Methodological study on pesticides in Alsatian groundwater

Fausta Musci, Concetta I. Giasi
Department of water engeneering and of Chemistry - Politecnico di Bari
Via Orabona 4, 70100 Bari, ITALY
f_musci@hotmail.com

Chantal de Fouquet
Centre de Géostatistique, Ecole des Mines de Paris, 77305, Fontainebleau, France

Abstract: The risk assessment conducted by many federal and state agencies have generally relied on deterministic approaches, that use single input/output values, generally selected to fulfill the goal of being health-protective. But, the presence of uncertainty and variability within the parameters of the procedure of risk assessment let them assume different values within a range of possible values, each with different probability of occurrence. In particular, the case study deals with groundwater contamination by agrochemical substances occurring in the French aquifer of Alsace. The regional supply of drinking water, water for irrigation and industrial water depends mainly on this water resource. A proper management of this area must consider, thus, the sustainability of a landscape capable of multiple uses and the overwhelming presence of censored data. For this reason, particular attention is given to the characterization of the extent and the chemico-physical distribution of the pollutant source for what concern the delimitation of the hazardous areas, to the determination of the probability density functions of the concerned variables and of the representative concentrations..

Keywords: groundwater contamination, geostatistics, estimation, non-linear methods.

1. Introduction

An instrument of high political and social importance is the risk assessment, or the evaluation of the risk associated with any event that can negatively affect the human health or the environment. Thus, the environmental impacts must be anticipated and prevented before they really happen and risk assessment has the logical structure to do it. The most immediate approach is therefore deterministic: by assigning to each of the input variables a single value, it gets a punctual value of risk. Every single value is generally selected to be reasonably certain that risk is not underestimated and to err on the side of overestimating risk. But, the presence of uncertainty and variability within the parameters of the procedure of risk assessment makes them actually random variables, as they are parameters that can assume different values within a range of possible values, each with different probability of occurrence. Therefore, these parameters can only be considered through a stochastic approach. In order to describe natural phenomena correlated in space and time and to quantify the uncertainty of the estimations of these phenomena carried on from a sampling generally very fragmentary, this work refers to the theory of the regionalized variables. It was developed by Matheron (1965) and then popularized by many others. In particular, the case study addresses the various method of linear and non-linear geostatistics for characterizing the exposure concentration through the inference of spatial structure from spot samples. Moreover, the overwhelming presence of censored data needs several statistical methods to be assessed, implemented and applied in order to characterize both variability and uncertainty of the exposure, effects and risk assessment.

2. Materials and Methods

2.1 Case study

In the Rhine valley, the alluvial formations create a large aquifer, one of the largest reservoirs of drinking water in Europe. In the Alsatian part, this reservoir has the order of 45 billion m^3 of water for an area of 2800 km^2. The shallow depth of the groundwater makes its exploitation easy, which is an economic advantage. In fact, the groundwater provides three quarters of the drinking water needs of the population and more than half of the industrial and agricultural water needs. But besides this, the lack of protective geological cover and the shallowness of the aquifer make it particularly vulnerable to contamination due to human activities. And so, pesticides, as Atrazine, have been detected in the Alsatian groundwater.

Atrazine is an herbicide of the triazine chemical family, with radical absorption. It has been used in France on the cultures of corn since 1962, thus its use was prohibited by the 30 September 2003. Because it does not absorb strongly to soil particles (K_{oc} = 100 g/ml) and it has a lengthy soil half-life (60 to 100 days), it is expected to have a high potential for groundwater contamination, even though it is only moderately soluble in water (33 µg/ml). The Drinking Water Directive (DWD), Council Directive 98/83/EC, defines the sanitary thresholds (0.1 µg/l) for the concentration of these contaminants in drinking water.

The chosen data set is composed by four months of measurement: September 2002, March 2003, September 2003 and March 2004. This choice is based on the available samples (September 2003 is largely sampled – heterotopic case), the continuity of information available in time for each station, as well as the significance from the hydrological point of view. In fact, these months represent the beginning and the end of the recharge period of the aquifer. The period is also in correspondence of the interdiction of Atrazine's use in France.

2.2 Methodology of analysis

Geostatistics is based on the study of the spatial behaviour of variables. Even the concept of variable is converted in its spatial context as the regionalized variable [Matheron, 1965]. The model of the regionalized random variable is the basis principle of such kind of science.

The proposed procedure carries on through a series of steps, which will be deliberately presented in a synthetic and intuitive manner. For further details it is possible to refer to Matheron (1965, 1970), Chilès & Delfiner (1999), Rivoirard (1995) and Chauvet (1999).

1st. Exploratory data analysis. It refers to a statistical study of the data sets, for getting a first idea about data, their distribution, significance and consistency.

2nd. Structural analysis. It concerns all the methodologies aimed to investigate the spatial structure of data and exploit it to build reasonable spatial models. A synthetic form for explaining the structural variability of data is the experimental variogram. By fitting a continuous mathematical function on raw variogram it is possible to exploit such powerful instrument in order to model the variability structure for the whole spatial domain (not only on the measured points) [Isaaks and Srivastava, 1989].

3rd. Validation of a structural model. In practice it is important to evaluate the performance of fitting a variogram model.

4th. Local estimation. It allows passing from a discrete information to a continuous description of the phenomenon. The geostatistical estimator used for the estimation process is called kriging. For each target point, the linear estimator $Z^*(x_0)$ is expressed as the linear combination of the known points $Z(x_i)$, together with the conditions of unbiasedness and of minimization of the error variance.

5th. Multivariate aspects. Several regionalized variables could be treated together, so it is possible to enjoy also joint information that would increase the degree of accuracy of the results. The conjoint spatial structure of the variables is described by their cross-variogram and coregionalization models are used, between which the linear model of coregionalization [Journel & Huijbregts, 1978] is the simplest one. Thus the estimation is performed by cokriging.

6th. Non linear methods. This work refers to two principal methods: the probability from conditional expectation and the indicator cokriging. The first approach is of parametric type and is based on the "conditional expectation" estimator. It requires that the variable is multigaussian;

thus, first of all, a transformation of the original variable *Z(x)* – called anamorphosis ¬ is necessary for obtaining a random function *Y(x)* with gaussian distribution. It can be shown [Goovaerts, 1997] that the conditional distribution of *Y(x)* is Gaussian-shaped, with mean equal to its simple kriging $Y(x)^{SK}$ from the available data and variance equal to the simple kriging variance $\sigma^2_{SK}(x)$. Therefore, the posterior or conditional cumulative distribution function (in short, ccdf) at location x is

$$\forall\, y \in \mathbf{R},\, F(x; y \mid data) = G\left(\frac{y - Y(x)^{SK}}{\sigma_{SK}(x)}\right) \qquad [1]$$

where G(.) is the standard Gaussian cdf. Its implementation relies on an assumption of strict stationarity and knowledge of the prior mean m, in order to express $Y(x)^{SK}$ [Emery, 2006]. In the second approach, of non parametric type, the exceedance or not of a given threshold s (a concentration risk, for instance) at a point x can be coded by the indicator variable. Therefore, a cokriging estimation of the indicators could be performed, in order to consider simultaneously the indicator variables associated to different thresholds.

3. Results

Univariate linear geostatistical techniques have allowed obtaining estimation maps of Atrazine. This was a preliminary study of the data, which took account of the data set as measured, so without any kind of transformation, despite the highly asymmetric and discontinuous variables. In this case, the undefined values were exactly considered equal to their instrument detection limit (IDL). Same assumption has been made for the estimation in multivariate conditions, where variables are treated together, thanks to their significant correlations.

But kriging and its extensions provide what might be called, by abuse of language, the most probable value of the pollutant concentration at any point in space, combined with the variance of the error. This has two consequences. The first is that the map erases the "peaks" and "hollows" of pollution and is "attracted" by the average pollution on the area of interest: the real variability in the space of the pollution is not reproduced when the data are interpolated (smoothing property). The second consequence is that the complete distribution of the error is not accessible: just the mean (zero by construction) and the variance are known [Deraisme et al., 2003]. Therefore, these maps provide only an image more or less accurate of the reality. While, the comparison to a regulatory threshold needs to take into account the estimation error in order to reproduce the spatial variability. This is the object of non-linear methods. The proposed approaches reflect both the conditional expectation and indicator cokriging. While this second method can solve the uncertainty due to censored data, because all values are encoded in a binary variable [0,1], according to a certain threshold value bigger than the IDL, the first method is a bit more complex to implement. Needing a multi-Gaussian distribution, firstly it requires a parametric approach, performing thus a normalizing transformation of the strongly asymmetrical original data. These transformed variables must be thus multigaussian, that is to say every linear combination of the gaussian values should follow a gaussian distribution. In practice, the multigaussian hypothesis cannot be fully validated because, in general, the inference of multiple-point statistics is beyond reach. Usually, only the univariate and bivariate distributions are examined [Goovaerts, 1997]. However, the uncertainty of censored data persists, even in the Gaussian field, making the obtained transformed distribution inaccessible, and virtually impossible to analyze. The study then solves this "inaccessibility" through the use of indicator variables associated with different thresholds of the Gaussian transformed. In fact this allows, on the one hand, testing the bivariate normality (instead of the multivariate) of the obtained variables and, on the other hand, having a model to use in the estimation phase. Moreover, again because of this "inaccessibility" of the obtained transformation, the mean of the distribution is unknown, therefore, an approach via ordinary multigaussian kriging is preferred to using the simple kriging. In figure 1 the maps of atrazine obtained by linear estimation methods and non-linear methods are reported, just for September 2003: tendentially they identify the same contaminated areas.

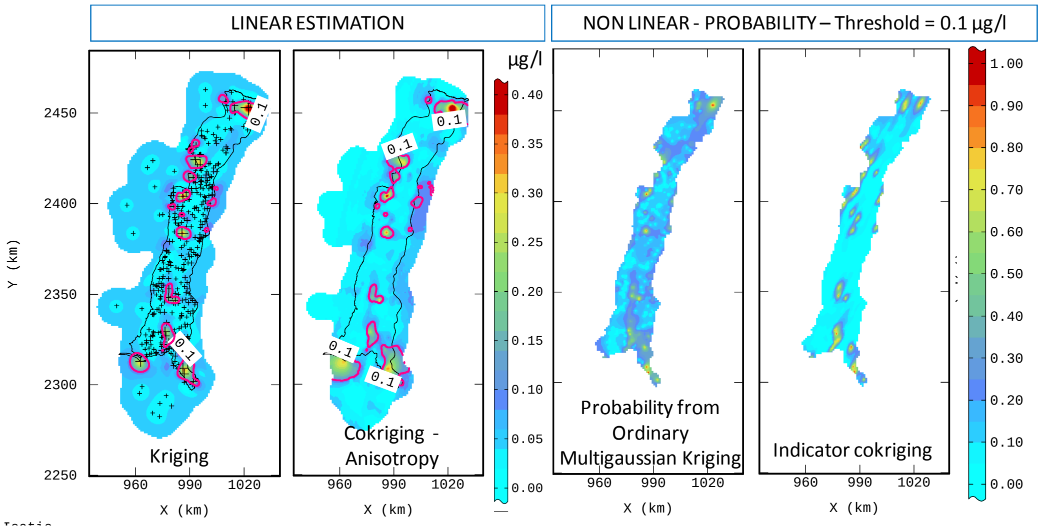

Fig. 1 Maps of atrazine obtained by estimation methods and non-linear methods – Sept 2003.

4. Concluding remarks

The proposed validation, through an ad hoc method of cross-validation, provided as result that the obtained probability by indicator cokriging is closer to the original data. The explanation for this result is that, probably, making a hypothesis of bivariate normality on the highly asymmetric and discrete available data sets is not unimportant in the estimation phase. Finally the risk results well characterized, also in function of the several assumptions and checks made during the analysis, and allows making considerations in terms of potential areas to be remediated and population potentially exposed to a hazard. Thus, the performed study is able to take into account uncertainty and variability related to the distribution of pesticides in groundwater in characterizing the scenario of contamination in the process of risk assessment. Moreover, the sensitivity analysis has allowed proceeding step by step in the study of the contamination by atrazine, considering limitations and advantages of the geostatistical methods, linear and non-linear. Finally, most of the methodologies presented in this study are also applicable in other field, as soil or air contamination.

References

Chauvet, P. (1999). «Aide-mémoire de gèostatistique linéaire». *Les Presses de l'Ecole des Mines*.

Chilès, J.P., Delfiner, P., (1999). «Geostatistics: Modeling Spatial Uncertainty». Wiley, New York.

Deraisme J., Bobbia M. (2003). « L'apport de la géostatistique à l'étude des risques liés à la pollution atmosphérique ». *Environnement, Risques & Santé* – vol.2, n°3, mai 2003.

Emery, X., (2006). «Ordinary multigaussian kriging for mapping conditional probabilities of soil properties». *Geoderma* 132 (1–2), 75–88.

Goovaerts, P., (1997). «Geostatistics for Natural Resources Evaluation». *Oxford University Press*, New York.

Isaaks, E.H. and Srivastava R.M. (1989). «An introduction to applied geostatistics». *Oxford University Press.*

Journel, A.G, Huijbregts, C.J. (1978). «Mining Geostatistics». *Academic Press*, London, 600 pp.

Matheron, G. (1965). «Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature». Masson, Paris.

Rivoirard, J. (1995). «Concepts et méthodes de la géostatistique». *Cours C-158*, Centre de Géostatistique, Ecole des Mines de Paris.