HAL
open science

Optimal spatial design for air quality measurement surveys

Thomas Romary, Laure Malherbe, Chantal de Fouquet

To cite this version:

Thomas Romary, Laure Malherbe, Chantal de Fouquet. Optimal spatial design for air quality measurement surveys. Environmetrics, 2014, 25, pp.16 - 28. 10.1002/env.2253 . hal-01089309

HAL Id: hal-01089309

https://minesparis-psl.hal.science/hal-01089309

Submitted on 2 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal spatial design for air quality measurement surveys

Thomas Romary * Laure Malherbe Chantal de Fouquet

December 1, 2013

Abstract

Measurement surveys using passive diffusion tubes are regularly carried out to elaborate atmospheric concentration maps over various areas. Sampling schemes must be designed to characterize both contaminant concentrations (of benzene or nitrogen dioxide for example) and their relations to environmental variables so as to obtain pollution maps as precise as possible. Here, a spatial statistical methodology to design benzene air concentration measurement surveys on the urban scale is exposed. In a first step, an a priori modeling is conducted that is based on the analysis of data coming from previous campaigns on two different agglomerations. More precisely, we retain a modeling with an external drift which consists of a drift plus a spatially correlated residual. The statistical analysis performed on available data leads to choose the most relevant auxiliary variables and to determine an a priori variogram model for the residual. An a priori distribution is also defined for the variogram parameters, whose values appear to vary from a campaign to another. In a second step, we optimize the positioning of the measuring devices on a third agglomeration according to a Bayesian criterion. Practically, we aim at finding the design that minimizes the mean over the urban domain of the universal kriging variance, whose parameters are based on the a priori modeling, while accounting for the prior distribution over the variogram parameters. Two global optimization algorithms are compared: simulated annealing and a particle filter based algorithm.

keywords Optimal Design, Geostatistics, External Drift Kriging, Air Quality

*thomas.romary@mines-paristech.fr

1 Introduction

Mapping air pollution as precisely as possible is a major issue for French Local Air Quality Monitoring Agencies (AASQAs) both for regulatory and information purposes and for public health concerns. Seasonal or annual average concentration maps can be obtained from passive sampling data collected at a large number of sites across the area of interest. In addition to continuous monitoring performed by automatic stations, sampling surveys are useful to evaluate the spatial distribution of atmospheric concentrations more precisely and provide information about air quality in any part of the territory, as required by the European legislation. Furthermore, spatial information provided by passive sampling surveys may be very helpful in addressing the question of the spatial representativeness of monitoring stations. This issue has been receiving increasing attention during those last years due to exposure concerns, cost constraints and air quality modeling considerations. Those reflections highlight the necessity of developing efficient sampling design methodologies which are able to integrate various data about the environment characteristics and so that the resulting concentration maps fulfill precision criteria.

With its internal quantification of spatial variability through the covariance function (or variogram), kriging methodology can produce maps of optimal predictions and associated prediction error variance from incomplete and possibly noisy spatial data, see *e.g.* Chilès and Delfiner (2012) and de Fouquet et al. (2007) for an application in atmospheric sciences. Through external drift (or universal) kriging, relevant environmental covariates can be taken into account to improve the prediction accuracy.

Various methods have been proposed in the literature for the problem of collecting spatial data, see Müller (2007) or De Gruijter et al. (2006) for a review. Specifically, several competing goals can be considered when sampling spatial data. The main interest can be in the quality of the resulting kriging map, in the sense of the prediction variance under known values of the covariance parameters, as in Van Groenigen et al. (1999) or Brus and Heuvelink (2007). Otherwise, we may want to estimate the covariance parameters accurately, as in Zhu and Stein (2005), or both covariance and drift parameters as proposed in Müller and Stehlík (2010). Information matrix based criteria have therefore been proposed. Providing sampling designs that satisfy both criteria is a particularly challenging task however. The key issue is how to incorporate the parameter uncertainty into design criteria to correctly represent the uncertainty in prediction. Several approaches have been proposed. Zhu and Stein (2006) consider to split the sample into two parts, each of them satisfying one of the criteria. Zimmerman (2006) proposes to minimize the empirical kriging variance, a criterion that endorses both aspects.

The purpose of this work is to design a sampling scheme for a fixed number of locations, taking into account the available information coming from previous surveys. The quality of the resulting kriging maps being our main goal, we also would like to control the accuracy of the parameter estimation as a small perturbation of a covariance parameter value may have a great impact on the prediction variance. The methodology developed relies on the definition of a quality criterion of the sampling design, based on the external drift kriging variance (see *e.g.* Chilès and Delfiner (2012)). A general family of criteria is proposed. In particular, a Bayesian criterion is introduced that allows to ac-

count for an *a priori* knowledge on the uncertainty of model parameters such as covariance parameters. The *a priori* distribution is defined from an analysis of the available data and our own knowledge about the phenomenon under study. The uncertainty about the parameters is accounted for afterwards. First a population of candidates designs is generated, then they are classified according to their performances in estimating the parameters of the covariance accurately.
Optimizing one of these criteria over a grid is a challenging combinatorial problem. Heuristic global optimization are needed and the adaptation of the simulated annealing algorithm (Kirkpatrick et al. (1983)) to the problem of optimal spatial design proposed by Van Groenigen et al. (1999) is generally used. In this paper, we describe a proposal kernel that helps to better explore the space of possible designs and hence to optimize the criterion faster. Besides the simulated annealing algorithm using the aforementioned kernel, we propose an interacting particles algorithm inspired by Cappé et al. (2004). The latter allows to generate a whole population of near optimal designs, whereas the simulated annealing proposes only one solution. This population can then be used to build a intensity map of the sampling process that can be used as a tool by the operator to effectively install the passive samplers. Indeed, at the agglomeration scale, downtown in particular, physical constraints may preclude from using exactly the location proposed by one optimal sampling design.
In Section 2, the modeling of the pollutant concentration is exposed. Within this settings different criteria are exposed and discussed. Section 3 presents and discusses the two global optimization algorithms that are proposed. The methodology is then applied in section 4 to the problem of designing a sampling scheme for a survey of the benzene concentration over the agglomeration of Bordeaux. Data from previous surveys were available to conduct the modeling step. Finally, we show how the proposed methodology can be used to dimension the network of passive samplers. This work completes and extends Romary et al. (2011).

2 Modeling and criterion definition

2.1 Modeling, prediction and inference

A classical spatial model is adopted for the pollutant concentration. The concentration is decomposed in a drift plus a spatially correlated residual:

$$Z(x) = \beta_0 + Y'(x)\beta + S(x), \tag{1}$$

where Z is the pollutant concentration variable, $x \in \mathcal{X} \subset \mathbb{R}^2$ is the spatial coordinate, Y is the matrix of covariates exhaustively known on $\mathcal{X}$, β is a vector of unknown parameters and $S(x)$ is a centered, spatially correlated residual with covariance C_θ, with parameters θ. No assumption is made over the distribution of S for the moment.
Once the covariance model has been fitted and given $\mathbf{Z} = (Z(x_1), \ldots, Z(x_n))$, $\mathbf{Y} = (Y(x_1), \ldots, Y(x_n))$ and the covariance matrix C_θ of S at locations $(x_1, \ldots, x_n)$, $Z(x_0)$ can be predicted by kriging with external drift (also called universal kriging):

$$\widehat{Z}(x_0) = \left(c_0 + \mathbf{Y}(\mathbf{Y}'\mathrm{C}_\theta{}^{-1}\mathbf{Y})^{-1}(\mathbf{Y}_0 - \mathbf{Y}'\mathrm{C}_\theta{}^{-1}c_0)\right)' \mathrm{C}_\theta{}^{-1}\mathbf{Z}, \tag{2}$$

where $c_0 = (C_\theta(x_0, x_1), \ldots, C_\theta(x_0, x_n))'$ and $\mathbf{Y}_0 = Y(x_0)$ is the vector of covariates at x_0. The associated prediction error variance, given by:

$$\begin{aligned}\mathbb{V}(Z(x_0) - \widehat{Z}(x_0)) = C(0) - c_0' \mathrm{C}_\theta^{-1} c_0 \\ + (\mathbf{Y}_0 - \mathbf{Y}' \mathrm{C}_\theta{}^{-1} c_0)' (\mathbf{Y}' \mathrm{C}_\theta{}^{-1} \mathbf{Y})^{-1} (\mathbf{Y}_0 - \mathbf{Y}' \mathrm{C}_\theta{}^{-1} c_0), \quad (3)\end{aligned}$$

accounts for the prediction error variance of the residual (second term of (3)) and for the prediction error of the drift as well (third term of (3)). Additionally, the prediction error variance depends on the sampling scheme η through c_0, C_θ and $\mathbf{Y}$ and not on the values of Z, when C_θ is known beforehand. It can be shown that the universal kriging predictor and the associated prediction variance corresponds to their Bayesian counterpart when Z is Gaussian, C_θ is known and an non informative prior is used for β, see Handcock and Stein (1993).

Throughout the above exposition it has been assumed that the covariance function $C_\theta(x, y)$ and its parameters are known. In practice, however, this is almost never the case. There exists a wide range of different techniques for estimating C from least squares based methods to maximum likelihood under Gaussian hypothesis through Bayesian methods, see *e*.g. Gelfand et al. (2011). We first describe the classical least squares approach in the case of a stationary covariance. This approach consists of different steps. First, the regression term is estimated by ordinary least squares (OLS). Then the empirical semivariogram of the residuals is computed by the following formula:

$$\widehat{\gamma(h)} = \frac{1}{2N(h)} \sum_{|x_i - x_j| \approx h} (\widehat{s}(x_i) - \widehat{s}(x_j))^2, \quad (4)$$

where $N(h)$ is the number of pair of points separated approximately by the distance h and the $\widehat{s}(x_i)$ are the residual from the fitted provisional mean function at location x_i. A parametric model $\gamma(h)$ is then fitted by weighted least squares to the empirical variogram. Then, the covariance matrix of the residuals is built upon the following formula: $C(x, y) = \widehat{C}(0) - \gamma(|x - y|)$, where $\widehat{C}(0)$ is given by the sill ($\max \gamma(h)$) of the fitted semivariogram. The regression parameters are consequently reestimated by generalized least squares (GLS) or EGLS (where E stands for estimated). If desired, the GLS residuals may be computed and the semivariogram reestimated from them. This procedure is necessary because the residuals obtained from the OLS estimation do not belong to the correct subspace, as the spatial correlation is not taken into account in OLS.

There exists two main problems with OLS estimation of covariance parameters however : the need of tuning the window size for the computation of the empirical variogram and the lack of uncertainty quantification of the estimates. The empirical variogram remains a necessary tool however to choose the parametric family of covariances that are to be used.

Under the Gaussian assumption and with particular assumption (Mardia and Marshall, 1984), the maximum likelihood approach allows to build an asymptotic estimate of the variance of the estimated covariance parameters. Indeed, under this model, the log-likelihood function of $\mathbf{Z}$ takes the form:

$$l(\beta, \theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log\left(\det(\mathrm{C}_\theta)\right) - \frac{1}{2} (\mathbf{Z} - \mathbf{Y}'\beta)' \mathrm{C}_\theta^{-1} (\mathbf{Z} - \mathbf{Y}'\beta). \quad (5)$$

By maximizing (5), we get the maximum likelihood estimators $\widehat{\beta}_n$ and $\widehat{\theta}_n$. Then the Fisher information matrix can be derived to provide an estimate of the inverse covariance matrix of the estimated covariance parameters:

$$\mathcal{I}(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} l(\beta, \theta) \left(\frac{\partial}{\partial \theta} l(\beta, \theta) \right)' \right). \tag{6}$$

Moreover, the (j, k)th element of $\mathcal{I}(\theta)$ takes the following form:

$$\mathcal{I}_{j,k}(\theta) = \frac{1}{2} \left(\mathrm{C}_\theta^{-1} \mathrm{C}_\theta^j \mathrm{C}_\theta^{-1} \mathrm{C}_\theta^k \right), \tag{7}$$

where $\mathrm{C}_\theta^j = \frac{\partial \mathrm{C}_\theta}{\partial \theta_j}$. Even if this result is theoretically correct under particular conditions and not in others, especially when the covariance parameters cannot be estimated consistently (see *e.g.* Zhu and Stein (2006)), it provides a reasonable indicator of the quality of the estimation of θ. We can see that (7) depends on the locations of the points through C_θ and its derivatives and not on the sampled values. The problem however is that (7) depends on the unknown value of θ. Therefore, it can only be used locally, say around an initial guess θ_0, or within a Bayesian framework when some prior information over the covariance parameters is known.

The information matrix of the regression parameters can also be derived and is obtained by a formula similar to (6). It is worth noticing that it is orthogonal to (6), as the terms $\frac{\partial l(\beta,\theta)}{\partial \theta \partial \beta}$ are null. Therefore, introducing the parameter estimation uncertainty in a design quality criterion can be done separately for θ and β.

Finally, when θ is estimated from the same dataset with which the prediction is performed, an additional uncertainty enters the prediction variance (3). It is generally approximated by

$$\mathrm{tr} \left(\mathcal{I}(\theta)^{-1} \left(\frac{\partial \lambda_\theta}{\partial \theta} \right)' C_\theta \frac{\partial \lambda_\theta}{\partial \theta} \right), \tag{8}$$

where $\lambda_\theta = \mathrm{C}_\theta^{-1} c_0$, that is the kriging weights, see Zimmerman and Cressie (1992). (8) accounts for the mean squared error of the difference between the predictions performed with the estimated θ and with the true value of θ.

2.2 Quality criteria of sampling designs

Equation (3) leads to a straightforward quality criterion of the sampling scheme when the aim of the study is to predict the pollutant concentration as accurately as possible:

$$O(\eta) = \frac{1}{|\mathcal{X}|} \int_{\mathcal{X}} \mathbb{V}(Z(x) - \widehat{Z}(x)) \mathrm{d}x, \tag{9}$$

where $|\mathcal{X}|$ is the area of $\mathcal{X}$. Here, the prediction error variance is averaged over the area under study. Therefore, sampling schemes obtained by minimizing this criterion will lead to accurate mapping of the concentration, performing a balance between the quality of the drift prediction and the spatial prediction. However, as shown in Romary et al. (2011), it may be of interest to target

preferentially some areas, for instance to map more precisely the pollutant concentration where it is likely to be high. This weight function can be designed to obtain a more accurate mapping in some areas, for instance as a function of the covariates, especially where we expect to encounter extreme values. Note that it will also be helpful for the subsequent variographic analysis once the survey has been performed. Indeed, it will make it possible to correct the undesirable effects of preferential sampling for the estimation of covariance parameters by assigning a weight to each datapoint.

Finally, when some parameters of the model cannot be fitted accurately or when their inference presents variations over the different available datasets, we can associate them an *a priori* distribution. Then a Bayesian version of (9) can be considered:

$$O_{Bayes}(\eta) = \frac{1}{|\mathcal{X}|} \int_{\Theta} \int_{\mathcal{X}} \mathbb{V}(Z(x) - \widehat{Z}(x)|\theta) w(x) p(\theta) dx d\theta, \tag{10}$$

where $\theta \in \Theta$ is the set of uncertain parameters and $p(\theta)$ is its *a priori* distribution. It is worth noticing that (3) is proportional to the point variance of Z and hence the optimal design (according to (9) or (10)) is independent of a multiplicative factor of the covariance function.

The estimation uncertainty of θ is not taken into account in (9) and (10). The error term (8) can be added but will increase the computational burden. A criterion based on the full Bayesian predictive distribution over $\mathcal{X}$ and (β, θ) could also be considered but remains infeasible, as each evaluation of such a criterion would involve a long MCMC run. As an alternative, we propose to consider the following criterion:

$$O_{param}(\eta) = -\int_{\Theta} \log \det(\mathcal{I}(\theta|\eta)) p(\theta) d\theta, \tag{11}$$

where $\mathcal{I}(\theta|\eta)$ is the Fisher information matrix associated to the design η. It can be seen as an average of the log variance of the likelihood estimates of the covariance parameters over the prior distribution. Minimizing this quantity will lead to optimal designs for the estimation of the covariance parameters. It seems therefore reasonable to think that a design that shows good performances in (11) will come along with a small value for (8) over $\mathcal{X}$, while reducing the computational burden as (11) do not involve an integral over $\mathcal{X}$.

The remaining problem is how to minimize (10) and (11) simultaneously. One could think about building a compound criteria as a weighted average or to split the design into two populations and perform the optimization sequentially on each. These solutions imply however to make arbitrary choices of the weights or the proportion of the population, otherwise to test a wide range of tuning parameters values and choose the one that produce the best results, assuming that the computational burden is not to heavy. Splitting the sampling design into two with varying proportion has been proposed in Zhu and Stein (2006). In their conclusions, they highlighted the fact that the best results where obtained when the proportion of additional points (added with respect to an estimation criterion) was below 10% of the overall population. Therefore, we propose to generate a population of candidate designs with good performances with respect to (10) and then to assign them a weight given by (11). This will makes more sense after the description of the interacting particles algorithm.

3 Optimization algorithms

Once the covariance model, the relevant auxiliary variables and the prior distribution for θ have been identified, the criterion (10) has to be optimised with respect to η. Its computation involves an integral over $\mathcal{X}$ that is evaluated on a grid discretizing the domain. When considering a fine grid, it amounts to choosing the best combination of n locations among all points of the grid. It is therefore a hard combinatorial problem. Considering all possible combinations is practically out of reach. Moreover, there is no closed-form expression of the gradient of (10) and a steepest descent approach is not feasible. Therefore, we have to rely on heuristic optimization algorithms to perform the global optimization: the standard Simulated Annealing (SA, Kirkpatrick et al. (1983)) scheme and an Interacting Particles Algorithm (IPA, inspired by Cappé et al. (2004)) are proposed. The variant of the SA algorithm proposed here is depicted in algorithm 1.

Algorithm 1 Simulated annealing

```
Generate η_0
for i = 1 ... N do
  Simulate ζ from q_{η_{i-1}}
  Simulate u ~ U_[0,1]
  if u < α = min(exp((O(η_{i-1}) − O(ζ))/T), 1) then
    η_i = ζ
  else
    η_i = η_{i-1}
  end if
  Decrease T
end for
```

Since the seminal work of Kirkpatrick et al. (1983), inspired itself by Metropolis et al. (1953), SA has been widely used and is well documented, see *e.g.* Müller et al. (2004) for an application to optimal design. The fundamental idea of this algorithm is that a change of scale, named temperature (T in algorithm 1) , allows larger moves on the surface to optimize and therefore avoid theoretically to remain trapped in a local minimum.

Practically, SA is an iterative optimization algorithm in which a sequence of states (η_i) is generated by deriving a new solution from randomly modifying the previous state by the inhomogeneous kernel $q_{\eta_{i-1}}$. Each time a new sampling scheme is generated, the criterion is evaluated and compared with the value of the previous state. If the criterion value has been improved by the change, the new state is systematically accepted. Otherwise, the simulated annealing algorithm may accept changes that worsen the criterion according to a given probability, given in algorithm 1. This helps to avoid remaining trapped in a local minimum.

The main shortcomings of simulated annealing are well known, particularly the need of a tedious tuning of the parameters and the uncertainty of finding the global optimum in practical applications. Indeed, theoretical results point out that in most cases a logarithmic decrease of the temperature is necessary to ensure the asymptotic convergence, a condition that is never met in practice.

One generally uses a geometric cooling schedule with a factor between 0.99 and 0.9. Furthermore, the propositional kernel needs also to be tuned as it may vary with respect to the temperature: at high temperatures SA authorizes the sampling scheme to vary widely while at low temperatures only slight modifications of the current state will be accepted at a satisfactory rate.
Finally, because of its stochastic nature, SA will not produce the same results from one run to another, especially if the objective function to optimize presents several local minima. It may be the case in sampling scheme optimization where the objective functions are generally flat with lots of local minima, see *e.g.* Müller et al. (2004) and Amzal et al. (2006). Therefore, there should not be *one* but *a population* of optimal designs, corresponding to different minima of the criterion. Going further, this population of optimal designs can be seen as the modes of a distribution over the space of the sampling designs. Simulating from this distribution would make it possible to generate a population of optimum designs or even a density representing the probability of a location to be chosen to position a sensor. Based on these tools, the practitioner can then choose the design that best fits with practical constraints
This reflection leads us to propose the algorithm 2 whose goal is to simulate from a distribution of sampling designs rather than to find one optimum design, while simplifying the parametrization with respect to SA. Indeed, IPA will produce a whole family of weighted solutions without using a temperature parameter. Here, instead of modifying sequentially one sampling design, a whole population of n particles (n sampling designs) is considered. At each iteration, each of the n particles is modified. Then the $n/2$ best particles are selected and resampled according to their weights. The weights used for the resampling are function of the objective function value and of the global distance $\Delta(\eta^{(j)})$ between the considered particle $\eta^{(j)}$ and all the others. This distance is introduced so as to hinder the degeneracy of the population and is defined as follows:

$$\Delta(\eta^{(j)}) = \exp\left(\frac{1}{\Omega}\sum_{i\neq j}\sum_{l=1}^{n}|\eta^{(j)}(l) - \eta^{(i)}(l)|\right),$$

where Ω is a normalizing constant and $|\eta^{(j)}(l)-\eta^{(i)}(l)|$ is the Euclidean distance between the points l of the designs j and i. This algorithm has the advantage to produce a whole population of candidate designs without any tedious parametrization. It is also straightforwardly parallelizable.

Algorithm 2 Interacting particles algorithm

Generate $\eta_0^{(1,\ldots,n)}$
for $i = 1\ldots N$ **do**
  for $j = 1\ldots n$ **do**
    Simulate $\tilde{\eta_i}^{(j)}$ from $q_{\eta_{i-1}^{(j)}}$
  end for
  Select the $n/2$ best particles $\tilde{\eta_i}^{(1,\ldots,n/2)}$
  Generate $\eta_i^{(1,\ldots,n)}$ by resampling $\tilde{\eta_i}^{(1,\ldots,n/2)}$ with weights $\exp\left(-O(\tilde{\eta}_i^{(j)})\right) \times \Delta(\tilde{\eta_i}^{(j)})$
end for

Once the population of candidate designs has been generated, the criterion for the estimation parameters (11) can be computed for a carefully chosen subsample. By subsampling or reweighting the design population with respect to the value of (11), we can produce the intensity map of the sampling process by a kernel density estimation.
In both algorithms, the following propositional kernel is used:

$$q = pq_1 + (1-p)q_2,\ p \in [0,1], \tag{12}$$

where q_1 replaces one or several uniformly selected points of the current design by one of its neighbors on the grid that is distant from less than h_i, and q_2 replaces one or several uniformly selected points by one of its n_i nearest neighbours in the covariates cloud. This kernel helps to explore faster the covariates cloud and therefore accelerates the convergence, with respect to Van Groenigen et al. (1999) and Brus and Heuvelink (2007) where only the kernel q_1 is used. In SA and IPA, the search radius h_i, the number of neighbors n_i and the number of points that are affected decrease along the algorithm.

4 Application

In this section, we apply the methodology described above to generate an optimal sampling design for a monitoring survey of the atmospheric benzene concentration by passive diffusion tubes over the agglomeration of Bordeaux. In our application, we would like to use all the information about the spatial behaviour of the air concentration of pollutants gathered in previous surveys, not necessarily on the same agglomeration the sampling design has to be generated. The estimation of β and θ is performed on each available dataset. For the covariance estimation step, it can be useful to consider a single parametric family. Indeed, in the approach described below, it will be easier to account for the uncertainty of a parameter value than that of a covariance model. Then the variation range of the estimated covariance parameters is considered to build a (informative) uniform prior. To sum up, the study of the available datasets makes it possible to select the auxiliary variables and to define an *a priori* distribution on the covariance parameters that will be used in the following. In practice, the estimation value of the regression parameters vary largely from one agglomeration to another, depending on the urban characteristics of each one. Therefore it makes sense to consider a non informative prior on β, which is taken into account by the universal kriging based criterion. The estimation step is described briefly prior to the results. For more information on the latter step, we refer the interested reader to Romary et al. (2011).

4.1 Model estimation

A geostatistical analysis of the data collected during previous surveys is performed in order to set up the model (covariates, covariance function and priors) to use when applying the optimization method to another agglomeration. Data from benzene concentration monitoring surveys conducted in two French cities (Lille and Reims) are used to fit the geostatistical model (1). For each survey of each agglomeration, a variable selection is performed among the translated logarithm of land cover variables, population density and their cross products.

Among a preselection obtained by LASSO due to the large number of possible predictors (see Tibshirani (1996) and Friedman et al. (2010) for the associated R package), most relevant variables are identified by stepwise backward elimination (see *e.g.* Saporta (2006)).

For Lille surveys, a heteroscedastic variance is fitted, accounting for the in-

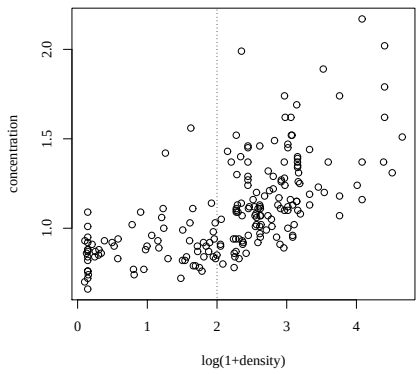


Figure 1: Mean benzene concentration as a function of the population density (Lille).

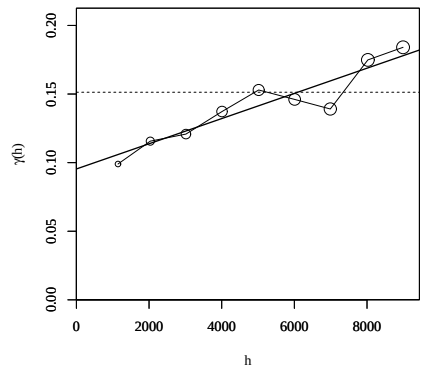


Figure 2: Experimental and fitted variogram of the residuals (Lille, average of two surveys).

crease of the benzene concentration variability as a function of the population density: the intraclass variance appeared to be two times larger for population density values higher than the median, see figure 1. Owing to this, weighted least squares were used to estimate the associated linear regression models, where the weights were chosen to be the inverse of the intraclass variance. Urban fabric proportion and population density were finally retained as prior covariates as they appeared relevant in most surveys. Then, for each survey, the variogram of the residuals has been fitted with the automatic variogram fitting function of the package Renard et al. (2010) for R. Note that the heteroscedasticity had been previously filtered out from the residuals by the weighted least squares estimation step. Therefore, a model consisting of an heteroscedastic nugget effect plus a linear spatial structure appeared to be the most adequate for all the surveys. Finally, the following covariance model has been retained:

$$C(x,y) = \theta_1(1 + \mathbb{1}_{\{Y_1(x)>m_1\}})\mathbb{1}_{\{|x-y|=0\}} + \theta_2(1 - \frac{|x-y|}{R}) \ , \ x,y \in \mathcal{X}, \quad (13)$$

where $R = \max_{(x,y)\in\mathcal{X}^2}(|x-y|)$, Y_1 stands for the population density, m_1 is its median over the domain, θ_1 and θ_2 are two varying parameters. The spatially varying nugget effect accounts for the heteroscedasticity related above. It has been observed on Lille but not on Reims. It is retained in the prior modeling as the typology of Bordeaux is closer to that of Lille than to that of Reims, in terms of size, population density and urban fabric proportion. The second term of (13) corresponds to a linear spatial structure and was relevant for all surveys. Hence, $S(x)$ can be decomposed in two terms : the first one accounting for a

heteroscedastic punctual variance and the second one accounting for a linear spatial structure.
Finally, note that (13) is a generalized covariance function: there is no covariance model associated with the linear variogram, as it corresponds to a non stationary model. However when considered on a bounded domain a generalized covariance function can be built to simplify the computations, see Chilès and Delfiner (2012) for more details on generalized covariances.
The estimates of θ_1 and θ_2 were found to vary among the different studied datasets. Therefore, we chose to consider them as uncertain and to use the Bayesian criterion (10). As noted in section 2.2, this criterion is proportional to a multiplicative factor of the covariance function. This allows us to set $C(0) = 1$, where $Y_1(x) < m_1$, and $\theta_1 = 1 - \theta_2$, so that only one parameter varies. This enables to simplify expression (13) to get:

$$C(x,y) = 1 - \frac{|x-y|}{R} + \theta_1(\mathbb{1}_{\{Y_1(x)>m_1\}}\mathbb{1}_{\{|x-y|=0\}} - \frac{|x-y|}{R}) \ , \ x,y \in \mathcal{X}. \quad (14)$$

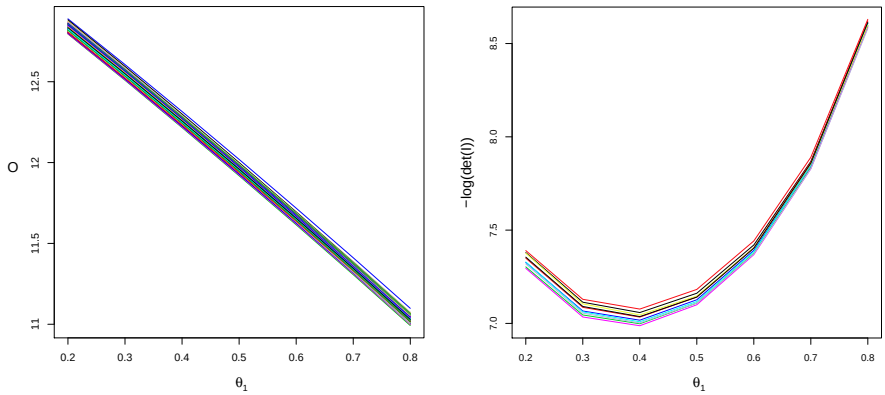
The *a priori* distribution of this parameter has been set to the uniform distribution over $[0.2, 0.8]$ so as to include all situations encountered in the model estimation step.
Concerning the regression parameters, they were find to vary widely from one survey to another, even on the same agglomeration. Therefore, the choice of a non informative prior, through the use of the universal kriging variance, is justified.

4.2 Results

The French agglomeration of Bordeaux is taken as an application case. The domain size is $38km \times 45.5km$, it is discretized on a 184×154 grid with a 250m step. The criterion (10) is optimized with the following weight function

$$w(x) = 1 + 9\mathbb{1}_{\{Y_1(x)>m_1 \cap \{Y_2(x)>m_2\}}(x),$$

where m_2 stands for the median of the urban fabric proportion. Such a criterion assigns a relative weight of 10 to points belonging to areas where both population density and urban fabric proportion exceed the respective median observed on the domain. Furthermore, to speed up the calculation, a coarser grid, with a 500m step, is used on areas where the weight function value is low. The justification for this is that since a low number of points will lie in those areas, it is unnecessary to offer a large choice of locations. This allows to reduce the number of grid points from 28336 to 11940, hence to reduce the computational burden by a factor 2.4. The weight function is of course modified so as to maintain the desired weights.
Moreover, the evaluation of the integral with respect to $\theta = \theta_1$ in (10) can be performed with a simple numerical integration scheme as the function that maps θ_1 to the mean universal kriging variance is linear over the interval $[0.2, 0.8]$, as shown in figure 3, although it is not obvious from (14). This allows to evaluate the integral by using solely the midpoint of the interval to evaluate it. It terms to compute (10) with $p(\theta_1) = \mathbb{1}_{\{\theta_1=0.5\}}$.
The determinant of the Fisher information matrix is not linear in θ_1 however. This implies that the numerical integration scheme to evaluate (11) requires



a. b.

Figure 3: Mean universal kriging variance (a.) and -log determinant of the Fisher information matrix as a function of θ_1

more than one single point. We therefore decide to compute (11) by the trapezoid formula over [0.2,0.8] discretized with a 0.1 step. The performances of both algorithms are compared, in terms of optimization quality and computing time on the problem of placing 105 sampling points. Finally, the criterion (11) is computed for the optimal designs generated. This number of samplers was chosen as the number of samplers used for previous surveys on the agglomeration of Bordeaux, taking into account the budget constraints.

The SA algorithm was run until no moves were accepted for 200 iterations, that is until the acceptation rate became less than 0.005 over the last iterations. The initial temperature was chosen such that the initial acceptation rate of worsening proposition was greater than 0.9, following the rule of thumb given in Kirkpatrick (1983). The temperature was lowered exponentially with a geometrical rate of 0.95. The search neighbourhood radii, in the physical space and in the covariates cloud, as well as the number of points moved, decreased along the algorithm together with the temperature. The initial design was generated randomly. The algorithm was implemented with R (R Development Core Team, 2010) and took about 4 hours to run on a desktop computer.

Figure 4 presents the results provided by the SA algorithm. It is made of 4 graphs. In the upper left corner is depicted the curve of the criterion value of each accepted design along the algorithm. The profile of this curve is typical of a SA algorithm: when the temperature is high, worsening states are more likely to be accepted than when it is low. Indeed, this curve becomes more regular along the iterations. In the upper right corner is plotted the empirical acceptation rate over all iterations: it starts from values close to 1 and decreases gently towards lower values. Note that over the last 200 iterations, its value is 1/200. The two lower figures represent the optimal sampling design obtained by SA. We can claim that the criterion (10) has been properly optimized. Indeed, the lower left plot, where the sampling design is plotted together with the urban fabric proportion, shows that most of the points of the resulting sampling scheme are

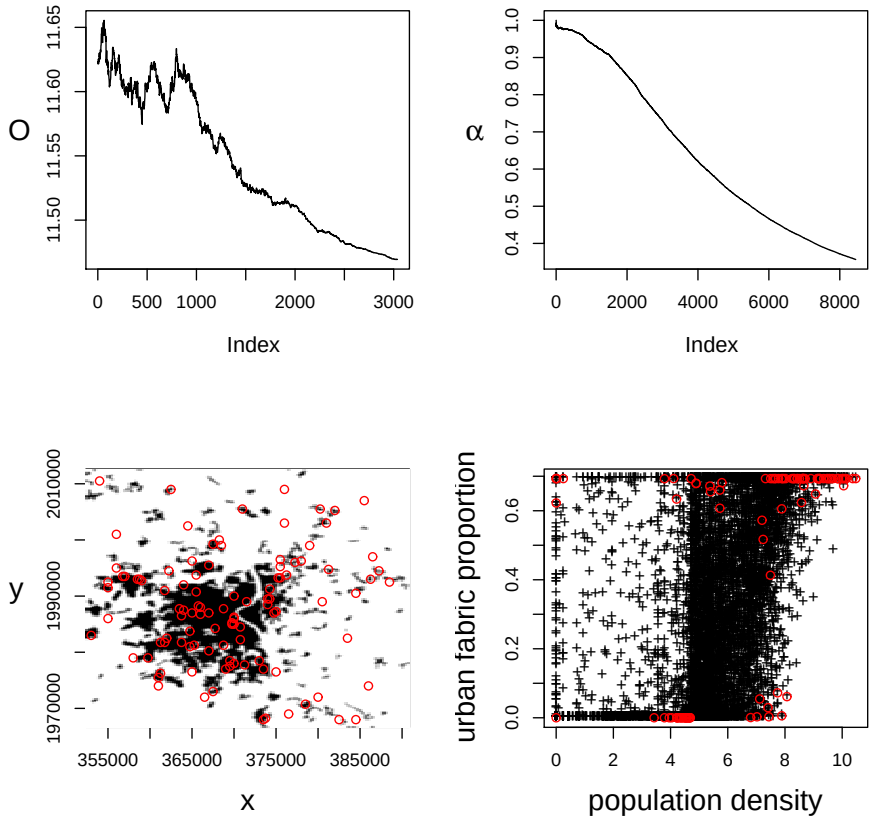


Figure 4: Optimal sampling design for Bordeaux (SA). The upper left corner panel depicts the decrease of the criterion value along the algorithm; the upper right, the decrease of the empirical acceptation rate of a transition. Lower left: sampling design in the physical space and urban fabric proportion. Lower right: sampling design in the covariates cloud.

placed in urban locations while the others cover uniformly the surroundings of the agglomeration. This indicates that the weighting has been respected while the infill effect of minimizing the average variance over the whole domain is present. The lower right picture represents the sampling points in the covariates cloud. Most of the points are located close to the bounds of each covariate and close on the left to the limit defining the nonstationary nugget effect which is the median. This behaviour was expected as it is well known that the best design to estimate a linear regression by ordinary least squares is to have one point at each extremity of the regression line. As the drift is implicitly estimated by *weighted* least squares, due to the nonstationary covariance model used here, additional points near the limits defining the non stationary variance have been generated with a number of points on both sides that depends on the weighting function, see Appendix A for more details. Therefore, the occurrence of this pattern indicates that the criterion has been properly optimized. Finally, SA has produced an optimal sampling scheme that makes the balance between a good spatial prediction over the domain and the estimation of the drift, while respecting the weighting imposed for the densely populated and constructed areas.

Finally, the computed value of (11) was found equal to 4.49, to be compared with the results obtained by the designs generated by IPA.

Figure 5 represents the results obtained with the IPA. It has been run for 1000 iterations with 50 particles and took about 8 hours to run on a dual processor computer. No acceptation rate is plotted since the transitions are always accepted in this algorithm. The graph on the upper left shows the criterion value curves (one for each of the 50 particles) as a function of the number of iterations while that on the upper right represents the *best* (in the sense of the Fisher information matrix based criterion (11) sampling scheme generated in the physical space, together with the urban fabric proportion over the agglomeration. The decreasing profile of the criterion curves is absolutely different from that observed with the SA (figure 4). It is very fast during the first 100 iterations, then less fast between 100 and 400 iterations and finally the curves seem to attain a floor beyond 400 where they exhibit a stationary behaviour. Moreover, the minimum value reached for the criterion is significantly greater than the one found by SA.

The *best* design that is represented in both physical and covariates spaces has a value of (10) of about 11.5. The corresponding value of (11) is slightly lower however (4.48). The representation in the space of covariates is given in the lower left corner. It is very similar to that of figure 4, except that the points are less concentrated near the bounds of the auxiliary variables values. If we look at the design that performs the best in terms of average kriging variance among the whole population however, it is slightly higher than the one found by SA (4.53). This value is actually one of the worse among the generated population. Overall, the value of (11) is comprised between 4.48 and 4.54.

The graph in the lower right corner of figure 5 depicts the kernel density estimate of the point locations computed with the *best* 1000 designs generated by the IPA. They have been subsampled from the whole population by selecting the best 1000 designs according to (11), with values between 4.48 and 4.50.

Although this density is estimated from only *near* optimal designs, it gives an insight on the likely locations of design points. Indeed, the optimal design generated by SA shares some patterns with this density: the high concentration of

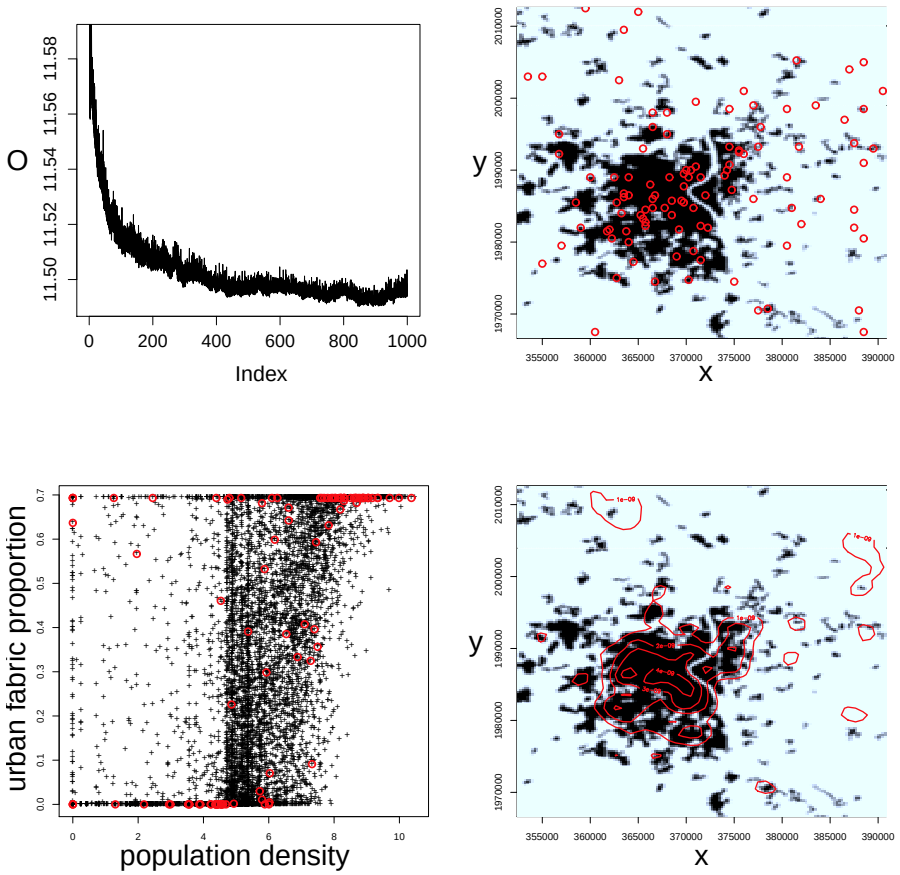
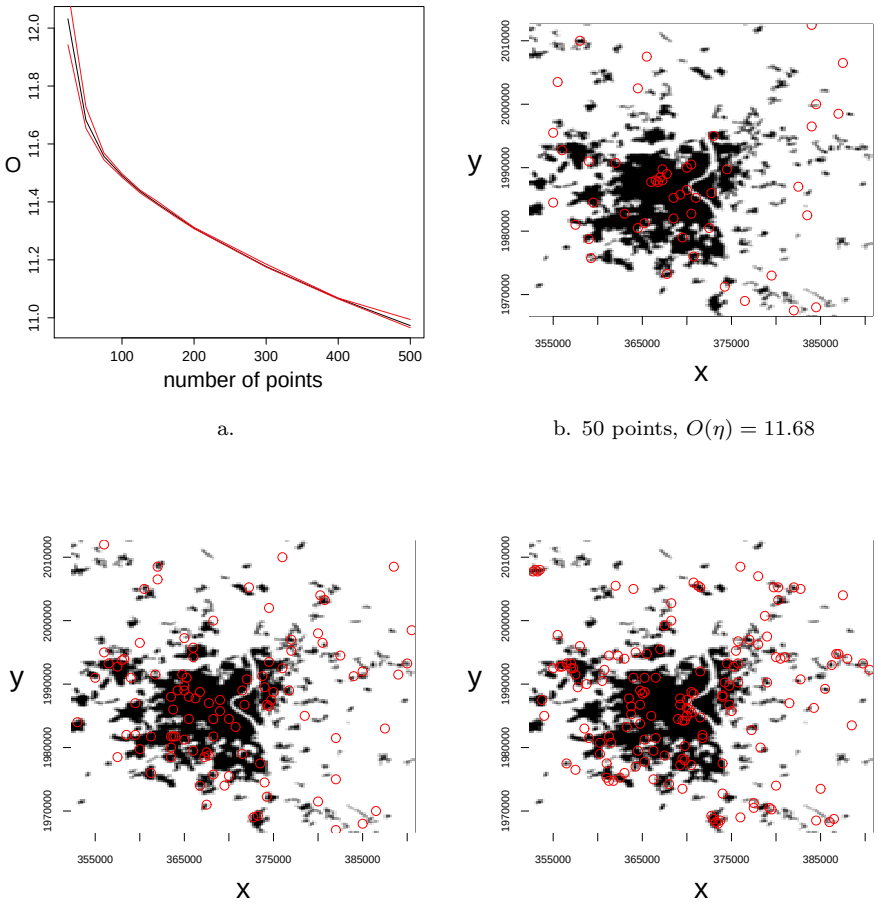


Figure 5: Optimal sampling design for Bordeaux (IPA). The upper left corner panel depicts the trajectories of the criterion value along the algorithm for each particle; the upper right, the best sampling design in the physical space. Lower left: sampling design in the covariates cloud. Lower right: Contours of the kernel density estimate of the sampling points over the subsampled population.

points in the center of the city and the almost uniform location of the points outside the urbanized areas. Some artifacts exist however: the mode situated on the north edge of the agglomeration seems odd. This may be due to the small size of the population used. A larger population may have given better results, as it would have performed a better exploration of the set of the designs, but the numerical burden would have been heavier. Nevertheless, this density estimate illustrates that there is not a single optimal sampling design but a whole family of near-optimal designs (as can be seen when running several times SA with different random seed), which shares the patterns of this estimated density.
The fast decrease of the objective function during the first iterations of IPA can be explained by the ability of the algorithm to explore quickly the space of possible designs and to select the best design. Its inefficiency to optimise the criterion properly however is certainly due to the fact that no constraint is imposed on the decrease of the criterion value, contrarily to SA. Even if this would drive away from the objective of simplifying the parametrization, the good properties of this algorithm could be exploited in a combination with SA. For instance, several near-optimal designs provided by IPA could be used as starting points for SA.

4.3 Dimension of the sampling design

By running several times the SA (with different seeds) with an increasing number of points, a decision support for the size of the sampling design can be determined. In figure 6 a., the criterion (10) is plotted as a function of the number of points, in black, while the red curves are made of the maximum and the minimum of the criterion among 10 different runs. The criterion decreases quickly from 25 to 100 points then shows an almost linear decrease. Moreover, the range of the variation interval of the criterion value decreases dramatically as the number of points increases except when considering 500 points. This illustrates a good stability of the results for this implementation of the SA. The variability observed for the last case is certainly due to the incomplete convergence of the algorithm: it reached the maximum number of iterations, set to 20000, prior to fulfill the stopping criterion. While adding points at the beginning decreases drastically the mean prediction variance, there seems to be a threshold (around 100) beyond which adding new points will only have a impact on the prediction variance at a scale inferior to the discretization step. Therefore, beyond this threshold, adding points does not affect so much the perceived prediction quality, the criterion decrease being explained by the mechanistic effect of subtracting one term to the sum that approximates (10). Finally, we can conclude from figure 6 a. that a necessary number of sampling points is to consider, which is about 100 in our case, to get high precision maps at a given resolution. Note that adding more points is still useful if the budget allows it. In that case, it may be interesting to reduce the discretization step. Three resulting designs are depicted in figure 6 b.,c. and d., respectively for 50, 100 and 150 points. The evolution of the design when the number of points increases is mainly characterized by two patterns: an improvement in the spatial coverage of the low-weighted areas and an increased density of points in high-weighted areas.
Drawing such a figure can be useful for future users. The size of the sampling scheme can be decided as a function of the desired level of precision the forth-



a.

b. 50 points, $O(\eta) = 11.68$

c. 100 points, $O(\eta) = 11.49$

d. 150 points, $O(\eta) = 11.39$

Figure 6: Criterion as a function of the size of the network (a.) and several optimal sampling designs of different sizes (b.,c.,d.)
generated by SA.

coming map should meet. Particularly, the responsible of an AASQA can see the level of precision that he can afford and decide to amend the budget devoted to the survey accordingly, especially in order to reach the necessary number of points defined above. The possible losses of sensors due to dysfunction or vandalism (around 10% in average) can also be taken into account a priori.

5 Conclusions

The current work exposes an optimal design methodology for the problem of generating sampling designs for measurement surveys of air quality, where the objective is to map the concentration of pollutant as precisely as possible at

the considered scale. This methodology allows to generate optimal sampling schemes based on available information, including known relationships between the pollutant and environmental variables as well as uncertainty about the covariance parameters. A discussion about the different possible criteria lead to choose a pragmatic solution: first optimize the Bayesian weighted average kriging variance over the field, then order several near optimal designs according to their performance in terms of covariance parameters estimation quality. A new proposal kernel for the global optimization algorithms and an adapted interacting particles algorithm have also been introduced. The methodology has been applied here to benzene sampling over urban areas but it is general and can be straightforwardly extended to the study of other spatial phenomena and to other spatial scales.

We considered the problem of the existence of possible multiple optima of the objective function through the use of an interacting particle algorithm. Estimating the density of the sampling points generated gave an insight of the patterns that a good sampling design should exhibit. It may also be more useful to the end-user to have a indicative map of the locations rather than a single "optimal" design, as practical installation of passive samplers requires flexibility due to physical constraints.

In particular, this map can be obtained by subsampling the population of designs according to different criteria. Here, we chose to select the design with respect to the quality of the forthcoming covariance parameters estimation. It is worth noticing however that the *best* design among the population produced by the IPA, chosen according to the Fisher information matrix based criterion, and the *best* design produced by the SA algorithm are very similar in terms of criteria values and spatial patterns. It seems that the prominent patterns that a good design (according to our criteria) should exhibit are: a domain covered uniformly overall with a high concentration of points within the urban centers with some close pairs, points well distributed in the covariates cloud, so as to perform a good estimation of the regression part of the model. These results recall those already shown in Diggle and Lophaven (2006). In that paper, the authors computed a Bayesian criterion based on the predictive posterior distribution for a selection of designs that were built upon geometrical considerations. The results showed that the best performances were obtained for designs that fills the space uniformly with additional close pairs. Optimizing such a criterion remains intractable, as a MCMC run is required for each evaluation. A promising solution to speed up the calculations however would be to use the recently developed SPDE-INLA methodology (see Lindgren et al. (2011) and Rue et al. (2009)) that decreases drastically the computational burden. The theoretical assumptions may be too restrictive for pollution however.

The main problem with the Bayesian approach indeed is that it relies on strong assumptions about the distribution of the data that may not be fulfilled. The data from the available surveys were indeed far from being normally distributed with possible extreme values. It remains to define a more suitable probabilistic model to the pollutant concentration distribution and to adapt the criteria accordingly. In this work, we found it reasonable to consider that the larger weight given to urban areas would force the points to better cover the areas where extreme values can occur but this approach can seem insufficient.

We finally studied the effect of the dimension of the sampling design on the quality criterion value of the optimal design generated by SA. The plot of the

curve, that represents the optimal value (found by SA) of the criterion as a function of the number of sampling points, leads us to define a *necessary* number of sampling points that corresponds to the number of points beyond which this curve becomes almost linear, indicating that the effect of adding new points becomes less important.
The proposed methodology has been developed with the aim of supplying scientific and technical support to the French local air quality monitoring agencies. For the moment it has been applied to benzene sampling over urban areas but it can be extended to other pollutants such as NO2, to larger spatial domains like regions or even other applications like soil sampling.

Acknowledgements This work has been carried out for the French central laboratory for air quality surveillance. It has been supported by the French ministry responsible for the ecology and sustainable development. We also thank the associations AIRAQ, ATMO Nord-Pas-de-Calais and ATMO Champagne-Ardenne for having supplied the data.
We are finally grateful to two anonymous reviewers who greatly helped to improve the manuscript.

A Optimal sampling for the linear regression parameters estimation in presence of a heteroscedasctic variance

We examine here the simple case of building an optimal design for weighted linear regression with one predictor assuming independent errors (residuals), with two levels for the residual variance (below and above some threshold for the predictor).
We consider the following model:

$$Y_x = \alpha x + \varepsilon_x,\ x \in [0,1],$$

where ε_x are independent, centered, random variables with $\mathbb{V}(\varepsilon_x) = 1 + \mathbb{1}_{x>a}$, $0 < a < 1$. We assume that we have n measures $Y = (y_1, \ldots, y_n)$ and we form the vector $X = (x_1, \ldots, x_n)$. We also form the matrix Σ such that

$$\Sigma_{ij} = (1 + \mathbb{1}_{x_i>a})\delta_{ij},$$

where $\delta_{ij} = 1$ if $i = j$, 0 otherwise.
In that framework, it is well known that the BLUE (best linear unbiased estimator) of α knowing X is given by:

$$\widehat{\alpha_X} = \left(X'\Sigma^{-1}X\right)^{-1} X'Y,$$

and its variance by:

$$\mathbb{V}(\widehat{\alpha_X}) = \left(X'\Sigma^{-1}X\right)^{-1}.$$

We can write down the prediction variance as follows:

$$\begin{aligned}\mathbb{V}\left(Y_x - \widehat{Y_x}\right) &= \mathbb{V}\left((\alpha - \widehat{\alpha_X})x + \varepsilon_x\right)\\ &= \mathbb{V}(\widehat{\alpha_X})x^2 + 1 + \mathbb{1}_{x>a}\end{aligned}$$

Minimizing the averaged prediction variance over $[0,1]$ terms to minimize:

$$\begin{aligned} O(X) &= \int_0^1 \mathbb{V}\left(Y_x - \widehat{Y_x}\right) dx \\ &= \frac{1}{3}\mathbb{V}(\widehat{\alpha_X}) + 2 - a. \end{aligned}$$

Therefore, in that case, minimizing the averaged prediction variance over $[0,1]$ terms to minimize $\mathbb{V}(\widehat{\alpha_X})$.
We now give a closer look to this term. We have:

$$\begin{aligned} (\Sigma^{-1}X)_i &= \frac{x_i}{1+\mathbb{1}_{x_i>a}} = x_i - \frac{x_i}{2}\mathbb{1}_{x_i>a} \\ \Rightarrow\ X'\Sigma^{-1}X &= \sum_{i=1}^n \left(x_i^2 - \frac{x_i^2}{2}\mathbb{1}_{x_i>a}\right) \\ \Rightarrow\ \mathbb{V}(\widehat{\alpha_X}) &= (X'\Sigma^{-1}X)^{-1} = \frac{1}{\sum_{i=1}^n \left(x_i^2 - \frac{x_i^2}{2}\mathbb{1}_{x_i>a}\right)} \end{aligned}$$

This quantity is not differentiable with respect to x_i because of the indicators. Using derivation in the sense of the distributions makes appear Dirac terms and the quantity obtained is difficult to interpret. Nevertheless, it can be seen that this quantity is minimized when the denominator takes large values, that is for values close to 1 (the maximum of x) or values close to a (but inferior).

References

Amzal, B., Bois, F. Y., Parent, E., Robert, C. P., 2006. Bayesian-optimal design via interacting particle systems. Journal of the American Statistical Association 101 (474), 773–785.

Brus, D. J., Heuvelink, G. B. M., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138, 86–95.

Cappé, O., Guillin, A., Marin, J.-M., Robert, C. P., 2004. Population Monte-Carlo. Journal of Computational and Graphical Statistics 13 (4), 907–929.

Chilès, J. P., Delfiner, P., 2012. Geostatistics, Modeling Spatial Uncertainty, 2nd Edition. John Wiley & Sons, New-York.

de Fouquet, C., Gallois, D., Perron, G., 2007. Geostatistical characterization of the nitrogen dioxyde concentration in an urban area. part i: spatial variability and cartography of the annual concentration. Atmospheric Environment 41, 6701–6714.

De Gruijter, J., Brus, D., Bierkens, M., 2006. Sampling for natural resource monitoring. Springer Verlag.

Diggle, P. J., Lophaven, S., 2006. Bayesian geostatistical design. Scandinavian Journal of Statistics 33 (1), 53–64.

Friedman, J. H., Hastie, T., Tibshirani, R., 2 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33 (1), 1–22.
URL http://www.jstatsoft.org/v33/i01

Gelfand, A. E., Diggle, P. J., Fuentes, M., Guttorp, P. (Eds.), 2011. Handbook of Spatial Statistics. Chapman & Hall/CRC.

Handcock, M., Stein, M., 1993. A Bayesian Analysis of Kriging. Technometrics 89 (426), 368–390.

Kirkpatrick, S., 1983. Optimization by Simulated Annealing: quantitative studies. Journal of Statistical physics 34 (975-986).

Kirkpatrick, S., Gelatt Jr., C. D., Vecchi, M. P., 1983. Optimization by Simulated Annealing. Science 220 (4598), 671–680.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B 73 (4), 423–498.

Mardia, K. V., Marshall, R. J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. biometrika 71 (1), 135–46.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. T. M., 1953. Equations of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1091.

Müller, P., Sansó, B., De Iorio, M., 2004. Optimal bayesian design by inhomogeneous Markov chain simulation. Journal of the American Statistical Association 99 (467), 788–798.

Müller, W. G., 2007. Collecting Spatial Data: Optimum Design of Experiments for Random Fields, 3rd Edition. Physica-Verlag, Heidelberg.

Müller, W. G., Stehlík, M., 2010. Compound optimal spatial designs. Environmetrics 21, 354–364.

R Development Core Team, 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL http://www.R-project.org

Renard, D., Bez, N., Desassis, N., Laporte, F., 2010. RGeoS: The Geostatistical package. Mines ParisTech, version 9.0.0.
URL http://cg.ensmp.fr/rgeos/

Romary, T., de Fouquet, C., Malherbe, L., 2011. Sampling design for air quality measurement surveys: An optimization approach. Atmospheric Environment 45 (21), 3613 – 3620.

Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. Journal of the Royal Statistical Society: Series B 71 (2), 319–392.

Saporta, G., 2006. Probabilités, analyses des données et statistiques, 2nd Edition. Technip.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B 58 (1), 267–288.

Van Groenigen, J. W., Siderius, W., Stein, A., 1999. Constrained optimization of soil sampling for minimisation of the kriging variance. Geoderma 87, 139–259.

Zhu, Z., Stein, M. L., 2005. Spatial sampling design for parameter estimation of the covariance function. Journal of Statistical Planning and Inference 134 (2), 583 – 603.

Zhu, Z., Stein, M. L., 2006. Spatial sampling design for prediction with estimated parameters. Journal of Agricultural Biological and Environmental Statistics 11, 24–44.

Zimmerman, D. L., 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17 (6), 635–652.
URL `http://dx.doi.org/10.1002/env.769`

Zimmerman, D. L., Cressie, N., 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. Annals of the Institute of Statistical Mathematics 44, 27–43.
URL `http://dx.doi.org/10.1007/BF00048668`