Cartography for air quality monitoring : the geostatistical approach

Abstract
Being able to provide a quick and accurate pollutant maps from readings at isolated measurement stations is becoming more important today in light of the European norms on air quality and the public's demand to be informed. Commonly used algorithms for cartography are quick but their accuracy remains to be determined. Firstly, the choice of method is arbitrary and based on user's subjective perception. Secondly, none of these methods can take relevant auxiliary information (like meteorological data, traffic density or industrial emissions) into account, thus limiting the realism of the final map. The geostatistical mapping algorithm, called kriging, is based on the specific spatial behaviour of the mapped pollutant via a spatial correlation function calculated from the sample measurements. Because the form of this correlation function determines the shape of the final map, the geostatistical algorithm adapts itself directly to the spatial characteristics of the data, thus eliminating the need for a subjective evaluation. Auxiliary information can also be incorporated in the final map by integrating any relationship (linear or non-linear) with the mapped pollutant into the kriging algorithm. The paper begins with a presentation of the concepts behind the geostatistical mapping algorithm. Two practical applications, mapping benzene (C6H6) over Rouen and nitrogen dioxide (NO2) over Paris, are used to highlight the capacity of the geostatistical approach to produce realistic maps that take traffic information and emissions into account when monitoring air quality in real-time. All calculations are made with the ISATIS geostatistical software package.

Key-words*: geostatistics, kriging, variogram, interpolation, cartography, monitoring.*

Introduction

Nowadays, the existing legislation (European norms) and the public demand to be informed, require air quality monitoring organisations to provide accurate pollutant maps from readings at isolated monitoring sites.

In order to map one pollutant, the data must be interpolated between monitoring sites on a regular grid that covers the studied area. Interpolated values can then be graphically presented as a map. To interpolate between monitoring sites means to set one estimated concentration at each point of the grid. At one specific grid node, the estimated concentration is a weighted average of the concentrations observed at the monitoring stations close to the grid node. Yet there are several weighting algorithms, each one leading to a different estimated concentration and, therefore, to a different pollutant map. Choosing one algorithm depends on how consistently it combines the experimental measurements and on how it integrates other information – such as traffic density or meteorological conditions.

1 – Commonly used algorithms

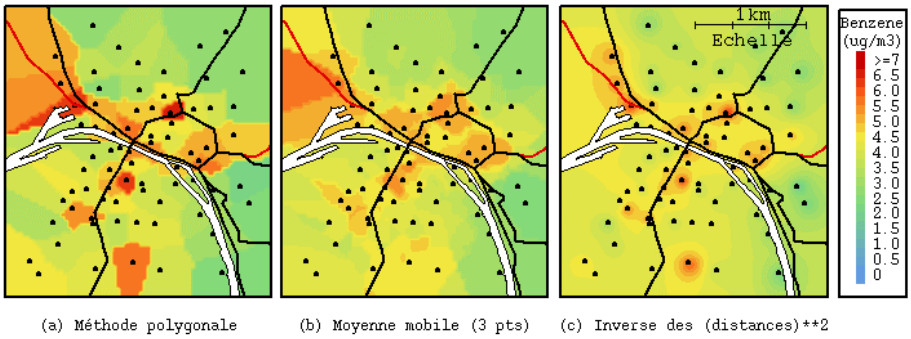

Figure 1: The maps obtained by (a) the polygonal, or nearest neighbour method, (b) by a moving average of the 3 nearest values and (c) by the inverse distance squared method

To show the impact of the method used to draw the final map, figure 1 displays three different maps obtained by commonly used methods. The data used are benzene concentrations measured by 73 diffusive samplers over Rouen city. The positions of the samplers are displayed as black points, the Seine river in white and main roads in black and red lines (the latter for motorways). The shape of the maps results from the method used which only depends on the location of the monitoring sites in relation to the point considered. The user notices the map that seems the most realistic, which means that the user's choice is necessarily based on subjective perception, that is to say the user's a priori ideas. In order to always get the most representative map, the method used must also take into account the pollutant's specific spatial features.

M. BOBBIA (Air Normand –21 av. de la Porte des Champs, 76000 Rouen - France) & F. MIETLICKI ,C.ROTH (Airparif - 7 rue Crillon, 75004 Paris -France)
& J.DERAISME (Géovariances – 49Bis av. Franklin Roosevelt, 77212 Avon -France)

2 – Geostatistical approach : kriging method

According to the geostatistical algorithm - called kriging -, the weighting rule and so the resulting map are directly determined by the spatial behaviour of the mapped pollutant. The variogram calculated from the values observed at the monitoring sites makes it possible to quantify the pollutant's spatial continuity. A mathematical function is then fitted to the experimental variograms to get a model which characterizes the spatial variability for any distance and any direction in the two-dimensional plan.

Figure 2 displays the experimental variograms and the variogram model fitted to (a) benzene data in Rouen, and (b) nitrogen dioxide data in Paris (see paragraph 3). The variogram quantifies the variability between two measures of the same pollutant according to the distance between them. The variogram can be interpreted in terms of continuity or spatial correlation, if one graphically reverses the variogram to obtain a decreasing function : the higher the distance between two sites, the lower the correlation between the measures. For instance, benzene in Rouen is less correlated in space than nitrogen dioxide in Paris. The pollutant is no longer correlated beyond the distance where the variogram model becomes stabilized at a specific level, 3 km for benzene and 35-40 km for nitrogen dioxide. There is no correlation between two measures distant from one another from more than that specific level

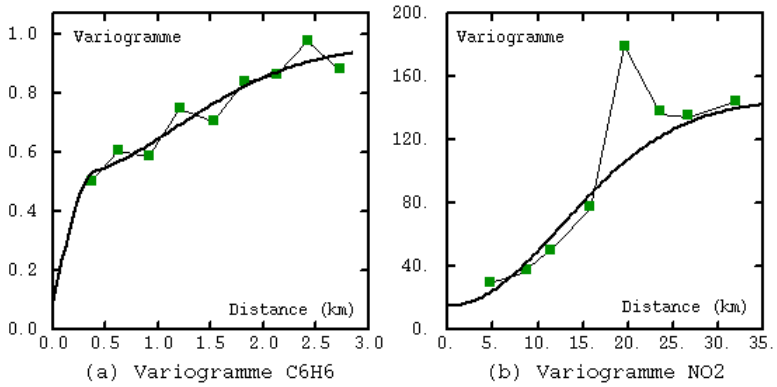

Figure 2: The experimental variograms (thin discontinuous line) and the variogram models (thick line) for (a) C6H6 over Rouen and (b) nitrogen dioxide (NO2) over Paris.

The variogram model that is specific to the pollutant is then used in a linear system of equations, the so-called kriging system, to calculate the optimal weight to be given to the data while interpolating. Doing so, a greater weight is given to the data that is better correlated with the concentration at the target grid point. At the same time, less weight is given to the data that is less correlated with the concentration. This means that geostatistical interpolation does not take into account the distance between monitoring sites and grid point, but the pollutant's spatial correlation between these sites and the concentration at the target grid point. By doing so, kriging adapts itself to the spatial features of the pollutant via the variogram model.

Moreover, when some auxiliary variables such as population density, traffic emissions or meteorological conditions are linked to the pollutant concentration, geostatistics enable the integration of these variables into interpolation. When the auxiliary variable and the pollutant are directly related, such link can be modelled too, thanks to the cross variogram that quantifies the spatial correlation between the pollutant and the auxiliary variable. The co-kriging technique integrates the spatial behaviour of both the pollutant and the auxiliary variable as well as the cross correlation between them to estimate the pollutant from concentration measures and auxiliary data. When the shape of the auxiliary variable displays the pollutant's global behaviour, the external drift kriging technique enables the modelling of the shape and the incorporation of this shape into the interpolated map so as to better reproduce the pollutant's spatial structure.

Because of the probabilistic scope of geostatistics, it is possible to quantify the uncertainty associated with the interpolated value by using the kriging variance. This variance - or its square root, the kriging standard deviation – represents the possible scattering of the real yet unknown concentration, around the value obtained by kriging. The smaller the scattering, the closer to reality the interpolated value will be – on average -, and so the map will be more accurate. Thus, the large values of the kriging variance may display under-sampled areas on the interpolated map.

The two examples introduced here show that geostatistical interpolation makes it possible to produce pollutant maps that are more realistic than those obtained by classical interpolation methods. Since kriging takes into account the spatial continuity of the observed data, the resulting map respects the pollutant's specific spatial behaviour. When the pollutant is linked to auxiliary variables, this additional indirect information is taken into account so as to produce a map that shows the link. Furthermore, because of the geostatistical approach, it is possible to quantify the map's uncertainty and to detect under-sampled areas.

3 – First practical application : Benzene mapping over Rouen

Six monitoring campaigns were carried out in Rouen in 1998. The measures were made with diffusive samplers located so as to measure background pollution over the whole urban area. Each campaign provides average concentrations over a week. The average of the six campaigns will be used to map the estimated annual mean background level. To do this, 73 monitoring sites were used. In order to improve the map obtained from benzene sampling, two auxiliary variables were chosen (for lack of anything better) : population density and relief. Usually benzene concentration is not directly linked to these variables, yet an indirect relationship might actually exist, if we consider the following argument : (a) town centre and main roads are located in the basin created by the river Seine – this would lead to an inverse relationship between pollutant concentration and topographical level rather due to the local lie of the land than to the scientific explanation of this phenomenon, and (b) the more dense the population, the more cars there will be, and thus the higher the benzene concentration might be.

To integrate these variables in a mapping process, one will build a function of population and relief to show the global trend - called external drift - of benzene concentration that is representative of the global pollutant behaviour (figure 3a). To see the drift impact, kriging maps obtained by using, or not, the external drift are displayed on figures 3b and 3c respectively. External drift input to benzene information clearly improves the estimate. The map is more detailed, and extrapolation areas or areas with less measures coverage give more accurate information.

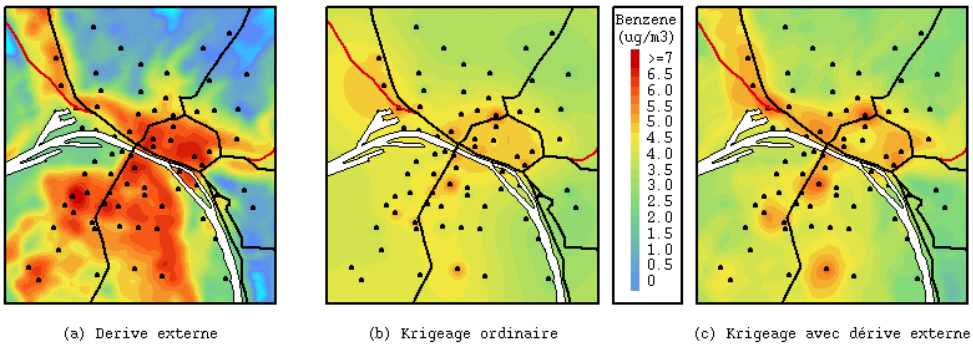

Figure 3: (a) The global trend of C6H6 concentrations over Rouen as modelled from the population density and the local relief, and the kriged C6H6 maps obtained (b) from only the measured C6H6 values and (c) after also including the global trend in the geostatistical map.

There are two methods that can help us assess if a map is reliable or not. Besides the formerly mentioned kriging variance, cross validation consists in estimating one point where the value is known by eliminating this point from input data. By doing so for all monitoring sites successively, one obtains a point cloud (real values versus estimated values) that is all the closest to the bisector line since the map is well estimated. Figure 4 displays kriging standard deviation maps (with and without external drift), as well as the result from cross validation for both estimates.

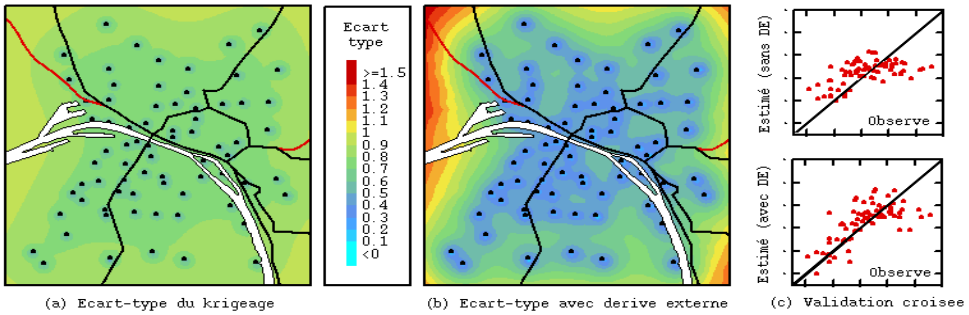

Figure 4: The maps of the kriging standard deviation for C6H6 over Rouen based on (a) kriging from only the measured C6H6 values and (b) kriging after also including the global trend in the kriging procedure.

Results from cross validation show that taking the drift into account, that is to say indirect information, improves the map's accuracy within the monitoring sites' convex hull. In extrapolation areas, kriging standard deviation with external drift displays the strong uncertainty associated to such areas in a more realistic way.

M. BOBBIA (Air Normand –21 av. de la Porte des Champs, 76000 Rouen - France) & F. MIETLICKI ,C.ROTH (Airparif - 7 rue Crillon, 75004 Paris -France)
& J.DERAISME (Géovariances – 49Bis av. Franklin Roosevelt, 77212 Avon -France)

4 – Second practical application: Nitrogen dioxide mapping over Paris

The study aims at mapping the median (percentile 50) statistical data for nitrogen dioxide measured by 20 analysers located in the Paris area between 1997 and 1999. Average NOx emissions were compiled in an inventory for a 29 km^2 neighbourhood in 1998. The inventory was used as an auxiliary variable because there is an indirect relationship between NOx emissions and nitrogen dioxide concentration. Indeed, nitrogen dioxide concentration at one specific point is non-linearly linked to NOx concentration because of chemical reactions in the atmosphere, the NOx concentration resulting itself from atmospheric diffusive and transport processes of NOx emissions located in the neighbourhood of the point.

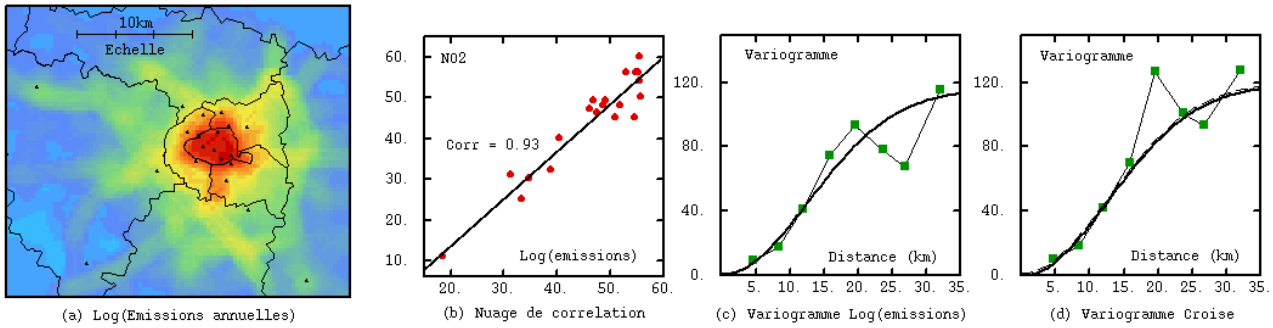

Figure 5: (a) The logarithm of the annual NOx emissions over the Paris region, (b) the scatter plot between NO2 and log(emissions), (c) the variogram of the log(emissions), and (d) the cross-variogram of NO2 and log(emissions)

Nitrogen dioxide concentration is not linked linearly with NOx emissions, but one can see experimentally that it is quite well linked with the emissions logarithm, from now on called log(emissions) – figure 5b. The emissions logarithm displayed on figure 5a can then be used as an auxiliary variable to map nitrogen dioxide over the Ile-de-France region. The studied area covers a 90 km x 75 km area around Paris. Departmental borders are displayed by black lines, monitoring sites by black points.

The direct linear relationship between nitrogen dioxide and log(emissions) can be taken into account for mapping nitrogen dioxide by applying the co-kriging method. To do that, one must also model the log(emissions) variogram as well as the cross variogram between nitrogen dioxide and log(emissions); see figure 5c and 5d respectively. Knowing the log(emissions) variable at every point and modelling the relationship between the log(emissions) variable and nitrogen dioxide make it possible to improve the mapping of nitrogen dioxide between monitoring sites. In order to assess the impact of the auxiliary variable, kriging and co-kriging maps are displayed in figures 6a and 6b respectively. One can note that adding an auxiliary variable obviously improves the map, especially in areas with very few monitoring sites and where extrapolation is involved.

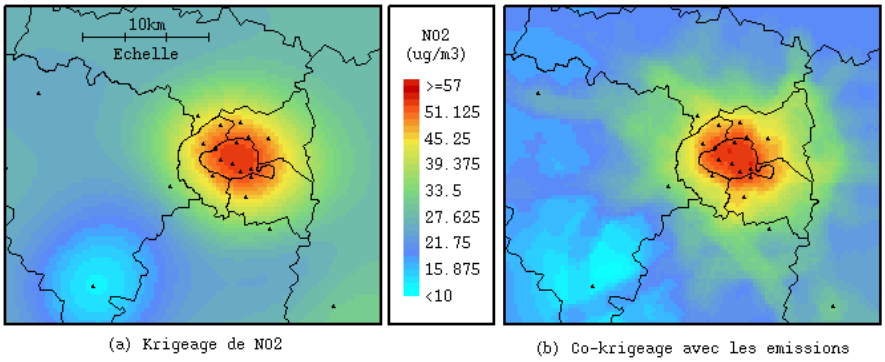

Figure 6: The maps for NO2 over the Paris obtained (a) from kriging only the measured NO2 values and (b) from cokriging that includes the logarithm of annual NOx emissions as auxiliary information.

Taking into account an auxiliary variable – such as the NOx emission logarithm for instance –improves the map's accuracy not only in the analysers' convex hull, but also in extrapolation areas.

M. BOBBIA (Air Normand –21 av. de la Porte des Champs, 76000 Rouen - France) & F. MIETLICKI ,C.ROTH (Airparif - 7 rue Crillon, 75004 Paris -France) & J.DERAISME (Géovariances – 49Bis av. Franklin Roosevelt, 77212 Avon -France)