Geostatistics and GIS: tools for environmental risk assessment[1]

Sabrina Maggio, Claudia Cappello, Daniela Pellegrino
Dip.to di Scienze Economiche e Matematico-Statistiche, Facolta' di Economia, University of Salento, Italy, sabrina.maggio@unisalento.it

Abstract: The environmental risk analysis involves the observation of complex phenomena. Different kinds of information, such as environmental, socio-economic, political and institutional data, are usually collected. In this paper, spatial-temporal geostatistical analysis is combined with the use of a Geographic Information System (GIS): the integration between geostatistical tools and GIS enables the identification of alternative scenarios and possible strategies for the environmental risk management. A case study on environmental data measured in the southern part of Apulia region (South of Italy), called Grande Salento, is discussed. Sample data (concentrations of PM_{10}, wind speed, temperature) taken at different air monitoring stations are used for stochastic prediction, through space-time indicator kriging.

Keywords: GIS, Geostatistics, PM_{10} pollution, space-time indicator kriging

1 Introduction

Environmental risk management involves the integrated use of several tools and techniques, including GIS, sample design, Geostatistics and data management. In particular, data management process requires the integration of several data divided into three categories: i) environmental data (land use, land cover, vegetation, geology, meteorology and measures of pollutants concentration); ii) socio-economic data (population and housing census data, community vulnerability data and data on utilities and access); iii) political and institutional data (Chen et al., 2003). Moreover, a spatial-temporal approach is often required for environmental risk assessment; hence, the interaction between space-time modeling of air pollution, adopted by the statistical community in environmental studies (De Iaco et. al, 2001; Kolovos et al., 2004; Spadavecchia and Williams, 2009, among others), and urban environment representation (traffic network, location of industrial facilities, emission sources and topographic conditions), easily managed in a GIS, is necessary. The aim of this paper is to combine the use of space-time geostatistical techniques and the GIS potential. A case study on an environmental data set, involving both atmospheric variables and air pollutant concentrations, measured in November 2009 at monitoring stations located in Grande Salento (Lecce, Brindisi and Taranto districts in the Apulia Region) is discussed. In particular, air pollution due to PM_{10} (Particulate matter) concentrations and atmospherical variables, such as wind speed and

[1]Supported by Fondazione Cassa di Risparmio di Puglia.

temperature in the same region, are considered. Exploratory Spatial Data Analysis for a deep understanding of the analyzed phenomenon is performed using the Geostatistical Analyst Tool of ArcGis. Structural analysis for space-time variogram estimating and modeling and space-time prediction, based on kriging, is computed by using modified *Gslib* routines. A $3D$ representation for the space-time evolution of the conditional probability associated with PM_{10} is produced by using *ArcScene* (an extension of ArcGis). The overlay between the probability map and relevant urban spatial data is shown for Brindisi Municipality.

2 Empirical framework and methods

The study of the evolution of PM_{10} is very important for the effects that this pollutant has on human health. Many studies have shown that exposure to PM_{10} increases the risk of mortality both in long and short term. According to National Laws concerning the human health protection, PM_{10} hourly average concentrations cannot be greater than 50 μg/m^3 for more than 35 times per year. During the month under study, the PM_{10} hourly values exceeded the threshold 80 times, especially on the 13rd, 14th, 23rd and 24th of November. In the present case study, the following steps have been considered: (1) defining the space-time indicator variables according to appropriate thresholds, computed from the observed data; (2) modeling space-time indicator variogram of the variables by using the generalized product-sum variogram model; (3) using space-time indicator kriging, over the area of interest and during the period 1-6 December 2009, in order to obtain: a) the joint probability that PM_{10} concentrations exceed fixed thresholds and the atmopheric variables take values not greater than the corresponding monthly means, b) the joint probability that the atmopheric variables take values not greater than the corresponding monthly means; (4) computation and $3D$ representation of the probability that PM_{10} concentrations exceed the fixed thresholds, conditioned to adverse atmospheric conditions (i.e. wind speed and temperature which are lower than the corresponding monthly mean values). In Geostatistics, observations are modelled as a partial realization of a spatio-temporal random function Z, which is decomposed into a sum of a trend component and a stochastic residual component. In the following case study, the formalism of a spatio-temporal indicator random function ($STIRF$),

$$I(\mathbf{u}, z) = \begin{cases} 1 & \text{in case of } Z \text{ not greater (or not less) than the threshold } z, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{u} = (\mathbf{s}, t) \in D \times T, z \in \mathbb{R}$ ($D \subseteq \mathbb{R}^2$ and $T \subseteq \mathbb{R}_+$), is considered. Spatio-temporal dependence of a $STIRF$ is characterized by the indicator variogram of I: $2\gamma_{ST}(\mathbf{h}) = \mathrm{Var}[Y(\mathbf{s}+\mathbf{h}_s, t+h_t) - Y(\mathbf{s}, t)]$, which depends solely on the lag vector $\mathbf{h} = (\mathbf{h}_s, h_t)$, $(\mathbf{s}, \mathbf{s}+\mathbf{h}_s) \in D^2$ and $(t, t+h_t) \in T^2$. The fitted model for γ_{ST} must satisfy an admissibility condition in order to be valid. Hence, the following generalized product-sum model (De Iaco et al. 2001) has been fitted to the empirical indicator space-time variograms:

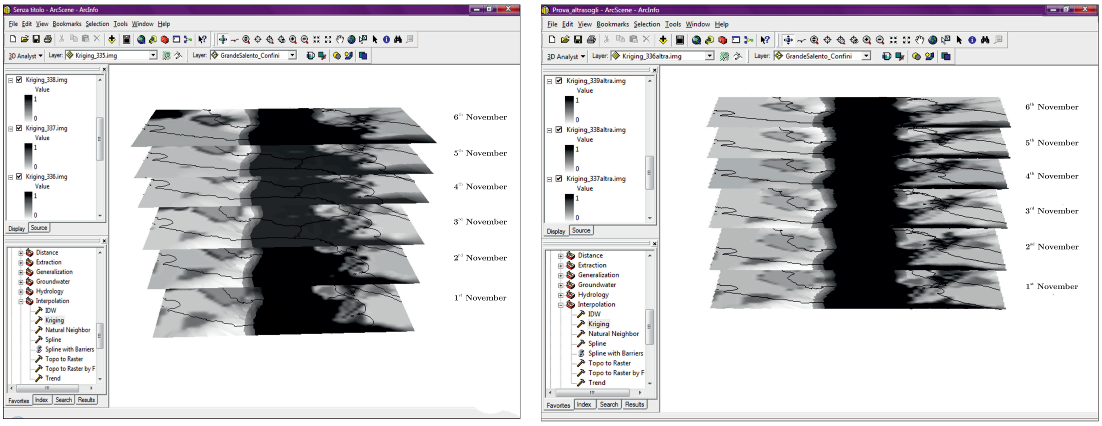
$$\gamma_{ST}(\mathbf{h}_s, h_t) = \gamma_{ST}(\mathbf{h}_s, 0) + \gamma_{ST}(\mathbf{0}, h_t) - k\gamma_{ST}(\mathbf{h}_s, 0)\gamma_{ST}(\mathbf{0}, h_t), \quad (1)$$

where $\gamma_{ST}(\mathbf{h}_s, 0)$ and $\gamma_{ST}(\mathbf{0}, h_t)$ are valid spatial and temporal bounded marginal variograms and $k \in]0, 1/max\{sill\gamma_{ST}(\mathbf{h}_s, 0), sill\gamma_{ST}(\mathbf{0}, h_t)\}]$. Basic theoretical results can be found in De Iaco et al. (2001), moreover recently it was shown that strict conditional negative definiteness of both marginals is a necessary as well as a sufficient condition for the product-sum (1) to be strictly conditionally negative definite (De Iaco et al., 2011).

3 Case study

In this analysis, the $STIRF$s associated with the spatial-temporal distributions of PM_{10}, as well as of temperature and wind speed, have been examined in the Grande Salento region during November 2009. The data set consists of daily averages of 3 variables, PM_{10}, temperature and wind speed, measured in November 2009 at 28 monitoring stations located in the Grande Salento.
After computing descriptive statistics, spatial-temporal indicator kriging using the generalized product-sum variogram model has been applied in order to predict, over the area of interest and for the period 1-6 December 2009, the probability that PM_{10} concentrations exceed the fixed limits, in the presence of adverse atmospherical conditions to the pollutant dispersion, i.e. temperature (T) and wind speed (WS), which are lower than the corresponding monthly mean values (12.54 °C and 2.11 *meters/second*, respectively). In this case study, the thresholds for the PM_{10} have been fixed equal to the 75th and 80th percentiles of samples data (37.804 and 40.57 $\mu g/m^3$, respectively), which can be considered critical with respect to the law limit. Hence, 3 indicator random fields have been defined: $I_1(\mathbf{u}; 37.804, 12.53, 2.11) = 1$, if $PM_{10} \geq 37.804$, $T \leq 12.53$, $WS \leq 2.11$, 0 otherwise, $I_2(\mathbf{u}; 40.57, 12.53, 2.11) = 1$, if $PM_{10} \geq 40.57$, $T \leq 12.53$, $WS \leq 2.11$, 0 otherwise, $I_3(\mathbf{u}; 12.53, 2.11) = 1$, if $T \leq 12.53$, $WS \leq 2.11$, 0 otherwise, with $\mathbf{u} \in D$.
Indicator sample space-time variograms for the indicator variables under study and their models have been determined first. The fitted space-time variogram model for the random fields I_1 is characterized by: $\gamma_{ST}(\mathbf{h}_s, 0) = 0.066\,[1 - \exp(-3\,\mathbf{h}_s/15000)]$, $\gamma_{ST}(\mathbf{0}, h_t) = 0.185\,[1 - \exp(-3\,h_t/6)]$, $k = 3.767$ and global sill equal to 0.205;
for I_2: $\gamma_{ST}(\mathbf{h}_s, 0) = 0.059\,[1-\exp(-3\,\mathbf{h}_s/15000)]$, $\gamma_{ST}(\mathbf{0}, h_t) = 0.169\,[1-\exp(-3\,h_t/6)]$, $k = 4.112$ and global sill equal to 0.187;
for I_3: $\gamma_{ST}(\mathbf{h}_s, 0) = 0.094\,[1-\exp(-3\,\mathbf{h}_s/20000)]$, $\gamma_{ST}(\mathbf{0}, h_t) = 0.235\,[1-\exp(-3\,h_t/6)]$, $k = 3.712$ and global sill equal to 0.247. Probability maps have been predicted over the area of interest for the period 1-6 December 2009. In particular, the indicator kriging has been used to estimate the joint probability that PM_{10} concentrations exceed fixed thresholds and the atmopheric variables take values not greater than the corresponding monthly means first, and secondly the joint probability that the atmopheric variables take values not greater than the corresponding monthly means.

Then, the probabilities that PM_{10} values do not exceed the fixed thresholds, conditioned to adverse atmospheric conditions, over the area of interest and during the period 1-6 December 2009, have been computed. From the obtained results it is evident that, in Brindisi and Lecce districts, the probability that PM_{10} daily concentrations exceed the fixed thresholds, under adverse atmospherical conditions, decreases from the 1st to the 6th of December 2009 and along the NorthWest-SouthEast direction (Fig. 1). Finally, the Brindisi Municipality is considered in



a) **b)**

Figure 1: conditional probability maps of PM_{10} concentrations, for the thresholds: a) 37.804 $\mu g/m^3$ (75th percentile), b) 40.57 $\mu g/m^3$ (80th percentile), during the period 1-6 December 2009.

detail. In this area, the concentrations of PM_{10} is compared with land use and traffic network. The probability that PM_{10} concentrations do not exceed the fixed threshold is higher in the city center; on the other hand it is much more likelihood that the PM_{10} concentrations exceed the limit in the SouthWest hinterland.

References

Chen K., Blong R., Jacobson C., (2003) Towards an integrated approach to natural hazard risk assessment using GIS: with reference to bushfires, *Environmental Management*, 31, 546-560.

De Iaco, S., Myers, D.E., Posa, D. (2001) Space-time analysis using a general product-sum model, *Statistics and Probability Letters*, 52, 1, 21-28.

De Iaco, S., Myers, D.E., Posa, D. (2011) On strict positive definiteness of product and product-sum covariance models, *Journal of Statistical Planning and Inference* 141, 1132-1140.

Kolovos A., Christakos G., Hristopulos D.T., Serre M.L., (2004) Methods for generating non-separable covariance models with potential environmental applications, *Advances in water resources*, 27, 815-830.

Spadavecchia L., Williams M., (2009) Can spatio-temporal geostatistical methods improve high resolution regionalisation of meteorological variables?, *Agricultural and Forest Meteorology*, 149, 6-7, 1105-1117.