Geostatistical assessment of long term human exposure to air pollution

Nicolas JEANNEE, V. NEDELLEC, S. BOUALLALA, J. DERAISME, H. DESQUEYROUX

geovariances
THE UNIVERSE OF GEOSTATISTICS

ADEME

VNConsultants

ISATIS COMPANY

Objectives

To illustrate the efficiency of geostatistics in providing the basic figures to perform an Health Impact Assessment (HIA) of ambient air pollution.

Key points:

- HIA requires the accurate assessment of the population exposure to air pollutants.
- Case Study from UNECE-WHO Pan European Program for Transport, Health and Environment: “Transport-related health impacts and their costs and benefits with a particular focus on children”.
- Poor efficiency of linear estimation techniques (kriging/cokriging) to solve non linear problems and perform risk analysis

  ⇒ Stochastic simulations of PM10 that integrate:
  - correlation between PM10 and more densely acquired NO_2 data,
  - more recent PM10 data that supplemented the PM10 monitoring network.

- Population exposure to different levels of concentrations, once air concentration results are coupled with geo-data from the last national census (1999).

Contents

Introduction

Health Impact Assessment (HIA)

- Scientific approach that allows to forecast impact of air pollution on public health.
- Epidemiological studies investigate the relationship between
  - temporal variation of pollutant air concentrations (data from air monitoring network) and
  - health outcomes in the population (data from hospitals, other public health institutions, measured in representative sample of the population).

⇒ Exposure response function (ERF): estimate the number of cases (morbidity or mortality) for a given atmospheric concentration of a given air pollution indicator.

Specific HIA on transport-related air pollution

- Accurate assessment of the population exposure to chemical compounds that are indicators of transport-related pollution.
- Numerous epidemiological study results established ERF between PM10 (particulate matter with an aerodynamic diameter less than 10 micron) air concentration and increased frequency in many health outcomes.
- Preliminary step of this HIA: the assessment, with data from the French air monitoring network, of PM10 ambient air concentrations.

Introduction

Presentation of the approach:

- Data analysis and spatial modeling of average annual PM10 concentrations from existing measuring stations in France in year 2000 (interest is put on long-term exposure effects).

- Significant increase in the reliability of the results by taking into account:
  - the correlation between PM10 and more densely acquired NO2 data,
  - more recent PM10 data that supplement the PM10 monitoring network in otherwise entirely non sampled areas.

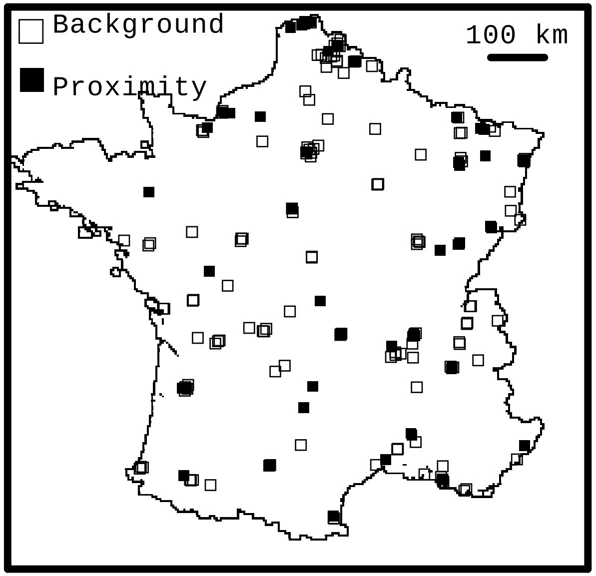
- Linear estimation techniques not adapted for non linear calculations and risk analysis $\Rightarrow$ conditional cosimulations of PM10 concentrations.

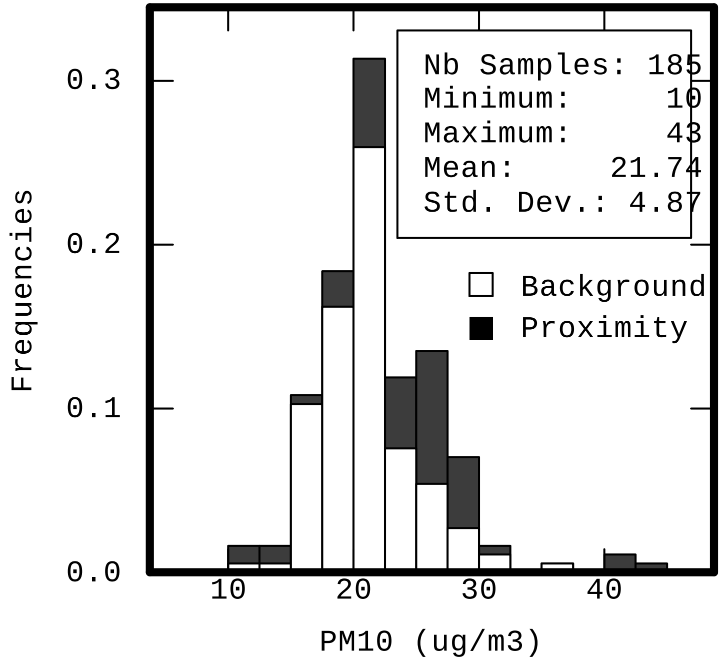
- Coupling air concentration results with geo-data from the last national census (1999), the population exposed to different levels of average annual concentrations is calculated.
- Statistical parameters from the resulting distributions are derived in the perspective of carrying out the HIA study on transport related air pollution.

- All geostatistical calculations performed using Isatis® software.

Spatial modeling

- Data analysis of average annual PM10 concentrations
  - 185 measured stations in year 2000 (54 proximity stations)

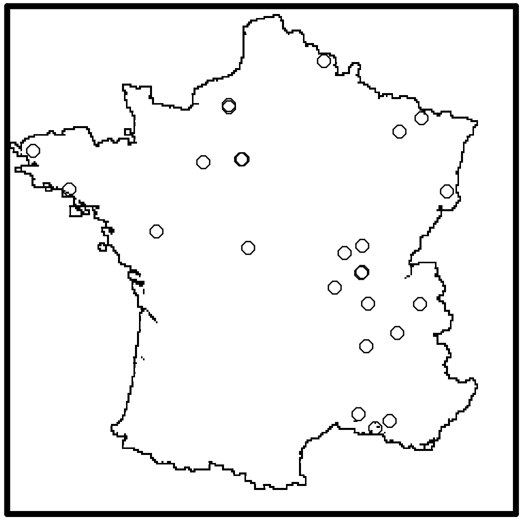


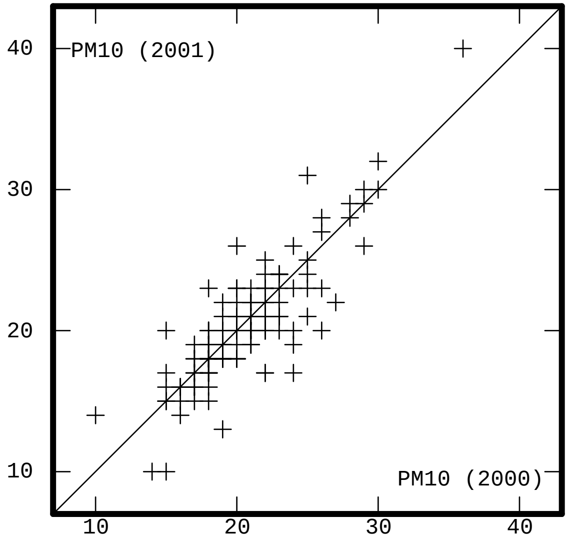


  - Proximity stations excluded from the analysis (lack of spatial representativity)

Spatial modeling

- Correlations:
  - PM10 measured in 2001:
    - 23 additional samples
    - Correlation coeff.: 0.84



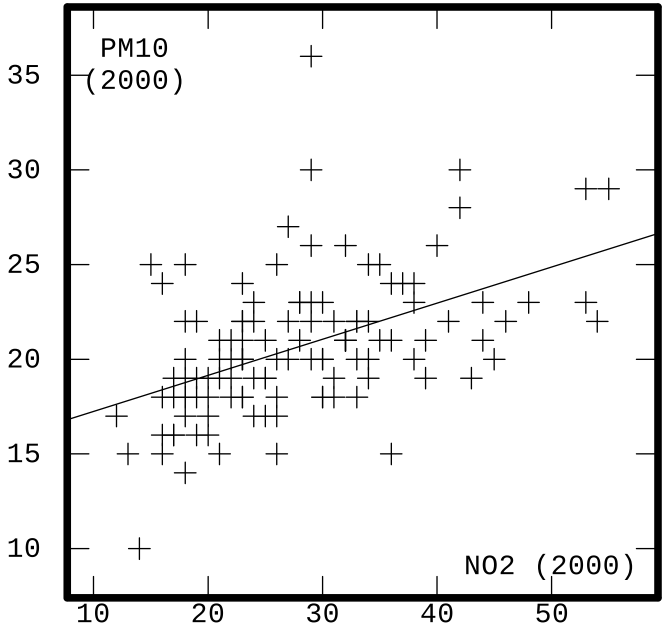


  - Possible models to integrate these additional data:
    - Standard cokriging between PM10 (2000) and PM10 (2001)
    - Kriging of PM10 (2000) completed by 2001 measures, the latter being penalized by a Variance of Measurement Error (equal to the variance of the residuals around linear regression)

Spatial modeling

Correlations:

- NO2 data:
  - Measured at 296 stations in 200(including 259 background stations
    (to be compared with PM10: 131)
  - Correlation coeff. with PM10: 0.49

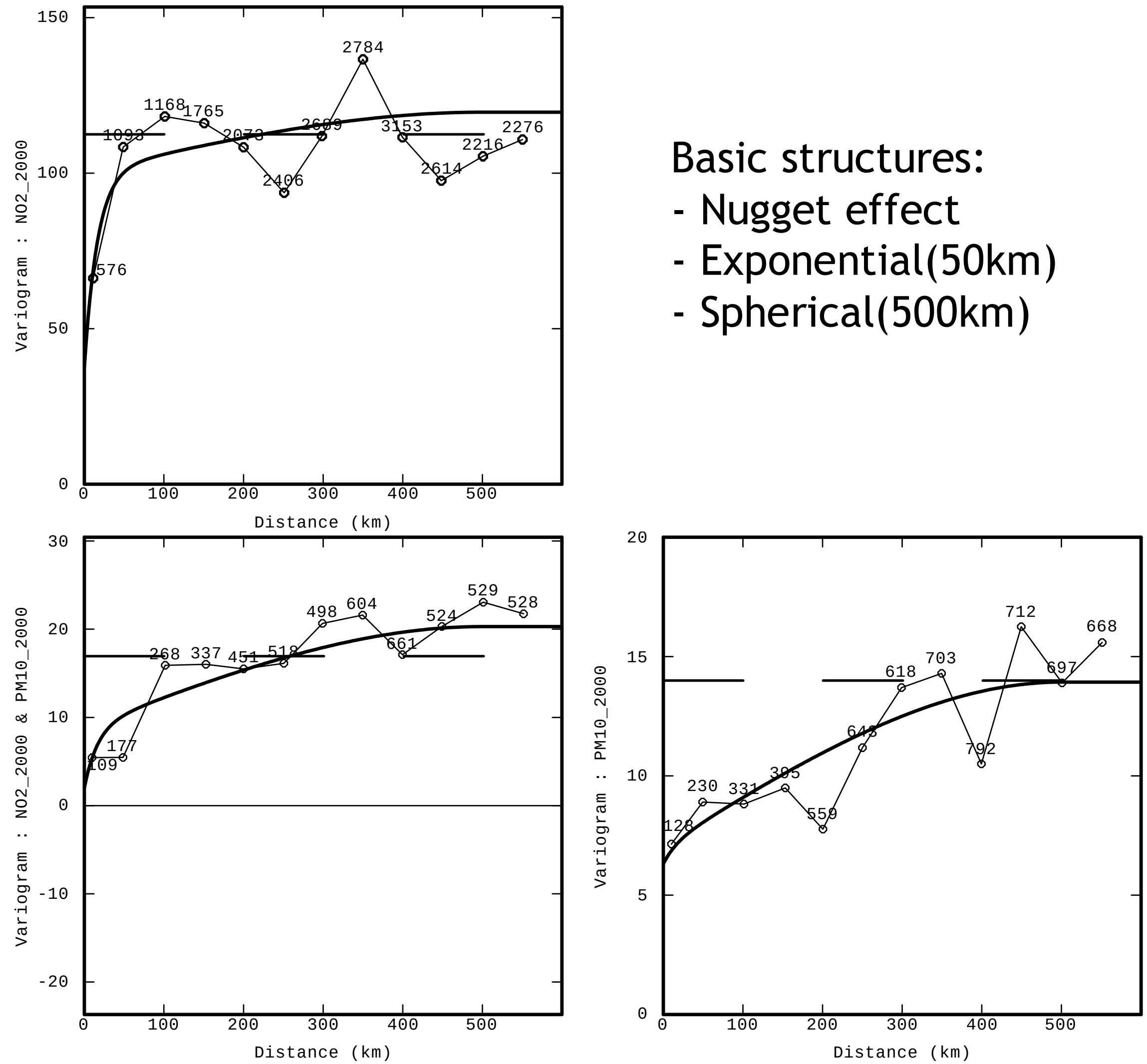


- Model: ordinary cokriging with PM10

Spatial modeling

- Variogram model

  (linear model of coregionalization)



Basic structures:
- Nugget effect
- Exponential(50km)
- Spherical(500km)

Spatial modeling

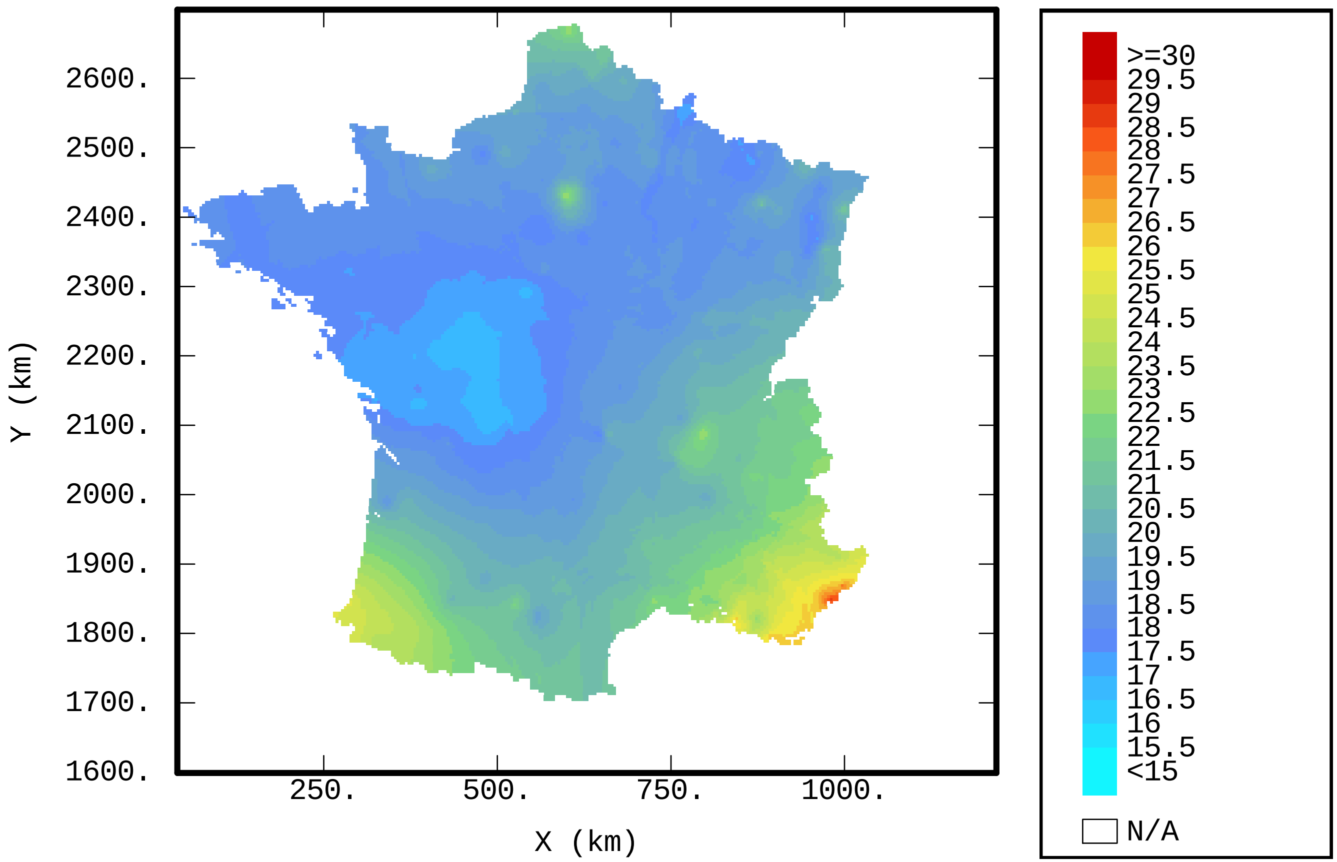
- Several models to estimate PM10 (2000):
  - Ord. kriging of PM10 (2000),
  - Ord. kriging of PM10 (2000 completed by 2001 with VME),
  - Ord. cokriging of PM10 (2000) and PM10 (2001),
  - Ord. cokriging of PM10 (2000 completed by 2001) and NO2 (2000).
- Validation:
  - Calculation of the Mean Quadratic Error (MQE) on five validation sets (22 to 30 datapoints per set):

	Validation set					
	1	2	3	4	5	Mean Rank
1/d2 PM10 2000	26,5	13,9	16,5	12,4	21,6	4,6
OK PM10 2000	**10,1**	7,4	15,4	**11,9**	6,8	2,8
OK PM10 2000 comp2001	11,8	**6,5**	12,2	14,9	4,2	2,8
OCK PM10 2000 / PM10 2001	**10,2**	6,8	12,4	13,4	6,4	3,0
OCK PM10 2000_comp / NO2 2000	10,5	6,7	**11,8**	12,2	**4,1**	1,8

(Ranked MQE, 1=best, 5=worst)
(Raw MQE)

Spatial modeling

- Resulting mapping of PM10 (2000)



Population exposure

- Poor efficiency of linear estimation techniques to solve non linear problems and perform risk analysis

  ⇒ **Stochastic simulations of PM10 using the model previously established.**

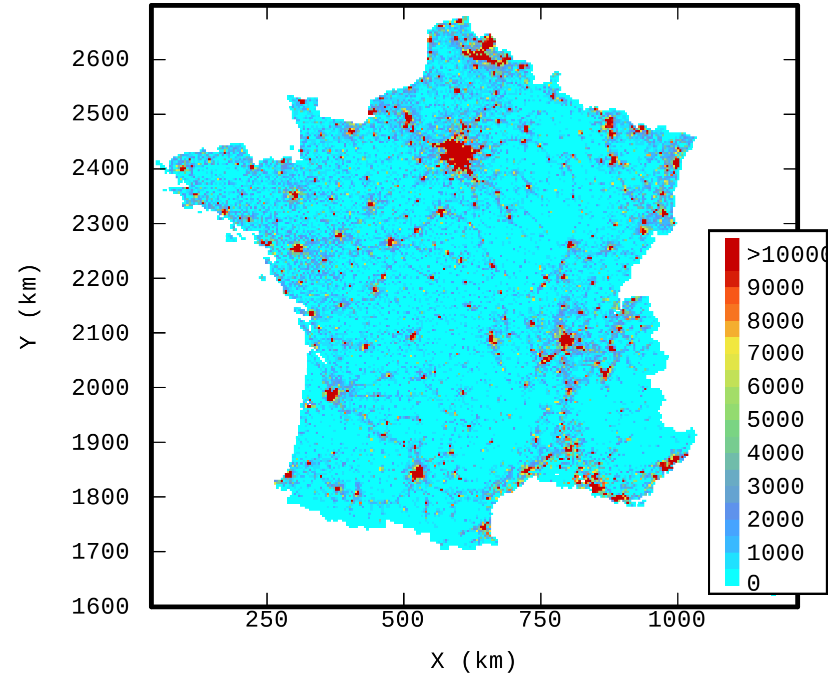
- Analysis of gaussian transforms (anamorphosis)
  - PM10 data have been transformed into gaussian data, and the anamorphosis function has been modeled (data clustering taken into account to avoid bias).
  - Though our interest is on PM10, NO2 has been transformed too:
    - correlation analysis and bivariable spatial structure between two gaussian transforms usually yields to better results and ensures the homogeneity of the process,
    - the Turning Bands co-simulation algorithm requires first the non conditional simulation of both variables, that should be gaussian.
  - Calculation and modeling of variograms of PM10 and NO2 gaussian transforms using the same basic structures as for the raw concentrations.

Population exposure

- Co-simulation of PM10 and NO2 concentrations
  - 200 conditional co-simulations have been performed using the Turning Bands technique, with 500 turning bands.
    - TB algorithm simplifies the 2D simulation in several 1D simulations along randomly generated lines, then reconstruct the 2D simulation by averaging the projected values from the 1D simulations (Matheron, 1973).
    - Number of turning bands: only parameter required to ensure the consistency of the resulting simulations (histogram and variogram reproduction).
    - Cosimulations obtained by simulating each basic structure, using the linear model of coregionalization decomposition.
  - Adequacy of this number of turning bands has been verified on a few simulations, in terms of histogram and variogram reproduction, before the back-transformation in raw scale.
  - Once simulations are obtained and validated, end of the geostatistical work…

Population exposure

- Population exposure:
  - Number of inhabitants known for each 4km x 4km grid cell (last national census of 1999).
  - For each PM10 simulation, computation of the population exposed to a given interval of pollution, for example: population exposed to PM10 concentrations between 5 and 10 μg/m³.



- Repeat for all simulations ⇒ distribution of the population exposed for example to an average annual PM10 concentration between 5 and 10 μg/m³.

- Characteristics about this statistical distribution derived for conducting the HIA:

	5-10 µg/m3	10-15 µg/m3	15-20 µg/m3	20-25 µg/m3	25-30 µg/m3	30-35 µg/m3	35-40 µg/m3
Mean	0,23	2,09	28,50	22,09	4,49	0,61	0,21
Std. Deviation	0,14	0,50	1,53	1,36	0,81	0,28	0,15
Quantile 2.5%	0,08	1,13	25,65	19,33	2,82	0,18	0,01
Quantile 97.5%	0,50	3,13	31,68	24,47	6,11	1,30	0,55

Conclusions (1/2)

- Efficiency of geostatistics to provide the basic figures of a specific HIA on air pollution, based on PM10:
  - The geostatistical framework offers the possibility to generate several realizations of the phenomenon of interest, here the annual PM10 concentrations.
  - Realizations obtained by means of conditional cosimulations between PM10 and NO_2 (Turning Bands algorithm)
  - Integration of complementary PM10 data from 2001 through a variance of measurement error approach.
  - Calculation of the population exposed to different levels of PM10 concentrations for each realization.
  - Statistical results are then used for carrying out the HIA.

Conclusions (2/2)

- ## Future work:
  - Part of the PM10 pollution specifically attributable to traffic.
  - Integration of auxiliary variables (like NO2, soil occupation, etc) does not replace information linked to the physico-chemical process of the pollution (obtained from detailed analysis of the emissions and transformation process, through a classical numerical simulation of transport).
  - The latter could be incorporated in the geostatistical method as an accurate cofactor (collocated cokriging, kriging with external drift).
  - Advantage of this model: integrate the actual data from the air monitoring network and the best knowledge on the pollution phenomenon.

- **Acknowledgments:** my co-authors, and the financial support of the French agency ADEME (Agence de l'Environnement et de la Maîtrise de l'Energie) through contracts nº 03 62 C0023 and nº 03 62 C0053.

Just in case…

- Kriging with variance of measurement error (VME)
  - Numeric values with varying level of precision might be available for the variable of interest. For example, the data may come from several surveys: old ones and new ones, the latter being more accurate due to advances in measurement techniques.
  - In such cases error variances albeit different for each sub-population may be known. Certain data might be assumed to have an error variance of 0, whilst some indirect or old measures are uncertain with a known error variance.
  - Assumption: instead of the "true" concentration value z_i we only know z_i+e_i with e_i a random error satisfying the following conditions for each sampling point *i*: $E[e_i]=0$, $Cov[e_i, e_k]=0$ for k ? i, $Cov[z_i, e_i]=0$ and $Var[e_i]=V_i$.
  - Kriging with variance of measurement error (VME) integrates these error variances. From a kriging system point of view, the VME approach simply consists in adding the V_i values to the diagonal covariance terms.