant relation between range and the probability of success.

In short, the categorization method not only allows significant functional stratification of average annual sound values to be carried out but also functional stratification of temporal sound variability. This could allow important savings in terms of the number and the time of measurements from spatial and temporal perspectives.

IV. CONCLUSIONS

The present study which was carried out in an agglomeration with more than three million inhabitants (Madrid) shows that the categorization method is an adequate tool for assessment of temporal and spatial noise, thus enabling the functional stratification of noise in cities to be identified. Therefore, this method has advantages in terms of the reduction of sampling points and measurement time.

The analysis of sound levels registered over a year in the 21 measurement stations located on roads with different functionality implies the following additional conclusions:

(1) The mean values of the analyzed sound indicators (L_{Ad}, L_{Ae}, L_{An}, L_{Adn}, L_{Aden}, and L_{A24}) decrease as the number of the category increases. A comparison of sound levels with the Kruskal-Wallis and Mann-Whitney U tests shows that the differences among values of functional

TABLE X. Results for Spearman's rho between types of range (R_{50}, R_{80}, R_{90}, R_{95}, and R_{99}) and probability of success (% $L_{\mathrm{Aeq,1h}}$ values in the interval $L_{\mathrm{A24}} \pm \varepsilon$ for $\varepsilon = 0.5$, 1, 2, and 3 dB).

	% $L_{\mathrm{Aeq,1h}}$ values in the interval $L_{\mathrm{A24}} \pm \varepsilon$			
Range	$\varepsilon = 0.5$	$\varepsilon = 1.0$	$\varepsilon = 2.0$	$\varepsilon = 3.0$
R_{50}	-0.94[a]	-0.95[a]	-0.98[a]	-0.98[a]
R_{80}	-0.92[a]	-0.95[a]	-0.95[a]	-0.96[a]
R_{90}	-0.89[a]	-0.92[a]	-0.93[a]	-0.95[a]
R_{95}	-0.87[a]	-0.89[a]	-0.91[a]	-0.92[a]
R_{99}	-0.58[b]	-0.64[b]	-0.70[a]	-0.77[a]

[a] Significant at $p \leq 0.001$.

[b] Significant at $p \leq 0.01$.

categories are statistically significant for a confidence interval of 95%. This finding demonstrates the applicability of the categorization method to spatial assessment, as it can be applied to all periods of the day.

(2) When analyzing the discrimination capacity of the categorization method using predictive ROC classification, we found that all the pairs of categories presented AUC values above 0.94, indicating the high precision of the method. These values are the result of sensitivity and non-specificity close to 100% and 0%, respectively. Also, ROC classification has a good predictive capacity for non-measured values. A 100% predictive capacity was found in categories A and B [except L_{Ae} (category A), L_{An} (category A), and L_{Adn} (category B) which have values of 90%] and 80% predictive capacity in category C for all sonorous indicators.

(3) The mean values of the analyzed range types (R_{50}, R_{80}, R_{90}, R_{95}, and R_{99}) decrease from category A to category C. A comparison of mean values with the Kruskal-Wallis and Mann-Whitney U tests showed that the differences among values of functional categories are statistically significant for a confidence interval of 95%. This finding demonstrates the applicability of the categorization method to temporal assessment. This significant functional stratification of temporal variability was mainly owed to the significant differences between average and low sound values (percentile $P_{\mathrm{x}} - \text{median } M_{\mathrm{e}}$). Also, the difference between diurnal and nocturnal sound levels ($L_{\mathrm{Ad}} - L_{\mathrm{An}}$) presented functional stratification in the three analyzed categories. This has never been achieved in previous studies.

(4) The highly significant relation among types of range as a measurement of temporal variability and the success probability of average annual sound value corroborate the advantages from temporal perspective of the traffic roads stratification according to their functionality.

ACKNOWLEDGMENTS

The authors wish to thank the Government of Extremadura, the Regional Ministry of Economy, Trade and Innovation and the European Social Fund for funding the project. This work was also partially supported by the Spanish Ministerio de Economía y Competitividad (Project TRA2012-37117) and the European Regional Development Fund (ERDF).

[1] G. Rey Gozalo, J. M. Barrigón Morillas, and V. Gómez Escobar, "Analysis of noise exposure in two small towns," Acta Acust. Acust. **98**, 884–893 (2012).

[2] G. Rey Gozalo, J. M. Barrigón Morillas, and V. Gómez Escobar, "Urban streets functionality as a tool for urban pollution management," Sci. Total Environ. **461–462**, 453–461 (2013).

[3] World Health Organization, "Burden of disease from environmental noise. Quantification of healthy life years lost in Europe" (World Health Organization, Geneva, 2011).

[4] W. Babisch, "Updated exposure-response relationship between road traffic noise and coronary heart diseases: A meta-analysis," Noise Health **16**, 1–9 (2014).

[5] M. Sørensen, Z. J. Andersen, R. B. Nordsborg, T. Becker, A. Tjønneland, K. Overvad, and O. Raaschou-Nielsen, "Long-term exposure to road traffic noise and incident diabetes: A cohort study," Environ. Health Persp. **121**, 217–222 (2013).

[6] J. Dratva, E. Zemp, D. F. Dietrich, P-O. Bridevaux, T. Rochat, C. Schindler, and M. W. Gerbase, "Impact of road traffic noise annoyance on health-related quality of life: Results from a population-based study," Qual. Life Res. **19**, 37–46 (2010).

[7] European Commission, "Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management of environmental noise (END)" (Official Journal L 189 12-26, European Parliament and the Council of the European Union, Brussels, Belgium, 2002).

[8] ISO 1996-2, *Description, Measurement and Assessment of Environmental Noise. Part 2: Determination of Environmental Noise Levels* (International Organization for Standardization, Geneva, Switzerland, 2007).

[9] M. Ausejo, M. Recuero, C. Asensio, I. Pavón, and J. M. López, "Study of precision, deviations and uncertainty in the design of the strategic noise map of the macrocenter of the city of Buenos Aires, Argentina," Environ. Model. Assess. **15**, 125–135 (2009).

[10] European Commission Working Group–Assessment of Exposure to Noise, "Good practice guide for strategic noise mapping and the production associated data on noise exposure, Version 2," European Commission, Brussels, Belgium, 2007.

[11] J. M. Barrigón Morillas, V. Gómez Escobar, J. Trujillo Carmona, J. A. Méndez Sierra, R. Vílchez-Gómez, and F. J. Carmona del Río, "Analysis of the prediction capacity of a categorization method for urban noise assessment," Appl. Acoust. **72**, 760–771 (2011).

[12] ISO 1996-2, *Description and Measurement of Environmental Noise. Part 2: Acquisition of Data Pertinent to Land Use* (International Organization for Standardization, Geneva, Switzerland, 1987).

[13] H. Doygun and D. K. Gurun, "Analyzing and mapping spatial and temporal dynamics of urban traffic noise pollution: A case study in Kahramanmaras, Turkey," Environ. Monit. Assess. **142**, 65–72 (2008).

[14] J. H. Ko, S. I. Chang, and B. C. Lee, "Noise impact assessment by utilizing noise map and GIS: A case study in the city of Chungju, Republic of Korea," Appl. Acoust. **72**, 544–550 (2011).

[15] J. Romeu, T. Genescà, T. Pàmies, and S. Jiménez, "Street categorization for the estimation of day levels using short-term measurements," Appl. Acoust. **72**, 569–577 (2011).

[16] E. Suárez and J. L. Barros, "Traffic noise mapping of the city of Santiago de Chile," Sci. Total Environ. **466–467**, 539–546 (2014).

[17] P. H. Zannin and D. Queiroz de Sant'Ana, "Noise mapping at different stages of a freeway redevelopment project—A case study in Brazil," Appl. Acoust. **72**, 479–486 (2011).

[18] C. Prieto Gajardo and J. M. Barrigón Morillas, "Stabilisation patterns of hourly urban sound levels," Environ. Monit. Assess. **187**, 1–16 (2014).

[19] C. Prieto Gajardo, J. M. Barrigón Morillas, V. Gómez Escobar, R. Vílchez-Gómez, and G. Rey Gozalo, "Effects of singular noisy events on long-term environmental noise measurements," Pol. J. Environ. Stud. **23**, 2007–2017 (2014). Available at http://www.pjoes.com/pdf/23.6/Pol.J.Environ.Stud.Vol.23.No.6.2007-2017.pdf.

[20] R. Makarewicz and M. Gałuszka, "Empirical revision of noise mapping," Appl. Acoust. **72**, 578–581 (2011).

[21] J. M. Barrigón Morillas and C. Prieto Gajardo, "Uncertainty evaluation of continuous noise sampling," Appl. Acoust. **75**, 27–36 (2014).

[22] R. E. De Vor, P. D. Shomer, W. A. Kline, and R. D. Neathamer, "Development of temporal sampling strategies for monitoring noise," J. Acoust. Soc. Am. **66**, 763–771 (1979).

[23] P. D. Schomer and R. E. De Vor, "Temporal sampling requirement for estimations of long-term average sound levels in the vicinity of airports," J. Acoust. Soc. Am. **69**, 713–719 (1981).

[24] E. Gaja, A. Gimenez, S. Sancho, and A. Reig, "Sampling techniques for the estimation of the annual equivalent noise level under urban traffic conditions," Appl. Acoust. **64**, 43–53 (2003).

[25] G. Rey Gozalo, J. M. Barrigón Morillas, and V. Gómez Escobar, "Analyzing nocturnal noise stratification," Sci. Total Environ. **479–480**, 39–47 (2014).

[26] INE, "Population, area and density by municipality," National Statistics Institute, Madrid, Spain, 2012.

[27] AEMET, "Summary guide of Spanish climate (1981–2010)," State Meteorological Agency, Madrid, Spain, 2012.

[28] J. M. Barrigón Morillas, V. Gómez Escobar, J. A. Méndez Sierra, R. Vílchez-Gómez, J. M. Vaquero, and J. Trujillo Carmona, "A categorization method applied to the study of urban road traffic noise," J. Acoust. Soc. Am. **117**, 2844–2852 (2005).

[29] ISO 9613-2, *Attenuation of Sound During Propagation Outdoors. Part 2: General Method of Calculation* (International Organization for Standardization, Geneva, Switzerland, 1996).

[30] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," Ann. Math. Stat. **18**, 50–60 (1947).

[31] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," J. Am. Stat. Assoc. **47**, 583–621 (1952).

[32] D. J. Hand and R. J. Till, "A simple generalisation of the area under the ROC curve for multiple class classification problems," Mach. Learn. **45**, 171–186 (2001).

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

[34] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology **143**, 29–36 (1982).

[35] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," Pattern Recogn. **30**, 1145–1159 (1997).

[36] J. A. Swets, "Measuring the accuracy of diagnostic systems," Science **240**, 1285–1293 (1988).

[37] C. G. Don and I. G. Rees, "Road traffic sound level distributions," J. Sound Vib. **100**, 41–53 (1985).

[38] G. Brambilla, L. Maffei, M. Di Gabriele, and V. Gallo, "The perceived quality of soundscape in three urban parks in Rome," J. Acoust. Soc. Am. **134**, 782–790 (2013).

[39] G. Brambilla, V. Gallo, F. Asdrubali, and F. D'Alessandro, "The perceived quality of soundscape in three urban parks in Rome," J. Acoust. Soc. Am. **134**, 832–839 (2013).

[40] S. Jiménez, M. Genescà, J. Romeu, and A. Sanchez, "Estimation of night traffic noise levels," Acta Acust. Acust. **94**, 563–567 (2008).

[41] IEC 61672-1, Electroacoustics – Sound level meters – Part 1: Specifications (International Electrotechnical Commission, Geneva, Switzerland, 2002).

[42] ANSI S1.4 (R2006), Specification for sound level meters (American National Standards Institute, New York, 1983).