MAPPING ANNUAL NITROGENE DIOXIDE CONCENTRATIONS ABOVE MULHOUSE URBAN AREA

C. de FOUQUET[(1)], D. GALLOIS[(1)], L. MALHERBE[(2)], G. CARDENAS[(2)]

[(1)]École des Mines de Paris - Centre de Géostatistique, 35 Rue Saint Honoré, 77305 Fontainebleau Cedex, France. [(2)]INERIS, Institut National de l'Environnement Industriel et des Risques. Parc technologique ALATA. BP2. 60550 Verneuil-en-Halatte, France.

Abstract

To measure urban background nitrogen dioxide concentrations, passive samplers are exposed to ambient air in urban background locations for several consecutive fortnights in winter and summer. Because of various technical constraints, the total number of samplers collected during a year is limited to a few tens, which is not sufficient for calculating precise maps.
Nitrogen dioxide comes mainly from the combustion of fossil hydrocarbons (road traffic, heating, some industrial processes). Auxiliary variables like emission inventories, population density or land use, give an approximate description of those emitters, and can be used as additional information.
For the old industrial agglomeration of Mulhouse, the relationships between the auxiliary variables and the seasonal concentrations prove to be different in winter and in summer, in accordance with the atmospheric physico-chemical phenomena. However, the seasonal concentrations show a high correlation.
Cokriging the annual or seasonal concentrations from the seasonal measurements ensures the consistency of those estimations. It allows adapting the sampling design, increasing the precision with the same number of measurements.

Keywords: external drift kriging, cokriging, air quality, nitrogen dioxide, sampling design, seasonal concentration, annual concentration.

Introduction

Nitrogen dioxide NO_2 is an urban air pollutant formed by reaction of oxygen and nitrogen produced by the combustion of fossil hydrocarbons. The main sources are

road traffic, heating and specific industrial activities. Due to complex meteorological and photochemical phenomena, NO_2 increases in winter and is lower in summer. As NO_2 can cause respiratory irritations, the European regulation has fixed an annual mean lower than 40 $\mu g/m^3$ as quality objective for 2010.
Nowadays, permanent stations measuring air pollution are only few per town, and it is not possible to obtain a precise cartography of the yearly or seasonal means from those measurements. Thus, monitoring campaigns have been conducted in some towns, first to characterize the concentration levels in relation to the main pollution sources (main road, industrial zones, etc.), secondly to map NO_2 yearly mean levels, as precisely as possible. During these campaigns, NO_2 is measured using "passive diffusion samplers", installed at sites carefully chosen as representative of the background pollution, and exposed for several successive fortnights in winter and in summer.

NO_2 measurements campaigns being expensive, the objective assigned to geostatistical studies is to improve the accuracy of the estimation by taking into account some available information, providing an approximate description of the emission sources. For example, road traffic and heating should be partially linked to the population density or the land use, e.g. residential, industrial… and the local density of building. For some agglomerations, emission inventories including an evaluation of the local road traffic, or the declaration of industrial emissions made by the firms, are available.
When this information is known at local scale, for example over a 1-km resolution grid, concentrations are estimated using this auxiliary information as external drift, or in a cokriging process. For example, Bobbia et al [1] presented an instructive comparison of the estimations obtained using or not this auxiliary information. At this stage, the remaining question was to choose for each case the "best auxiliary variables", among sometimes a lot of information.

As measurements campaigns last several fortnights, it occurs for some samplers that, due to technical problems, only one "seasonal" measurement is available. In this case, it is well known [2] that cokriging the yearly concentration from the seasonal measurements allows using all the available seasonal data and ensures the consistency of the estimations, provided the multivariate variogram model between seasonal and yearly concentrations is consistent. In this case, cokriging the yearly concentration is equivalent to cokrige each seasonal concentration and then calculate their average. In the first method, the cokriging variance of the yearly mean is obtained directly. When each seasonal or annual concentration is estimated by kriging, the consistency between seasonal and yearly estimations is no more ensured, except in some very particular cases. The interest of cokriging will then be shown on an example.

1 Brief literature review

The European Framework Directive on Ambient Air Quality Assessment defines a regulatory framework for monitoring and evaluating air quality. Air pollution mapping at a relevant temporal scale is a valuable tool for providing the required information. In that context, geostatistical methods have been receiving particular attention for a few years and are now commonly applied by the French air quality monitoring network (AASQAs).

Kriging techniques, which were rather used to interpolate concentrations in areas equipped with a relatively dense monitoring network ([3], [4], [5], [6], [7]), have been implemented to process data from passive sampling campaigns at an urban or regional scale. Pollutants under study are especially ozone, nitrogen dioxide and benzene. In view of the significance of the obtained results, further efforts have been made to provide more precise maps. From a spatial point of view, interest has been paid to multivariate kriging methods, such as external drift or cokriging, to introduce auxiliary information in the estimation process ([8], [9], [10]). Such methods may significantly improve the results provided that auxiliary variables are properly chosen.

2 Experimental relationships of NO2 seasonal concentrations in urban environment

In the following, focusing on the spatial interpolation, we assimilate seasonal measurement during three fortnights with the seasonal concentration, and the average of the seasonal measurements with the yearly concentration.

2.1 Urban context and sampling

Mulhouse is an old industrial agglomeration of more than 110 000 inhabitants, located near Germany and Switzerland at the south of Alsace plain (north-east of France), under continental climate. About seventy urban or peri-urban sites were monitored for NO_2 measurements during three fortnights in winter and then in summer 2001. Due to technical problems, 62 seasonally measurements are available in winter, 59 in summer, and only 50 for the “yearly mean” that is about 25% of the time (Figure 1).

Land use is given on a grid with 200m resolution. Among other variables, only the value of dense building has been retained. Population density and nitrogen oxide emissions inventory, notated NO_x, including road traffic and industry, are given on a grid with 1km resolution.

2.2 Relations between winter and summer concentrations

Because of the spatial correlation, elementary statistics are indicative, the sampling sites being preferentially implanted toward the center of the agglomeration. For the “yearly” sites, the winter average, 28.4 μg/m^3, is sharply higher than the summer one, 16.2 μg/m^3.

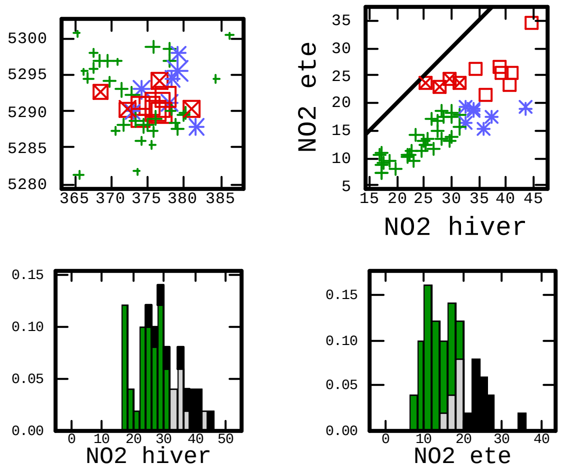

Figure 1. Location (top left), scatter diagram (top right) and histograms of seasonal NO2 measurements. High concentrations (in black for the histograms) are marked as squares on the map and the scatter diagram, and intermediate concentrations (light grey) as stars. Stations marked with $\boxed{\text{X}}$ have a high summer concentration and an intermediate winter one. On the map distances are in km.

Table 1. Elementary statistics of the NO_2 concentration measurements on the common 50 annual sites. Min. minimum, max maximum.

	min.	max.	mean	standard deviation	variation coefficient
Winter	16.98	44.70	28.37	7.20	0.25
Summer	7.25	34.64	16.20	5.99	0.37
Year	12.23	39.67	22.29	6.30	0.28

The associated standard deviations, respectively 7.2 and 6.0 μg/m^3, show that the variability increases with the concentrations, whereas the relative variability, given by the dispersion coefficient (the ratio of the standard deviation to the mean) is higher in summer than in winter (respectively 0.37 and 0.25). Clearly, annual statistics are intermediate between winter and summer values (Table 1).

Low concentrations are located on the same samplers in winter and in summer (Fig. 1), mainly around the agglomeration. High summer concentrations correspond either to some winter concentrations higher than 35μg/m^3, or to intermediate winter concentrations located at the close suburb, near an important traffic infrastructure. Apart from some of these high seasonal values, the scatter diagram between winter and summer concentration appears rather linear, with a high correlation coefficient equal to 0.82.

2.3 Relations with auxiliary information

Taking the translated logarithm of the auxiliary variables ($\log\left(1+\frac{z}{m}\right)$, z being the variable and m a normative factor, for example the mean) is a robust way to linearise the relationships with concentration, as the scatter diagrams show (Figure 2). This linearity will be useful for external drift (co)kriging, which assumes a local linear relationship between the main variables and the auxiliary information.

The contribution of heating in winter, and the consequences of summer road traffic explain the different location of the high seasonal concentrations. Indeed, the correlation coefficients between winter concentration and the three auxiliary variables are equivalent whereas for summer concentrations, the correlation increases with NO_x emissions and decreases with the dense building land use (Figure 2 and Table 2). Note that all these coefficients are lower than the one between seasonal concentrations.

Table 2. Correlation coefficients between the NO2 seasonal concentrations and the translated logarithm of explicative variables.

	Dense building	NO_x emissions	Population density
winter	0.69	0.68	0.69
summer	0.57	0.73	0.67

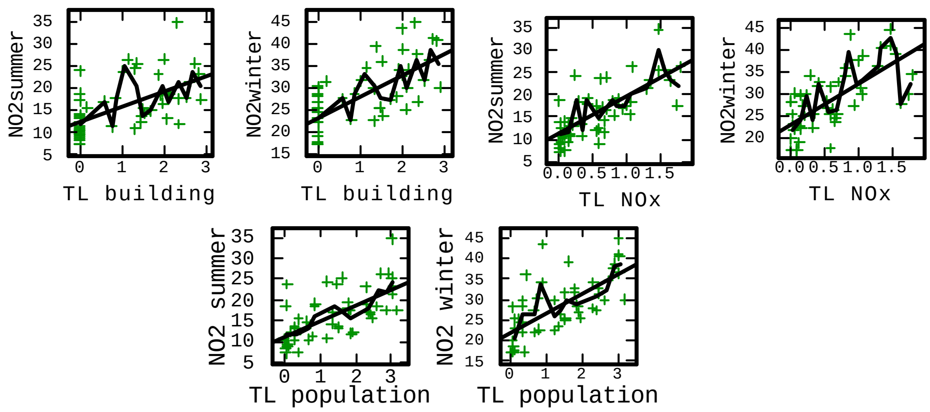

Figure 2. Scatter diagrams of NO_2 seasonal concentrations and auxiliary variables. TL: translated logarithm.

The two instructive following observations:
- a high correlation level between winter and summer NO_2 concentrations;
- a good correlation between seasonal concentration and some auxiliary variables;

appear equally valid for another agglomeration, Montpellier, in a rather different geographical context [11].

2.4. Experimental variograms of seasonal concentrations

Simple and cross variograms can be fitted using the classical “linear model of coregionalisation” (Figure 3), in which any spatial structure present in the cross variogram, is simultaneously present on each of the two simple variograms. The experimental cross variogram of seasonal concentrations is located on the maximal correlation envelop in the model, showing the importance of the joint spatial variability.

Experimental global sills values, close to 65 $(\mu g/m^3)^2$ for winter, and 40 for summer are consistent with the higher variability of winter concentrations. At the 15km scale, a hypothesis of quasi-stationary seems acceptable for both seasons, without contradiction with the global non-stationarity over all the agglomeration, with rather high concentrations in the centre and northward.

The intermediate 5km spatial component and the large scale component (conventionally chosen at 20km) are common to both seasons, with higher sills for winter (respective sills are 35 and 25 for the 5km range, and 38 and 9 for the 20km one), whereas in this fitting short range structures are present only for summer concentrations.

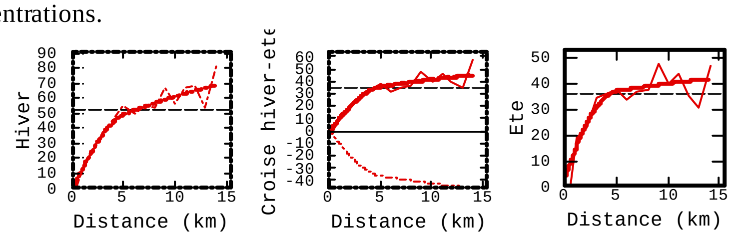

Figure 3. Fitting of simple and cross variograms of NO_2 seasonal concentrations using the linear model of coregionalisation (Hiver: winter, été: summer). The dotted lines on the cross variogram indicate the maximum envelop in the model.

3 Cokriging of seasonal and annual NO2 concentrations

As mentioned before, each seasonal concentration is approximated by the average concentration sampled during 45 days, and the yearly concentration by their mean. Despite not to be neglected [12] the time component of the estimation variance is not considered in this paper, focusing only on the spatial estimation.

In all this part, the classical cross-validation method is used to quantify the actual gain of precision brought by the auxiliary variables used as external drift: each point is successively removed from the data and its value estimated from all the other data.

3.1 Usefulness of auxiliary information for estimating seasonal concentrations

In a first step, we examine the gain of precision brought by auxiliary information, given on a regular grid on the whole area.

For the external drift estimation with moving neighborhood, the residuals being not available, the variograms are indirectly fitted by cross validation. For both seasonal concentrations, a linear variogram was retained. Auxiliary information retained is the following:

- winter : density of population and dense urban land use i.e. density of the buildings ;
- summer: all three variables, i.e. the two previous ones and NO_x emissions inventory.

Comparing separately (ordinary) kriging and kriging with external drift for each season separately shows the following results (Table 3.):

- in a coherent way with the experimental dispersion variances of seasonal measurements, the variance of the estimation error for (ordinary) kriging is a little smaller for summer concentrations ; but the variance of the *relative* error (i.e. the error $z - z^*$ divided by the measured value z) is clearly smaller for winter concentrations.
- Correlation coefficient between seasonal measurement z and its estimation z^* are consistent with the results on the variance of the relative error: introducing auxiliary information as external drift always improve the precision of the estimation, with a larger gain in winter.
- External drift kriging: with a standard deviation of 16% for the relative error, and a correlation coefficient greater than 0.80 between estimation and measured value, the precision can be considered as acceptable for winter concentrations. The results are less good for summer concentrations, with a standard deviation of the relative error of 30% and a correlation coefficient between estimation and measured value lower than 0.70.

The scatter diagram between estimated and measured concentrations (not shown) being effectively linear, the corresponding correlation coefficient reflects the association level between the two variables.

Clearly, with a greater variability, winter concentrations are relatively better spatially structured and better “explained” by auxiliary information than the summer ones.

As the correlation coefficient between seasonal concentrations is greater than those between concentrations and auxiliary information, the link between these concentrations can be used to improve the estimation, mainly for estimating summer concentrations or at sites where only one seasonal measurement is available.

Table 3. Cross validation results for seasonal concentrations. Ordinary (OK) and external drift (EDK) Kriging.

season	model	r (Z,Z*)	Var. error	Var. relative error
winter	OK	.76	19.6	0.031
	EDK	.83	14.4	0.026
summer	OK	.62	18.3	0.092
	EDK	.67	17.0	0.088

3.2. Cokriging the yearly concentration

Cross-validation tests (not developed here) show that cokriging seasonal concentrations instead of kriging them separately effectively improves the precision.

Another interest of cokriging is that it ensures the consistency of linear relations between estimated quantities, at least when the variographic model itself is consistent. In the present case, if each seasonal or annual concentrations are estimated separately by kriging (or any other monovariate estimator), nothing assures that the estimated annual concentration is the mean of the two seasonal ones.

As the European reglementation imposes a limit value on the annual concentration, we focus in the following on the estimation of this value.

The following estimations of the yearly concentration are compared:

- kriging with external drift, from the annual measurements only;
- average of the two seasonal kriging with external drift estimations;
- cokriging with external drift from the seasonal measurements, which is equivalent to averaging the two seasonal estimations made by cokriging.

The non bias conditions for external drift cokriging of the annual mean $\frac{1}{2}\left(Z_1(x)+Z_2(x)\right)$ from seasonal data $Z_i, i=1,2$ are easily obtained. Putting $Z_i(x)=R_i(x)+\sum_{\ell} a_\ell^i f_i^\ell(x)$, where f_i^ℓ are the auxiliary variables (eventually after transformation as log translation), and a_ℓ^i the unknown deterministic coefficients, the estimator is written:

$$\frac{1}{2}\left(Z_1(x)+Z_2(x)\right)^* = \sum_{\alpha} \lambda_x^\alpha Z_1(x_\alpha) + \sum_{\beta} \kappa_x^\beta Z_2(x_\beta)$$

Supposing the a_ℓ^i without fixed relationships between them and between values for winter and summer (i.e. for $i = 1,2$). The non bias condition

$$E\left[\left(\frac{Z_1(x)+Z_2(x)}{2}\right)-\left(\frac{Z_1(x)+Z_2(x)}{2}\right)^*\right]=0$$

imposes

$$\forall\ell, \sum_\alpha \lambda_x^\alpha f_1^{\ \ell}(x_\alpha) = \tfrac{1}{2} f_1^{\ \ell}(x) \text{ and } \forall\ell, \sum_\alpha \kappa_x^\beta f_2^{\ \ell}(x_\beta) = \tfrac{1}{2} f_2^{\ \ell}(x).$$

In practice, the function f_i^1 is taken as constant, and the two conditions for $\ell = 1$, also valid for cokriging with unknown means and without external drift, simply imply that the sum of the weights relative to each measurement period is equal to $\frac{1}{2}$.

It is not necessary to use the same auxiliary variables as drift for both seasons.

3.3. Cross-validation results

The classical cross-validation method, taking away successively each data, is used to quantify the improvement of precision brought by auxiliary variables or by cokriging from the seasonal concentrations (Table 4).

The correlation coefficient of cross validation estimation and data, and the experimental variance of the estimation errors indicate the same rank for the quality of the estimation. For the direct kriging of annual concentration, the relative variance, calculated as the variance of the ratio of the estimation error to the data, give slightly different results from the previous criteria when comparing two sets of external drifts.

When available, the other seasonal data is retained at test point for cokriging. The main results are the following:

- As for seasonal estimation (not shown), different auxiliary variables used as external drift give almost the same cross validation result. In practice, other criteria should be considered, mainly the resulting maps, to choose the most suited, following the practical knowledge of the air pollution phenomenon [12].

- Using only the 50 yearly measurements for the direct kriging of the annual value considerably reduces the precision, in comparison to the average of both seasonal kriging estimations, which take into account all seasonal measurements.

- Because of the high correlation level between seasonal concentrations, cokriging considerably improves the cross validation results. As a seasonal measurement is not available on each cell of the estimation grid, this last cross validation result could be regarded as too optimistic. In fact, it has some very interesting practical consequences.

Table 4. Cross validation results for "yearly" concentration. Kriging (K) and cokriging (CK) with external drift. 49 test points, because of moving neigborhood. r denotes the correlation coefficient between estimated and measured values and Var. the variance.

model	r (Z,Z*)	Var. error	Var. relative error
EDK Population, dense building	.81	12.5	0.032
EDK NOx	.82	11.8	0.035
Mean of seasonal EDK	.85	10.2	0.030
EDCK	.95	3.4	0.011

4 Optimizing the sampling design

The high correlation of seasonal concentrations makes the information partly redundant when a same site is sampled during both seasons. Keeping the same number of data, sampling different sites in winter and summer would allow increasing the number of measured sites, so as to improve the estimation. Another possibility consists in removing one of the two measurements at some sites, to decrease the sampling cost.

To check the feasibility of a lighted sampling, 30% of the data are removed, keeping 30% of the sites sampled only in winter, 30% only in summer, the remaining 40% sites being sampled during both seasons. Spatially the removed data were sorted at random.

Keeping some sites common to both seasons is necessary to draw the model, and in particular the cross variogram. To evaluate the actual influence of the loss of information, a cross validation of the cokriging is now realised by suppressing and re-estimating successively a quarter of the stations. The model used (variograms, external drift…) is that drawn with all the information.

For the reduced sampling, the experimental mean quadratic error on the yearly concentration is 11.7, and 10.1 for the whole sampling (on exactly the same yearly sites). Hence the precision of the estimation is only little diminished compared to the saving of 30% of measurements. Moreover, cokriging maps are very analogous, computed with complete or reduced data. Two others reduced samplings were computed in the same way, changing the sets of removed data, and giving roughly same results.

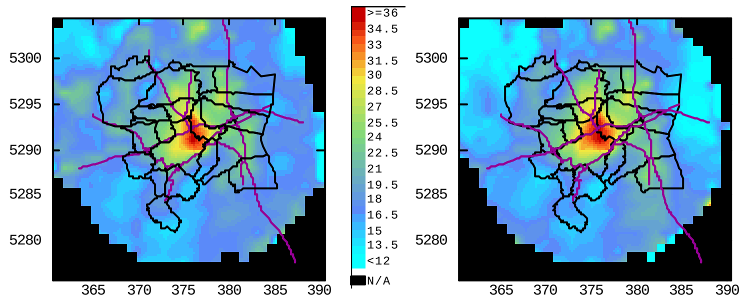

Figure 4. Cokriging of the yearly concentration from seasonal measurements, using all data (left) or 70% of them (right).

The previous cross validation used the variogram model drawn with all data. But to be valid, the model should be drawn again with the reduced sampling. A new variogram was fitted using the 70% retained measurements, a little different from the previous model. The new mean quadratic error on the yearly concentrations becomes equal to 10.0: the estimation seems as good as using all measurements. In fact, re-drawing the model allows making it more appropriate to the reduced sampling. The map calculated with this new model is a bit different from that calculated using all measurements (Fig. 4), but the major differences are located only on the edge of the map, where estimations are unreliable because made in complete extrapolation from the data.

As a conclusion, taking into account not only auxiliary information but also the high correlation of seasonal concentrations allows to significantly reduce sampling costs, without diminishing the precision level. In practice, it is necessary to keep enough samplers common to both seasons to verify the validity of the multivariate variogram model, distributing them regularly not only in the geographic space but also in the space of auxiliary variables.

Acknowledgments

Authors thank very gratefully Gilles Perron from ASPA (Association pour la Surveillance et l'Etude de la Pollution Atmosphérique en Alsace) for his advices and comments.

References

[1] BOBBIA M. MIETLICKI F.ROTH C. 2000. Surveillance de la qualité de l'air par cartographie : l'approche géostatistique. Prix du meilleur poster INRETS, 5-8 juin Avignon, France.

[2] MATHERON G. 1970. La théorie des variables régionalisées, et ses applications. Les Cahiers du Centre de Morphologie Mathématique, fascicule 5. Ecole des Mines de Paris, Fontainebleau.

[3] CASADO L.S. ROUHANI S. CARDELINO C.A. FERRIER A.J. 1994. Geostatistical analysis and visualization of hourly ozone data. Atmospheric Environment 28, n°12, 2105-2118.

[4] LEFOHN A.S. KNUDSEN H.P. MCEVOY L.R. 1988. The use of kriging to estimate monthly ozone exposure parameters for the Southern United States. Environmental Pollution, 53, 27-42.

[5] LIU S. L.-J. ROSSINI A.J. 1996. Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto - A pilot study. Environment International, N° 6, 667-692.

[6] NIKIFOROV S.V. AGGARWAL M. NADAS A KINNEY P.L. 1998. Methods for spatial interpolation of long-term ozone concentrations. Journal of Exposure Analysis and Environmental Epidemiology, vol. 8, n°4, 465-481.

[7] TAYANÇ M. 2000. An assessment of spatial and temporal variation of sulfure dioxide levels over Istanbul, Turkey. Environmental Pollution, 107, 61-69

[8] BOBBIA M. PERNELET V. ROTH C. 2001. L'intégration des informations indirectes à la cartographie géostatistique des polluants. Pollution Atmosphérique n° 170 - Avril-Juin.

[9] 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.

[10] PHILLIPS D.L. LEE E. H. HERSTOM A. A. HOGSETT W. E. TINGEY D.T. 1997. Use of auxiliary data for spatial interpolation of ozone exposure in southearn forests. Environmetrics, 8, 43-61.

[11] GALLOIS D., de FOUQUET C., LE LOC'H G., MALHERBE L., CARDENAS G. Mapping annual nitrogene dioxide concentrations in urban areas. Submitted to the 7. International geostatistics Congress, Banff, 2004.

[12] GALLOIS D. 2004. Optimisation des estimations spatio-temporelles par des méthodes géostatistiques. Rapport de stage ASPA. Ecole des Mines de Paris.