

Contribution of Geostatistics to the Study of Risks Related to Air Pollution

Jacques Deraisme¹, Michel Bobbia²
and Chantal de Fouquet³

¹*Geovariances*

²*Air Normand*

³*Mines Paristech
France*

1. Introduction

The diffusion of a pollutant in a geographical space, is a phenomenon whose main characteristics are the following: it is driven by complex physical and chemical laws influenced by local environmental parameters, it spreads out in a three dimensional space and, finally it is of dynamic nature. Two main approaches may be proposed to allow a spatial estimation of population exposure: (1) a “deterministic” approach which transforms estimated emissions derived from inventories into atmospheric concentrations (Hauglustaine (1998), and (2) a “probabilistic” approach which consists of spatializing concentrations obtained from an air quality monitoring network. Predicting the pollutant concentration at a given point in space at a given time and assigning to that value an indication of its reliability is out of the scope of entirely deterministic models. By choosing a probabilistic framework geostatistics opens the way to models that can be characterized by actual observations for providing practical answers to a large variety of questions specifically related to risk analysis. Typical questions addressed by this approach are: “what is the probable value of the pollutant concentration at a specific location without measurement?”, “what is the probability that the admissible pollution level is reached at the same point?”. We must pay attention that unlike deterministic modelling, a sufficient number of measures is required to ensure reliable mapping.

This framework aims at considering the phenomenon under observation as an outcome of a random process (Matheron,1989). This process is represented by a random function $Z(x, y, z, t)$ where x, y and z are the coordinates in a geographical three dimensional space and t is the time axis. For simplicity we will illustrate the intention of this article by reducing the work space to two dimensions, hence the random function is now noted $Z(x,y)$. By reintroducing the third dimension of the space and especially the temporal component, new difficulties arise for specifying completely the models from actual data. These will be discussed later.

Geostatistics generally proposes that the whole set of possible random functions is restricted to stationary random functions, particularly regarding their mean and variance values.

Even with this important restriction the geostatistical framework is richer than other possible models, like, for instance, the purely statistical models which suppose that

measures taken at two different points are not correlated. Clearly the pollution at two different location is likely to be more similar when these points are close than at a larger distance. Characterizing that correlation is at the centre of the geostatistical approach.

The methods developed under the geostatistical framework are many and depend on the nature of the question. Simple issues can be easily solved by linear geostatistical methods. When the questions are more complex, particularly when dealing with risk analysis, non linear geostatistical methods are applicable.

The outline of the chapter is:

- reminders of the most important concepts underlying the geostatistical methods.
- necessity of non linear methods
- discussion on what is achievable to estimate without resorting to simulations
- example of the estimation of the population exposed to a pollution
- example dealing with a temporal issue

2. Reminders on the models in geostatistics

Characterising a spatial distribution in the general case is complex, because of the numerous parameters that would be necessary to infer from the available measures. In order to get a map of a property measured at some points over a given area, geostatistics proposes a simple method called kriging¹ named after Dr Krige, who first pointed out the drawbacks of traditional estimation methods of gold deposits in South Africa. The method aims to interpolate the value at any point by assigning weights depending on the correlation with the data measures (Chilès, 1999). The correlation is quantified by means of the variogram² that can be estimated from bi-point statistics of the data (Cressie, 1998; Goovaerts, 1997). By introducing other measured parameters or factors that may have an influence on the parameter of interest the result of the interpolation can be obtained as a linear combination of all parameters at data point referred to as cokriging (Jeannée, 2006). Like any linear interpolator, kriging fails in providing more information related to confidence levels or risk. In order to go ahead in that direction non linear geostatistics based on Gaussian models has been developed. The particular case of data whose distribution is Gaussian is intensively used in statistics as well as in geostatistics for the simple reason that the density of probability only depends on two parameters, i.e. the mean and the variance. Introducing the idea of Gaussian spatial distribution, the additional knowledge of the variogram will entirely characterise the probabilities at several points simultaneously. When the distribution of values is not Gaussian, which is the case most of the time, we come back to the Gaussian case by means of a transform of the data measures, by a process called normal score transform. The variable of interest $Z(x)$, x designating the location of a point in the space, can be written as the transform by the Gaussian anamorphosis function, noted φ , of a random function $Y(x)$, that has a Gaussian spatial law centred and normalized: $Z(x) = \varphi(Y(x))$.

¹Kriging is a linear combination of data with weights chosen for minimizing the variance of the estimation error.

² The variogram represents the variance of the difference between two measures as a function of the distance between these measures.

When the anamorphosis function φ is bijective, hence strictly increasing on its domain of definition, the knowledge of the measures $Z_\alpha = Z(x_\alpha)$ or of their normal score transforms $Y_\alpha = \varphi^{-1}(Z_\alpha)$ is equivalent. At all locations x , the a priori probability to overcome a threshold « s » can be easily calculated on the gaussian transforms by using the classical cumulated gaussian density function noted G. This can be written :

$$\begin{aligned} P(Z(x) \geq s) &= P(\varphi(Y(x)) \geq s) \\ &= P(Y(x) \geq \varphi^{-1}(s)) \\ &= 1 - G(\varphi^{-1}(s)) \end{aligned}$$

The probabilistic model gives also access to statistical quantities like quantiles. The order ω quantile of a variable is the value of the random variable for the cumulated probability ω . Thus that quantile of the normal variable $Y(x)$, noted q^ω is $P(Y(x) < q^\omega) = \omega$. Since the anamorphosis φ is increasing, the order ω quantile of $Z(x)$ can be directly deduced from the Gaussian quantile of the same order.

In an analogous way the confidence interval for the statistical risk ε of a random variable $W(x)$ is defined as $P(q^{\varepsilon/2} \leq W(x) < q^{1-\varepsilon/2}) = \varepsilon$. For a normal variable that interval is symmetrical: $q^{\varepsilon/2} = -q^{1-\varepsilon/2}$. For instance the confidence interval at the risk level 5% is defined by difference between the quantiles at the order 97.5% and 2.5%. For a gaussian variable those quantiles are equal to the mean value plus or minus 1.96 standard deviations. In practice we will be interested to the distribution law of a Random Function in a given point, conditioned by the available measures at the experimental stations. For a Random Function of Gaussian spatial law, that conditional distribution is simply:

$$Y^*(x) + \sigma_K(x).W(x)$$

$Y^*(x)$ is for the result of a kriging with zero mean of $Y(x)$ by the data Y_α , $\sigma_K(x)$ is for the standard deviation of the associated error and W is a Random Function independent of Y^* . In that case the conditional law of the anamorphosed variable Z of Y is directly deduced of that of the Gaussian transform, i.e. $\varphi(Y^*(x) + \sigma_K(x).W(x))$. The probability of overcoming a threshold is then:

$$\begin{aligned} P(Z(x) \geq s | Z_\alpha) &= P\left(W(x) > \frac{\varphi^{-1}(s) - Y^*(x)}{\sigma_K(x)}\right) \\ &= 1 - G\left(\frac{\varphi^{-1}(s) - Y^*(x)}{\sigma_K(x)}\right) \end{aligned} \tag{1}$$

In absence of any measurement error, at any data location α we have $Y^*(x_\alpha) = Y_\alpha$, and the kriging variance is null. That probability is then either 0 or 1, depending on whether $Y_\alpha = \varphi^{-1}(Z_\alpha)$ is inferior or superior to $\varphi^{-1}(s)$. Far from the experimental data, $Y^*(x)$ tends towards the a priori expectation, equal to 0 and $\sigma_K(x)$ tends towards the a priori variance, equal to 1. We then come back to the a priori probability $1 - G(\varphi^{-1}(s))$. If we replace in the calculations the kriging with known mean (also called simple kriging) by the kriging with unknown mean (also called ordinary kriging) the results will be in line with the local mean

of the concentrations. In the case of urban pollution (by NO_2 in the example presented hereafter) it allows to account for the higher pollution at the town centre than at the periphery.

In an analogous way, the confidence interval at the statistical risk level ε of the actual concentration $Z(x)$ is deduced from the confidence interval of the gaussian transform $\left[\varphi(Y^*(x) - \sigma_K(x).q^{1-\varepsilon/2}), \varphi(Y^*(x) + \sigma_K(x).q^{1-\varepsilon/2}) \right]$. Because of the non linearity of the anamorphosis φ , that interval is no more symmetrical to the conditional expectation of Z .

3. Pros and cons of linear estimation methods

Let's take the example of the annual mean concentration in NO_2 in the city of Rouen (France). If we imagine we had a perfect knowledge of the actual pollution, the map would resemble that of figure 1 (which is in fact a simulation), showing the high spatial variability of NO_2 from one point to another.

The available data consist of samples measured on passive diffusive tubes (Roth, 2001), that have been located in the Rouen city and suburbs during the year 2000 for 6 fortnights in total. These measures have been made on 89 sites spread over an area of 300 km^2 . On the maps (Fig.1, 2, 5, 6 and 7) only the central area of $9 \times 9 \text{ km}^2$ with 59 sites equipped with diffusive tubes is represented. Public health regulations require experts to evaluate the probability to overcome the annual threshold of $40 \mu\text{g}/\text{m}^3$ as well as the number of inhabitants potentially exposed to excessive pollutant concentrations. We will assume that the threshold applies to the concentration at any point of the space and that the measuring period is large enough to neglect the temporal term of the estimation error.

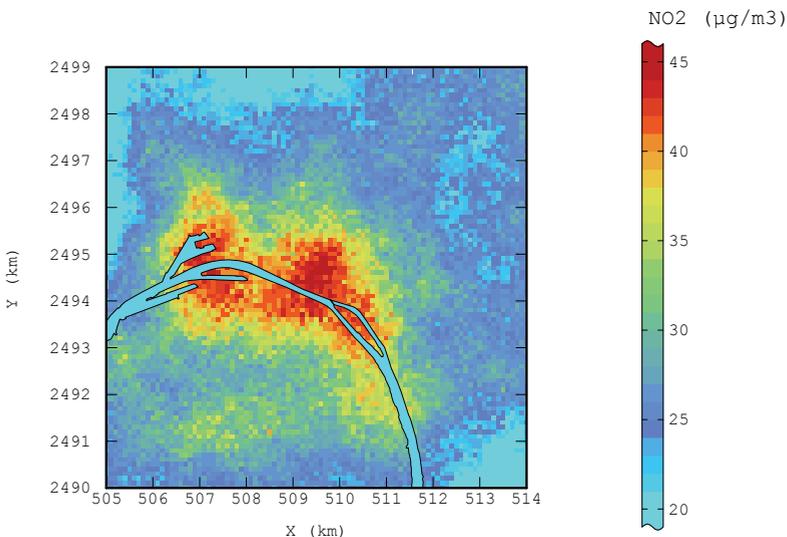


Fig. 1. Map of the true annual mean concentration of NO_2 at the centre of the Rouen city.

A first solution would be to predict the pollution concentrations at unsampled locations. Among different interpolators, the techniques based on kriging provide the best linear

interpolator in the sense that they minimise the variance of the estimation error. To be specific we have used in this example a co-kriging (extension of kriging for processing several correlated variables) with a co-factor combining information related to pollution sources and the spatial distribution of the inhabitants. The resulting map (Fig. 2.) provides a simplified image of the real phenomenon. If it reproduces the main tendencies of the pollution phenomenon, it fails in describing the variations of the NO_2 concentration at a scale of a few hundred metres, i.e. inferior to the spacing between the measures (1km).

By looking at both maps (Fig. 1 and 2) it is immediately visible that applying a threshold (here $40\mu\text{g}/\text{m}^3$) on the estimated concentrations would lead to an under-estimation of the "contaminated surface" with respect to the right surface obtained from the threshold applied on the real, but unknown, values. The bias cannot be neglected: the contaminated area is estimated as 352 ha whereas its real surface equals 569 ha. The underestimation is observed because the threshold is relatively high with respect to the average level of the measured concentrations. At the contrary for a low threshold, we would have observed an over-estimation. The figure 3 shows that the histogram of the estimated concentrations is more squeezed around the average values, that is known as the smoothing effect of kriging, and that the bimodal distribution is not kept on the estimated values.

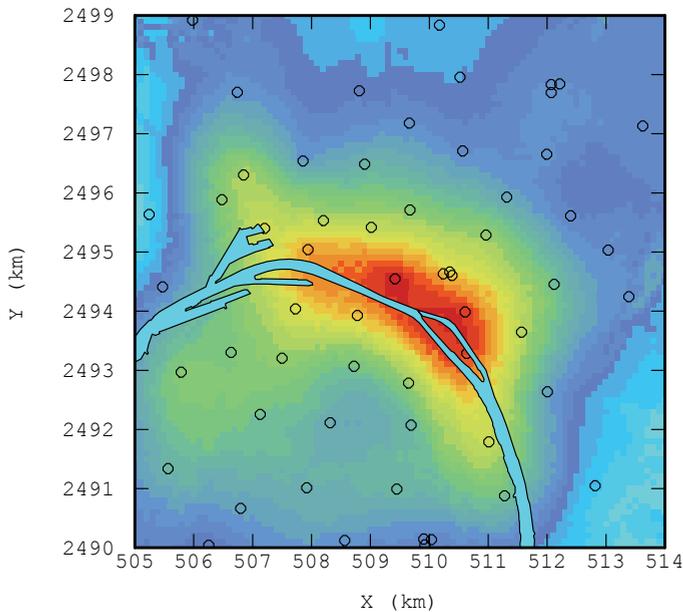


Fig. 2. Map of the co-kriging of annual mean concentration of NO_2 at the centre of Rouen.

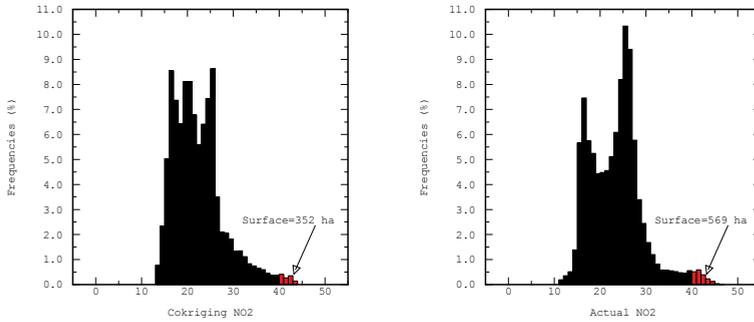


Fig. 3. Histograms of the average annual concentrations of NO_2 at the centre of Rouen for the estimated and actual concentrations and thresholding at $40\mu\text{g}/\text{m}^3$.

The scatter diagram (Fig. 4) between the “reality” and its estimation illustrates the presence of an error due to the interpolation procedure. This error can only be diminished by adding information, i.e. other measures. The uncertainty resulting from that error can be represented by the notion of statistical risk.

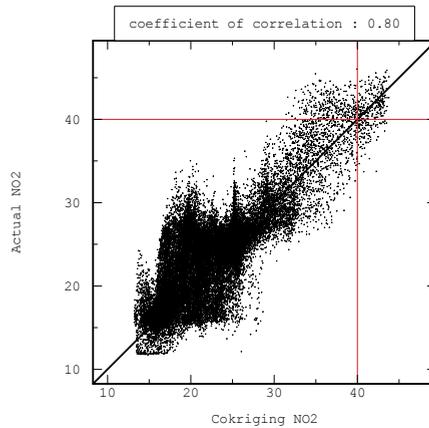


Fig. 4. Scatter diagram of the actual NO_2 concentration and its estimate at the centre of Rouen.

The methods derived from kriging are by nature inadequate for reproducing the actual variability, because the kriging methods aim to minimise the difference between the actual value and its interpolation: the interpolated values by kriging include necessarily less extreme values and more intermediate values, conveying the so-called smoothing effect of kriging. What is at stake is the confusion between the unknown map of the actual concentration and the available maps that are only estimates of the reality. Contrary to a topographical map (which can be considered as “exact”) every map of a pollutant calculated from a sampling will never be the reality, is it resulting from geostatistical methods or from physic-chemical models. Those maps will only provide an image more or less close to the reality.

4. Non linear estimation or simulations?

In real situations we will never have access to the actual phenomenon which is unique and largely unknown. The geostatistical framework proposes to consider that phenomenon as a realisation – among other equally possible – of a random phenomenon characterised by its spatial structure. In our example it amounts to say that the real phenomenon could resemble one of the representations of figure 5, analogous to the initial representation from a general point of view while differing notably on detailed local features.

Reality being unknown at most locations, we use a model of a random function, from which it is possible to derive statements correct in probability. We will favour the models that, after non linear operators (like applying a threshold) produce non biased results. The maps of figure 5 have been obtained by simulating outcomes of a Gaussian random function back-transformed by an anamorphosis function. The principle of geostatistical simulations consist in drawing outcomes of the random function, keeping the statistical distribution and the spatial structure. The conditional simulations will recreate the measured values at the experimental data points, i.e. is an exact interpolator. When many simulations are available, the probability of exceeding a threshold can be empirically calculated at every point by computing the proportion of the outcomes that exceed the threshold.

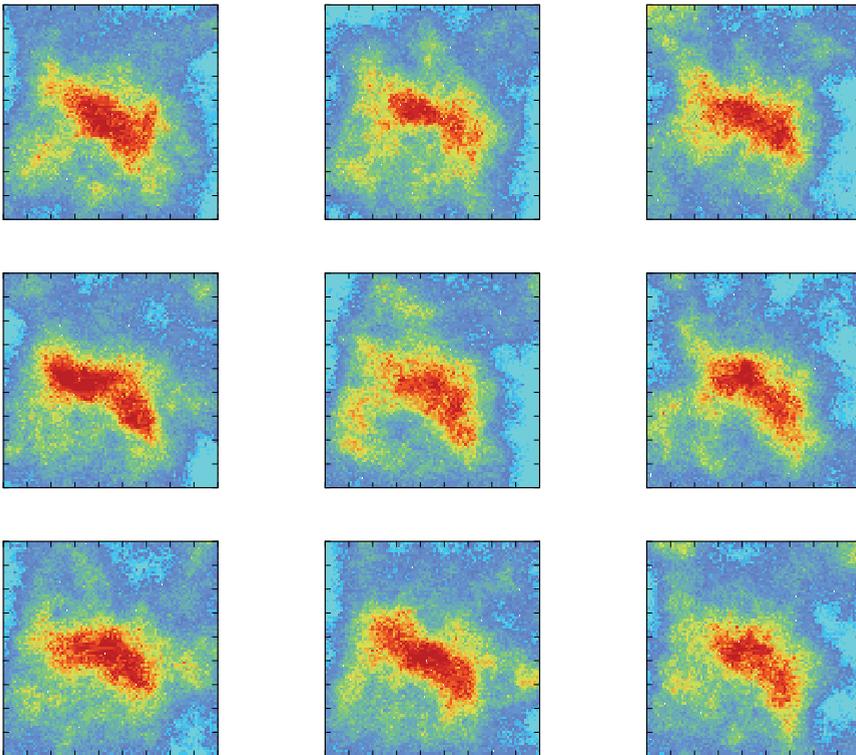


Fig. 5. Possible maps of the average annual concentration in NO₂ at the centre of Rouen (the legend is the same as that of Fig. 1).

Nevertheless in the framework of the model, the probability of exceeding any threshold can be directly calculated without resorting to simulations, by just applying the distribution law of the annual concentration Z conditional to the data according to equation (1). The advantage of this approach is that the calculation is analytical and straightforward.

In the example of the pollution by NO_2 in Rouen, the Gaussian model provides at any location the probability of exceeding the threshold of $40\mu\text{g}/\text{m}^3$. The studied area has a surface S of 81km^2 (i.e. 8100 ha). By multiplying the spatial mean of the probability by S we get an unbiased estimate of the surface probably exposed to the pollution. We get then a surface of 583 ha, close to the true surface of 569 ha (Fig.3.), instead of 352 ha calculated from the kriged map.

If the estimated concentration is less than the threshold, it does not mean that the risk of exceeding the threshold is nil, because of the estimation error. Conversely it may happen that the estimated concentration is above the threshold while the actual concentration is less. In practice the estimation error is unknown but the probabilistic model gives access to its mean (zero because of the unbiasedness condition applied in kriging) and its variance. The calculation of the probability of exceeding a threshold takes into account that variance or more precisely the estimation variance $\sigma_k^2(x)$ of the gaussian transform. Particularly this variance is increasing with the distance of the current location x to the measured points. On the figure 6 the iso-value lines of the probability of exceeding threshold are reported. The area where the estimated concentration is above the threshold corresponds to high values of the probability, even if it is not equal to 1, particularly in locations where the estimation is of poor quality in terms of precision. Conversely if the estimated concentration is less than the

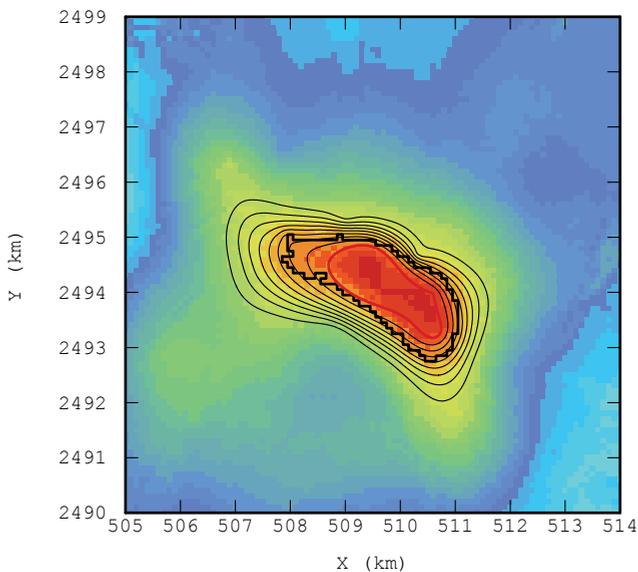


Fig. 6. Map of the cokriging of the annual mean concentration of NO_2 at the centre of Rouen. The curves represent the iso probability by steps of 0.1. Each point inside the red curve has a probability above 0.9 to exceed the threshold of $40\mu\text{g}/\text{m}^3$. The bold black outline surrounds the zone where the cokriging has given an estimate above the threshold.

threshold and the estimation is not precise the probability of exceeding threshold tends to increase. An intermediate probability reveals the uncertainty due to the estimation error, particularly when the estimated value is close to the threshold and the estimation is not precise. The map of the probability can be used to classify the different zones of the urban area with regard to the levels of risks and taking into account the uncertainty in the knowledge of the actual pollution. Such maps are a useful and can help in making decisions related to the protection of the population from pollution.

5. Evaluation of the population exposed to the pollution

More complex questions, like evaluating functions depending on several variables at several points simultaneously, require a simulations approach. Let's take as an example the case of the computation of statistics on the population exposed to a threshold. For sake of simplification we will not consider the movement of the population during the year.

The number of inhabitants b_i is supposed to be known on a fixed grid, the numbers concerned by the exceeding of threshold s is $B(s) = \sum_i b_i I_{Z(x_i) \geq s}$.

For the risk δ related to the local exceeding of the threshold, the exposed population is estimated by $B_\delta^*(s) = \sum_i b_i I_{P[Z(x_i) \geq s | Z_\alpha] \geq \delta}$. It can be calculated from the map of the population (Fig.7) and the map of the conditional probabilities (Fig.6) cut at δ : it is the sum of the inhabitants living in the cells for which the probability of exceeding threshold is above δ .

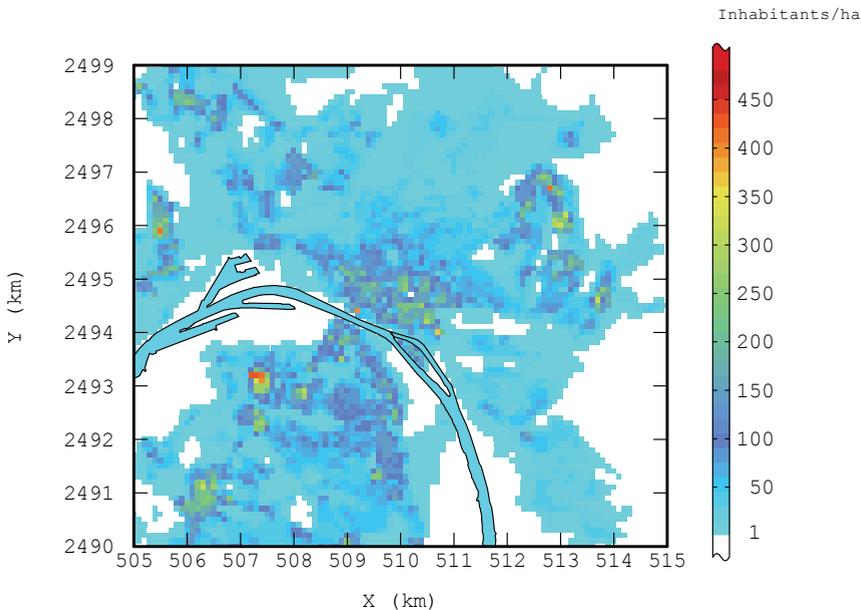


Fig. 7. Map of the density of the population living in Rouen urban area.

This calculation is based on adding up the population of each cell independently. This is a naïve way to proceed.

In fact the quantiles of the exposed population $B(s)$ require the spatial law of the concentration. They must be calculated by means of simulations (Deraisme, 2002) because it is mandatory to take into account the probability to exceed the threshold simultaneously in different locations. The spatial distribution of the exposed population is characterized by the simulation of the concentration on a regular grid. For a given simulation the exposed population is simply obtained by adding the inhabitants living in the cells where the simulated concentration exceeds the threshold of $40\mu\text{g}/\text{m}^3$. By repeating this computation on each of the 100 simulations we get a reliable representation of the distribution of the exposed population (Fig. 8). It should be noted that carrying out a large number of simulations is leading to computationally heavy.

For the statistical risk of 10%, the exposed population is estimated as 33 480 inhabitants: the probability that the exposed population is more than 33 480 inhabitants is 0.1. In other words, there is a 90% chance that 33 480 inhabitants are exposed to the excess of the threshold. Another way to say it is that this value represents the quantile at 90% of the exposed population.

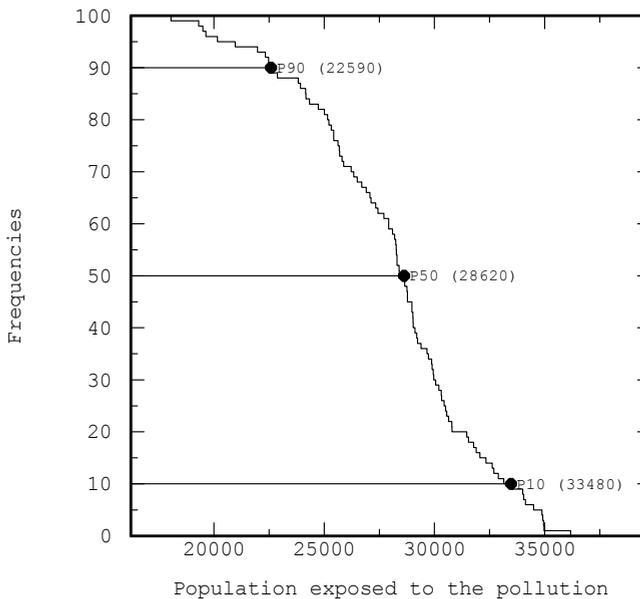


Fig. 8. Distribution of the population of Rouen urban area exposed to a pollution by NO_2 above the threshold of $40\mu\text{g}/\text{m}^3$ obtained from 100 simulations. The quantiles at 10%, 50% and 90% (noted respectively P90, P50 and P10) are reported.

The first “naïve” calculation, multiplying the population map with the probability map truncated at the risk of 10%, would have given 47 880 inhabitants, all living in zones where the probability of exceeding the threshold is above 0.9. The significant difference is due to the fact that, in the first case, the risk is applied to the quantity we want to evaluate, i.e. the

exposed population, while in the “naïve” way the statistical risk is applied on the NO_2 concentration independently of the population. The quantile of the exposed population depends on the probability to exceed the threshold together at different locations, which is ignored by the “naïve” calculation.

For more simple calculations, such as the average population exposed to the pollution, simulations are not necessary. We get the same result, i.e. 28150 inhabitants, by multiplying the population by the probability map (without any cut) and summing up the contribution of each cell. Averaging simulations and calculating directly the conditional expectation are equivalent.

6. Introducing the temporal factor

In the previous example only the spatial variability has been taken into account in the risk evaluation. Yet the variability in time is an important factor inherent to the definition of some official regulations.

In the following example we will examine the calculation of the quantile of daily mean values of SO_2 concentrations on the urban area of Le Havre. For that study, twenty four analysers gave daily records of the SO_2 concentration for a year. The aim is to identify the zones where the daily concentration exceeds the level of $250 \mu\text{g}/\text{m}^3$, prescribed by the regulations, for more than 4 days per year, i.e. 1% of the year. In other words we will look at the spatial distribution of the quantile at 99% of the daily SO_2 concentration.

The evaluation of a quantile of high order requires that the measures recorded every day are complete for all data locations and all day of the year. At least we must be certain that the missing values are inferior to the threshold.

A simple method consists in calculating the quantile at 99% on each analyser and in modelling its spatial distribution, by means of geostatistical simulations of that quantile. (Fig.9).

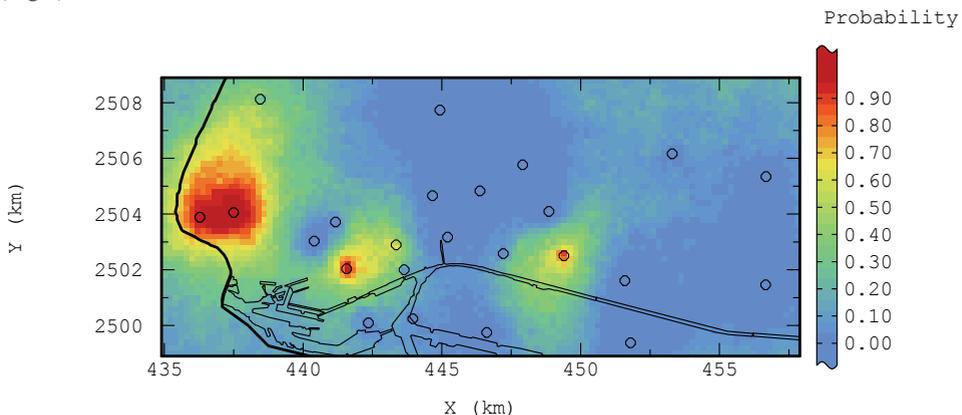


Fig. 9. Map of the probability of exceeding the threshold of SO_2 concentration of $250 \mu\text{g}/\text{m}^3$ for more than 4 days in the urban area of Le Havre.

This simplified method ignores the variability in time: for a given quantile we will get the same results whether the pollution peaks are synchronous between the different analysers or not. Without fully characterising all complex correlations in space and time, an

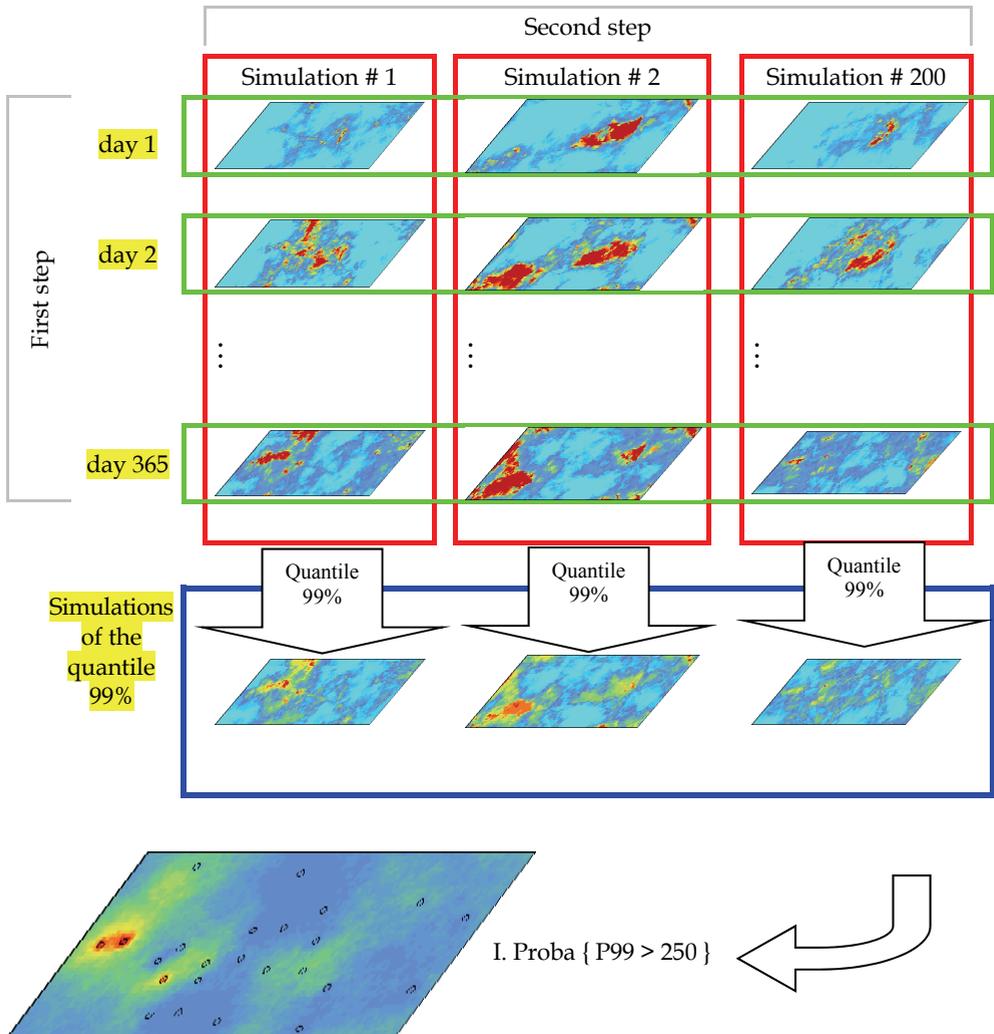


Fig. 10. Flow chart of the procedure used for simulating the daily concentrations in SO₂ in Le Havre for a year.

intermediate solution consists of simulating the daily concentrations for all days of a year, independently for each day (see first step of the figure 10). An automatic procedure providing 200 simulations has been used. We have obtained in total $365 \times 200 = 73000$ draws of the daily concentration at regular locations of a grid covering the urban area. The results have been validated by analysing the distributions and spatial correlations for several specific situations.

A map of the quantiles at the 99% frequency has been calculated for each simulation of the 365 days as shown in the second step of the scheme of figure 10. By comparing the 200

quantiles to the threshold we can then get the probability of exceeding the threshold (Fig.11): at each location it is the number of occurrences where the simulated quantile is above the threshold divided by the total number of simulations.

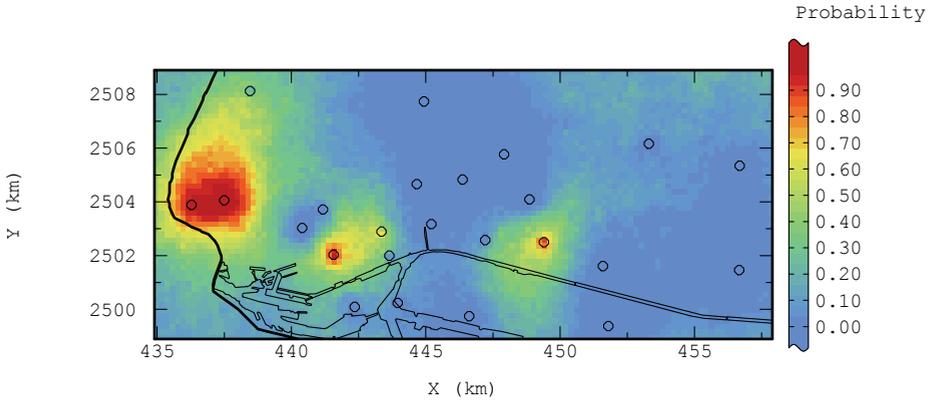


Fig. 11. Probability map of exceeding a threshold of $250\mu\text{g}/\text{m}^3$ of SO_2 for more than 4 days in a year in the Le Havre urban area.

Both maps of the figures 9 and 11 look, at first glance, rather close, however by computing the difference between both calculations, important discrepancies can be observed (see Fig. 12).

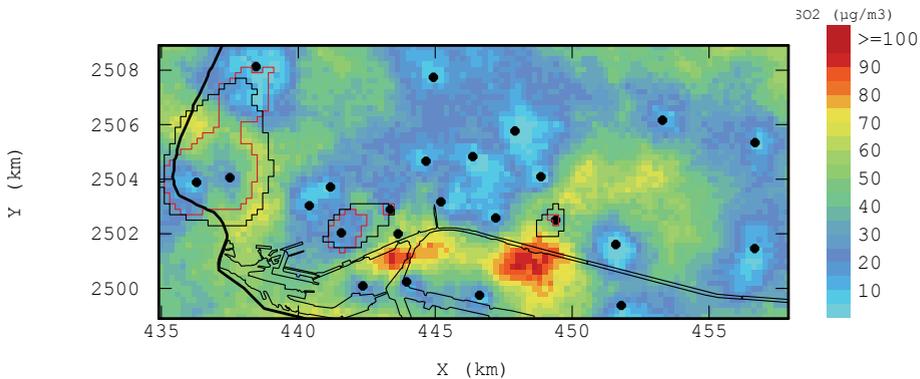


Fig. 12. Map of the differences of SO_2 quantiles at 99% obtained "directly" and from daily simulations. The outlines delineate the zones where the quantile at 99% is above $250\mu\text{g}/\text{m}^3$, from the direct method (black) and from the simulations (red).

The correlation between the mean quantile calculated from 200 simulations and the average of 20 direct simulations of the quantile (Fig.13) shows clearly the bias due to the implicit hypothesis that the pollution events occur simultaneously: the direct simulation of the quantile gives almost systematically values above the quantile calculated from the daily simulations of the SO_2 concentration.

This schematic example illustrates the risk of important bias generated by excessive simplifications. It still remains that the correlation in time should also be introduced.

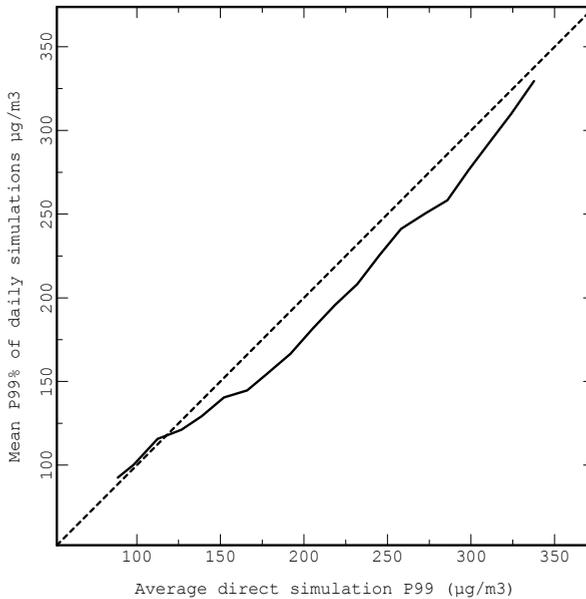


Fig. 13. Conditional expectation of the quantile at 99% calculated from 200 simulations of the daily concentrations of SO_2 versus the average of 200 direct simulation of the quantile at 99%. The first bisector is shown as a dashed line.

7. Conclusion

Geostatistics provides a rigorous conceptual framework for modelling phenomena spreading in space and time. The application of the geostatistical methods requires some care because, if they are not adapted for answering appropriately the problem, the consequences on the results may be very significant. An effective limitation lies in the use of kriging techniques which provide an optimal solution to estimate quantities depending linearly on the variable of interest. However for evaluating the risk of atmospheric pollution, applying a threshold on the pollutant concentration is typically a non linear operation. If the threshold is applied on the kriged map the calculations will result in a potentially important bias. More complex models are then required, they are enhanced by non linear geostatistical methods.

The calculation of the conditional probability using the Gaussian model makes the solution of simple problems quite straightforward: delineation of the geographical area concerned by the pollution or evaluation of the average of the number of inhabitants exposed to it.

The conditional simulations provide a consistent solution and the most adequate results for the evaluation of the population exposed to pollution. These techniques are used to evaluate health risks. When dealing with simulations a sufficient number of outcomes must be generated, particularly to derive high order quantiles. The problems become still more complex when the time dimension of the pollution events has to be taken into account, such as the control of the number of days when the pollutant concentration exceeds a threshold.

The simulations provide a correct solution but are heavy in computations. The example on SO₂ in Le Havre is favourable because the data are available at the same locations all over the year. By performing the simulations independently for each day, the most important bias can be avoided. Nevertheless a complete spatial and temporal model would have the theoretical and practical advantage of handling globally the whole set of measures. Further work is required before being able to apply such models. The present regulations in European countries often refer to complex criteria combining different time scales. The power of simulation techniques is not, in theory, limited regarding this complexity, as soon as the model correctly represents the pollutant concentrations at the origin of the pollution events.

The applications illustrated here have been successfully carried out because the criteria were simple (a threshold applied on one pollutant for a given period of time). When the criteria are multiple (several pollutants, several time scales) the simulations will definitely be the only possible method, but computationally heavy. The need for better models, particularly dealing with the temporal aspect, is still important.

8. Acknowledgement

The data used for illustrating applications of geostatistical methods have been kindly provided by Air Normand, an organization in charge of the air quality control in the Normandy region of France. The computations have been performed using the Isatis software marketed by Geovariances.

9. References

- Chilès J.-P., Delfiner P. (1999). *Geostatistics Modeling Spatial Uncertainty*. New York: Wiley & Sons
- Cressie N, Kaiser MS, Daniels MJ, Aldworth J, Lee J, Lahiri SN, Cox LH (1998) Spatial Analysis of particulate matter in an urban environment. In: *Second European Conference on Geostatistics for Environmental Application*, eds JJ Gomez-Hernandez, A Soares & R Froidevaux, Kluwer Academic Publishers, pp. 41-51
- Deraisme J, Jaquet O, Jeannée N (2002) Uncertainty management for environmental risk assessment using geostatistical simulations. In: *Fourth European Conference on Geostatistics for Environmental Application*, eds X Sanchez-Villa, J Carrera & JJ Gomez-Hernandez, Kluwer Academic Publishers, pp. 139-150
- Geovariances (2009) *Isatis technical References Version 9*, 222 p.
- Goovaerts P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York
- Hauglustaine, D.A., Brasseur, G.P., Walters, S. *et al.* (1998). MOZART: a global chemical transport model for ozone and related chemical tracers, Part 2. Model results and evaluation. *J Geophys Res* 1998; 103: 28291-28335.
- Jeannée N., L. Mosqueron, V. Nedellec, Ch. Elichegaray, S. Bouallala, H. Desqueyroux, B. Guillaume, C. Liousse, R. Lagache. Assessment of urban population exposure to atmospheric pollution: existing methods and application to PM10 in France. *Pollution Atmosphérique*, 190, Avril-Juin 2006, 197-209.

-
- Matheron G. (1989). *Estimating and Choosing - An Essay on Probability in Practice*, Springer, Berlin
- Roth C, Bournel-Bosson C (2001) Mapping diffusive sampling results: including uncertainty and indirect information. In: *Int. conf. Measuring Air Pollutants by Diffusive Sampling*, Montpellier, Sept. 26-28 2