STATISTICAL ANALYSIS ON ANTE-OPERAM DATA COLLECTED IN THE QUIET URBAN AREAS SELECTED BY THE LIFE+2010 QUADMAP PROJECT

Chiara Bartalucci, Francesco Borchi and Monica Carfagni

Department of Industrial Engineering, University of Florence, via di S. Marta 3, 50139 Firenze, Italy
e-mail: francesco.borchi@unifi.it

Alessandra Petrucci and Maria Silvana Salvini

Department of Statistics, Informatics, Applications "G. Parenti", University of Florence, Viale Morgagni 59, 50134 Firenze, Italy

QUADMAP (QUiet Areas Definition and Management in Action Plans) is a LIFE+2010 Project on Quiet Urban Areas which aims to deliver a method regarding identification, delineation, characterization, improvement and managing of Quiet Urban Areas (QUAs) as meant in the Environmental Noise Directive 2002/49/EC. The project will also help to understand the definition of a QUA, the meaning and the added value for the city and their citizens in terms of health, social safety and lowering stress levels. At the beginning of 2013 the first version of a methodology to select, analyze and manage QUAs has been produced and subsequently applied in ten pilot areas chosen in Firenze, Bilbao and Rotterdam. During the analysis phase, quantitative (noise maps and acoustic measurements) and qualitative (end-users questionnaires, general and non-acoustic information) data have been collected and examined. Once the ante-operam phase of analysis has been completed, the interventions' realization in the pilot areas started and was followed by post-operam surveys.
In this paper, results of the carried out statistical analysis based both on quantitative and qualitative data acquired in the city of Florence during the ante-operam phase are presented.
In particular, the use of logistic regression and ordinal logistic models is investigated. The use of multidimensional techniques aims to evaluate the net effect of single variables.

1. Introduction

The European Directive 2002/49/EC on the Assessment and Management of Environmental Noise (further abbreviated as END) was adopted to define a common approach to avoid, prevent or reduce the harmful effects due to noise exposure and to preserve the environmental noise quality where it is good. One of the main environmental problems targeted in this sense is the need of improvement in the definition of QUAs (Quiet Urban Areas). The END defines "Quiet Area in an agglomeration" as "an area, delimited by the competent authority, which is not exposed, for instance, to a value of Lden or of another appropriate noise indicator greater of a certain threshold (set by the Member State) from any noise source". From this definition it is not clear enough how to allow an appropriate assessment and management (action planning) of QUAs in urban environment [1-3]. A further issue concerns the fact that areas where citizens expect to find a quiet environment often do not meet the noise limits associated to them by the national law, assuming noise limits are established and a law exists. In those countries where national and local factors in identifying and protecting QUAs are considered, significantly different and mainly very general approaches have been taken into account [4]. The consequence of this "freedom of choice" results in non-homogeneous collections of data as well as in a divergent approach across the EU.

QUADMAP Project started in September 2011 with the final aim of developing a complete, practical and demonstrated methodology to select, analyse and manage QUAs. At the beginning of 2013 a first version of the methodology was drafted and, as a consequence, it started to be tested in the pilot cases identified by the Project partners [5]. To this aim the following typologies of case studies have been chosen: six schoolyards in Florence, one square and a peri-urban green corridor in Bilbao and two public parks in Rotterdam. The application of the methodology to the pilot cases during the ante-operam phase led to its updating, while the final optimization has been achieved after the interventions' realization and the post-operam phase have been carried out at the beginning of 2015 [6]. As a consequence, guidelines to facilitate the application of the methodology have been developed [7].

In Fig. 1 the final version of the methodology proposed by QUADMAP is schematically illustrated. A further description of each section (selection, analysis and management) of the methodology and related tools can be found in previous publications [6, 7].

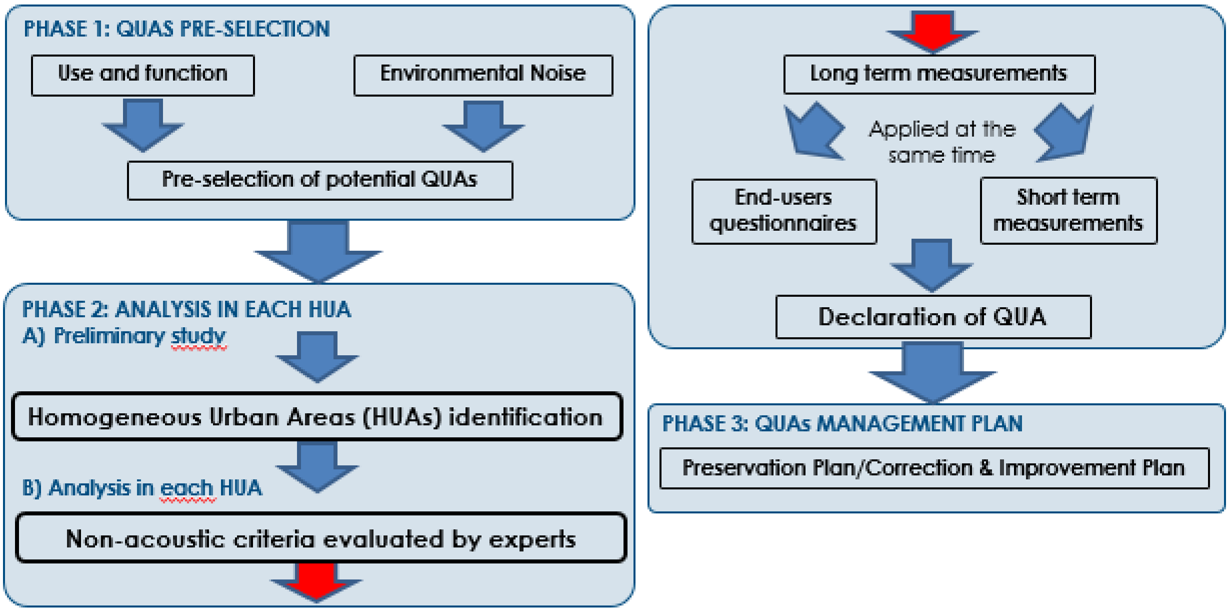

Figure 1. QUADMAP methodology-flowchart.

One of the main topic of the QUADMAP Project is the systematic use of the soundscape concept in the qualitative acoustic analysis and in the public opinion collection about present and desired features of selected QUAs. Soundscape is defined as "the acoustic environment as perceived, experienced and understood by people, group or community" [8]. As a direct application of this concept, QUADMAP Project promoted the application of a participatory approach, in terms of questionnaires submitted to end-users, aimed at integrating the people's perception with objective acoustic measurements.

2. Data

By applying tools described in the methodology [6, 7] the following typologies of data have been collected by QUADMAP in each pilot area:
- quantitative data (noise maps, short and long term measurements, wave recordings);
- qualitative data (end-users questionnaires, general and non-acoustic information).

For the purpose of this study, short term measurements and end-users questionnaires collected in the six pilot cases located in Florence during the ante-operam phase are deeper evaluated by using statistical analysis and models with the main aim to understand the effect of each considered parameter.

The format of the end-users questionnaire has been developed by the project partners, it is very broad and includes more than one hundred questions, structured with open and closed questions regarding both the acoustical and general perception of QUAs [6, 7].

Short term measurements are based on the Time History of sound pressure levels. They are carried out at the same time as each end-users questionnaire and have a duration of 15-30 minutes.

Collected data have been considered in a comprehensive manner, without taking into account the division by school and age of the respondent, in order to have an overview of the selected cases.

3. Descriptive statistical analysis

After collecting data and before performing synthetic models, firstly a descriptive analysis of the sample has been carried out by using the statistical software SAS, version 9.1. According to this analysis, the structure of the examined group results to be not completely consistent with the overall population. The majority of respondents is female (61.6%) and the age distribution is highly skewed; in fact, over 60% are under 20 years and about 35% have more than 30. Students, children and adolescents under 20 years old who attend school gardens form the sample size for almost two-thirds, while the remaining part is divided into various trades. This particular composition, dominated by children, can have an impact on the responses to the questionnaire which are in fact more suitable to older age than for young people and in some cases may have been misled. As a consequence, the structure of the sample may also influence the results of the statistical models that have been applied to the data set and that will be illustrated in the section 4.

Secondly, several contingency tables have been produced in order to explore the relationship between the variables. Contingency tables are frequency tables in multiple dimensions, in which each cell contains the number of cases for each level of the considered variables. As one of the most interesting of the obtained results, from Fig. 2 it can be observed that there is a similar frequency for those respondents who deem the acoustical environment as noisy and for those who consider it as calm both when noise levels associated to short term measurements are relatively low (55÷60 dB(A)) and when they are high (⩾65 dB(A)).

A further confirmation of this result is given in section 4.

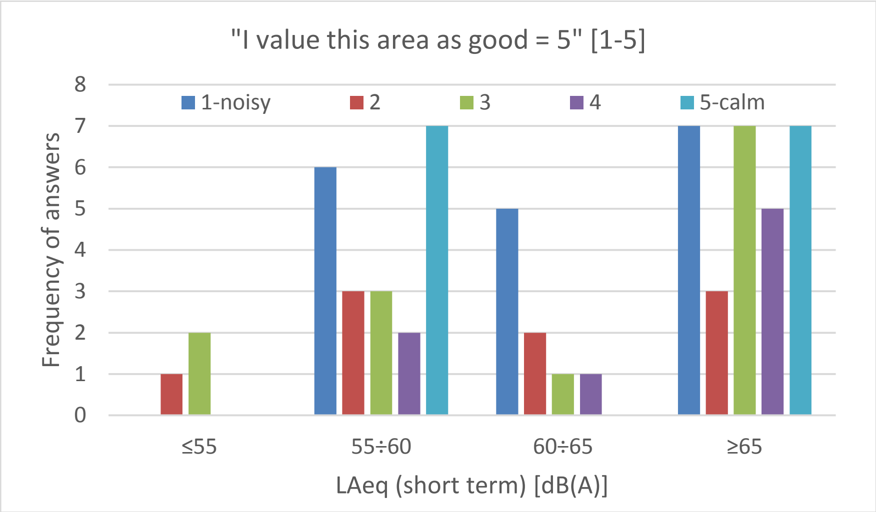

Figure 2. How the perception of the sound environment (noisy-calm) is affected by noise levels.

4. Statistical models and analysis

4.1 Model's variables

Independent variables considered for the models illustrated in section 4.2 are both quantitative and qualitative, while the only considered dependent variable is a qualitative one, i.e. the addressed question is *"I value this area in general as good"*, with possible answers from 1 to 5 (Fully disagree, Disagree, Neutral, Agree, Fully agree).

Concerning qualitative independent variables, they have been selected among the questions of the end-user questionnaire and are *"Referring to this area, I perceive each of the following items as pleasant: Air quality, Safety, Well-maintenance, Services and equipment (benches, playing areas ..), Accessibility, Acoustic environment, Natural elements (green areas, water, birds ..), Climate (humidity, brightness, wind), Visual aspects, Smells"*. The answer to these questions is to express an opinion on the different aspects of the quality of the area, translated into a score, on a scale from 1 to 5 (not pleasant at all, quite unpleasant, pleasant enough, pleasant, very pleasant).

Information acquired from the noise measurements constitute the objective part of the database.

Based on short term measurements, parameters such as LAeq (equivalent continuous sound pressure level), L10-L90 (sound level exceeded respectively for 10% and 90% of the measurement time) and LA50 (sound level exceeded for 50% of the measurement time) are evaluated and taken into account as independent variables for the statistical models.

4.2 Logistic regression and ordinal logistic models

Among available statistical models, logistic regression and ordinal logistic ones have been evaluated.

In case of the logistic model, unlike the classic linear prediction curve, the probability is no longer a linear function of the parameters, but it is the logit of the probability. The logit function is defined as the logarithm of the ratio of the probability of success on the failure, that is:

(1) $$logit(p) = \log(\frac{p}{1-p}).$$

In simple logistic regression, thus:

(2) $$logit(p_i) = \alpha + \beta \cdot x_i.$$

It follows that the probability can be expressed as:

$$p_i = \frac{e^{\alpha+\beta\cdot x_i}}{1+e^{\alpha+\beta\cdot x_i}}. \quad (3)$$

What differentiates the ordinal logistic model, with respect to the linear one, lies in the different type of the response variable. As reported before, in the logistic model responses must be binary, that means that the dependent variable can assume only two values, whether in the ordinal model the answer is structured on several ordered categories and this is perfectly suited to the typologies of available data in which the evaluation of the area is expressed on a scale from 1 to 5 in the end-user questionnaire. The solution, in this case, lies in the construction of a model capable of providing a "cumulative probability". The cumulative probabilities are the sum of the probability of obtaining a response corresponding to a category less than or equal to that reference. To take an example, if there's the necessity of evaluating the cumulative probability for the score equal to 3, then: P (Y ≤ 3) = P (Y = 1) + P (Y = 2) + P (Y = 3). Clearly, the cumulative probability for the higher score is equal to 1.

Using the logit of the probability, the model can be reformulated as in Eq. 4.

$$logit[P(Y \le j)] = \alpha_j - \beta * x. \quad (4)$$

Which, translated, is equivalent to:

$$\log\left[\frac{P(Y \le j)}{P(Y > j)}\right] = \alpha_j - \beta * x. \quad (5)$$

The probability, therefore, is expressed as:

$$P(Y \le j) = \frac{e^{\alpha_j-\beta x}}{1+e^{\alpha_j-\beta x}}. \quad (6)$$

Note that the subscript j of the intercept α represents the reference category.

4.3 Results

In this section results obtained from statistic models, implemented to understand how respectively quantitative and qualitative variables influence the general perception of a QUA, are illustrated.

The implementation of both models has been performed with the statistical software SAS.

Firstly, a simple logistic regression model has been applied, using as dependent binary variable the general perception of area and as quantitative explanatory factor only the LA50. Results suggest that, among those evaluated from short term measurements, the most appropriate parameter to describe the perception of users is the LA50 [6].

Afterwards, more complex logistic regression and ordinal regression models have been implemented, considering as dependent factor still the general perception of the area and as independent variables all the qualitative ones indicated in section 4.1 and, according to results obtained with the simpler models, the LA50 as unique quantitative factor. Before running the logistic model, according to its structure, answers from 1 to 5 (Fully disagree, Disagree, Neutral, Agree, Fully agree) to the question chosen as dependent variable have been reclassified in two modalities: agreement (correspondent to: Neutral, Agree, Fully agree), disagreement (correspondent to: Fully disagree, Disagree).

In Table 1 outputs of this last model are reported, stressing that only a few variables show statistically significant influence on the general perception of the area. In particular, variables evaluated as significant are all qualitative and show signs of the coefficients in the expected sense (e.g. the air quality tends to increase the positive perception of the area when evaluated as pleasant or very pleasant), while the LA50 appears to be not statistically significant when considered together with the qualitative variables.

Table 1. Logistic regression models-parameters estimation.

Analysis of Maximum Likelihood Estimates						
Parameter	*Modalities*	*Degree of Freedom (DF)*	*β*	*Standard Error (SE)*	*Wald Chi-Square*	*Pr > χ2*
Air quality	3	1	0.7373	0.4387	2.8246	0.0928
Air quality	4	1	1.1032	0.6462	2.9146	0.0878
Maintenance	5	1	1.5989	0.8694	3.3820	0.0659
Climate	2	1	-1.3714	0.6131	5.0029	0.0253
LA50		1	0.0101	0.0456	0.0490	0.8249

Legend: Air quality 3= Neither pleasant nor unpleasant, 4=Pleasant; Maintenance 5= very pleasant; Climate 2= unpleasant.

Much more explicit are the results of the ordinal logistic model, where the dependent variable is left in five modalities (see Table 2).

Table 2. Ordinal logistic models-parameters estimation.

Analysis of Maximum Likelihood Estimates						
Parameter	*Modalities*	*Degree of Freedom (DF)*	*β*	*Standard Error (SE)*	*Wald Chi-Square*	*Pr > χ2*
Intercept**	5 very pleasant	1	-3.8362	1.0763	12.7047	0.0004
Intercept	4 pleasant	1	-1.8951	1.0619	3.1847	0.0743
Intercept	3 quite pleasant	1	0.1074	1.0591	0.0103	0.9192
Intercept	2 not so pleasant	1	2.5527	1.1026	5.3596	0.0206
Smells	s*	1	0.2649	0.1427	3.4452	0.0634
Visual aspects	s*	1	0.2183	0.1346	2.6312	0.1048
Climate	s*	1	0.2520	0.1868	1.8197	0.1773
Natural elements	s*	1	0.4896	0.1676	8.5367	0.0035
Acoustic environment	s*	1	0.4252	0.1369	9.6395	0.0019
Availability	s*	1	-0.1009	0.1685	0.3584	0.5494
Services	s*	1	0.3412	0.1158	8.6834	0.0032
Maintenance	s*	1	0.3812	0.1354	7.9246	0.0049
Security	s*	1	0.3179	0.1614	3.8782	0.0489
Air quality	s*	1	0.4488	0.1581	8.0599	0.0045
LA50		1	0.00907	0.0178	0.2589	0.6109

Legend:*s=sufficient (scores 3, 4 and 5), reference category: insufficient (scores 1 and 2) **Reference category: not pleasant at all (score 1).

According to values assumed by the β coefficient (β represents the difference between the expected value of getting a high evaluation of the area given a positive opinion on a particular aspect and the expected value of the same evaluation given a negative opinion), the output of ordinal logistic model seems to be similar to that obtained in the previous model, for qualitative variables considered as significant in both cases.

As a first conclusion, while the LA50 turns out to be statistically significant if evaluated alone, when the subjective variables are included the effect of this quantitative factor loses significance, according to both the considered statistical models.

In Table 3 the adaptability of both models to the database is evaluated.

Table 3. Models' adaptability.

Model	*N° of significant variables*	*Pseudo R-square*	*H&L GOF (p-value)*
Ordinal logistic	7	0.3947	
Logistic regression	4	0.3545	0.6062

In both cases the adaptation of the model, measured by the adjusted Pseudo R-square, is not completely satisfactory, because the value of the index is less than 0.40. Nevertheless, in social studies values of this coefficient close to 0.40 are recurrent, since subjective and perceptions variables lead to lower values of the R-square with respect to experimental and economic ones.

The GOF (Goodness Of Fit) Hosmer & Lemeshov statistic test has as null hypothesis to have a good fit against the alternative of having a poor foresight (lack of fit) (Table 3). This type of test, which is not applied to multinomial models, for the logit regression model allows to accept the null hypothesis of good fit.

5. Conclusions

The Environmental Noise Directive (END) provides to Member States a very generic definition of QUA. The main aim of the LIFE+2010 QUADMAP Project is to develop and share a complete methodology to select, analyze and manage QUAs. At the beginning of 2015 the final version of this methodology, successfully tested in ten pilot areas respectively located in Florence, Bilbao and Rotterdam, has been delivered together with practical guidelines.

In this paper descriptive analysis and statistical models have been evaluated, starting from qualitative and quantitative data achieved during the ante-operam phase in the pilot cases selected in Florence.

According to both descriptive analysis and statistical models, the users of the six school gardens in Florence discriminate the area mainly on the basis of the perception of the air quality and of the well-maintenance. The role of the quantitative variables is found to be, however, quite marginal. Also in the literature [9-11] examples are present of how quantitative information, although they are the most obvious to detect, can be misleading if they are the only considered ones. In particular, it can be seen that the measured noise levels can hardly be associated to the users' perception of the external area. In fact, it may happen that the sound environment is negatively judged in case of low noise levels and vice versa. This is probably explained by the fact that, often, even if they perceive high levels of noise, they do not generate annoyance because they are unconsciously perceived as an integral part of the garden itself. In addition, as indicated in section 3, since the sample was prevalently composed by children sometimes also under six years old, it has been quite difficult to ask them especially about the perception of the acoustic environment. As a consequence, during the post-operam surveys a simplified version of the questionnaire has been developed and submitted in schoolyards.

About the comparison between the two typologies of tested models, it can be concluded that their adaptability is similar and, although not completely satisfying, it is acceptable for the current data set. Nevertheless, the ordinal logistics models show a higher number of significant coefficient if compared to logistic regression ones and, in addition, in the first case many variables can be interpreted as determinants of the general perception of the area as good.

ACKNOWLEDGEMENTS

The authors would like to thank all who sustained them with this research, especially the European Commission for its financial contribution to the Project into the LIFE+2010 program.

REFERENCES

1. Final Report on Task 1 "Review of the Implementation of Directive 2002/49/EC on Environmental Noise", May, (2010), performed by Milieu Ltd, Risk and Policy Analysis Ltd (RPA) and TNO, contracted by DG Environment of European Commission in dec. 2008.
2. Final Report on Task 2 "Inventory of Potential Measures for a Better Control of Environmental Noise", May, (2010), performed by Milieu Ltd, Risk and Policy Analysis Ltd (RPA) and TNO, contracted by DG Environment of European Commission in dec. 2008.
3. Final Report on Task 3 "Impact Assessment and Proposal of Action Plan", May, (2010), performed by Milieu Ltd, Risk and Policy Analysis Ltd (RPA) and TNO, contracted by DG Environment of European Commission in dec. 2008.
4. EEA, Good practice guide on quiet areas, April, (2014).
5. Bartalucci, C., Bellomini, R., Borchi, F., Carfagni, M., Governi, L., Luzzi, S. and Natale, R. LIFE+2010 QUADMAP project (Quiet Areas Definition and Management in Action Plans): the proposed methodology and its application in the pilot cases of Firenze, *Proceedings of the 42st International Congress on noise*, Innsbruck, Austria, 15-18 September, (2013).
6. Carfagni, M., Bartalucci, C., F. Borchi, L. Governi, Petrucci, A., Weber, M., Aspuru, I., Bellomini, R., Gaudibert, P. LIFE+2010 QUADMAP Project (QUiet Areas Definition and Management in Action Plans): the new methodology obtained after applying the optimization procedures, Proceedings of the 21st International Congress on Sound and Vibration, Beijing, China, 13-17 July, (2014).
7. QUADMAP Project, (2015). *Guidelines for the identification, selection, analysis and management of quiet urban areas*. [Online.] available:http://www.quadmap.eu
8. ISO/CD 12913-1 Acoustics -- Soundscape -- Part 1: Definition and conceptual framework. (Draft).
9. Curcuruto, S., Asdrubali, F., Brambilla, G., Silvaggio, R., D'Alessandro, F., Gallo, V. Socio acoustic survey and soundscape analysis in urban parks in Rome, Rome, Italy, (2011).
10. Carvalho, A.P.O., Morgado, A.E.J., Henrique, L. Relationships between subjective and objective acoustical measures in churches, Poythecnic institute of Porto, Porto, Portugal, (1991).
11. Engel, M.S., et al. Statistical analysis of a combination of objective and subjective environmental noise data using factor analysis and multinomial logistic regression, Laboratory of Environmental and Industrial Acoustics and Acoustic Comfort, Federal University of Paranà, Curitiba, PR, Brazil, (2014).