Automatic Mapping Algorithm of Nitrogen Dioxide Levels from Monitoring Air Pollution Data Using Classical Geostatistical Approach: Application to the French Lille City

G. Cardenas and E. Perdrix

Abstract This work aims to test a new method of automatic mapping of gaseous nitrogen dioxide in an urban area. These maps have to be realised based on the data provided by automatic on-site monitoring stations, taking into account their scarcity that could hamper a correct spatial interpolation of the pollution levels. In the first part of this study, we propose a new methodology to generate additional data, based on several previous field campaigns performed by passive sampling. Among these passive sampling sites some of them (henceforth called "virtual stations") are time-correlated to a given fixed station (called "reference station"). In the second part of this study, we have tested the suitability of our method for the automated generation of variograms on a case study as well as the quality of the estimations calculated based on these data. For mapping, geostatistical methods were applied, particularly the cokriging one. This multivariable method exploits the additional information given by auxiliary variables; in the case of nitrogen dioxide, variables depicting the area, such as the population density or the emissions inventory, may therefore be used. In order to take into account the uncertainty of the data generated in the virtual stations, we included in the variance-covariance kriging matrix an additional component called the variance of measurement error (VME); a methodology to calculate this component is described. Finally, the resulting maps are well detailed and do show the main features of the pollution due to nitrogen dioxide on the considered domain.

1 Introduction

In France, the air quality monitoring networks have to regularly publish the levels of pollutants in their respective zone. Several methods of cartography have been developed, since maps have proved to be an efficient way to present information about air quality to the public.

G. Cardenas
Institut National de l'Environnement Industriel et des Risques, Parc Technologique ALATA, BP2, 60550 Verneuil-en-Halatte, France
e-mail: giovanni.cardenas@ineris.fri

A. Soares et al. (eds.), *geoENV VI – Geostatistics for Environmental Applications*, 95–101

In order to produce them, geostatistical methods of interpolation (Kriging) are often used. Their main advantage is to integrate the spatial correlation of the pollutant through the calculation of the spatial covariance function based on the available data.

Usually, maps are realised with measurements provided by passive samplers. Because of their low cost, these data are abundant and allow a large spatial coverage. However, they provide a time average pollutant's concentration over quite a long period (7 or 14 days).

On another hand, the data given by fixed automatic monitoring stations consisting of every 15-minutes measurements are used to monitor the alert thresholds of given pollutants such as ozone or nitrogen dioxide. However the number of fixed automatic stations is generally insufficient to allow a good spatial interpolation of the pollution levels.

In order to produce a daily pollution maps, it is compulsory, firstly, to add additional data to the measurements given by the fixed monitoring stations and, secondly, to conceive an automatic mapping algorithm able to integrate both type of data.

2 Methodology

The aim is to set up a method in order to i) generate additional data and ii) build up an algorithm for automatic mapping, based on the automatic modelling of the spatial covariance of NO2.

2.1 Temporal Generation of Additional Data

Our method to produce additional data is based on the existence of several previous field campaigns done by passive sampling on an urban area.

These field campaigns are used in the way that sites of passive sampling (called "virtual stations") are time-correlated to a given fixed station (called "reference station") and put together.

In these determined "virtual stations" it is possible, by applying this method, to estimate the concentration of nitrogen dioxide, at any given time, from the value measured at the correlated reference station.

This method is an alternative to generate additional data. However for the urban areas lacking of sufficient campaigns done by passive sampling, other methods must be applied. Another possibility may be, for instance, the generation of additional data from mobile campaigns and forecast models (models such as ANACOVA), etc.

2.2 Automatic Mapping

The previously explained method enables to estimate the concentration of nitrogen dioxide at the virtual stations; this dataset is then used to run a geostatistical interpolation of the concentrations, in order to get a map of the pollution levels.

As a reminder, among the mostly used geostatistical methods in the field of air quality, one may cite: the monovariable methods, made up of ordinary kriging and simple kriging or with known mean, and the multivariable methods which aim to make the most of the additional information brought by auxiliary variables.

The multivariable methods are an adaptation of kriging to different assumptions in order to take into account the auxiliary variable; more precisely, for the estimation of the nitrogen dioxide concentrations, the consideration of auxiliary variables depicting the area such as the relief, the population density or the emission inventory allow to produce more detailed maps of the spatial distribution of nitrogen dioxide.

In the field of air quality, the most often used multivariable methods are cokriging, kriging with external drift and a variation of this latter which is kriging of the residues (Cardenas and Malherbe, 2002; Chilès and Delfiner, 1999). In order to select the better approach, the classical approach consists in comparing the results from the different methods. Table 1 shows the steps of a geostatistical study aiming to map nitrogen dioxide concentrations.

The main steps of the procedure are related to the search of the spatially-correlated auxiliary function with pollution and (subsequently) to the calculation of variograms and fitting of the respective models. With robustness in mind, pre-existing configurations have been tested to build the mapping algorithm, therefore taking advantage of the available knowledge of the phenomenon.

1. Search of the auxiliary function: the goal of this step is to find a known function in the whole domain correlated with the nitrogen dioxide concentration. This auxiliary function may be constituted of variables describing the area such as the population density, the relief and the emission inventory. We propose to primarily study all the possible configurations of these auxiliary variables and to only retain the best-correlated ones to the concentrations. Then this selected function will be used daily in the mapping algorithm.

Table 1 Steps of a geostatistical study

Steps of a geostatistical study	
Action	Results
1. Study of the available auxiliary data: relief, emissions, population density, meteorological data, etc.	Choice of the best correlated auxiliary function to the concentrations.
2. Calculation of the anisotropic and isotropic variograms of the concentration and the auxiliary variable.	Search of possible anisotropies, analysis of the auxiliary variable quality.
3. Fitting of the variogram's model(s)	Choice of the basic structures and the range of the models.
4. Test of the models by cross-validation.	Validation of the models and choice of the estimating method: ordinary kriging, cokriging, kriging with external drift or kriging of the residues (Gallois et al., 2005).

Table 2 Automation of the procedure

Automation of the procedure	
Action	**Result**
1. Creation of the dataset based on the data from the "Temporal generation of additional data method" (reference stations and the correlation parameters: slope and Y-axis intercept).	Dataset of the daily concentrations of nitrogen dioxide.
2. Creation of the algorithm enabling the automatic execution of daily maps of the estimation of nitrogen dioxide.	The algorithm must fulfill the following tasks: 1. Calculation of the correlation cloud with the pre-determined auxiliary variable. 2. Calculation of the experimental variograms. 3. Based on these experimental variograms: fitting of the sills of the pre-selected basic structures. 4. Cross-validation. 5. Using the pre-selected kriging method: interpolation of the concentrations at a grid cell. 6. Realisation of estimated maps and of the map of estimation variance using a pre-determined color scale. 7. Export of the results: • Export of the resulting maps in an image format bmp or jpeg. • Export of the results files in ascii format or grid format for a sub-sequent use in a GIS.
3. Results analysis	Here are some useful information to determine the quality of the obtained estimations. 1. Correlation cloud between the concentration and the auxiliary variable (correlation coefficient). 2. Experimental variograms and fitted models. 3. Cross-validation statistics: error and relative error variance. 4. Correlation cloud: concentrations at the virtual stations *versus* estimated concentrations (correlation coefficient). 5. Estimation and estimation variance statistics.

However the function may differ with seasons (one for summer and another for winter). Indeed several studies have shown that the correlation between nitrogen dioxide and the other variables describing the area are not as good in summer as for other seasons. On another hand, the existence of a decreasing relationship between the temperature and the concentration of NO2 is well-known, which is why the cartography of NO2 is more interesting in winter than in summer where concentrations are lower, from the point of view of air quality monitoring.

2. To fit the variograms, the basic structures (nugget effect, choice of the model: spherical, exponential or gaussian, etc.) have been pre-selected from the analysis of the available information such as the existing passive sampling campaigns .
3. The best method of estimation, among the four available ones (ordinary kriging, cokriging, kriging with external drift and kriging of the residues (Gallois et al., 2005) has been draw out from the previous estimations of pre-existing campaigns by passive sampling.

   Table 2 shows the built-up procedure to develop the mapping algorithm.

3 Case study

These methods have been tested on data from the city of Lille and its suburbs. The dataset consists of 16 fixed NO2 monitoring stations distributed on the estimation area and 15 measurement campaigns by passive sampling, lasting 14 days each, done during the years 1998/1999 and 2003/2004. Moreover, the following auxiliary variables are available: NOx emission inventory and population density.

3.1 Selection of the Virtual Stations

The virtual stations have been selected applying the "Temporal generation of additional data method". They represent 65 passive sampling sites which are very satisfactorily time-correlated to the measurements done at two fixed reference stations: the suburban station "Halluin caillou" and the urban station "Roubaix Château".

3.2 Calculation of the VME

The data extracted from the virtual stations are, by construction, marred with errors. Geostatistics are able to take into account this uncertainty through the variance of measurement error (VME), term that can be added to the variance-covariance matrix during kriging. MEV is calculated from the difference or residue between the estimated values and the ones measured during the 15 campaigns.

$$R(x) = NO_2\text{Estimated} - NO_2\text{Measured}$$

Finally for each virtual station, variance of residues is calculated, showing the "dispersion" of the residues around their mean value. Figure 1 shows the main statistics of this variable. The detailed analysis of this figure leads to the existence of a group of 10 passive sampling sites (virtual stations) related to a clear "urban environment" where it is more difficult to get accurate estimations; these are the stations 62, 132, 137, 54, 14, 356, 94, 144, 161 and 50 where the mean value of the 15 campaigns is superior to 30 $\mu g/m^{3.}$. These stations may be representative of a more local environment, influenced, for example, by traffic emissions.

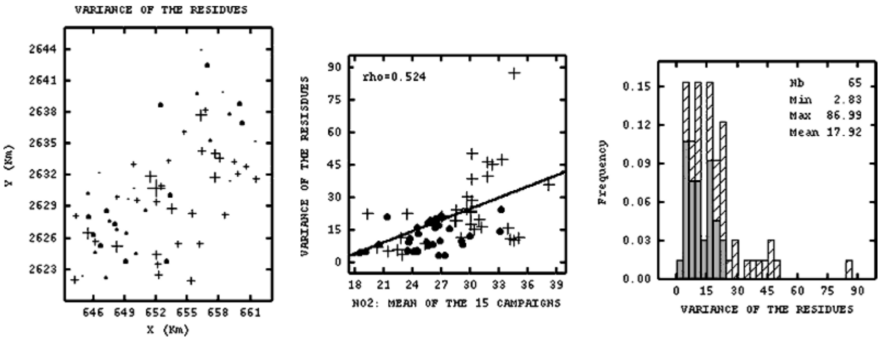

Fig. 1 Statistics of the Variance of the Residues or MEV
Note: Figure on the left: Geographical localisation of the virtual stations (the symbols size are proportional to the variance of the residues).
Figure on the center: Correlation cloud between the variance of the residues and the mean of the 15 nitrogen dioxide campaigns (the straight line represent the linear regression).
Dots represent the virtual stations linked to the periurban fixed station Halluin Caillou.
Cross in green represent the virtual stations linked to the urban fixed station Roubaix Château.
Figure on the right: histogram of the variance from the residues at the virtual stations.
Grey bars represent the virtual stations linked to the periurban fixed station Halluin Caillou.
Hatched bars represent the virtual stations linked to the urban fixed station Roubaix Château.

3.3 Choice of the Estimation Method

After a detailed study of 15 campaigns by passive sampling, we decided to perform the daily estimation by means of the collocated cokriging method with VME of the virtual stations included, and with a combination of emission inventory and population density as auxiliary variable.

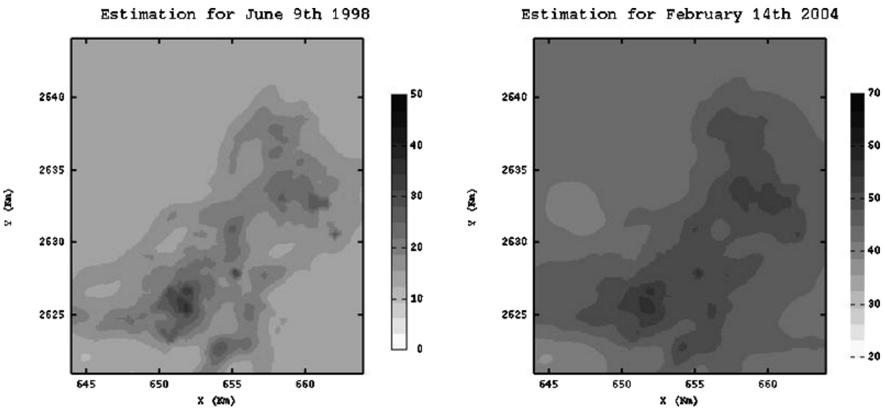

Fig. 2 Daily estimation maps of NO_2 for June 9th 1998 and February 14th 2004
Note: The scale is μgm^{-3}. The maximum values are located in the town center of Lille and Roubaix cities

Moreover, a different variogram model is fitted for each season: for summer, an isotropic model is constituted by a nugget effect and a gaussian structure with a 5 km range and, for winter a nugget effect and a gaussian structure with a 8 km range. The sill of the structures are fitted automatically.

3.4 Results

As an example, two daily estimations are shown: one in June 9th 1998 (summer period) and the other one in February 14th 2004 (winter period).

In order to evaluate the quality of our estimations, the daily means estimations relative to the winter period from February 5th to 18th 2004 and to the summer period from May 27th to June 9th 1998 were calculated and therefore, compared to the estimations got by kriging from the passive samplers measurements.

In spite of the use of auxiliary variables one may notice, on the example, a smoothing of the estimated results: high values have been under-estimated by 27% in winter and 37% in summer; in the same way, low values have been over-estimated by 35% in winter and 47% in summer. (The quality criteria is fixed to 50% for this kind of method).

4 Conclusions

A way to improve our method seems to be the reduction of the smoothing effect, maybe by working on “Temporal generation of additional data method” (as they do a first smoothing of the high values) and by creating more virtual stations.

Finally from a methodological point of view, we conclude that in order to carry out of daily maps of nitrogen dioxide, one has to perform as a starting point, a manual analysis of the pre-existing campaigns in order to pre-define some variables and parameters (best correlated auxiliary variable, structure and range of the variogram).

References

Cardenas G. Malherbe L. (2002). Application des méthodes géostatistiques à la cartographie de la qualité de l’air, STIC et Environnement, Rouen, 19–20 juin 2003

Chilès JP, Delfiner P (1999) Geostatistical: modelling spatial uncertainty, Wiley Series in Probability and Mathematical Statistics, p. 695

Gallois D, de Fouquet C, Le Loc’h G, Malherbe L, Cardenas G (2005) Mapping annual nitrogene dioxide concentrations in urban areas. Geostatistics Banff 2004. Springer, 2005, pp. 1087–1096