Statistical analysis of acoustic data. Combining objective and subjective measures.

C. Bartalucci, F. Borchi, M. Carfagni[1], M.S. Salvini and A. Petrucci[2,*]

[1] *Department of Industrial Engineering, University of Florence, via di S. Marta 3, 50139 Firenze, Italy;* chiara.bartalucci@unifi.it , francesco.borchi@unifi.it, monica.carfagni@unifi.it
[2] *Department of Statistics, Computer Sciences, Applications "G. Parenti", University of Florence, Viale Morgagni 59, 50134 Firenze, Italy* mariasilvana.salvini@unifi.it, alessandra.petrucci@unifi.it
**Corresponding author*

Abstract. *This paper presents a statistical approach to analyze a set of data collected by QUADMAP (QUiet Areas Definition and Management in Action Plans), a LIFE+2010 Project on Quiet Urban Areas (QUAs). This Project aims to define a method regarding identification, outlining, characterization, improvement and managing of 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. Logistic models are applied to part of the quantitative (noise measurements) and qualitative (results of interviews about noise perception in an urban setting) data collected in the city of Florence during the ante-operam phase. The results underline the importance of the survey's design, in order both to obtain information combining the different type of data and to be able to evaluate the net effect of single variables.*

Keywords. *Logistic models, Acoustic data, Objective measures, Subjective measures*

1 Introduction

The European Directive 2002/49/EC on the Assessment and Management of Environmental Noise (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.
QUADMAP Project started in September 2011 with the final aim of developing a complete, practical and demonstrated methodology to select, analyse and manage QUAs (Quiet Urban Areas). 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 [1]. 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 [2]. Consequently, guidelines to facilitate the application of the methodology have been developed [6]. 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 [2, 6], 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 [2, 6].
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.
Among available statistical models, logistic regression and ordinal logistic models have been evaluated.

3 Methods

After collecting data and before performing synthetic models, firstly a descriptive analysis of the sample has been carried out. 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. Therefore, the structure of the sample may also influence the results of the statistical models that have been applied to the data set. Independent variables considered for the models 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, etc.), 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 Results

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 [2]. 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 above cited 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.

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

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

Table 1. Logistic regression models: parameters estimation

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	s*	1	0.00907	0.0178	0.2589	0.6109

Table 2. Ordinal logistic models: parameters estimation

In table 3 the goodness of fit of both models to the database is evaluated.

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

Table 3. Goodness of fit of the models

5 Conclusions

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 to analyze noise objective and subjective variables.

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 [3, 4, 5] 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, 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. Consequently, 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 the ordinal logistics models show a higher number of significant coefficients 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.

References

[1] Bartalucci, C., Bellomini, R., Borchi, F., Carfagni, M., Governi, L., Luzzi, S. and Natale, R. (2013). 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.

[2] Carfagni, M., Bartalucci, C., F. Borchi, L. Governi, Petrucci, A., Weber, M., Aspuru, I., Bellomini, R., Gaudibert, P. (2014). 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.

[3] Curcuruto, S., Asdrubali, F., Brambilla, G., Silvaggio, R., D'Alessandro, F., Gallo, V. (2011). Socio acoustic survey and soundscape analysis in urban parks in Rome, Proceedings of the 11th ICBEN Congress, London, United Kindom, 24-28 July, 2011.

[4] Carvalho APO, Morgado AEJ, Henrique L. Relationship between subjective and objective acoustical measures in churches. (1997). *Build Acoust* **4**, 1–20.

[5] Engel, M.S., Hochsteiner De Vasconcelos Segundo E., Trombetta Zannin P. H. (2014). Statistical analysis of a combination of objective and subjective environmental noise data using factor analysis and multinomial logistic regression, *Stoch Environ Res Risk Assess* **28**, 393–399.

[6] QUADMAP Project, (2015). Guidelines for the identification, selection, analysis and manage-ment of quiet urban areas: http://www.quadmap.eu