Spatial sampling design with skew distributions: The special case of trans-Gaussian kriging

Gunter Spöck

Abstract So far, methodologies for spatial sampling design assume the investigated random field to be Gaussian. Most often the minimization of the kriging variance averaged over the investigated spatial design region is considered as a design criterion. The actual advantage of using this design criterion is that the kriging variance is independent from the actual data values but dependent only on their relative locations. The independence from data values is a result of the Gaussian assumption for the considered random field. If the data follow a skew distribution, like for example data whose Box- Cox transformation is multivariate Gaussian, the assumption of independence of the design criterion from data values can no longer be held. Kriging with Box-Cox transformed data is also known as trans-Gaussian kriging. We consider as design criterion the average of the expected lengths of 95% predictive intervals from trans-Gaussian kriging and show how sampling designs may be calculated efficiently using recent results of the author on the approximation of random fields by mixed linear models. To make the computations of such sampling designs faster NVIDIA CUDA technology is used and the design algorithms are implemented in parallel on fast NVIDIA graphical processing units (GPUs). Moreover, both, design criteria taking covariance function estimation by REML into account and not, are investigated. All theoretical findings are illustrated by a practical example taken from a rainfall monitoring network.

Gunter Spöck,
Alpen-Adria Universität Klagenfurt,
Universitätsstrasse 65-67,
9020 Klagenfurt, Austria,
e-mail: gunter.spoeck@aau.at

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

1 Introduction

We consider an isotropic random field $\{Y(x) : x \in \mathbf{X} \subset \Re^2\}$. The task of geostatistics is to produce a prediction map of all $y(x_0), x_0 \in \mathbf{X}$, based on available data $y(x_1), y(x_2), \ldots, y(x_n)$ and to report on the accuracy of these predictions. The best known methodology for this task of interpolation is kriging, also known as best linear unbiased prediction. The so-called universal kriging predictor dependends on the data $\mathbf{y} = (y(x_1), y(x_2), \ldots, y(x_n))^T$, the covariance matrix $\mathbf{K} = Cov(\mathbf{y})$ and the covariance vector $\mathbf{c}_0 = (Cov(Y(x_0), Y(x_1)), \ldots, Cov(Y(x_0), Y(x_n)))^T$. The mean squared error (MSEP) of this unbiased predictor is given by

$$\sigma^2(x_0) = \mathbf{E}(Y(x_0) - \hat{Y}(x_0))^2 = C(\mathbf{0}) - \mathbf{c}_0^T \mathbf{K}^{-1} \mathbf{c}_0 + \mathbf{g}(x_0)^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{g}(x_0),$$

where

$$\mathbf{g}(x_0) = \mathbf{f}(x_0) - \mathbf{F}^T \mathbf{K}^{-1} \mathbf{c}_0.$$

Obviously, the mean squared error $\sigma^2(x_0)$ is dependent on the arrangement of the sampling locations $x_1, x_2, \ldots, x_n$ via the covariance matrix $\mathbf{K}$, the covariance vector $\mathbf{c}_0$ and the design matrix $\mathbf{F}$.
As with any statistical prediction method, our aim is to make best use of the available data, possibly under certain budget constraints or constraints on the number of locations that can be sampled. Thus, the goal is to obtain as good predictions as possible for the complete area of investigation $\mathbf{X}$. One possibility to formalize this goal is to try to select the sampling locations $x_1, x_2, \ldots, x_n$ in such a way that the sum of all kriging MSEP's

$$\int_{\mathbf{X}} \sigma^2(x_0) dx_0 \tag{1}$$

becomes a minimum over the area of investigation $\mathbf{X}$. Notably, this is a very complicated optimization problem and becomes still more complicated by the fact that the covariance matrix $\mathbf{K}$ enters the kriging MSEP in its inverse form $\mathbf{K}^{-1}$.
The importance of (optimal) spatial sampling design considerations for environmental applications and soil science has been demonstrated in quite a few papers and monographs: Brus and de Gruijter (1997); Brown et al. (1994); Cox Jr. (1999); Groenigen et al. (1999); US-EPA (2002); Lark (2002); Caeiro et al. (2003); Hengl et al. (2003); Diggle and Lophaven (2006); Le and Zidek (2006); Brus and Heuvelink (2007); Dobbie et al. (2008); Delmelle and Goovaerts (2009).

2 The spatial mixed linear model

This section presents our approach to spatial sampling design. No stochastic search algorithms like simulated annealing to optimize the design criterion are needed in

this approach because we make use of the mathematical structure of the investigated design criteria. We consider a mean square continuous (m.s.c.) and isotropic random field $\{Y(x) : x \in \mathbf{X} \subseteq \Re^2\}$ such that

$$Y(x) = \mathbf{f}(x)^T \beta + \varepsilon(x), \quad \mathrm{E}\varepsilon(x) = 0, \tag{2}$$

where $\mathbf{f}(x)$ is a known vector of regression functions, $\beta \in \Re^r$ a vector of unknown regression parameters and denote

$$\mathrm{Cov}(Y(x), Y(y)) = C(||x-y||); \quad x, y \in \mathbf{X}.$$

Then, according to Yaglom (1987), knowing the covariance function $C(.)$, the spectral distribution function can be obtained from the inversion formula

$$\frac{G(\omega^+) + G(\omega^-)}{2} = \int_0^\infty J_1(t\omega)\omega C(t)dt,$$

where $G(\omega^+)$ and $G(\omega^-)$ denote the right- and left-hand side limits at ω and J_1 denotes the Bessel function of first kind and order 1. The spectral distribution function is positive, monotonically increasing and bounded from above. Approximating $G(.)$ by means of a step function with positive jumps $a_i^2 = G(\omega_{i+1}) - G(\omega_i)$ at preselected points ω_i, $i = 0, 1, , \ldots, n-1$, and changing to polar coordinates $(t, \varphi) = (radius, angle)$, the polar spectral representation theorem for m.s.c. isotropic random fields tells us that the error process may be approximated as

$$\varepsilon(t, \varphi) \approx \sum_{m=0}^{\infty} \{\cos(m\varphi) \sum_{i=1}^{n} J_m(\omega_i t) U_{m,i}\} + \sum_{m=1}^{\infty} \{\sin(m\varphi) \sum_{i=1}^{n} J_m(\omega_i t) V_{m,i}\}, \tag{3}$$

where all the random variables $U_{m,i}$ and $V_{m,i}$ are uncorrelated, have mean zero, and their variances are $\mathrm{var}(U_{m,i}) = \mathrm{var}(V_{m,i}) = d_m a_i^2$; and $d_m = 1$ for $m = 0$ and $d_m = 2$ for $m \geq 1$. By truncating the above series at a sufficiently large $m = M$, we get an approximation of our random field in form of a mixed linear model

$$Y(x) \approx \mathbf{f}(x)^T \beta + \mathbf{g}(x)^T \alpha + \varepsilon_0(x). \tag{4}$$

From (3) it becomes clear that the components of the additional regression vector $\mathbf{g}(\cdot)$ are made up of cosine-sine-Bessel surface harmonics. The idea to approximate the spatial random field by means of a large regression model with random coefficients was first proposed by Fedorov (1996), who approximated the random field by the so-called Karhunen-Loeve-Eigen-expansion, which is in general much more complicated to calculate than our approach via the polar spectral approximation.

3 Classical Bayesian experimental design problem

Starting from our spatial mixed linear model (4) we may gain further flexibility with a Bayesian approach incorporating prior knowledge on the trend. To this we assume that the regression parameter vector β is random with

$$\mathrm{E}(\beta) = \mu \in \Re^r, \quad \mathrm{Cov}(\beta) = \Phi.$$

This is exactly in the spirit of Omre (1987) who introduced Bayesian kriging this way. He used physical process knowledge to arrive at "qualified guesses" for the first and second order moments, μ and Φ. On the other hand, the state of prior ignorance or non-informativity can be modelled by setting $\mu = \mathbf{0}$ and letting Φ^{-1} tend to the matrix of zeroes, thus passing the "Bayesian bridge" to universal kriging, see Omre and Halvorsen (1989).
Now, combining (4) with Bayesian prior knowledge, we arrive at the **Bayesian spatial linear model** (BSLM)

$$Y(x) = \mathbf{h}(x)^T \gamma + \varepsilon_0(x), \tag{5}$$

where

$$\mathbf{h}(x) = \begin{pmatrix} \mathbf{f}(x) \\ \mathbf{g}(x) \end{pmatrix}, \gamma = \begin{pmatrix} \beta \\ \alpha \end{pmatrix},$$
$$\mathrm{E}\gamma = \begin{pmatrix} \mu \\ 0 \end{pmatrix} =: \gamma_0, \mathrm{Cov}(\gamma) = \begin{pmatrix} \Phi & 0 \\ 0 & \mathbf{A} \end{pmatrix} =: \Gamma.$$

Here $\varepsilon_0(x)$ is white-noise with variance σ_0^2 and $\mathbf{A}$ denotes the covariance matrix of α, resulting from the polar spectral approximation of the random field. Spöck and Pilz (2010) demonstrate that Bayesian linear trend estimation in the above BSLM actually approximates Bayesian linear kriging in the original model abitrarily closely. The same is true for the total mean squared error (TMSEP) of the trend prediction and the TMSEP of Bayesian kriging.
Thus taking the TMSEP of the trend prediction in the approximating model as a substitute for the Bayes kriging TMSEP we arrive at the following classical experimental design problem for so-called I-optimality:

$$\int_{\mathbf{X}} \mathbf{h}(x)^T (\mathbf{H}^T(d_n)\mathbf{H}(d_n) + \sigma_0^2 \Gamma^{-1})^{-1} \mathbf{h}(x) dx \to \min_{d_n}. \tag{6}$$

Here $d_n = \{x_1, x_2, \ldots, x_n\}$ collects either the design points to be added to the monitoring network or in the case of reducing the network the design points remaining in the monitoring network. $\mathbf{H}(d_n)$ expresses the dependence of the design matrix $\mathbf{H} = (\mathbf{h}(x_i)^T)_{i=1,2,\ldots,n}$ on the design points in the set d_n.
At this point we advise the reader not familiar with Bayesian experimental design theory to read the Appendix of Spöck and Pilz (2010). The key point in this theory is that the above so-called concrete design problem, which does not have a convenient

mathematical structure, may be extended to a so-called continuous design problem that has the nice feature to be a convex optimization problem. Thus, the whole apparatus of convex optimization theory is available to approximately solve the above design problem for I-optimality. In particular, directional derivatives may be calculated and optimal continuous designs may be found by steepest descent algorithms. Continuous designs are just probability measures ξ on $\mathbf{X}$ and may be rounded to exact designs d_n. Defining the so-called continuous Bayesian information matrix

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x)\mathbf{h}(x)^T \xi(dx) + \frac{\sigma_0^2}{n} \Gamma^{-1} \tag{7}$$

and

$$\mathbf{U} = \int_{\mathbf{X}} \mathbf{h}(x)\mathbf{h}(x)^T dx, \tag{8}$$

it may be shown that the set of all such information matrices is convex and compact and that the extended design functional

$$\Psi(\mathbf{M}_B(\xi)) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(\xi)^{-1}) \tag{9}$$

is convex and continuous in $\mathbf{M}_B(\xi)$. The above design functional $\Psi(.)$ thus attains its minimum at a design $\xi^* \in \Xi$, where Ξ is the set of all probability measures defined on the compact design region $\mathbf{X}$, see Pilz (1991). The minimization of $\Psi(\cdot)$ is the continuous analogue to our original design problem (6). The closeness of exact designs d_n to the optimal continuous design ξ^* may be judged by means of a well-known efficiency formula, see Appendix of Spöck and Pilz (2010).

4 The Smith and Zhu design criterion

In real world applications the isotropic covariance function $C_\theta(t)$ is always uncertain and must be estimated. The kriging predictor used is then based on this estimated covariance function $C_{\hat{\theta}}(t)$. Thus, the kriging predictor is always a plug-in predictor and the reported (plug-in) kriging variance underestimates the true variance of this plug-in predictor.
Smith and Zhu (2004) consider spatial sampling design by means of minimizing the average of the expected lengths of $1-\alpha$ predictive intervals:

$$\int_{\mathbf{X}} \mathrm{E}(\text{length of predictive interval at } x_0) dx_0. \tag{10}$$

Their predictors of the $\alpha/2$ and $1-\alpha/2$ quantiles of the predictive distributions are selected in such a way that the corresponding predictive intervals have coverage probability bias 0. The predictors of the mentioned quantiles are essentially the plug-in kriging predictors based on restricted maximum likelihood (REML) estimation of

the covariance function plus/minus a scaled plug-in kriging standard error term that is corrected to take account of REML estimation. Based on Laplace approximation it can be shown that this design criterion, up to the order $O(n^{-2})$, where n is the number of data, is equivalent to:

$$\int_{\mathbf{X}} [\sigma^2_\theta(x_0) + \mathrm{tr}(\kappa_\theta^{-1}\{\frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}\}^T \mathbf{K}_\theta \frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}) + \\ z^2_{1-\alpha/2}\{\frac{\partial \sigma_\theta(x_0)}{\partial \theta}\}^T \kappa_\theta^{-1} \frac{\partial \sigma_\theta(x_0)}{\partial \theta}] dx_0 \rightarrow \underset{d_n=\{x_1,\ldots,x_n\}}{Min} \tag{11}$$

Here

$$(\kappa_\theta)_{i,j} = (\mathrm{tr}(\mathbf{W}_\theta \frac{\partial \mathbf{K}_\theta}{\partial \theta_i} \mathbf{W}_\theta \frac{\partial \mathbf{K}_\theta}{\partial \theta_j}))_{i,j} \tag{12}$$

is the Fisher information matrix for REML,

$$\mathbf{W}_\theta = \mathbf{K}_\theta^{-1} - \mathbf{K}_\theta^{-1}\mathbf{F}(\mathbf{F}^T\mathbf{K}_\theta^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{K}_\theta^{-1},$$

$z_{1-\alpha/2}$ is the $1-\alpha/2$-quantile of the standard normal distribution, $\sigma^2_\theta(x_0)$ is the universal kriging variance at x_0 and $\lambda_\theta(x_0)$ is the universal kriging weights vector for prediction at x_0. This design criterion takes account of both prediction accuracy and covariance uncertainty.
Sections 2 and 3 have demonstrated that by using the BSLM (5) as approximation to the true isotropic random field the design criterion of I-optimality can be completely expressed in terms of the (concrete) Bayesian information matrix

$$\mathbf{M}_B = \mathbf{H}^T\mathbf{H} + \sigma_0^2\Gamma^{-1}.$$

Going from this information matrix to its continuous version $\mathbf{M}_B(\xi)$ according to (7), the extended design functional $\Psi(\mathbf{M}_B(\xi)) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(\xi)^{-1})$ becomes continuous and convex on the compact and convex set of all such information matrices $\mathbf{M}_B(\xi)$. This was the reason why classical convex experimental design algorithms could be used to find optimal spatial sampling designs minimizing the criterion (9). In Spöck et al. (2012) it is shown that the Smith and Zhu design criterion has also some favourable properties, so that classical convex experimental design theory can be applied to this design criterion, too:

- Expression (11) can be expressed completely in terms of the Bayesian information matrix $\mathbf{M}_B$.
- The design functional is continuous on the convex and compact set of all $\mathbf{M}_B(\xi)$ and has some advantageous properties according to which classical experimental design algorithms may be used in order to find spatial sampling designs.

Assuming the BSLM (5) the covariance function is actually parametrized through the diagonal matrix $\mathbf{A}$ and the nugget variance σ_0^2. Since the Smith and Zhu design criterion assumes the covariance parameters to be estimated by restricted maximum likelihood we actually estimate this diagonal matrix $\mathbf{A}$ and σ_0^2 by this methodology.

The a priori covariance matrix $\Phi = \mathrm{cov}(\beta)$ must be given almost infinite diagonal values because the Smith and Zhu (2004) approach assumes the trend parameter vector β to be estimated by generalized least squares and $\Phi \to \infty$ bridges the gap from Bayesian linear to generalized least squares trend estimation. The a priori mean $\mu = \mathrm{E}(\beta)$ can be set to $\mathbf{0}$ then.
According to the polar spectral representation (3) several values in the diagonal matrix $\mathbf{A}$ are identical:

$$\mathbf{A} = \mathrm{diag}(\{d_m a_i^2\}_{m=0,\ldots,M;i=1,\ldots,n;k=1,2}), \tag{13}$$

where the definitions of d_m and a_i^2 and the indexing derive from the polar spectral representation (3). For restricted maximum likelihood estimation of $\mathbf{A}$ we have two possibilities:

- We can leave the a_i's unspecified: This approach is almost nonparametric because a lot of a_i's and corresponding frequencies w_i are needed to get the isotropic random field properly approximated, and corresponds to a semiparametric estimation of the spectral distribution function via a step function.
- We can specify a parametric model for the a_i^2's: The polar spectral density function for an isotropic random field over $\Re^2$ possesing for example an exponential covariance function $B(h) = C\exp(-\frac{3h}{\alpha})$ is given by

$$g(w) = \frac{C\frac{3}{\alpha}w}{((\frac{3}{\alpha})^2 + w^2)^{3/2}}. \tag{14}$$

  The polar spectral density function is defined just as the first derivative of the polar spectral distribution function $G(w)$. A possible parametrization for the a_i^2's then is

$$a_i^2 = \frac{g(w_i) + g(w_{i-1})}{2}(w_i - w_{i-1}), i = 1, 2, \ldots, n, \tag{15}$$

  where $0 = w_0 < w_1 < \ldots, w_n$ are fixed frequencies.

For the optimization of the Smith and Zhu design criterion we make use of the same exchange design algorithms as described in Section 6.1 . We only have to replace $\Psi(\mathbf{M}_B(\xi))$ by the Smith an Zhu design functional given in Spöck et al. (2012).

5 Spatial sampling design for trans-Gaussian kriging

In trans-Gaussian kriging the originally positive valued data $Z(x_i), i = 1, 2, \ldots, n$ are transformed to Gaussianity by means of the Box-Cox transformation

$$g_\lambda(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda} & : \quad \lambda \neq 0 \\ \log(z) & : \quad \lambda = 0 \end{cases}.$$

Let $\mathbf{Z} = (Z(x_1), Z(x_2), \ldots, Z(x_n))^T$ be the vector of original data and

$$\mathbf{Y} = (g_\lambda(Z(x_1)), g_\lambda(Z(x_2)), \ldots, g_\lambda(Z(x_n)))^T \tag{16}$$

be the vector of transformed data. The predictive density for trans-Gaussian kriging at a location x_0 then may be written:

$$\varphi(g_\lambda(z); \hat{Y}_{OK}(x_0), \sigma^2_{OK}(x_0)) * z^{\lambda-1}, \tag{17}$$

where $\varphi(.; \hat{Y}_{OK}(x_0), \sigma^2_{OK}(x_0))$ is the Gaussian density with mean equal to the ordinary kriging predictor $\hat{Y}_{OK}(x_0)$ at x_0 and based on the transformed variables $\mathbf{Y}$, and variance equal to the ordinary kriging variance $\sigma^2_{OK}(x_0)$, $z^{\lambda-1}$ is the Jacobian of the Box-Cox transformation.
For spatial sampling design we can consider again the average expected length of $(1-\alpha)$-predictive intervals. In order to make the expected lengths of predictive intervals also dependent on REML-estimation of the covariance function, we can consider instead of the Gaussian density $\varphi(.; \hat{Y}_{OK}(x_0), \sigma^2_{OK}(x_0))$ that unique Gaussian density φ whose 0.025- and 0.975-quantiles are given by the Smith and Zhu (2004) 95% predictive interval

$$\begin{aligned} &\hat{Y}_{OK}(x_0) \pm \\ &1.96 \quad \sigma_\theta(x_0)\{1 + \frac{1}{2\sigma^2_\theta(x_0)}[\mathrm{tr}(\kappa_\theta^{-1}\{\frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}\}^T \mathbf{K}_\theta \frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}) + \\ &1.96^2 \quad \{\frac{\partial \sigma_\theta(x_0)}{\partial \theta}\}^T \kappa_\theta^{-1} \frac{\partial \sigma_\theta(x_0)}{\partial \theta}]\}. \end{aligned} \tag{18}$$

Last but not least, to get expected predictive intervals we must replace in the statistic $t(\mathbf{Y}) = \hat{Y}_{OK}(x_0)$ every variable $Y(x_i)$ for which we do not have data by its ordinary kriging predictor based on the available data. Furthermore, we note that in the above approach we have not taken into account the fact that the transformation parameter λ itself is estimated too, i.e. by maximum likelihood, and then is plugged-into the ordinary kriging predictor. In a future paper we will take account of this additional uncertainty.

6 The spatDesign toolbox

The spatial sampling design and geostatistics toolbox spatDesign has been developed since 2003. It can be run in both MATLAB and Octave and can be downloaded from:

- http://wwwu.uni-klu.ac.at/guspoeck/spatDesignMatlab.zip
- http://wwwu.uni-klu.ac.at/guspoeck/spatDesignOctave.zip

The toolbox underlies the GNU Public Licence Version 3 or higher and thus is freely available. In MATLAB (www.mathworks.com) the toolbox is fully func-

tional but assumes that also the MATLAB Optimization and Statistics toolboxes are installed. In Octave the Smith and Zhu criterion is not implemented. The spatial sampling design functions corresponding to the Smith and Zhu design criterion need (on a standard PC) a lot of computing time. For this reason this part of the toolbox has been parallelized to work with NVIDIA GPU's and the freely available MATLAB parallelization package GPUmat (www.gp-you.org). If you have a CUDA (www.nvidia.com) compatible graphics card installed then this will be automatically detected and the parallelized algorithms for the Smith and Zhu design criterion will be used.

6.1 Spatial sampling design software

spatDesign V.2.2.0 implements three design criteria for Bayesian linear kriging, where the first two are criteria for prediction only, with the covariance function assumed to be certain. The third criterion is the Smith and Zhu (2004) criterion taking account of also the fact that the covariance function is estimated. The Smith and Zhu design criterion has also been implemented for transformed-Gaussian kriging. The implemented criteria for prediction only are:

- I-optimality:
$$\Psi(\mathbf{M}_B(d_n)) = \mathrm{tr}(\mathbf{U}\mathbf{M}_B(d_n)^{-1}) \to \min_{d_n} \tag{19}$$
$$\mathbf{U} = \sum_{i,j=1}^{m} \mathbf{h}(x_{i,j})\mathbf{h}(x_{i,j})^T; \tag{20}$$
where the integral in (8) has been replaced by the sum over a fine grid of locations $x_{i,j} \in \mathbf{X}$.
- D-optimality:
$$\Psi(\mathbf{M}_B(d_n)) = |(\mathbf{H}^T(d_n)\mathbf{H}(d_n) + \sigma_0^2\Gamma^{-1})^{-1}| \to \min_{d_n} \tag{21}$$

The Smith and Zhu design criterion can be expressed (according to Sections 4, 5) also in the form
$$\Psi(\mathbf{M}_B(d_n)) \to \min_{d_n}, \tag{22}$$
The interested reader is referred to Spöck et al. (2012).
The basic algorithm for calculating spatial sampling designs is an exchange algorithm from experimental design theory going back to Fedorov (1972). Contrary to the construction of optimal discrete designs, here we cannot prove convergence of the exact designs to the functional value $\Psi(d^*)$ of an optimal exact design d^*; we can only guarantee stepwise improvement of a given exact starting design, i.e. the sequence of functional values $\Psi(d_{n,s})$ decreases monotonically with increasing iteration index s. The algorithm is an exchange algorithm improving n-point designs and starting from an initial design.

6.1.1 Exchange algorithm

Step 1. Use some initial design $d_{n,1} = \{x_{1,1}, \ldots, x_{n,1}\} \in \mathbf{X}^n$ of size n.
Step 2. Beginning with $s = 1$ form the design $d_{n+1,s} = d_{n,s} + (x_{n+1,s})$ by adding the point

$$x_{n+1,s} = \arg \min_{x \in \mathbf{X}} \Psi(\mathbf{M}_B(d_{n,s} + (x)))$$

to $d_{n,s}$.
Then form $d_{n,s}^j = d_{n+1,s} - (x_{j,s}), j = 1, 2, \ldots, n+1$ and delete that point $x_{j^*,s}$ from $d_{n+1,s}$ for which

$$\Psi(\mathbf{M}_B(d_{n,s}^{j^*})) = \min_{j \in \{1,\ldots,n+1\}} \Psi(\mathbf{M}_B(d_{n,s}^j))).$$

Repeat Step 2 until the point to be deleted is equivalent to the point to be added.

6.1.2 Generation of an initial design

The initial design is a one-point design which minimizes the design functional among all designs of size $n = 1$. Note that such a design exists since the Bayesian information matrix is positive definite even for designs of size $n = 1$.

Step 1. Choose $x_1 \in \mathbf{X}$ such that
$x_1 = \arg \min_{x \in \mathbf{X}} \Psi(\mathbf{M}_B((x)))$, and set $d_1 = (x_1)$.
Step 2. Beginning with $i = 1$, find x_{i+1} such that
$x_{i+1} = \arg \min_{x \in \mathbf{X}} \Psi(\mathbf{M}_B(d_i + (x)))$ and form $d_{i+1} = d_i + (x_{i+1})$.
Continue with i replaced by $i+1$ until $i+1 = n$.
Step 3. If $i+1 = n$ then stop and take
$d_{n,1} = \{x_1, \ldots, x_n\}$ as an initial design.

6.1.3 Combination of the above two algorithms

It is a good idea to combine the initial design algorithm and the exchange algorithm in the following way:

Step 1. Start with the initial design algorithm and find a design with one first design point.
Step 2. Having found a design with $m \geq 1$ design points apply the exchange algorithm to this design to improve it.
Step 3. Add to the design from Step 2 one further design point by means of the initial design algorithm to get $m+1$ design points.
Step 4. Go back to Step 2 and iterate Step 2 and Step 3 until you have found n desired design points.

6.1.4 Reduction of experimental designs

Often it is desired to reduce a given experimental design $d = \{x_1, x_2, \ldots, x_n\}$ to one including only $m < n$ design points from d:

Step 1. Delete that design point x_{j^*} from d for which
$x_{j^*} = \arg\min_{x_j \in d} \Psi(\mathbf{M}_B(d-(x_j)))$, and set
$d := d-(x_{j^*})$.
Step 2. Iterate Step 1 until the design d contains only m design points.

Also this algorithm may be combined with an improvement step similar to the exchange algorithm.

6.1.5 Inverse of the information matrix

The calculation of exact designs requires in every step the calculation of the inverses of the information matrices $\mathbf{M}_B(d_{n,s})$ or $\mathbf{M}_B(d_{n+1,s})$. In the next Sections we will see that these information matrices can have a quite high dimension of about 3000×3000. So, how can one invert such large matrices in affordable time? There is computationally no need to make explicit use of numerical matrix inversion algorithms, when one considers the update formulas (13.26) and (13.28) in Pilz (1991):

$$\mathbf{M}_B(d_{n,s}+(x))^{-1} = \frac{n+1}{n}\{\mathbf{M}_B(d_{n,s})^{-1} - \frac{\mathbf{M}_B(d_{n,s})^{-1}\mathbf{h}(x)\mathbf{h}(x)^T\mathbf{M}_B(d_{n,s})^{-1}}{n+\mathbf{h}(x)^T\mathbf{M}_B(d_{n,s})^{-1}\mathbf{h}(x)}\},$$

$$\mathbf{M}_B(d^j_{n,s})^{-1} = \frac{n}{n+1}\{\mathbf{M}_B(d_{n+1,s})^{-1} + \frac{\mathbf{M}_B(d_{n+1,s})^{-1}\mathbf{h}(x_{j,s})\mathbf{h}(x_{j,s})^T\mathbf{M}_B(d_{n+1,s})^{-1}}{n+1-\mathbf{h}(x_{j,s})^T\mathbf{M}_B(d_{n+1,s})^{-1}\mathbf{h}(x_{j,s})}\}$$

Obviously, only matrix- and vector multiplications are needed in these update formulae.

6.1.6 Basic sampling design functions

The basic spatial sampling design functions are:

- OPTIMALLY_DELETE_N_LOCATIONS_FROM_POOLDELETE.m
- OPTIMALLY_ADD_N_LOCATIONS_FROM_POOLCOMPLETE.m
- OPTIMALLY_ADD_N_LOCATIONS_FROM_POOLADD.m
- OPTIMALLY_IMPROVE_POOLDELETE_FROM_POOLCOMPLETE.m
- OPTIMALLY_IMPROVE_POOLDELETE_FROM_POOLADD.m

The names of these functions are self-explanatory: “Pooldelete” is the discrete pool of locations that are allowed to be deleted from the design. “Poolcomplete” is the complete compact area of points from $\mathbf{X}$ allowed to be added to the design.

"Pooladd" is the discrete pool of locations that are allowed to be added to the design. "Improve" means the exchange algorithm, where locations from "Pooldelete" may either be exchanged by locations from "Poolcomplete" or from "Pooladd" and the total number of sampling locations remains constant.

7 An example session

The purpose of this section is to demonstrate some capabilities of the spatDesign toolbox. The most important Matlab function calls related to sampling design with the Smith and Zhu design criterion are given. The data set considered is a rainfall

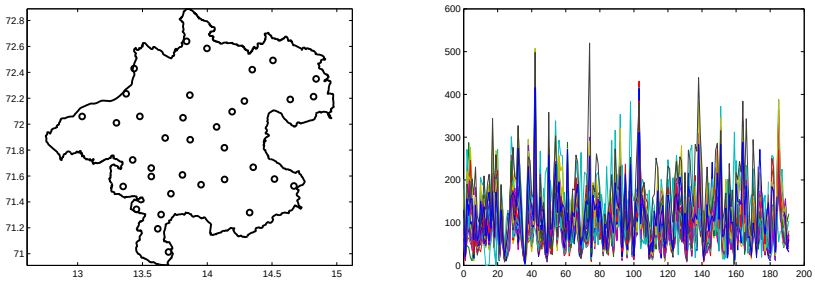

Fig. 1 Left: The 36 sampling locations of the Upper Austria rainfall data set set. Right: The 36 time series of average monthly rainfall at each station.

data set from Upper Austria. The monitoring network comprises 36 locations. Average monthly rainfall has been measured at each location starting in January 1994 and ending in December 2009. In Fig. 1 we see that there are obviously areas in the design region that look very empty, having no sampling locations.

7.1 Preparatory calculations

Next let us calculate from the above time series for each station the mean rainfall over the years, as well as the residual rainfall, for each of the 12 months, Fig. 2. We proceed by calculating from the rain residuals for each of the 12 months the empirical semivariogram and the semivariogram standardized by the variance of each month, Fig. 3. We next calculate for each station the mean of rain residuals, the empirical semivariogram corresponding to these means and a fitted exponential semivariogram, Fig. 3. The fact that the standardized semivariograms are almost the same for all months means that the space-time random field is separable and that we can use one and the same semivariogram (the grey one in the Right of Fig. 3)

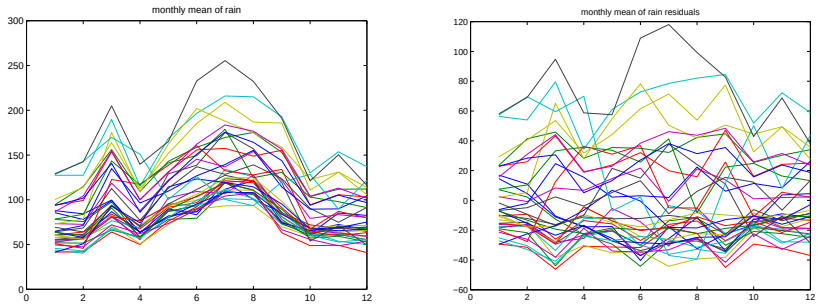

Fig. 2 Left: The average monthly rainfall over the years at each of the 36 stations, for each of the 12 months. Right: The residual rainfall at each of the 36 stations, for each of the 12 months.

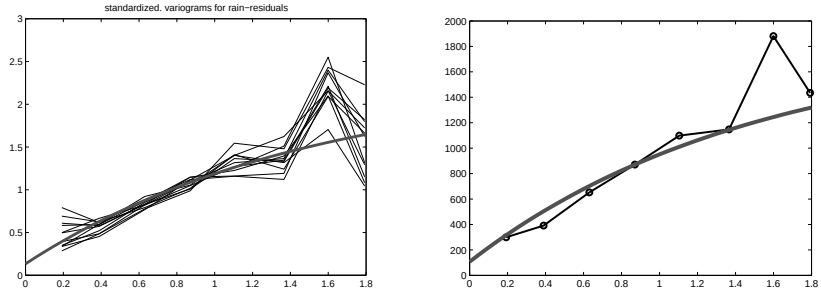

Fig. 3 Left: Standardized empirical semivariograms for all of the 12 months and a fitted exponential semivariogram (grey). Right: Empirical semivariogram for the mean of rain residuals and a fitted exponential semivariogram (grey).

for doing spatial sampling design for each month. In the next step we calculate the polar spectral distribution function corresponding to this semivariogram. Obviously, this spectral distribution function almost attains its maximum of 1735.2 at $w = 47$. We now select the frequencies $w_i, i = 1,2,\ldots,34$, calculate an approximation to the spectral distribution function via a step function (the steps are the a_i^2) and check whether this approximation to the spectral distribution function provides a good fit to the original covariance function , Fig. 4.

```
load wscaled
plotspectraldist(0:0.5:47,delta0rain)
w=wscaled*47;
[wrain,deltarain]=expstep(w,delta0rain(3),delta0rain(2)); % the discrete spectrum
hold on, approxspectraldist(wrain,deltarain)
```

```
plotcovarianceapprox(wrain,45,deltarain,delta0rain,12.5:0.05:17,73,12.5,12.5,
15,70.8,73); %the worst approximating covariance function
```

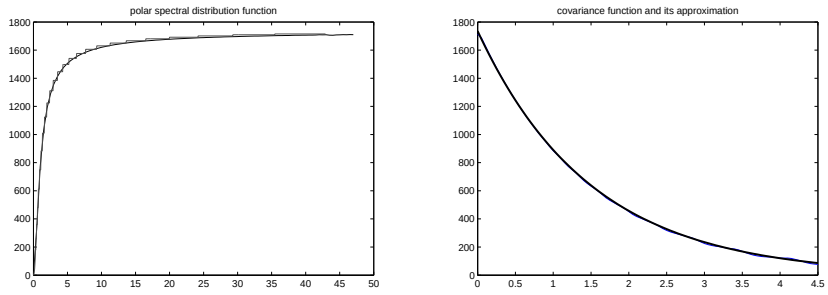

Fig. 4 Left: Polar spectral distribution function and its approximation (grey). Right: True covariance function (black) and its worst approximation (grey).

A look at the approximating covariance function in Fig. 4. shows that at the origin the difference between the true covariance function and the approximating covariance function is 20. This is small scale variation that the approximating covariance function does not take into account. Later in spatial sampling design we will add this value of 20 to the nugget effect 106.8 of the true covariance function. Thus, $20+106.8$ is the variance of the uncorrelated error process $\varepsilon_0(x)$ in our BSLM (5). We now have all the quantities that we need in order to do spatial sampling design on the basis of our Bayesian spatial linear model (5), corresponding to the assumption of Gaussianity of observations.

7.2 Optimal design for the BSLM

We consider to add 14 additional sampling locations from the complete design region **X** to the available grid of 36 sampling locations, Fig. 5.

```
[xadd,yadd,avglengthpredint]=optimally_add_n_locations_from_poolcomplete(
{},.....................................% no external drift
[],.....................................% no need to specify the matrix U
xyelevation(:,1),xyelevation(:,2),ones(36,1),..% the available data locations
20+delta0rain(1),.....................% the variance of the uncorrelated error
.....................................% process ε0(x)
```

```
[1000000,0,0;0,0.0000001,0;0,0,0.0000001],..% the a priori variance of
.....% the constant trend must be given
.....% almost infinite variance; no linear drift
wrain,.....% the frequencies of the Bessel harmonics
45,.....% the largest angular frequency
deltarain,.....% the a_i^2
delta0rain,.....% the parameters of the
.....% exponential covariance function
1,.....% the Box-Cox transformation parameter
.....% (no transformation, Gaussian kriging)
14,.....% we want to add 14 samples
12.5,15,70.8,73,.....% the size of the design region
10,.....% maximally iterate 10 times in the exchange
.....% step
boundary,.....% the polygonal design region
17,17,.....% discretization in Easting and Northing
.....% when considering new samples
'z',.....% we apply the Smith and Zhu (2004) design
.....% criterion to ordinary kriging
0.....% no graphical output
);
```

Fig. 5. shows the optimal 8- and 14 point designs. Obviously, certain locations have been selected with multiplicities larger than 1. The reason is that the Smith and Zhu design criterion does not only take account of best prediction but also of covariance estimation; in order to get the nugget effect and the behaviour of the covariance function close to its origin well estimated locations are needed in the optimal design which are close to each other. Fig. 6 plots the decrease of the average of the expected lenghts of the 95% predictive intervals when adding up to 14 design locations.

The calculation of the optimal 14 point design takes about one week, on an Intel i7 8 Core CPU and a NVIDIA 580 GPU with 1.6 Gb RAM. Similar calculations for the simpler design functional (19) take, without NVIDIA CUDA support, one day.

7.3 Design for the trans-Gaussian kriging

In the above example session we have assumed the data to be Gaussian and have intended to use ordinary kriging for prediction, although, as is visible from Fig. 2, the data are not Gaussian. Thus, we will consider now the assumption that the rainfall residuals can be transformed to Gaussianity by means of a Box-Cox transformation. Because the Box-Cox transformation works only for positive valued data we have

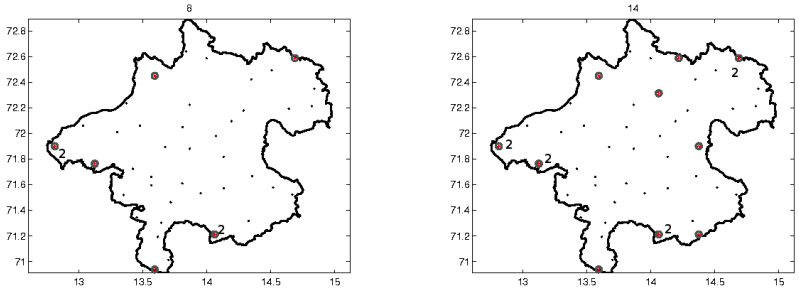

Fig. 5 Left: Optimal 8 point design for Gaussian kriging. Right: Optimal 14 point design for Gaussian kriging. Certain locations have been selected with multiplicities larger than 1.

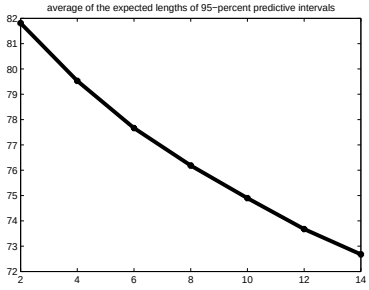

Fig. 6 Average of the expected lengths of 95% predictive intervals, when adding up to 14 design locations.

to add a positive offset to the 12 monthly sets of rainfall residuals. To identify the appropriate offset and optimal Box-Cox transformation parameter λ_0 we perform a sequence of Lilliefors tests for Gaussianity on the transformed rainfall residuals. We then retain this offset and the corresponding Box-Cox transformation parameter λ_0, where the sum of the 12 p-values from the Lilliefors tests attains its maximum. Fig. 7 gives the corresponding surfaces of the sum of p-values and number of rejected hypotheses for Gaussianity at the 10% significance level. According to these figures the optimal parameters are chosen as:

offset=53
$\lambda_0 = -0.25$

Obviously, for these parameters only one hypothesis of Gaussianity is rejected at the 10% significance level.
We now proceed as in the Gaussian case. Fig. 8 shows standardized empirical semivariograms of the Box-Cox transformed rain residuals and the semivariogram of the Box-Cox transformed means of rain residuals. Interestingly, the fitted semivariograms in this figure are just scaled versions of the fitted semivariograms from Fig.

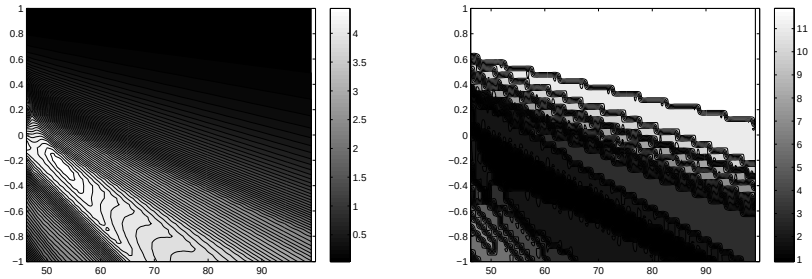

Fig. 7 Left: Sum of the p-values, depending on the transformation parameter λ and the offset. Right: Number of rejected hypotheses of Gaussianity at 10% significance level, depending on the transformation parameter λ and the offset.

3. The parameters of the semivariogram on the Right of Fig. 8 are delta0Box=[0.0047, 0.0768, 4.5000, 0, 0]. By means of

```
load wscaled
figure
plotspectraldist(0:0.5:50,delta0Box)
wBox=wscaled*50;
[wBox,deltaBox]=expstep(wBox,delta0Box(3),delta0Box(2));
hold on, approxspectraldist(wBox,deltaBox)
figure
plotcovarianceapprox(wBox,45,deltaBox,delta0Box,12.5:0.05:17,73,12.5,
12.5,15,70.8,73)
```

we go on to calculate the spectral distribution function and its step-wise approximation and the worst approximating covariance function. We note that close to the origin $h = 0$ the difference between the true covariance function and its approximation is 0.001. Finally, the function call

```
[xxresrainBox,yyresrainBox,avglengthpredintBox]=
optimally_add_n_locations_from_poolcomplete({},[],xyelevation(:,1),
xyelevation(:,2),meanresrainmeanmonth,delta0Box(1)+0.001,
[10000000000000,0,0;0,0.00000000000001,0;0,0,0.00000000000001],
wBox,45,deltaBox,delta0Box,lambda0,14,12.5,15,70.8,73,
10,boundary,17,17,'z',0);
```

calculates the optimal 14 point design. Fig. 9 visualizes optimal 8- and 14-point designs for trans-Gaussian kriging. Fig. 10 gives the expected lengths of 95% predictive intervals.

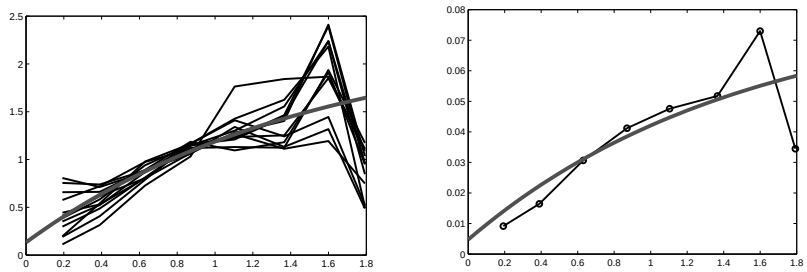

Fig. 8 Left: Standardized empirical semivariograms of the transformed rain residuals for all of the 12 months and a fitted exponential semivariogram (grey). Right: Empirical semivariogram for the transformed mean of rain residuals and a fitted exponential semivariogram (grey).

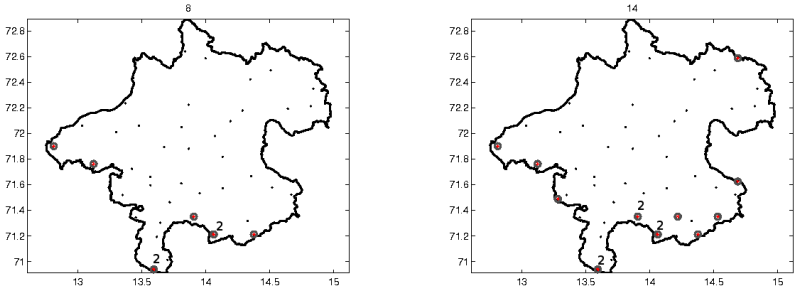

Fig. 9 Left: Optimal 8 point design for trans-Gaussian kriging. Right: Optimal 14 point design for trans-Gaussian kriging. Certain locations have been selected with multiplicities larger than 1.

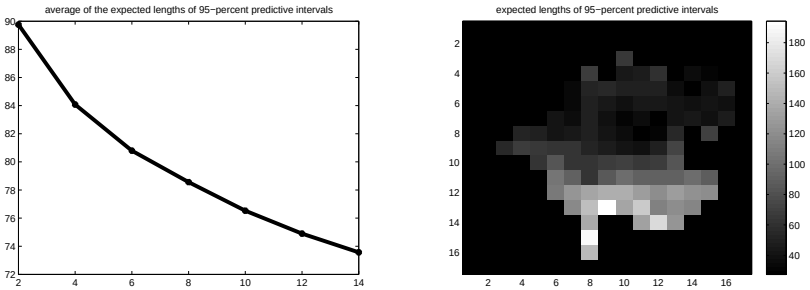

Fig. 10 Left: Decrease of the average expected lengths of 95% predictive intervals, when adding up to 14 design locations. Right: Expected lengths of 95% predictive intervals corresponding to the optimal 14 point design.

8 Conclusion

Obviously, the designs for trans-Gaussian kriging in Fig. 9 are completely different from the designs for Gaussian kriging in Fig. 5. Whereas the designs in Fig. 5 are

much more space-filling the design locations in Fig. 9 have been selected in areas that have high average value of rainfall. This fact becomes more clear when we compare Fig. 9 to Fig. 11, where the median rainfall from the predictive distributions of trans-Gaussian kriging is visualized. Obviously, in areas with high average rainfall the average expected length of 95% predictive intervals can be most decreased. This fact results from a fundamental difference between designs for Gaussian random fields and designs for trans-Gaussian kriging: Designs for trans-Gaussian kriging are dependent also on the data, through the ordinary kriging predictor in formula (18).

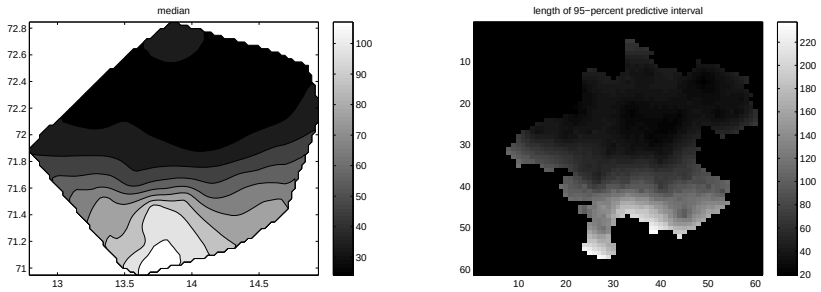

Fig. 11 Left: The median rainfall field+53 calculated from the predictive distributions of trans-Gaussian kriging applied to the 36 means of rainfall residuals. Right: Expected lengths of 95% predictive intervals corresponding to the trans-Gaussian kriging from the left figure.

References

Brown, P.J., Le, N.D. and Zidek, J.V. (1994). Multivariate spatial interpolation and exposure to air pollutants, *The Canadian Journal of Statistics*, 2, pp. 489-509.

Brus, D.J. and de Gruijter, J.J. (1997). Random sampling or geostatistical modeling? Chosing between design-based and model-based sampling strategies for soil (with discussion), *Geoderma*, 80, pp. 1-44.

Brus, D.J. and Heuvelink, G.B.M. (2007). Optimization of sample patterns for universal kriging of environmental variables, *Geoderma*, 138, pp. 86-95.

Caeiro, S., Painho, M., Goovaerts, P., Costa, H. and Sousa, S. (2003). Spatial sampling design for sediment quality assessment in estuaries, *Environmental Modelling and Software*, 18, pp. 853-859.

Cox Jr., L.A. (1999). Adaptive spatial sampling of contaminated soils, *Risk Analysis*, 19, pp. 1059-1069.

Delmelle E.M. and Goovaerts, P. (2009). Second-phase sampling designs for non-stationary spatial variables, *Geoderma*, doi:10.1016/j.geoderma.2009.08.007.

Diggle, P. and Lophaven, S. (2006). Bayesian Geostatistical Design, *Scandinavian Journal of Statistics*, 33, pp. 53-64.

Dobbie, M.J., Henderson, B.L. and Stevens Jr., D.L. (2008). Sparse sampling: Spatial design for monitoring stream networks, *Statistics Surveys*, 2, pp. 113-153.

Fedorov, V.V. (1996). Design of spatial experiments: model fitting and prediction. In Gosh, S. and Rao, C.R., editors, *Handbook of Statistics*, 13. Elsevier, Amsterdam, pp. 515-553.

Fedorov, V.V. (1972). *Theory of optimal experiments*, transl. and ed. by W.J. Studden and E.M. Klimko, Academic Press, New York, (Russian Original: Nauka, Moscow 1971), 292 p..

Groenigen, J.W. van, Siderius, W. and Stein, A. (1999). Constrained optimisation of soil sampling for minimisation of the kriging variance, *Geoderma*, 87, pp. 239-259.

Hengl, T., Rossiter, D.G. and Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps, *Australian Journal of Soil Research*, 41, pp. 1403-1422.

Lark. R.M. (2002). Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood, *Geoderma*, 105, pp. 49-80.

Le, N.D. and Zidek, J.V. (2006). *Statistical Analysis of Environmental Space-Time Processes*, Springer, New York, 356 p..

Omre, H. (1987). Bayesian kriging - merging observations and qualified guess in kriging, *Mathematical Geology*, 19, pp. 25-39.

Omre, H. and Halvorsen, K. (1989). The Bayesian bridge between simple and universal kriging, *Mathematical Geology*, 21, pp. 767-786.

Pilz, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models*, Wiley, New York, 306 p..

Smith, R.L. and Zhu, Z. (2004). Asymptotic theory for kriging with estimated parameters and its application to network design, www.stat.unc.edu/postscript/rs/supp5.pdf, 21 p..

Spöck, G. and Pilz, J. (2010). Spatial sampling design and covariance-robust minimax prediction based on convex design ideas, *Stochastic Environmental Research and Risk Assessment*, 24, pp. 463-482.

Spöck, G., Zhu, Z. and Pilz, J. (2012). Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design, accepted by *Stochastic Environmental Research and Risk Assessment*, 20 p.

US-EPA (2002). *Guidance on choosing a sampling design for environmental data collection*, EPA QA/G-5S, United States Environmental Protection Agency, 70 p..

Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions I*, Springer-Verlag, New York, 526 p..