Modeling environmental quality: a case study.

La modellizzazione della qualità ambientale: un caso di studio

De Iaco S. and Maggio S. and Palma M. and Posa D.

Abstract Assessing environmental quality usually requires the observation of two or more correlated variables, which are measured in several points of the study area. Sometimes, the characteristic of interest is sparsely sampled over the area, then it is convenient to incorporate some auxiliary variables, correlated with the variable of interest, into the estimation procedure. Indeed they carry relevant information for the variable being estimated, especially if they are more densely available over the domain. In this paper three different spatial interpolation approaches have been used in order to obtain spatial predictions of the variable of interest characterized by a severe lack of data.

Abstract *Generalmente, la valutazione della qualità ambientale necessita della misurazione di due o più variabili tra loro correlate, i cui valori sono rilevati in diversi punti dell'area d'indagine. Se le osservazioni per la variabile di interesse (variabile primaria) sono scarsamente disponibili sull'area d'indagine, le variabili ausiliarie, che risultano essere correlate con quella primaria e misurate in numerosi punti del dominio, dovrebbero essere inserite nella procedura di stima, dal momento che esse forniscono informazioni rilevanti per la stima della stessa variabile primaria. Nel presente articolo, la stima della variabile in esame, osservata su un*

De Iaco Sandra
Dept. of Management, Economics, Mathematics and Statistics - University of Salento (Italy) e-mail: sandra.deiaco@unisalento.it

Maggio Sabrina
Dept. of Management, Economics, Mathematics and Statistics - University of Salento (Italy) e-mail: sabrina.maggio@unisalento.it

Palma Monica
Dept. of Management, Economics, Mathematics and Statistics - University of Salento (Italy) e-mail: monica.palma@unisalento.it

Posa Donato
Dept. of Management, Economics, Mathematics and Statistics - University of Salento (Italy) e-mail: donato.posa@unisalento.it

ridotto numero di localizzazioni sul dominio d'indagine, viene effettuata mediante l'utilizzo di tre differenti metodi di interpolazione spaziale.

Key words: soft and hard data, cokriging, kriging with non-constant mean, BME kriging, GIS

1 Introduction

In environmental science, the behaviour of a natural process is often the result of several correlated variables. In this case, data are usually derived from an appropriate spatial sampling based on the observation of two or more variables, at some localizations of the area of interest and/or available environmental maps (sometimes hand-drawn) regarding different aspects connected with natural conditions (such as geology, land cover, land use) and anthropic activities (such as road network, location of industrial plants, urbanization). This kind of analysis requires the use of properly geostatistical techniques which take into account the spatial relationships among the available variables, as well as the use of a Geographical Information System (GIS) to manage large geo-referenced data sets. This integration between GIS and geostatistical tools can certainly support the evaluation of alternative scenarios and planning for possible intervention strategies.

In this paper, the spatial distribution of Radon-222 (Rn) concentrations in soil gas (the ^{222}Rn isotope, with a half-life of 3.8 days), registered during a monitoring campaign in 2012, at some points over the Lecce district (Italy), has been analysed. The collected data have been analysed taking into account the factors that favour the natural process of radon-based radioactivity exhalation. In particular, the local geology, concerning the specific kinds of lithology (such as karstification), as well as underground permeability levels, which permit the transition of a large amount of gas from great depths. Various statistical methods based on the properties of the log-normality of the data [9, 7] have been proposed in the literature; however the indicator kriging or disjunctive kriging [10] have been also extensively applied to calculate the probability that the chosen reference level is exceeded and produce the map of this probability. In the present paper, three different interpolation approaches have been applied in order to predict Rn concentrations over the study area. In particular, by considering the underground permeability levels (soft data) which characterized the study area, Rn concentrations (hard data) have been:

a) cokriged by using an indicator variable associated to medium-high/low underground permeability levels,
b) kriged with non-constant means at different levels of soil permeability,
c) estimated by using the Bayesian Maximum Entropy (BME) approach.

2 Geo-lithological features in the study area

Rn is an inert radioactive gas of natural origin, highly carcinogenic, classified by the IARC (International Agency for Research on Cancer) as the second leading cause of lung cancer after smoking. In fact, once inhaled, Rn reaches the respiratory system and, through the emission of radioactive particles, damages the cells bronchopulmonary irreversibly, often with lethal consequences for the human body.
Although Rn is produced in the depths of the earth, it easily reaches the surface because it is present in gaseous phase and because it interacts with other elements. The nature of the rocks, their mineralogical composition, the concentration of the precursor elements (uranium, thorium and radio), the porosity, permeability, water content, the presence of faults and fractures and the presence of cavities, are the elements that most affect the concentration of Rn in the soil and its transport. For these reasons, Rn risk assessment requires the knowledge of geologic and lithological features of the study area, as well as the permeability of the soil [8].

Lecce district, located at 199 metres above sea level, is characterized by a limestone substrate and the presence of sink-holes, caves and underground drainages. Karstification significantly affects the process of exhalation Rn from the ground, through the formation of an underground network of Rn diffusion that is transported by water and other soil gases; then, it is easily released to the outside because of the presence of numerous faults. By considering the lithological profile of Lecce district, the area is characterized by 6 different lithotypes, with predominance of sandstone (45.7%), limestone and dolomitic limestone (29.4%), as well as siltstone and sandstone (23.9%). Because of this particular lithology in Lecce district, the study area is almost all characterized by a medium-high level of permeability. The large presence of limestone and sandstone in the soil makes the underground highly permeable. It is well known that these two factors, i.e. high permeability and the predominance of sandstone, represent critical issues for the migration of Rn from the soil.

3 Soft and hard data

During a campaign conducted in July 2012, outdoor Rn concentrations (hard data) have been measured at 32 locations over Lecce district (Fig. 1) by using an active sampling-based system, in which the Rn and its decay products are conveyed in proximity of the radiation detector through a mechanical pumping. In particular, Rn concentrations have been recorded by a special instrument, equipped with an ionization chamber, connected to a probe, introduced into the ground up to 1 meter deep, and to a pump for the aspiration of Rn gas at fixed intervals of time (usually every minute for 20 minutes-periods).

Moreover, the underground permeability map for Lecce district has been considered and a regular grid with node separation distance equal to 2 km has been defined in order to cover the area under study (Fig. 1). Then, by using the software

Arcmap of ArcGIS and the shapefile concerning the underground permeability level in Lecce district, a value (1, 2 or 3) has been assigned to each point of the grid and each survey station, according to the different permeability levels (low, medium and high, respectively).

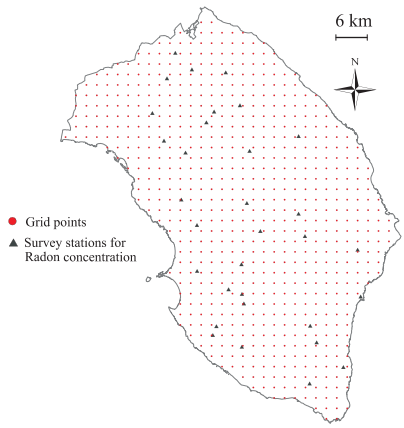

Fig. 1 Grid points with node separation distance equal to 2 km and survey stations for outdoor Rn concentrations

Evidently the auxiliary information, concerning the underground permeability of Lecce district, is more widespread over the study area with respect to the hard data, hence it has been used to improve predictions of the Rn concentrations in soil gas. In particular, this information has been introduced into the interpolation procedure as it will be shown in the following sections.

4 Spatial prediction techniques using both hard and soft data

The main objective of this paper is to propose and compare different spatial prediction techniques which take into account both hard and soft data. In particular, cokriging with indicator variable, kriging with non-constant mean and BME kriging are briefly reviewed and applied for Rn prediction. It is well-known that measurements of environmental variables can be characterized by a high coefficient of variation and/or by different level of variability, hence traditional prediction methods, such as ordinary kriging, do not account for these features. For example, the Rn concentrations in soil gas usually present a negative asymmetric distribution and more than one level of uncertainty depending on the equipment and the lithology.

The methods described in this paper utilize both image-derived (soft) data that are generally exhaustive and higher accuracy (hard) data, which are typically sparse. Hard data, or reference data, are often obtained through ground-based surveys, as in the present case study, but they may be obtained from higher accuracy remotely-sensed land cover products.

4.1 Cokriging with an indicator variable

This method allows the use of a double source of information, such as the primary variable (Rn concentrations), as well as the secondary data, to predict the variable of interest with higher accuracy throughout the study area. In particular, the auxiliary data, which represent the permeability classes at given for any node of the grid in Fig. 1, have been derived by an available geo-lithological map.

In this case, the routine Cokb3d of the GSLib package [5] has been suitably applied in order to perform the required predictions.

After modeling the sample direct and cross semivariograms of the analysed variables, the following *Linear Coregionalization Model* [12, 4] has been fitted:

$$\Gamma(\mathbf{h}) = \mathbf{B}_1\, g_1(\mathbf{h}) + \mathbf{B}_2\, g_2(\mathbf{h}) \tag{1}$$

where g_1 is a nugget model and g_2 is a gaussian model [6] with unit sill and range equal to 20 *km*; while $\mathbf{B}_l, l = 1,2,$ are the following coregionalization matrices:

$$\mathbf{B}_1 = \begin{bmatrix} 80 & 2 \\ 2 & 0.125 \end{bmatrix}, \qquad \mathbf{B}_2 = \begin{bmatrix} 263 & 4.8 \\ 4.8 & 0.088 \end{bmatrix}. \tag{2}$$

Finally, cokriging predictions of Rn concentrations in soil gas have been obtained as shown in Fig. 2-a).

4.2 Kriging with non-constant mean

The application of the interpolation techniques known as kriging with non-constant mean, is based on the effect of permeability on the Rn concentrations. In particular, the Rn concentrations grouped by the three permeability classes, have been appropriately analysed and then the hypothesis of the presence of different mean levels of Rn concentrations in soil gas has been assumed.

Computational aspects have been faced through the use of the GSLib package [5]. In particular, in the kriging prediction an exponential model [6] with sill equal to 170 and range equal to 15 km has been used; hence predictions for Rn concentrations in soil gas have been obtained for the study area (Fig. 2-b)).

4.3 BME kriging

BME accounts for measurements characterized by significant variability, i.e., soft data, in addition to measurements with little or no variability, i.e., hard data. These equations incorporate, in the prediction procedure, soft data for which an asymmetric distribution, such as the triangular distribution (used in this case study), can be assumed. BME is usually introduced as an alternative method in order to utilize soft data in prediction, especially when these last are skewed.

The numerical implementation of this approach has been performed by the Spatiotemporal Epistemic Knowledge Synthesis-Graphical User Interface (SEKS-GUI). A thorough review of the BME theory and techniques, which is clearly beyond the aim of this paper, can be found in the literature [1, 2, 11, 3]. The sample standardized semivariogram for hard and soft data has been fitted by a unit sill exponential model [6] with range equal to 9.683 km. This model has been used in order to obtain BME kriging predictions of Rn concentrations in soil gas (Fig. 2-c)).

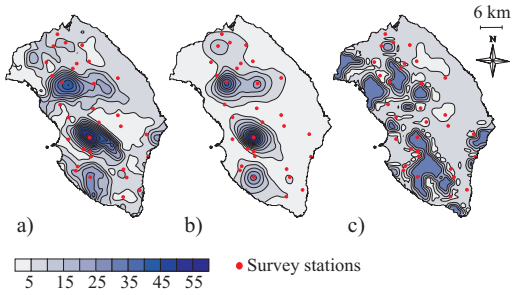

Fig. 2 Estimated Rn concentrations using a) cokriging interpolation with indicator variable, b) kriging with non-constant mean, c) BME kriging

Acknowledgements This research has been supported by the Interreg project "Development of an Environmental Meteorological Stations Network on the Ionian Islands and Southern Italy-DEMSNIISI" Greece-Italy 2007-2013.

References

1. Christakos, G.: A Bayesian/maximum-entropy view to the spatial estimation problem. Math. Geol. **22**, 7, 763–777 (1990)
2. Christakos, G.: Modern Spatiotemporal Geostatistics, 3rd ed., Oxford Univ. Press, New York (2000)
3. Christakos, G., Kolovos, A., Serre, M.L., Vukovich, F.: Total Ozone Mapping by Integrating Databases From Remote Sensing Instruments and Empirical Models, IEEE Transaction on Geoscience and Remote Sensing **42**, 991–1008 (2004)
4. De Iaco, S., Maggio, S., Palma, M., Posa, D.: Advances in spatio-temporal modeling and prediction for environmental risk assessment. In: *Air Pollution - A Comprehensive Perspective*, Budi Haryanto (Ed.), InTech, (2012) doi: 10.5772/51227
5. Deutsch, C.V., Journel, A.G.: *GSLib*: Geostatistical Software Library and User's Guide, Oxford Univ. Press, New York (1998)
6. Journel, A.G., Huijbregts, C.J.: Mining geostatistics, Academic Press, London (1981)
7. Miles, J.C.H.: Mapping the portion of the housing stock exceeding a radon reference level. Radiat. Prot. Dosim. **56**, 14, 207–210 (1994)
8. Miles, J.C., Appleton, J.D.: Mapping variations in radon potential both between and within geological units. J. Radiol. Prot. **25**, 257–276 (2005)
9. Miles, J.C.H., Green, B.M.R., Lomas, P.R., Cliff, K.D.: Radon in U.K. homes. Defining the areas affected. Radiat. Prot. Dosim. **36**, 117–120 (1991)
10. Raspa, G., Salvi, F., Torri, G.: Probability mapping of indoor radon-prone areas using disjunctive kriging. Radiat. Prot. Dosim. **138**, 1–17 (2010)
11. Serre, M.L., Christakos, G., Li, H., Miller, C.T.: A BME solution of the inverse problem, Stoch. Environ. Res. Risk Assess. **17**, 354–369 (2003)
12. Wackernagel, H.: Multivariate geostatistics: an introduction with applications, 3rd Ed., Springer, Berlin (2003)