Atti del Convegno intermedio Società Italiana di Statistica 2011, Università di Bologna, 8-10 giugno 2011

Statistics in the 150 years from Italian Unification, Quaderni di Dipartimento di Scienze Statistiche, Serie Ricerche 2011, n.2 Alma Mater Studiorum Università di Bologna
ISSN 1973-9346

New Technologies and Statistics: Partners for Environmental Monitoring and City Sensing

Giovanni Borga, Rina Camporese, Niccolò Iandelli and Antonella Ragnoli

Abstract Urban space is interconnected thanks to a myriad of technological devices whose data can be aggregated in a geographic database thereby providing a representation of what is happening around us. City Sensing is an "immersive sensing" and a new opportunity to survey the territory and the environment, by the means of low cost sensors small enough to be wearable. Advantages of such a framework are the widespread and numerous measurements at lower unit cost and also the near real-time friendly communication, together with an interaction with citizens. There are, of course, some limits: low-cost sensors' measurements are affected by a greater error; the huge amount of data produced can result in a sort of data overload; pressure for real time can lead to hasty elaborations. Statistics can offer some help to reduce the impact of the drawbacks related to measurement quality control and error estimates and it can also offer possible solutions for significant data synthesis and representation.

1 City sensing and New Technologies

A new strategy for environmental monitoring is outlined by the rapid development of sensors and computer networks: a great number of data acquisition instruments, distributed and interconnected, provide near real-time data flows. A wide spread monitoring network displaces the traditional paradigm based on the use of few stand-alone stations, focusing on the pervasiveness of low cost nodes, equipped with light sensors in order to get a small gridded representation of the territory. Urban space can be interconnected thanks to a myriad of technological devices whose data are aggregated in a geographic database, providing a relevant representation of what is happening around us. Having this in mind, City Sensing becomes an *immersive sensing*

[1] Giovanni Borga giovanni@borga.it Rina Camporese rina.camporese@poste.it
Niccolò Iandelli niccogeo@gmail.com Antonella Ragnoli aragnoli@libero.it
Iuav University of Venice - New Technologies and Information on Territory and Environment Research Group

and a new exciting opportunity to survey the territory. In combination with the Web 2.0 opportunities, City Sensing can be defined as a Sensor Web, to perform environmental monitoring in the style of social networking and from a cooperative perspective [1].

Recent technological research has produced sensors (mainly based on Micro-Electro-Mechanical System) integrated into commonly used instruments (i.e. smart-phones or devices small enough to be wearable and at low cost) that can measure various quantities by translating variations of physical parameters into electrical impulses (e.g. acceleration, temperature, humidity, concentration of gases, magnetic fields, …), or transform built-in microphone in a noise detector. Each mobile phone could become an *environmental station* and a node of a larger monitoring network. The spread of these technologies has opened the door to new research experiences made by the Iuav - NT&ITA Doctorate School and Research Group, such as the design of an integrated system of sensors for environmental and road traffic monitoring (widespread in the territory and based on WSN-Wireless Sensor Network), and the test of a prototype wearable multi-sensor with blue-tooth transmission.

2 Potentials, drawbacks and possible solutions

The main advantages of such a framework are the widespread and numerous measurements at lower unit cost (versus the traditional precise, expensive and few in number measures), the near real-time friendly communication and the possible interaction with citizens. There are, of course, some limits. Firstly, data coming from actual low-cost sensors are affected by a greater error as compared to certified expensive instruments. Secondly, a huge amount of data can be easily and quickly produced; this can result in a sort of data overload, which is difficult to manage and interpret. Furthermore, the pressure for real time data can lead to hasty and un-meditated elaborations. Statistics can offer some help to limit those drawbacks with regard to measurement quality control and error estimates [4]. The main advantages of using a statistical approach could be:

- rationalise the numerous and enthusiastic data collection processes, so as to make them more significant and representative, e.g. in terms of sampling strategies
- raise awareness of measurements' quality control and evaluation of errors
- keep uncertainty of the results into consideration
- enhance the essential role of metadata.

In cooperation with Information Design, statistics can also develop innovative solutions in favour of a significant data synthesis and representation [6,7], especially when multidimensional data have to be considered along with both space and time.

3 Air and noise pollution examples

Here are two examples of how new technologies modify the traditional approach to environmental monitoring. The Framework for the Development of Environment Statistics defined by the United Nations [8,9,10], which is currently under revision, has been taken into consideration as a reference frame for the following reflections.

As to air pollution, UNSD Environmental Indicators essentially regard emissions, while indicators on ambient concentrations of selected pollutants are not present, mainly because they lack quality, coverage across countries and international comparability. As spatial patterns of air pollutants concentration vary significantly across territories and being usually monitored with very few, precise and expensive stations, it often happens that national environmental statistics describe the characteristic of the monitoring network (air monitoring stations: number, type and locations), instead of pollutants concentration. The following quotation comes from a UN document dated 1991: “The cost of environmental monitoring has inhibited the development of statistically valid space/time sampling frames” [10]; it clearly explains the reason behind the state of the art. Low cost sensor networks open a new scenario, where challenges are no more related to the costs of measurement, but to instruments calibration, proper time and space dependent sample strategies, ascertainment of statistical validity, and significant data reduction of massive data sets.

With regard to noise pollution, in 1998 the population exposed to excessive noise (i.e. noise levels exceeding national standard) has been selected by the UN as a suitable indicator [9]. Furthermore, the 2002 EU Directive on Environmental Noise required Member States to draw harmonised *strategic noise maps*. Despite that, actual national statistics on noise pollution often show only the responses to noise pollution, in terms of actions and policies adopted to reduce noise pollution effects. In Italy, for example, noise barriers, low noise pavements, noise zoning are the selected indicators on “noise pollution” in “Urban Environment Indicators” statistical national report [5].

Currently, low-cost noise sensors guarantee a better measurement quality (provided that they are calibrated), as compared to sensors measuring concentration of gases, for which the output measurements are more controversial. Therefore, a hypothetical sample strategy to assess environmental noise in Italy is proposed below. It aims at obtaining noise exposure maps along the roads in urban environments, using two indicators quoted in the EU Directive: day-evening-night level in decibels and night-time noise indicator (obtained through A-weighted long-term average sound levels, determined over all the day periods of a year) [2].

The proposed sample strategy requires stratification according to space and time. As to space, road segments of urban environment could be stratified by techno-functional characteristics related to speed limits and traffic flow (highways, suburban, urban, local). Such information is available in the Catasto Strade (Roads Register), required by law and usually available, in some form, at least for principal towns. Another spatial stratification variable could be the land cover class, such as the one provided by the GSE Land European Urban Atlas Services (part of the European Earth Observation Programme - GMES). It comes from a very high-resolution hot spot mapping of urban functional areas and it allows for their stratification according to different urban fabric density (continuous, dense, medium, low, sparse) and to functional characteristics (residential, industrial, etc.). As to time stratification, the sample strategy could resemble the one adopted for HETUS – Harmonised European Time Use Survey, which covers an entire 12 months period - 24hrs - 7days, with stratification based on month and type of day (Mon-Fry, Sat, Sun). The characteristics of small noise sensors in terms of cost and transportability would easily adapt to such a sample. If a noise map has to be the output, estimates of noise indicators derived from sampled locations would then be used as expected values for the road segments that have not been surveyed, on the basis of spatial stratification variables.

In this view there comes a proposal which could make vein and pulses of an orthodox statistician tremble: to *contaminate* the traditional sampling approach with a wiki component, in the style of collaborative mapping - www.OpenStreetMap.org - and collaborative research - www.GalaxyZoo.org. The first experience shows how Web 2.0 collaborative activities can produce a valid map of the territory. The second is probably less known: a data set made up of a million galaxies images collected by the robotic telescope of the Sloan Digital Sky Survey have been made available on the web and the morphological classifications of galaxies, which enables scientists to understand how galaxies form and evolve, have been carried out by a network of registered web users, after a brief on-line tutorial phase. This experience shows how common citizens are open to follow simple guidelines to contribute to a scientific project, in the aim of creating a wide knowledge framework. Another emblematic experience in this field is NoiseTube.net: a research project, which aims at developing a new participative approach to noise pollution monitoring by involving the general public.

Traditional sampling measurements could be integrated with spontaneous contributions of citizens [3], capturing data with smart-phones applications, in order to cover non-sampled areas and periods. The final estimates would be produced through ex-post weight calibration and proper weighted averages of both structured and wiki components of the sample. It sounds complicated, but challenging, too.

References

1. Calabrese, F., Kloeckl, K., Ratti, C.: WikiCity: Real-Time Location-Sensitive Tools for the City. In: Foth , M. (ed.) Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City, Information Science Reference. Hershey, New York (2009)
2. European Parliament and Council: Directive 2002/49/EC on Assessment and Management of Environmental Noise. Official Journal of the European Communities (18.7.2002)
3. Goodchild, M.F.: Citizens as sensors: web 2.0 and the volunteering of geographic information (Editorial). GeoFocus International Review of Geographical Information Science and Technology, Volume 7, pages 8–10 (2007)
4. Goodchild, M.F.: Statistical Perspectives on Geographic Information Science. Geographical Analysis, Volume 40, pages 310–325, Blackwell Publishing Inc (2008)
5. Istat: Indicatori Ambientali Urbani. Istituto Nazionale di Statistica, Roma (2009)
6. Tufte, E.R.: Visual explanations. Images and quantities, evidence and narrative. Graphic press, Cheshire, Connecticut (2005)
7. Tufte, E.R.: The visual display of quantitative information, Second ed., Graphic press, Cheshire, Connecticut (2007)
8. United Nations: A Framework for the Development of Environment Statistics. Statistical Papers, Series M No. 78, United Nations, New York (1984)
9. United Nations: Concepts and Methods of Environment Statistics: Human Settlements Statistics? A Technical Report. Studies in Methods, Series F No. 51, United Nations, New York (1988)
10. United Nations: Concepts and Methods of Environment Statistics - Statistics of the Natural Environment. Studies in Methods, Series F No. 57, United Nations, New York (1991)