

Spatial Sampling Design and Soil Science

Gunter Spöck
University of Klagenfurt
Austria

1. Introduction

Spatial statistics is the science of the analysis of geo-referenced data and loosely speaking may be divided into the three sub-areas analysis of point processes, analysis of areal data and geostatistics. Point processes naturally arise for example in geophysics, when the locations of earthquakes are noticed, in epidemiology, where new illness cases of certain epidemics are mapped geographically or in biology, where cell centers of a certain tissue are mapped under the microscope. Areal data are data that are attached to areas like number of illness cases in certain medical districts, percent of grassland in a certain county or number of votes in certain political districts. The topic of this article is but geostatistics, the science of continuous stochastic processes or so-called random fields that are defined over some region in 2- or 3-dimensional geographic space \mathbf{X} or in space-time. A random field $\{Y(x) : x \in \mathbf{X}\}$ is a set of random variables or so-called regionalized variables $Y(x)$ that are attached to every location $x \in \mathbf{X}$. The probability law of the random field is uniquely determined by its so-called projective family, the set of all finite dimensional distributions of any finite set of $Y(x)$ obeying symmetry and consistency with marginal distributions although we will see next that geostatistics most often deals only with the first and second order characteristics of random fields, the trend or mean function $m(x) = E(Y(x))$ and the covariance function $C(x, y) = \text{Cov}(Y(x), Y(y))$. Most often a linear trend function $m(x) = \mathbf{f}(x)^T \beta$, where $\mathbf{f}(x)$ is a fixed vector-valued function and β is a regression parameter vector to be estimated, is sufficient for modelling purposes. For $x \in \mathbb{R}^2$, $\mathbf{f}(x)$ could be for example a vector of polynomials in the coordinates $x = (x_1, x_2)$. The covariance function $C(\cdot, \cdot)$ must be positive semidefinite, meaning that it must give any linear combination of $Y(x_1), Y(x_2), \dots, Y(x_n)$ positive variance. Most often an additional assumption of second-order stationarity must be met, meaning that $C(x, x+h) = C(h)$ is dependent only on lag h and not on the locations x and $x+h$ themselves. The next stronger assumption is the assumption of isotropy, meaning $C(x, x+h) = C(\|h\|_2)$, where $\|h\|_2$ is the Euclidean length of h . Both the assumption of second order stationarity and of isotropy are met in order to make the covariance function estimable from only single realized variables or data $y(x_1), y(x_2), \dots, y(x_n)$. The task of geostatistics is to produce a prediction map of all $y(x_0), x_0 \in \mathbf{X}$ based on the available data $y(x_1), y(x_2), \dots, y(x_n)$ and to report on the accuracy of these predictions. The best known methodology for this task of interpolation or map drawing is kriging, also known as best linear unbiased prediction. The so-called universal kriging predictor is dependent on both the data $\mathbf{y} = (y(x_1), y(x_2), \dots, y(x_n))^T$, the covariance matrix \mathbf{K} of the corresponding random variables and the covariance vector \mathbf{c}_0 between these random variables and the random

variable $Y(x_0)$ to be predicted. It can be written

$$\hat{Y}(x_0) = \mathbf{f}(x_0)^T \hat{\beta} + \mathbf{c}_0^T \mathbf{K}^{-1} (\mathbf{y} - \mathbf{F} \hat{\beta}), \quad (1)$$

where

$$\hat{\beta} = (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{K}^{-1} \mathbf{y} \quad (2)$$

is the generalized least squares estimate of the regression parameter vector β and \mathbf{F} is the design matrix corresponding to $\mathbf{f}(\cdot)$ and the locations x_1, x_2, \dots, x_n . The mean squared error (MSEP) of this unbiased predictor is given by

$$\mathbf{E}(Y(x_0) - \hat{Y}(x_0))^2 = \sigma^2(x_0) = C(\mathbf{0}) - \mathbf{c}_0^T \mathbf{K}^{-1} \mathbf{c}_0 + \mathbf{g}(x_0)^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{g}(x_0), \quad (3)$$

where

$$\mathbf{g}(x_0) = (\mathbf{f}(x_0) - \mathbf{F}^T \mathbf{K}^{-1} \mathbf{c}_0). \quad (4)$$

Obviously, the mean squared error $\sigma^2(x_0)$ is dependent on the arrangement of the sampling locations x_1, x_2, \dots, x_n via the covariance matrix \mathbf{K} , the covariance vector \mathbf{c}_0 and the design matrix \mathbf{F} .

Like with any statistical prediction method one aim of the statistician in kriging is to make best use of the available data, maybe under certain budget constraints or constraints on the number of locations that can be sampled. Thus, the aim is to get as good predictions as possible for the complete area of investigation \mathbf{X} . One possibility to formalize this aim is to try to select the sampling locations x_1, x_2, \dots, x_n in such a way that the sum of all kriging MSEP's

$$\int_{\mathbf{X}} \sigma^2(x_0) dx_0 \quad (5)$$

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} . It has been tried to solve this sampling design problem besides other criteria that also measure the accuracy of kriging predictions in a number of recent research papers. The aim of the next section is to discuss some of these papers.

In soil science especially the remediation of contaminated waste soils is of importance. Not only the area of contamination must be best detected but the contamination concentrations must be optimally predicted. Most often then soil with contamination level above a certain threshold must be remediated. Sampling design for contamination detection in soil science most often works in three phases:

1. The area of contaminated soil must be detected: Here most often samples on a coarse regular rectangular grid are taken to detect the contaminated area.
2. The spatial variability or the covariance function must be best estimated: Here the design from stage 1.) is updated by some nearby locations in order to get also some good estimates of the nugget effect and the small scale variability (the appearance of the covariance function near the origin).
3. Contamination concentrations must be best predicted and their uncertainties be documented.

This work especially deals with phase 2.) and phase 3.). We consider as design criterion for 3.) first the integrated kriging variance averaged over the area of contamination. Next as a combined design criterion for 2.) and 3.) the averaged expected lengths of predictive intervals are considered. Finally it is shown how these design criteria can be generalized to spatial variables having skewed distributions. In this context a criterion for spatial sampling design with Box-Cox transformed spatial variables is proposed.

Mathematically speaking we will approximate the investigated spatial random field by means of a large regression model consisting of cosine-sine Bessel surface harmonics with random amplitudes (Sections 3, 4). This approximating regression model is a direct result of the polar spectral representation theorem of isotropic random fields (Section 3). We show that kriging prediction in the original random field model is equivalent to trend prediction in this approximating Bayesian linear regression model (Section 4). Thus, standard convex experimental design theory for Bayesian linear regression may be used to find spatial sampling designs. Based on this theory we propose two algorithms for the calculation of exact designs (Section 7.3):

- An algorithm for removing redundant design locations.
- An algorithm for adding sampling locations to an existing network.

In Section 5 we generalize our approach to the case that the covariance function is estimated by restricted maximum likelihood (REML) and thus is uncertain. The result is that the actual predictive intervals are wider than when not taking into account this uncertainty. The Smith and Zhu design criterion adjusts to this uncertainty by considering the average of the expected lengths of predictive intervals as a design criterion. Section 6 again is a generalization, because here we assume the data distribution to be skewed and to be transformable to a Gaussian distribution by means of a Box-Cox transformation. Once again the average of the expected lengths of predictive intervals is considered as a design criterion. The proposed methodologies are illustrated by a true data example from soil science (Section 8). Finally, one from the author freely available MATLAB and Octave toolbox is described which can perform all the tasks of spatial sampling design and spatial interpolation (Section 7).

2. A review on spatial sampling design

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; Brus and Heuvelink, 2007; Caeiro et al., 2003; Cox Jr., 1999; Delmelle and Goovaerts, 2009; Diggle and Lophaven, 2006; Dobbie et al., 2008; Groenigen et al., 1999; Hengl et al., 2003; Lark, 2002; US-EPA, 2002). The papers on spatial sampling design may be divided into several categories of which some are overlapping. First of all we may differentiate between design criteria for spatial prediction and for estimation of the covariance function and between combined criteria for both desires at once. Works falling into the category of criteria for prediction are (Brus and Heuvelink, 2007; Fedorov and Flanagan, 1997; Müller and Pazman, 1998; 1999; Müller, 2005; Pazman and Müller, 2001). Criteria for the estimation of the covariance function are considered by (Lark, 2002; Müller and Zimmerman, 1999; Zimmerman and Homer, 1991). Combined criteria we can find in the article (Zhu and Stein, 2006), who consider the minimization of the average expected length of predictive intervals. Further papers falling into this category of combined criteria are Bayesian articles specifying a priori distributions over covariance functions like (Brown et al., 1994); Diggle and Lophaven, 2006; Fuentes et al.,

2007; Müller et al., 2004). Actually (Brown et al., 1994)) and (Fuentes et al., 2007) consider the covariance function to be non-stationary and deal with an entropy based design criterion according to which the determinant of the covariance matrix between locations to be added to the design must be maximized. Both make use of simulated annealing algorithms to find optimal designs obeying their criteria.

At this stage we are at a second distinguishing feature of optimal design algorithms. We can distinguish between stochastic search algorithms like simulated annealing (Aarts and Korst, 1989) or evolutionary genetic algorithms and deterministic algorithms for optimizing the investigated design criteria. With the exception of the works of (Fedorov and Flanagan, 1997; Müller and Pazman, 1998; 1999; Müller, 2005; Pazman and Müller, 2001; Spöck and Pilz, 2010) almost all algorithms for spatial sampling design optimization use stochastic search algorithms for the finding of optimal configurations of sampling locations x_1, x_2, \dots, x_n . The term spatial simulated annealing (SSA) finds its first manifestation in the work of (Groenigen et al., 1999). (Trujillo-Ventura and Ellis, 1991) consider multiobjective sampling design optimization.

Available free software for spatial sampling design optimization is quite rare. Up to the knowledge of the author there are only 6 sampling design toolboxes freely available including an own one: (Gramacy, 2007) has implemented sampling design for treed Gaussian random fields in the R-package *tgpp*. In treed Gaussian random fields the area X of investigation is partitioned by means of classification trees into rectangular sub-areas with sides parallel to the coordinate axes. This software is especially useful for the design of computer simulation experiments, where parameters guiding the computer simulation output are identified as spatial coordinates. Another software especially useful for computer simulation experiments and sequential design is the DAKOTA package (<http://dakota.sandia.gov>). Further papers falling into the category of computer simulation experiments are (Chen et al., 2006; Gramacy and Lee, 2010; Kleijnen et van Beers, 2004; Lim et al., 2002; Mitchell and Morris, 1992; Morris et al., 1993; Schonlau, 1997). Software freely available upon request for research purposes and monitoring network optimization is (Le and Zidek, 2009). This software implements the entropy based design criterion mentioned above. (Gebhardt, 2003) implements a branch and bound algorithm for designing with the criterion (5). (Baume et al., 2011) compare different greedy algorithms for spatial design. Maybe this small list of software for spatial sampling design is not complete and hopefully more software can be obtained from the different authors of research papers upon request. Since for the practitioner there is a strong need for spatial sampling design and almost no software is freely available this was an impetus for the author of this chapter to program a toolbox for spatial sampling design. The *spatDesign* Matlab toolbox of the author is freely available online at

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

An Octave version not implementing as many design criteria as the Matlab toolbox is available at

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

3. The spatial mixed linear model

This section starts our own 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 $\{Y(x) : x \in \mathbf{X} \subseteq \mathbb{R}^2\}$ such that

$$Y(x) = \mathbf{f}(x)^T \beta + \varepsilon(x), \quad E\varepsilon(x) = 0, \tag{6}$$

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

$$\text{Cov}(Y(x), Y(y)) = C(\|x - y\|); \quad x, y \in \mathbf{X}. \tag{7}$$

Then, according to (Yaglom, 1987), the covariance function can be represented in the form

$$C(t) = \int_0^\infty J_0(t\omega) dG(\omega), \quad t \geq 0, \tag{8}$$

where $J_0(\cdot)$ is the Bessel function of the first kind and order 0, $t = \|x - y\|$ is the Euclidean distance between x and y , and $G(\cdot)$ is the so-called (polar) spectral distribution function associated with $C(\cdot)$. As such $G(\cdot)$ is positive, monotonically increasing and bounded from above. On the other hand, knowing $C(\cdot)$ its spectral distribution can be obtained from the inversion formula

$$\frac{G(\omega^+) + G(\omega^-)}{2} = \int_0^\infty J_1(t\omega) \omega C(t) dt, \tag{9}$$

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. Approximating $G(\cdot)$ by means of a step function with positive jumps $a_i^2 = G(\omega_{i+1}) - G(\omega_i)$ at preselected points $\omega_i, i = 0, 1, \dots, n - 1$, and changing to polar coordinates $(t, \varphi) = (\text{radius}, \text{angle})$ the polar spectral representation theorem for m.s.c. isotropic random fields tells us that the error process may be approximated as

$$\varepsilon(t, \varphi) \approx \sum_{m=0}^\infty \{\cos(m\varphi) \sum_{i=1}^n J_m(\omega_i t) U_{m,i}\} + \sum_{m=1}^\infty \{\sin(m\varphi) \sum_{i=1}^n J_m(\omega_i t) V_{m,i}\}, \tag{10}$$

where all the random variables $U_{m,i}$ and $V_{m,i}$ are uncorrelated, have mean zero, and their variances are $\text{var}(U_{m,i}) = \text{var}(V_{m,i}) = d_m a_i^2$; 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

$$Y(x) \approx \mathbf{f}(x)^T \beta + \mathbf{g}(x)^T \alpha + \varepsilon_0(x) \tag{11}$$

From above it becomes clear that the components of the additional regression vector $\mathbf{g}(\cdot)$ are made up of the following radial basis functions (cosine-sine-Bessel-harmonics)

$$g_{m,i}(t, \varphi) = \cos(m\varphi) J_m(\omega_i t); \tag{12}$$

$$m = 0, \dots, M; i = 1, \dots, n$$

$$g_{m,i}(t, \varphi) = \sin((m - M)\varphi) J_{m-M}(\omega_i t);$$

$$m = M + 1, \dots, 2M; i = 1, \dots, n.$$

4. Classical Bayesian experimental design problem

Starting from our spatial mixed linear model (11) 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

$$E(\beta) = \mu \in R^r, \quad \text{Cov}(\beta) = \Phi. \quad (13)$$

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 (11) and (13), we arrive at the **Bayesian spatial linear model** (BSLM)

$$Y(x) = \mathbf{h}(x)^T \gamma + \varepsilon_0(x), \quad (14)$$

where

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

Here $\varepsilon_0(x)$ is white-noise with variance σ_0^2 and \mathbf{A} denotes the covariance matrix of α , resulting after 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 arbitrarily 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:

$$\int_{\mathbf{X}} \mathbf{h}(x_0)^T (\mathbf{H}^T(d_n) \mathbf{H}(d_n) + \sigma_0^2 \Gamma^{-1})^{-1} \mathbf{h}(x_0) dx_0 \rightarrow \min_{d_n}. \quad (15)$$

Here $d_n = \{x_1, x_2, \dots, x_n\}$ 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 $\mathbf{H} = (\mathbf{h}(x_i)^T)_{i=1,2,\dots,n}$ 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 seemingly showing no mathematical structure may be expanded 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

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x) \mathbf{h}(x)^T \xi(dx) + \frac{\sigma_0^2}{n} \Gamma^{-1} \quad (16)$$

and

$$U = \int_{\mathcal{X}} \mathbf{h}(x_0)\mathbf{h}(x_0)^T dx_0, \tag{17}$$

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

$$\Psi(\mathbf{M}_B(\zeta)) = \text{tr}(\mathbf{U}\mathbf{M}_B(\zeta)^{-1}) \tag{18}$$

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

5. The Smith and Zhu design criterion

In real world applications the isotropic covariance function $C_\theta(t)$ is always uncertain and estimated. The kriging predictor used is then based on this estimated covariance function $C_\theta(t)$. 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:

$$\int_{\mathcal{X}} E(\text{length of predictive interval at } x_0) dx_0. \tag{19}$$

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 they show that this design criterion up to order $O(n^{-2})$, where n is the number of data, is equivalent to:

$$\int_{\mathcal{X}} [\sigma_\theta^2(x_0) + \text{tr}(\kappa_\theta^{-1} \{ \frac{\partial \lambda_\theta(x_0)}{\partial \theta^T} \}^T \mathbf{K}_\theta \frac{\partial \lambda_\theta(x_0)}{\partial \theta^T}) + z_{1-\alpha/2}^2 \{ \frac{\partial \sigma_\theta(x_0)}{\partial \theta} \}^T \kappa_\theta^{-1} \frac{\partial \sigma_\theta(x_0)}{\partial \theta}] dx_0 \tag{20}$$

$\rightarrow \underset{d_n = \{x_1, \dots, x_n\}}{\text{Min}}$

Here

$$\kappa_{\theta,ij} = \text{tr}(\mathbf{W}_\theta \frac{\partial \mathbf{K}_\theta}{\partial \theta_i} \mathbf{W}_\theta \frac{\partial \mathbf{K}_\theta}{\partial \theta_j}) \tag{21}$$

is the Fisher information matrix for REML,

$$\mathbf{W}_\theta = \mathbf{K}_\theta^{-1} - \mathbf{K}_\theta^{-1} \mathbf{F}(\mathbf{F}^T \mathbf{K}_\theta^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{K}_\theta^{-1}, \tag{22}$$

$z_{1-\alpha/2}$ is the $1 - \alpha/2$ -quantile of the standard normal distribution, $\sigma_\theta^2(x_0)$ 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 both prediction accuracy and covariance uncertainty into account. Sections 3 and 4 have demonstrated that by using the BSLM (14) as approximation to the true isotropic random field the I-optimality design criterion can be completely expressed in terms of the Bayesian information matrix

$$\mathbf{M}_B = \mathbf{H}^T \mathbf{H} + \sigma_0^2 \mathbf{\Gamma}^{-1}. \quad (23)$$

Going from this information matrix to its continuous version

$$\mathbf{M}_B(\xi) = \int_{\mathbf{X}} \mathbf{h}(x) \mathbf{h}(x)^T \xi(dx) + \frac{\sigma_0^2}{n} \mathbf{\Gamma}^{-1}, \quad (24)$$

where ξ is a probability measure on the design space \mathbf{X} , the extended design functional

$$\Psi(\mathbf{M}_B(\xi)) = \text{tr}(\mathbf{U} \mathbf{M}_B(\xi)^{-1}) \quad (25)$$

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 (18).

In (Spöck et al., 2011) it is shown that also the Smith and Zhu design criterion has some favourable properties, so that classical convex experimental design theory can be applied to this design criterion, too:

- Expression (20) 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 (14) the covariance function actually is parametrized in 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 $\Phi = \text{cov}(\beta)$ 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 $\Phi \rightarrow \infty$ bridges the gap from Bayesian linear to generalized least squares trend estimation. The a priori mean $\mu = E(\beta)$ can be set to $\mathbf{0}$ then.

According to the polar spectral representation (10) several values in the diagonal matrix \mathbf{A} are identical:

$$\mathbf{A} = \text{diag}(\{d_m a_i^2\}_{m=0, \dots, M; i=1, \dots, n; k=1, 2}), \quad (26)$$

where the definitions of d_m and a_i^2 and the indexing derive from the polar spectral representation (10). 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 \mathbb{R}^2 possessing for example an exponential covariance function $B(h) = C \exp(-\frac{3h}{\alpha})$ is given by

$$g(w) = \frac{C \frac{3}{\alpha} w}{((\frac{3}{\alpha})^2 + w^2)^{3/2}}. \tag{27}$$

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

$$a_i^2 = \frac{g(w_i) + g(w_{i-1})}{2} (w_i - w_{i-1}), i = 1, 2, \dots, n, \tag{28}$$

where $0 = w_0 < w_1 < \dots, w_n$ 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 7.3 . We must only replace $\Psi(\mathbf{M}_B(\xi))$ by the Smith and Zhu design functional given in (Spöck et al., 2011).

6. Spatial sampling design for trans-Gaussian kriging

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

$$g_\lambda(z) = \begin{cases} \frac{z^\lambda - 1}{\lambda} & : \lambda \neq 0 \\ \log(z) & : \lambda = 0 \end{cases}. \tag{29}$$

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

$$\mathbf{Y} = (g_\lambda(Z(x_1)), g_\lambda(Z(x_2)), \dots, g_\lambda(Z(x_n)))^T \tag{30}$$

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

$$\Phi(g_\lambda(z); \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0)) * z^{\lambda-1}, \tag{31}$$

where $\Phi(\cdot; \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0))$ is the Gaussian density with mean the ordinary kriging predictor $\hat{Y}_{OK}(x_0)$ at x_0 and based on the transformed variables \mathbf{Y} , and variance the ordinary kriging variance $\sigma_{OK}^2(x_0)$. $z^{\lambda-1}$ is the Jacobian of the Box-Cox transformation.

There is a nice relationship between the α -quantiles z_α of the predictive density (31) and the Gaussian density $\Phi(\cdot; \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0))$ which makes spatial sampling design an easy task: It can be shown that the transformed α -quantiles $g_\lambda(z_\alpha)$ are actually equivalent to the α -quantiles y_α of the Gaussian density $\Phi(\cdot; \hat{Y}_{BK}(x_0), \sigma^2(x_0))$ that is truncated at 0 and where we only take the positive interval $[0, \infty)$ for calculating these y_α -quantiles.

For spatial sampling design we can consider again the average expected length of $1 - \alpha$ -predictive intervals. Once we have calculated the Gaussian density $\Phi(\cdot; \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0))$, truncated it at 0, and have determined its $\alpha/2$ - and $1 - \alpha/2$ -quantiles $y_{\alpha/2}$ and $y_{1-\alpha/2}$ the corresponding quantiles $z_{\alpha/2}$ and $z_{1-\alpha/2}$ can be calculated easily by means of applying the inverse Box-Cox transformation to $y_{\alpha/2}$ and $y_{1-\alpha/2}$. In order to make the expected length of predictive intervals also dependent on REML-estimation of the covariance function, we can

consider instead of the Gaussian density $\Phi(\cdot; \hat{Y}_{OK}(x_0), \sigma_{OK}^2(x_0))$ that unique Gaussian density $\hat{\Phi}$ whose 0.025- and 0.975-quantiles are given by the (Smith and Zhu, 2004) 95% predictive interval

$$\begin{aligned} & \hat{Y}_{OK}(x_0) \pm \\ & 1.96 \quad \sigma_{\theta}(x_0) \left\{ 1 + \frac{1}{2\sigma_{\theta}^2(x_0)} \left[\text{tr}(\kappa_{\theta}^{-1} \left\{ \frac{\partial \lambda_{\theta}(x_0)}{\partial \theta^T} \right\}^T \mathbf{K}_{\theta} \frac{\partial \lambda_{\theta}(x_0)}{\partial \theta^T}) \right] + \right. \\ & 1.96^2 \quad \left. \left\{ \frac{\partial \sigma_{\theta}(x_0)}{\partial \theta} \right\}^T \kappa_{\theta}^{-1} \frac{\partial \sigma_{\theta}(x_0)}{\partial \theta} \right\}. \end{aligned} \quad (32)$$

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. A future paper will take account of also this additional uncertainty.

7. 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 functional but assumes that also the MATLAB Optimization and Statistics toolboxes are installed. In Octave functions like weighted-least-squares that make in MATLAB use of the Optimization Toolbox and especially of the function FMINCON do not work. But in near future also an Octave workaround for the function FMINCON will become available from the author. The spatial sampling design functions corresponding to the Smith and Zhu design criterion need on a standard PC a lot of computation 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. Unfortunately, almost no efforts have been undertaken to date to parallelize also Octave, so the Smith and Zhu part of the Octave toolbox will not work unless you have a lot of time to wait for the results in the unparallelized version. Wolfgang Nowak from the Institute of Hydraulic Engineering (IWS), University of Stuttgart, has added his FFT-Kriging Toolbox to the software package. This is really fast MATLAB and Octave code for Bayesian linear kriging with external drift and works in both 2- and 3-dimensional Euclidean space. The original kriging code of the spatDesign toolbox is not so fast as this code and implements only 2D interpolation. The reason is that the original code works with local kriging neighborhoods and as a consequence for every new prediction an often high dimensional covariance matrix \mathbf{K} between given local observations must be inverted.

The spatDesign Toolbox provides code for all three areas of geostatistics,

- covariance estimation and variography

- spatial interpolation and kriging
- spatial sampling design and planning of monitoring networks.

But the largest emphasis of the toolbox surely is on spatial sampling design and the planning of monitoring networks.

7.1 Covariance estimation and variography software

For covariance estimation and variography the empirical semivariogram estimator of (Matheron, 1962)

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (Y(x_i) - Y(x_j))^2, \quad (33)$$

where

$$N(h) = \begin{cases} \{(i, j) : x_i - x_j \approx h\} : \text{stationarity} \\ \{(i, j) : \|x_i - x_j\|_2 \approx h\} : \text{isotropy} \end{cases} \quad (34)$$

is implemented in both forms the isotropic one and the general stationary one. The corresponding MATLAB and Octave functions are EMPVARIOGRAM.m and EMPVARIOGRAMANISO.m.

Having calculated the empirical semivariogram a theoretical semivariogram model can be fitted to the empirical semivariogram by means of weighted-least-squares. As theoretical semivariogram model a nested model or convex combination of an exponential and a Gaussian semivariogram model with different ranges and nugget is implemented. The corresponding MATLAB and Octave functions are WEIGHTEDLEASTSQUARES.m for the isotropic case and WEIGHTEDLEASTSQUARESANISO.m for the geometrically anisotropic case.

The toolbox implements also transformed-Gaussian kriging. For this reason a maximum likelihood estimation procedure has been implemented that estimates both the Box-Cox transformation parameter λ and the eventually geometrically anisotropic covariance function at once by means of a profile likelihood approach. Iteratively any kind of parameters are fixed and the likelihood function is maximized in the rest of the parameters until convergence. This function is quite slow since the likelihood function is iteratively maximized either in λ , the covariance parameters or the anisotropy parameters. The function performing these tasks is called ESTIMATE_TRANSFO_COV_ML.m.

In the up to here mentioned functions for covariance estimation the trend of the random field is assumed to be constant. For the case that the trend is of the form

$$m(x) = \mathbf{f}(x)^T \boldsymbol{\beta} \quad (35)$$

it must be eliminated before the covariance function or the semivariogram can be estimated. For this case the function CALCULATE_RESIDUALS_TREND_COVARIANCE.m has been implemented. Starting with a least squares estimate of $\boldsymbol{\beta}$ and the corresponding residuals iteratively empirical semivariograms of residuals, corresponding weighted least squares fits to the empirical semivariogram, generalized least squares estimates of $\boldsymbol{\beta}$ and again residuals and semivariograms.....are calculated until convergence.

7.2 Spatial interpolation and kriging software

As simplest interpolation routine Voronoi interpolation is implemented in the function `VORONOI POLYGONAL INTERPOLATION ONGRID.m`. Ungauged locations are given the same value as the closest datum location. Besides Voronoi interpolation also Bayesian linear kriging with external drift and transformed-Gaussian kriging based on the Box-Cox transformation are implemented.

The corresponding functions are named `KRIGELINEAR BAYES ONGRID.m` and `TRANSGAUSSIAN KRIGING ONGRID.m`. The output of the first function are the Bayesian kriging predictions and corresponding TMSEPs on a grid. The output of the second function are the skew predictive distributions from transformed-Gaussian kriging on a grid. Because the complete predictive distributions are calculated several statistics may be derived from them: The function `VISUALIZE POST QUANTILE.m` allows to calculate and visualize several quantiles of the predictive distributions as well as the means, medians and modal values of these distributions. The function `VISUALIZE PROB GREATER.m` allows to calculate maps that give the probabilities that certain thresholds are exceeded. Finally the functions `CROSS VALIDATION.m` and `VISUALIZE CROSS VALIDATION.m` calculate crossvalidation statistics like mean absolute errors, percents of actual data below the quantiles of the predictive distribution and percent actual data above threshold vs. expected percent data above threshold.

7.3 Spatial sampling design software

`spatDesign` 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 also account of the fact that the covariance function is estimated. Concerning trans-Gaussian kriging also the Smith and Zhu design criterion is implemented. The actual version of the toolbox is `spatDesign V.2.2.0`. The implemented criteria for prediction only are:

- I-optimality:

$$\Psi(\mathbf{M}_B(d_n)) = \text{tr}(\mathbf{U}\mathbf{M}_B(d_n)^{-1}) \rightarrow \min_{d_n} \quad (36)$$

$$\mathbf{U} = \sum_{i,j=1}^m \mathbf{h}(x_{i,j})\mathbf{h}(x_{i,j})^T; \quad (37)$$

where the integral in (17) has been replaced by the sum over a fine grid of locations $x_{i,j} \in \mathbf{X}$.

- D-optimality:

$$\Psi(\mathbf{M}_B(d_n)) = |(\mathbf{H}^T(d_n)\mathbf{H}(d_n) + \sigma_0^2\mathbf{\Gamma}^{-1})^{-1}| \rightarrow \min_{d_n} \quad (38)$$

The Smith and Zhu design criterion can be expressed according to Section 6 also as

$$\Psi(\mathbf{M}_B(d_n)) \rightarrow \min_{d_n}, \quad (39)$$

although because of place restrictions we cannot give this expression in detail here. The interested reader is referred to (Spöck et al., 2011).

The basic algorithm for calculating spatial sampling designs is an exchange algorithm from experimental design theory for calculating exact designs and 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.

7.3.1 Exchange algorithm

Step 1. Use some initial design $d_{n,1} = \{x_{1,1}, \dots, x_{n,1}\} \in \mathbf{X}^n$ 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 $d_{n,s}^j = d_{n+1,s} - (x_{j,s}), j = 1, 2, \dots, n + 1$ and delete that point $x_{j^*,s}$ from $d_{n+1,s}$ for which

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

Step 3. Repeat Step 2 until the point to be deleted is equivalent to the point to be added.

For the design functional (18) Step 2 is determined as follows:

$$x_{n+1,s} = \arg \max_{x \in \mathbf{X}} \frac{\mathbf{h}(x)^T \mathbf{M}_B(d_{n,s})^{-1} \mathbf{U} \mathbf{M}_B(d_{n,s})^{-1} \mathbf{h}(x)}{n + \mathbf{h}(x)^T \mathbf{M}_B(d_{n,s})^{-1} \mathbf{h}(x)}$$

$$j^* = \arg \min_{1 \leq j \leq n+1} \frac{\mathbf{h}(x_{j,s})^T \mathbf{Q}_B(d_{n+1,s}) \mathbf{h}(x_{j,s})}{n + 1 - \mathbf{h}(x_{j,s})^T \mathbf{M}_B(d_{n+1,s})^{-1} \mathbf{h}(x_{j,s})},$$

where

$$\mathbf{Q}_B(d_{n+1,s}) = \mathbf{M}_B(d_{n+1,s})^{-1} \mathbf{U} \mathbf{M}_B(d_{n+1,s})^{-1}.$$

For the Smith and Zhu design criterion no such simplification exists and the complete design functional $\Psi(\cdot)$ must be recalculated in every step.

7.3.2 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

$$x_1 = \arg \min_{x \in \mathbf{X}} \Psi(\mathbf{M}_B((x))), \text{ and set } d_1 = (x_1).$$

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 $d_{n,1} = \{x_1, \dots, x_n\}$ as an initial design.

7.3.3 Combination of the algorithms 7.3.1 and 7.3.2

It is a good idea to combine the initial design algorithm 7.3.2 and the exchange algorithm 7.3.1 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.

7.3.4 Reduction of experimental designs

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

- Step 1. Delete that design point x_{j^*} from d for which
 $x_{j^*} = \arg \min_{x_j \in v} \Psi(\mathbf{M}_B(d - (x_j)))$, and set
 $\bar{d} := d - (x_{j^*})$.
 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 7.3.1. In algorithm 7.3.1 merely the calculation of $x_{n+1,s}$ has to be replaced by

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

where d is the initial design that has to be reduced. This improved algorithm has the advantage that design points once deleted can reenter the design in the exchange step.

7.3.5 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):

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

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

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

7.3.6 Computation

The I-optimality and D-optimality criteria have the additional advantage that the optimizations in algorithms 7.3.1 - 7.3.4 can be computationally further simplified. What is more, the averaging over the design area \mathbf{X} corresponding to eq. (17) has to be done only once before actual sampling design by means of computing the matrix \mathbf{U} . The Smith and Zhu design criterion and the design criterion for trans-Gaussian kriging no longer have this advantage and the averaging over the design area \mathbf{X} corresponding to eq. (20) and (32) must be done whenever $\Psi(\mathbf{M}_B(d))$ is calculated. Both eq. (17) and (20) have been simplified by means of replacing the integrals by discrete sums over a fine grid of locations $x_{i,j} \in \mathbf{X}$. Also the calculation of the minima in algorithms 7.3.1 - 7.3.4 has been simplified by means of searching for the minimum over a fine grid of locations. When the minimum is found over the discrete grid we further iterate with a line search algorithm with this minimum as a starting value to the actual global minimum. For all that, the averaging for calculating the integral in the Smith and Zhu design criterion and the design criterion for trans-Gaussian kriging is computationally too intensive to be calculated on a standard PC.

Recently CUDA technology (www.nvidia.com) has been developed for NVIDIA graphic cards. This allows to make use of the parallel performance of these graphic cards and to put intensive floating point operations to these GPU's. Thus, we have invested in such a graphics card and now do the averaging operations corresponding to expression (20) and (32) in parallel on a multiprocessor NVIDIA GTX 580 GPU. To this, we have installed GPUmat (www.gp-you.org) a free software for MATLAB (www.mathworks.com) that automatically can thread operations to the NVIDIA GPU. The usage of this software is quite easy. We just have to specify MATLAB objects that we want to calculate on the GPU as GPUdouble or GPUsingle. Performance is immense. Processing on the NVIDIA GTX 580 is between 100 to 200 times faster than on a standard Pentium 8 Core 3.06 Ghz CPU.

7.3.7 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 investigation \mathbf{X} of points 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 to locations from "Poolcomplete" or from "Pooladd" and the total number of sampling locations remains constant.

8. 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 for trans-Gaussian kriging are given. The data set considered is the so-called Gomel data set. It comprises 591 sampling

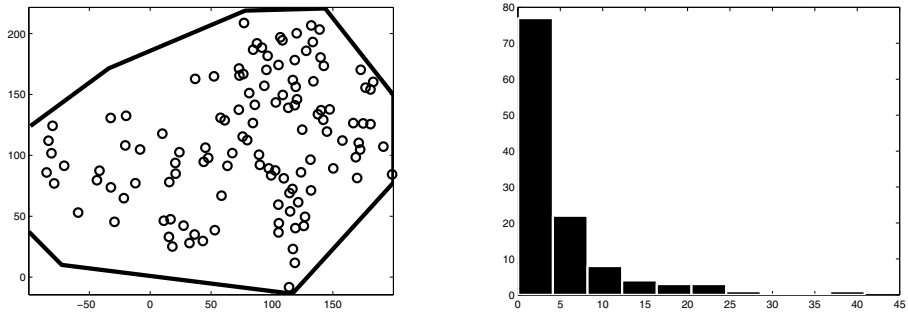


Fig. 1. Left: The 119 sampling locations of the Gomel data set. Right: The histogram of the Cs137 concentrations.

locations in the region of Gomel, Belarus, where ten years after the Chernobyl accident the concentration of Cs137 in soil has been measured. For demonstration purposes we do not use the full data set but only every fifth location, so that we have to deal with only 119 data. The following Matlab code plots the data locations and calculates a histogram of the Cs137 concentrations (see Fig. 1.).

- `x=Gomel.x(1:5:end);`
- `y=Gomel.y(1:5:end);`
- `z=Gomel.zaver(1:5:end);`
- `figure()`
- `plot(x,y,'o')`
- `hold on`
- `plot(polyBoxGomeliso.x,polyBoxGomeliso.y)`
- `figure()`
- `hist(z)`

In Fig. 1. we see that there are obviously areas in the design region that look very empty having no sampling locations. To the East the sampling grid is more dense than to the West. From the histogram of Fig. 1. it becomes obvious that the distribution of Cs137 is very skewed to the right. Thus, this data set is very suitable for testing the trans-Gaussian kriging spatial sampling design routines of the `spatDesign` toolbox. In Fig. 2. a Voronoi polygonal interpolation of the Cs137 concentrations is given. There are obviously 3 regions having very high Cs137 concentration, but the rest of the Gomel region has average to low concentrations. The 3 regions having high concentrations are regions that had heavy rainfall during the Chernobyl accident so that the radioactive cloud was washed out into the soil there.

- `grid=generategrid2(-100,200,-50,250,150,150,polyBoxGomeliso);`
- `prediction=voronoipolygonalinterpolationongrid(x,y,z,grid);`
- `a=reshape(prediction.prediction,150,150);`

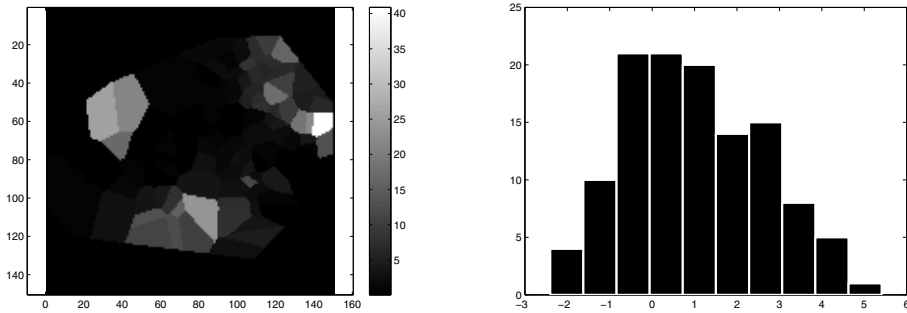


Fig. 2. Left: Voronoi interpolation of the Cs137 values. Right: Histogram of the Box-Cox transformed Cs137 concentrations.

- figure()
- imagesc(a(end:-1:1,:))
- colorbar

Next, let us estimate the Box-Cox transformation parameter λ_0 and an exponential isotropic covariance function for the corresponding Gaussian random field. δ_0 comprises the parameters of this estimated exponential covariance function.

- `options=optimset('LargeScale','on','MaxFunEvals',200,'TolFun',1e-3,'MaxIter',100);`
- `[\lambda_0,\delta_0]=estimate_transfo_cov_ml(x,y,z,0,[1,0,0,1],[0,2.4,80,0,0],0:15:150,10,-20,20,0.01,1,[1,1,1,0,0],0,0.001,0.001,options);`

The estimated parameters are $\lambda_0=0.1919$ and $\delta_0=[0.1644, 3.3601, 149.3302]=[nugget, sill, range]$. To check whether λ_0 truly transforms the Cs137 concentrations to Gaussianity we calculate the histogram of the transformed data (Fig. 2.). As a next step

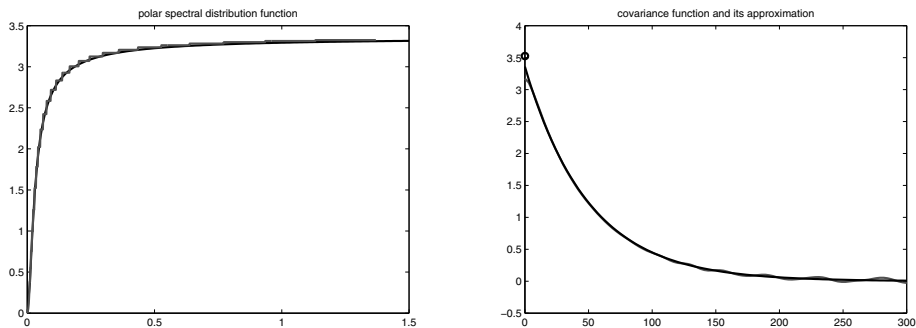


Fig. 3. Left: Polar spectral distribution function and its approximation (grey). Right: True covariance function (black) and its worst approximation (grey).

we calculate the polar spectral distribution function (Fig. 3.).

- `plotspectraldist(0:0.01:1.5,\delta_0);`

Obviously this spectral distribution function almost attains its maximum of 3.36 at $w_0 = 1.5$. We now select the frequencies $w_i, i = 1, 2, \dots, n$, calculate an approximation to the spectral distribution function via a step function (the steps are the a_i^2) and check whether with this selected approximation to the spectral distribution function the original covariance function becomes well approximated (Fig. 3).

- load wscaled
- w=wscaled*1.5;
- [w,deltastep]=expstep(w,delta0(3),delta0(2));
- approxspectraldist(w,deltastep)
- plotcovarianceapprox(w,45,deltastep,delta0,-100:3:200,250,-100,-100,200,-50,250);

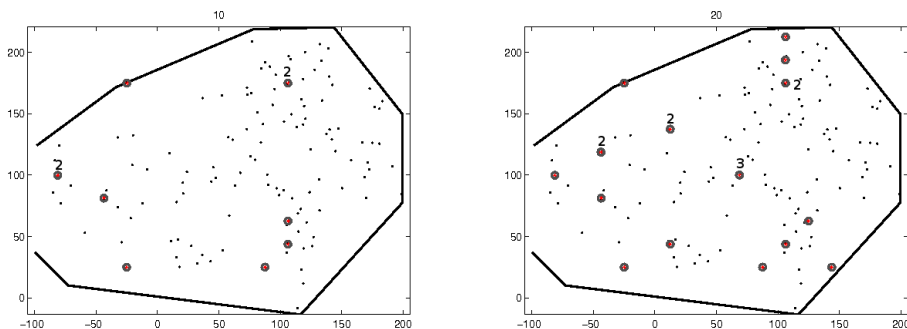


Fig. 4. Left: 10 locations added to design (circles) and their multiplicities. Right: 20 locations added to design (circles) and their multiplicities.

A look at the approximating covariance function in Fig. 3. shows that at the origin the difference between the true covariance function and the approximating covariance function is 0.17. 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 0.17 to the nugget effect 0.1644 of the true covariance function. Thus, $0.17+0.1644$ is the variance of the uncorrelated error process $\epsilon_0(x)$ in our BSLM (14).

We now have everything that we need in order to do spatial sampling design for trans-Gaussian kriging. We consider to add 20 additional sampling locations from the complete design region X to the available grid of 119 sampling locations (see Fig. 4.).

- [xadd,yadd,zadd,avglengthpredint]=optimally_add_n_locations_from_poolcomplete(
 - {}, % no external drift
 - [], % no need to specify the matrix U
 - x,y,z, % the Gomel data
 - 0.17+delta0(1), % the variance of the uncorrelated error process $\epsilon_0(x)$
 - [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
 - w, % the frequencies of the Bessel harmonics
 - 45, % the largest angular frequency
 - deltastep, % the a_i^2

delta0, % the parameters of the exponential covariance function
 lambda0, % the Box-Cox transformation parameter
 20, % we want to add 20 samples
 -100,200,-50,250, % the size of the design region
 100, % maximally iterate 100 times in the exchange step
 polyBoxGomeliso, % the polygonal design region
 17,17, % discretization in x an y when considering new samples
 'z' % we apply the (Smith and Zhu, 2004) design criterion to trans-Gaussian kriging
);

Fig. 5. shows the expected lengths of the 95% predictive intervals for the optimal 10 points and 20 points design, respectively. Fig. 6. plots the decrease of the average of the expected lengths of the 95% predictive intervals when adding up to 20 design locations. The calculation of

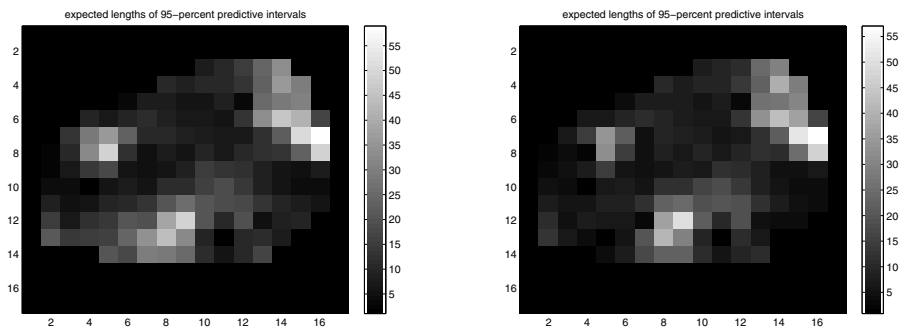


Fig. 5. Left: Expected lengths of 95% predictive intervals. 10 locations added to design. Right: Expected lengths of 95% predictive intervals. 20 locations added to design.

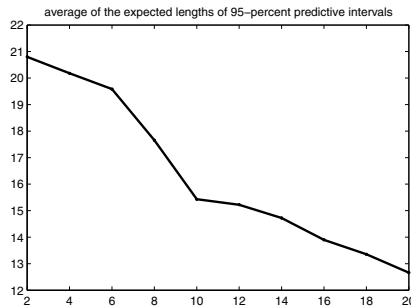


Fig. 6. Average of the expected lengths of 95% predictive intervals.

the optimal 20 point design takes on an Intel i7 8 Core CPU and a NVIDIA 580 GPU with 1.6 Gb RAM 2 weeks. Similar calculations for the simpler design functional (36) take without NVIDIA Cuda support 2 days. From Fig. 4. it is obvious that the first design locations are added to the margin of areas that look empty. In order to get also the covariance function (especially close to its origin) well estimated design locations are selected close to existing

locations or they are selected with multiplicities greater than 1. There is a fundamental difference between designing for Gaussian random fields and our designs for trans-Gaussian kriging. Contrary to the designs for Gaussian random fields designs for trans-Gaussian kriging are dependent also on the data by means of the ordinary kriging predictor in eq. (32).

9. Conclusions

Our approach to spatial sampling design is motivated by the fact that up to date most sampling design implementations use stochastic search algorithms like simulated annealing. The approach to spatial sampling design proposed in this Chapter makes use of the mathematical structure of the investigated design functionals. Based on the polar spectral approximation of the isotropic random field we are able to use classical experimental design theory for the calculation of spatial sampling designs. One only missing link for the Smith and Zhu design criterion and for the design criterion of trans-Gaussian kriging is the fact that up to date we were not able to show convexity properties of the corresponding design functionals. This is surely a topic for future research. The data example giving radioactivity measurements in soil shows that trans-Gaussian kriging and sampling design for it are relevant also to soil science.

A topic for future research is to extend the proposed approach also to non-stationary random fields. Some work towards this direction has already been done in (Spöck, 2008). The approach discussed therein locally maintains isotropy but globally is non-stationary and is also based on the approximation of a so-called locally isotropic random field by means of a cosine-sine-Bessel surface-harmonics regression model with random amplitudes.

10. References

- Aarts, E.H. and Korst, J. (1989). *Simulated Annealing and Boltzman Machines*, Wiley, New York, 284 p..
- Baume, O.P., Gebhardt, A., Gebhardt, C., Heuvelink, G.B.M., Pilz, J. (2011). Network optimization algorithms and scenarios in the context of automatic mapping, *Computers & Geosciences*, 37, (3), pp. 289-294.
- 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? Choosing 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.
- Chen, V.C.P., Tsui, K.L., Barton, R.R. and Meckesheimer, M. (2006). A review on design, modeling and applications of computer experiments, *IIE Transactions*, 38, pp. 273-291.
- 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. (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..
- Fedorov, V.V. and Flanagan, D. (1997). Optimal monitoring network design based on Mercer's expansion of the covariance kernel, *Journal of Combinatorics, Information and System Sciences*, 23, pp. 237-250.
- Fuentes, M., Chaudhuri, A. and Holland, D.M. (2007). Bayesian entropy for spatial sampling design of environmental data, *Journal of Environmental and Ecological Statistics*, 14, pp. 323-340.
- Gebhardt, C. (2003). *Bayesian Methods for Geostatistical Design*, Ph.D. Thesis, University of Klagenfurt, Austria, 139 p..
- Gramacy, R.B. (2007). tgp: An R package for Bayesian nonstationary semiparametric nonlinear regression and design by treed Gaussian Process models, *Journal of Statistical Software*, 19, Issue 9., 47 p..
- Gramacy, R.B. and Lee, H.K.H. (2010). Optimization under unknown constraints, *Bayesian Statistics 9*, (ed.) Bernardo, J.M. et al., Oxford University Press, 18 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.
- Kleijnen, J.P.C. and Beers, W.C.M. van (2004). Application driven sequential design for simulation experiments: Kriging metamodeling, *Journal of the Operational Research Society*, 55, pp. 876-893.
- 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. (2009). *Statistical Analysis of Environmental Space-Time Processes*, Springer, New York, 356 p..
- Lim, Y.B., Sacks, J. and Studden, W.J. (2002). Design and analysis of computer experiments when the output is highly correlated over the input space, *Canadian Journal of Statistics*, 30, pp. 109-126.
- Matheron, G. (1962). *Traite de geostatistique appliquee*, Tome I. Memoires du bureau de recherches geologiques et minieres principes, Paris, Editions Technip, 14, 333 p..
- Mitchell, T.J. and Morris, M.D. (1992). Bayesian design and analysis of computer experiments: Two examples, *Statistica Sinica*, 2, pp. 359-379.
- Morris, M.D., Mitchell, T.J. and Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction, *Technometrics*, 35, pp. 243-255.
- Müller, P., Sanso, B. and De Iorio, M. (2004). Optimal Bayesian Design by Inhomogeneous Markov Chain Simulation, *Journal of the American Statistical Association*, 99, pp. 788-798.
- Müller, W.G. and Pazman, A. (1998). Design measures and approximate information matrices for experiments without replications, *Journal of Statistical Planning and Inference* 71, pp. 349-362.

- Müller, W.G and Pazman, A. (1999). An algorithm for the computation of optimum designs under a given covariance structure, *Journal of Computational Statistics*, 14, pp. 197-211.
- Müller, W.G. (2005). A comparison of spatial design methods for correlated observations, *Environmetrics*, 16, pp. 495-505.
- Müller, W.G. and Zimmerman, D.L. (1999). Optimal designs for variogram estimation. *Environmetrics*, 10, pp. 23-37.
- 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.
- Pazman, A. and Müller, W.G. (2001). Optimal design of experiments subject to correlated errors, *Statistics and Probability Letters*, 52, pp. 29-34.
- Pilz, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models*, Wiley, New York, 306 p..
- Schonlau, M. (1997). *Computer experiments and global optimization*, Dissertation, University of Waterloo, Canada 130 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. (2008). Non-stationary spatial modeling using harmonic analysis, in J.M. Ortiz and X. Emery (ed.), *Proceedings of the Eight International Geostatistics Congress*, Gecamin, Chile, pp. 389-398.
- 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. (2011). Simplifying objective functions and avoiding stochastic search algorithms in spatial sampling design, accepted by *Stochastic Environmental Research and Risk Assessment*, 20 p.
- Trujillo-Ventura, A. and Ellis, J.H. (1991). Multiobjective air pollution monitoring network design, *Atmospheric Environment*, 25, pp. 469-479.
- 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..
- Zhu, Z. and Stein, M.L. (2006). Spatial sampling design for prediction with estimated parameters, *Journal of Agricultural, Biological and Environmental Statistics*, 11, pp. 24-49.
- Zimmerman, D.L. and Homer, K.E. (1991). A network design criterion for estimating selected attributes of the semivariogram, *Environmetrics*, 2(4), pp. 425-441.



Principles, Application and Assessment in Soil Science

Edited by Dr. Burcu E. Ozkaraova Gungor

ISBN 978-953-307-740-6

Hard cover, 394 pages

Publisher InTech

Published online 22, December, 2011

Published in print edition December, 2011

Our dependence on soil, and our curiosity about it, is leading to the investigation of changes within soil processes. Furthermore, the diversity and dynamics of soil are enabling new discoveries and insights, which help us to understand the variations in soil processes. Consequently, this permits us to take the necessary measures for soil protection, thus promoting soil health. This book aims to provide an up-to-date account of the current state of knowledge in recent practices and assessments in soil science. Moreover, it presents a comprehensive evaluation of the effect of residue/waste application on soil properties and, further, on the mechanism of plant adaptation and plant growth. Interesting examples of simulation using various models dealing with carbon sequestration, ecosystem respiration, and soil landscape, etc. are demonstrated. The book also includes chapters on the analysis of areal data and geostatistics using different assessment methods. More recent developments in analytical techniques used to obtain answers to the various physical mechanisms, chemical, and biological processes in soil are also present.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Gunter Spöck (2011). Spatial Sampling Design and Soil Science, Principles, Application and Assessment in Soil Science, Dr. Burcu E. Ozkaraova Gungor (Ed.), ISBN: 978-953-307-740-6, InTech, Available from: <http://www.intechopen.com/books/principles-application-and-assessment-in-soil-science/spatial-sampling-design-and-soil-science>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.